Heart Rate and Heart Rate Variability From Single-Channel Video and ICA Integration of Multiple Signals

Riccardo Favilla, Veronica Chiara Zuccalà, and Giuseppe Coppini, *Member, IEEE*

***Abstract*—Unobtrusive monitoring of vital signs is relevant for both medical (patient monitoring) and non-medical applications (e.g., stress and fatigue monitoring). In this paper, we focus on the use of imaging photoplethysmography (iPPG). High frame rate videos were acquired by using a monochrome camera and an optical band-pass filter (560 ± 20 nm). To enhance iPPG signal, we investigated the use of independent component analysis (ICA) pre-processing applied to iPPG signal from different regions of the face. Methodology was tested on 30 healthy volunteers. Heart rate (HR) and standard time and frequency domain descriptors of heart rate variability (HRV), simultaneously extracted from videos and ECG data, were compared. A mean absolute error (MAE) about 3.812 ms was observed for normal-to-normal intervals with or without ICA pre-processing. Smaller MAE values of frequency domain descriptors were observed when ICA pre-processing was used. The impact of both video frame rate and video signal interval were also analyzed. All the results support the conclusion that proposed ICA pre-processing can effectively improve the HR and HRV assessment from iPPG.**

***Index Terms*—Vital signs, heart rate, heart rate variability, imaging photoplethysmography, independent component analysis.**

I. Introduction

HEART rate (HR) and heart rate variability (HRV) are crucial indicators of the psychophysical status of an individual and are useful clues for detecting risky conditions. HR is the mean number of heartbeats per minute. It is basic indicator of cardiovascular homeostasis. HR varies according to the body's physical needs, changes being observed in a variety of conditions including physical exercise, sleep, anxiety, stress, illness, and assumption of drugs. Heart rate monitoring is therefore important in both normal and disease conditions. In illness, an association exists between HR and outcome in heart failure and high baseline heart rate is considered a cardiovascular risk factor [1]. HRV is the fluctuation in the time intervals between adjacent heartbeats. It is an index of the adaptation of heart to circumstances by detecting and readily responding to unpredictable stimuli. HRV is mainly modulated by the sympathetic and parasympathetic components of the autonomic nervous system [2]. In particular, the sympathetic stimulation is activated in response to stress, exercise, and heart disease and this causes an HR increase [2]. Parasympathetic activity is the result of the function of internal organs, reaction to trauma, allergic reactions, and the inhalation of irritants. This activity determines a decrease of HR [2]. HRV is altered in several cardiac diseases [3]. In addition, studies have also shown that smoking reduces the HRV due to the increase of the sympathetic activity and reduction of the vagal activity [4]. As a matter of facts, HRV is an indicator of health status in the general population [5], of adaptation to stress in athletes [6], and of fatigue in drivers [7]. Furthermore, HRV is important to measure mental stress and, coupled with HR, can be used to monitor individual well-being in behavioural research [8].

Non-contact measurement of HR and HRV could greatly simplify data acquisition, making such measurements easily available in non-clinical scenarios (e.g., driver monitoring [9], man-machine interaction monitoring [10]). Furthermore, non-contact systems could allow effective remote monitoring of patients. Several methods have been studied with this aim: HR from speech [11], thermal imaging [12], microwave Doppler effect [13], [14], and imaging photoplethysmography (iPPG) [9], [15]–[17]. Imaging photoplethysmography is based on similar principles as the detection of finger pulse amplitudes. The heartbeat initiates the pulse wave and it travels through the arterial vascular system reaching various parts of the body. Here, the pulse wave determines a short-termed change of blood volume in the observed skin region: the intensity of the absorbed light depends on this volume, likewise the finger pulse measurement (standard plethysmography - PPG) [18]. It is therfore not surprising that several studies have been performed to evaluate whether heart rate can be assessed from video streams [19], thus avoiding the use of wearable sensors.

Concerning the signal acquisition in iPPG, video sequences are usually taken from a subject's face due to the high blood supply and the imaging simplicity. Both webcams [16], [20]–[23] and conventional video cameras [24]–[27] have been used

Manuscript received March 26, 2018; revised September 20, 2018 and October 29, 2018; accepted October 29, 2018. Date of publication November 7, 2018; date of current version November 6, 2019. This work was supported in part by the European Community in the frame of the FP7-Project SEMEOTICONS under Grant 611516. *(Corresponding author: Veronica Chiara Zuccalà.)*

R. Favilla and G. Coppini are with the CNR Institute of Clinical Physiology, 56124 Pisa, Italy (e-mail: riccardo.favilla@ifc.cnr.it; giuseppe.coppini@ifc.cnr.it).

V. C. Zuccalà is with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi," University of Bologna, 40126 Bologna, Italy, and also with the CNR Institute of Clinical Physiology, 56124 Pisa, Italy (e-mail: veronica.zuccala2@unibo.it).

Digital Object Identifier 10.1109/JBHI.2018.2880097

2168-2194 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

for this task. Webcams provide low-cost and easily available setups, whereas standard cameras are expected to produce better quality signals with higher spatial and temporal resolution, along with extended spectral capabilities. Tayibnapis *et al.* [9] utilized an infra-red (IR) camera in order to capture the driver's facial images. Zhao *et al.* [28] utilized a near-IR camera in order to extract HR and respiratory rate (RR) in both day and night light conditions. Authors also conclude that HR and RR can be extracted using single channel images. Typically, the strongest plethysmography signal is contained in the green channel [29]. This result is due to the fact that (oxy-) hemoglobin absorbs green light better than that red. Moreover, green light penetrates deeper into the skin than the blue light [25]. It stands to reason that using spectrally tuned band can significantly enhance iPPG signal.

Concerning image processing, several methods have been proposed to extract blood volume pulse (BVP) signal from face movies. A widely adopted framework is based on Blind Source Separation (BSS) techniques. The video signal is modeled as a mixture of contributes including BVP, motion artifacts, and external illumination changes. Poh *et al.* [16] processed the RGB video components by Independent Components Analysis (ICA) to enhance BVP. An alternative approach based on BSS by Principal Components Analysis (PCA) was suggested by Lewandowska *et al.* [23]. BSS was applied to RGB video components providing acceptable BVP estimation under the assumption of very small face motion and low image noise. To overcome such BSS (ICA- and PCA-based) limitations, several researchers investigated alternative processing methods. Wang *et al.* [24] exploits image redundancy to counteract the effect of face movement. Feng *et al.* [22] adopted a simplified model of the optical properties of the skin to compensate for head motion. Tarassenko *et al.* [26] proposed an iPPG system exploiting autoregressive (AR) modeling of video time series to compute HR together with respiratory rate and Oxygen saturation. Though video signal intensity is the most utilized source of information to detect BVPs, a different method based on head motion related to BVP propagation is reported in [30].

The assessment of HRV from video is usually more demanding than measurement of HR. In fact, HR estimation only requires BVP detection so as to compute the average number of pulse per minute, HRV assessment requires a precise temporal localization of pulses. In principle, most methods for HR assessment from video can be adapted to estimate HRV. Video sequences are usually processed to detect blood volume pulses and obtain a tachogram (i.e., the time series of inter-beat interval duration), similarly to what done in ECG-based analysis. In particular, the tachogram can be analyzed both in time domain and/or in frequency domain [31]. For example, in [16] iPPG tachogram is used to compute standard HRV descriptors both in time and frequency domain. High correlation with parameters derived by standard PPG on 15 subjects is reported. Another solution based on Zero-phase Component Analysis (ZCA) has been reported by Iozzia *et al.* [32] to evaluate the suitability of iPPG to assess autonomic response. Tayibnapis *et al.* [9] applied the PCA to estimate HR and HRV signals from a RGB video. These signals were then used with a set of facial features to detect fatigue through a Support Vector Machine.

In general, HR estimation from iPPG is in good agreement with reference techniques, high correlation coefficients being reported by many authors. Standard PPG, due to its simplicity, is the most common reference in iPPG studies. Though PPG provides accurate HR measurements, the gold standard for HRV assessment remains ECG recording that allow a fine localization of the heart beat [31]. PPG seems a viable surrogate of ECG for healthy subject at rests, but its performances tends to worsen in exercise and in diseased people [33], [34]. In addition, it must be pointed out that, up to now, experimental results are from small-sized populations, data acquisition usually occurring in highly idealized conditions. In particular, short-term analysis is considered by most researchers with video recordings lasting non longer than 60 s, instead of the usual 5 min recommend in [31], usually captured with a subject staying still in front of a camera. Obviously, when acquisition constraints are relaxed, several factors can alter the iPPG signal and degrade the performances of processing algorithms. For example subject movements can be a source of troublesome artifacts. Rigid motions can be somehow compensated by proper tracking of the region of interest but the effects of non-rigid facial movements can be hardly removed. Rapid changes of environmental lighting can also interfere with iPPG and this need to be taken into account when setting up an iPPG system. Imaging parameters such as spatial and temporal sampling and the sensor spectral response impact on image quality. Though spatial resolution does not seem a major iPPG problem (mostly because iPPG signal is usually averaged on large facial areas), temporal resolution is expected to affect BVP localization with a possibly relevant impact on HRV analysis. In this respect, it is worth remembering that in standard ECG-based HRV measurements sampling rates lower than 100 Hz are discouraged [31]. Though iPPG is based on different principles than ECG, temporal sampling requirements should be more deeply investigated.

In this paper we report on a novel iPPG methodology to monitor HR and HRV of normal subjects [35]. Based on the previous considerations, this work has a twofold aim: a) investigate the use of ICA pre-processing of spectrally-tuned single-channel video streams to enhance iPPG signal by exploiting multiple spatial measurements, b) analyze the impact of imaging parameters on HR and HRV descriptors estimation. In next sections, after describing the experimental setup, we report on the estimation of HR and HRV parameters (computed both in time and frequency domain) in varying experimental conditions. The iPPG-derived parameters are compared with corresponding descriptors derived from simultaneous ECG recording.

II. Blood Volume Pulse From Video

The propagation of blood volume pulses makes skin reflectance changing in time, which, in turn, changes the color of video recordings and this is the primary source of information we consider to detect HR from face videos. However, color changes can be produced by other physiological processes, voluntary subject movement as well as ambient factors. In this work, we refer to the experimental setup sketched in Fig. 1, where an individual is sitting in front of a camera with the face uniformly lit by a stable white LED lamp with no other (natural

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

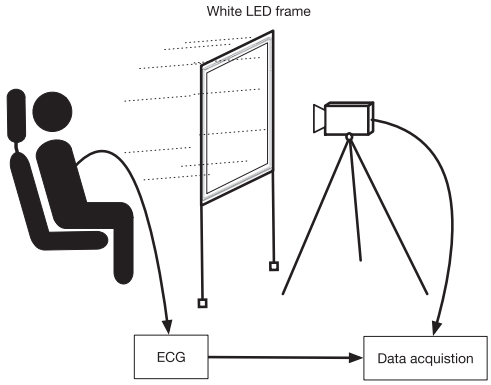

Fig. 1. Data acquisition setup.

or artificial) light sources. On this ground, we will focus mainly on subject-related effects.

Among these, respiration is responsible of small cyclic head and neck movements that can be detected by imaging face regions including motion sensitive features (e.g., the borders of lips). In addition, involuntary movements (e.g., muscular tremor and eye blinking) should be taken into account as potential source of interference. Voluntary head motion that may be related to a wide variety of activities (e.g., speaking, tracking moving objects) can be expected to introduce relevant picture changes interfering with the recording of vital signs. In this work, the BVP signal is extracted from a set of observations of the iPPG video taken in different facial regions. Let assume that, for a fixed wavelength, a iPPG signal $\mathrm{g}(t)$ is obtained at time t by spatially averaging the image intensity in a region of interest (ROI). We assume that averaging makes camera noise negligible.

According to a widely accepted model for iPPG [36], $\mathrm{g}(t)$ is produced by specular and diffuse reflections of the incident light $\mathrm{I}_0(t)$. Separating continuous and time varying, zero mean, contributes, one can write:

$$\mathrm{g}(t) = \mathrm{I}_0(1 + \mathrm{i}(t))(\kappa + \mu\mathrm{m}(t) + \sigma\mathrm{p}(t)) \tag{1}$$

where $\mathrm{I}_0(t) = \mathrm{I}_0(1 + \mathrm{i}(t))$, κ accounts for all contributes to constant reflection, $\mathrm{m}(t)$ describes changes of specular reflection, and $\mathrm{p}(t)$ relates to changes of diffuse reflection. The coefficients μ and σ quantify the change of skin reflection and the strength of BVP pulse, respectively. It is worth noting that the term $\mathrm{p}(t)$ is the only contribute pertaining BVP. Equation (1) can be simplified by assuming that the time varying components are small compared to continuous ones and that their cross products can be neglected:

$$\mathrm{g}(t) \approx \mathrm{I}_0\kappa + \mu\mathrm{I}_0\kappa\mathrm{i}(t) + \mathrm{I}_0\mu\mathrm{m}(t) + \mathrm{I}_0\sigma\mathrm{p}(t) \tag{2}$$

We can therefore conclude that the iPPG signal is, approximately, a linear combination of three contributes: $\mathrm{i}(t)$ related to illumination changes that, in our case, is related to subject motion with respect to the light source,[1] $\mathrm{m}(t)$ that describes specular changes of skin surface due to subject motion, and $\mathrm{p}(t)$ that must be imputed to diffused reflection and accounts for BVP contribute. The signal $\mathrm{i}(t)$ is often estimated from RGB videos which provides three different combinations of the underlying signals.

In the following we will consider monochrome images acquired with a narrow-band filter centered on hemoglobin absorption peaks. Multiple iPPG signals are generated from different ROIs in regions having high vascularization. Assuming that $i(t)$, $m(t)$, and $p(t)$ are approximately uniform among the ROIs we rewrite (2) for each region ROI:

$$\mathrm{g}_j(t) = \mathrm{I}_0\kappa_j + \mathrm{I}_0 c_j\mathrm{i}(t) + \mathrm{I}_0\mu_j\mathrm{m}(t) + \mathrm{I}_0\sigma_j\mathrm{p}(t) \tag{3}$$

where κ_j, μ_j, and σ_j are the constant reflection, the specular and the diffuse reflection in the j ROI, respectively. Based on (3) we exploit the spatial dependence of $g(t)$ to enhance BVP estimation by means of ICA.

III. Experiments

A. Study's Participants

Thirty healthy participants (12 females and 18 males with mean age of 39.17 years, range from 22 to 61) were recruited for voluntary participation in this study. The presented examination was part of a larger study aiming at evaluating the possibility of using imaging techniques for individual self-assessment and self-monitoring of cardio-metabolic risk [35]. The experimental protocol, conducted in Pisa, Italy, was approved (September 10th, 2015) by the local Ethical Committee of Pisa, Italy (Study 213/2014) and received the Ethical Clearance certification (0086129, November 11th, 2014) by the Italian National Commission for Research Ethics and Bioethics. Written informed consent was obtained from all subjects included in this study. The study protocol is compliant with the European Union General Data Protection Regulation [37].

B. Experimental Setup and Video Acquisition

The subjects were sitting still in front of the camera at a distance of about one meter (see Fig. 1). The chair had a headrest to contain head motion and making the video recording comfortable for the volunteer. After a three-minutes rest the subject's face was recorded for five minutes. The subjects were illuminated by a white LED light source and a Gigabit Ethernet camera with a CMOS monochrome sensor (UI-55240SE-NIR-GL, IDS GmbH, DE) was used. The camera was operated at 133 fps with 352×224 pixel, 8 bits/pixel. Video images were saved in real time in raw format by a Mac mini computer (Intel Core i7 dual-core processor with 3 GHz, 16 GB RAM, and 512 GB solid state disk). In order to enhance the iPPG signal, the camera had a band-pass filter centered at 560 nm with a bandwidth of 40 nm. Such a band is expected to contain a strong plethysmographic signal [25]. The ECG lead I was recorded at 300 Hz simultaneously with video through an ECG Five Channel Module (EG05000, Medlab GmbH, DE).

All videos and ECG recordings were acquired and analyzed through a custom software written in C++.

[1] In principle, small cyclic head movements due to Newtonian reaction to blood influx may results in related subtle cyclic changes of skin reflectance. However, their impact on $\mathrm{m}(t)$ in a uniform skin patch, not including motion-sensitive features, can be neglected.

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

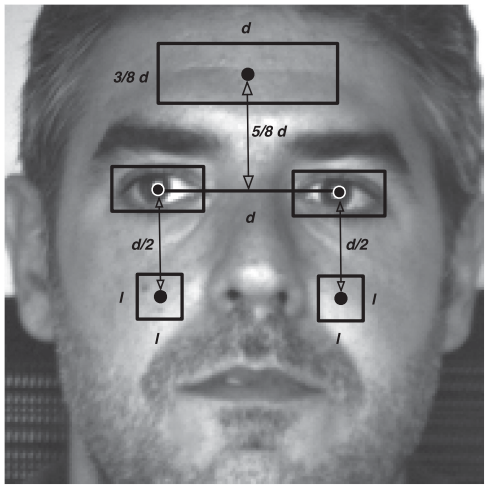

Fig. 2. Dimensions and positions of the ROIs (ROI1 on the forehead area, ROI2 on the right cheek, and ROI3 on the left cheek). d is the distance between the centers of the rectangles around the eyes and l is the side of the square that is set to 20 pixel.

C. ECG Signal Analysis

Reference data were obtained by extracting the tachogram from the ECG signal. To this end, Inter Beat Intervals (IBIs) were defined following QRS complex detection [38]. To remove possible artifacts, the IBIs were processed by the NC-VT algorithm [39] with a tolerance of 30%, all the IBIs with a duration less than 200 ms being removed.

The signals were analyzed both in time domain and in frequency domain. Concerning time domain, we calculated the average time NN between adjacent normal heartbeats and its standard deviation SDNN. Concerning frequency domain, analysis of HRV was performed by power spectrum density (PSD) estimated by the Lomb-Scargle periodogram [40]. This method is able to cope with non-uniformly sampled data even in presence of large gaps, which makes tachogram interpolation unnecessary. According to the standard definition of HRV frequency bands [31], low frequency (LF) and high frequency (HF) were calculated as the area under the PSD curve corresponding from 0.04 Hz to 0.15 Hz and from 0.15 Hz to 0.4 Hz, respectively [39]. The LF component reflects both sympathetic and parasympathetic actions, the HF component reflects parasympathetic action, and the LF/HF ratio is a measure of the sympatho/vagal balance [39].

D. ROI's Selection

For each video sequence, the Cascading Classifiers algorithm (OpenCV v. 3.0.0) was used to detect the face of the subject and locating the center of each eye in the first frame of the recorded video, the related distance being named d. Three different regions of interest were selected (Fig. 2). ROI definition was driven by anatomical and empirical considerations. In particular, ROIs were placed in highly vascularized regions of the face. We also tried to have a large integration area so as to reduce the impact of image noise on BVP estimation. In addition, ROI sizes were constrained to reduce interference with regions interested by marked involuntary movement, such as eyes and lips. The first ROI (ROI1) was a rectangle of the forehead area and the others (ROI2 and ROI3) were squares on the right and left cheek, respectively. ROI on the forehead was $d \times 3/8\, d$ placed at $5/8\, d$ above the inter-pupillary line. The two ROIs on the cheeks had a fixed side $l = 20$ pixel (about 20 mm on the subject face) placed at $d/2$ under the pupil.

E. Video Signal Analysis

In order to extract HR and HRV parameters from video, for each frame, grey levels were averaged in each ROI, which provided three iPPG signals ($v_1(t)$, $v_2(t)$, and $v_3(t)$). Then, two procedures were implemented to recover the pulse signal:

- M1 Each $v_i(t)$ was filtered using a FIR band-pass filter implemented via Hamming window, with lower cut-off at 0.75 Hz (45 bpm) and upper cut-off at 2 Hz (120 bpm). Such a band is well suited for normal subjects at rest or in during moderate activity, bandwidth can be adapted during intense physical activity or in presence of disease. The signals $x_1(t)$, $x_2(t)$, and $x_3(t)$ were so obtained.
- M2 As illustrated in Fig. 3, each $v_i(t)$ was detrended by subtracting a 2 s time average and the signals $y_1(t)$, $y_2(t)$, and $y_3(t)$ were obtained. Following whitening, the three detrended signals were jointly processed through Independent Component Analysis as implemented in FastICA algorithm [41]; three new sequences $y_A(t)$, $y_B(t)$, and $y_C(t)$ were so obtained, each representing a different signal contribution. The spectra of $y_A(t)$, $y_B(t)$, and $y_C(t)$ were computed via Fast Fourier Transform. The BVP-related component was individuated as the one with the highest peak in the range from 0.75 Hz to 2.0 Hz and was then filtered as in M1 and $y_S(t)$ was so obtained.

In both procedures, after filtering, BVP peaks were detected using a multi-scale algorithm [42] and local quadratic interpolation. Finally, the tachogram for each BVP peak sequence was computed. As described for ECG analysis, possible artifacts were removed by the NC-VT algorithm. Similarly to the ECG signal, NN and SDNN values were extracted from the tachogram. Frequency domain analysis was performed by PSD estimation by Lomb-Scargle periodogram and the LF, HF, and LF/HF values were calculated from the PSD.

F. Recording Length

Videos with different duration were compared with corresponding ECG portions. In particular, pieces spanning one, two, three, four, and five minutes (entire recording) were extracted from the available sequence.

G. Video Frame Rate

Original videos were recorded at 133 fps. To test the effect of lower sampling rates as those often reported in literature, each video was down-sampled from the native 133 fps to 66 fps and 33 fps. ROIs signals computed from down-sampled videos were interpolated with a cubic spline function and resampled at 133 Hz.

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

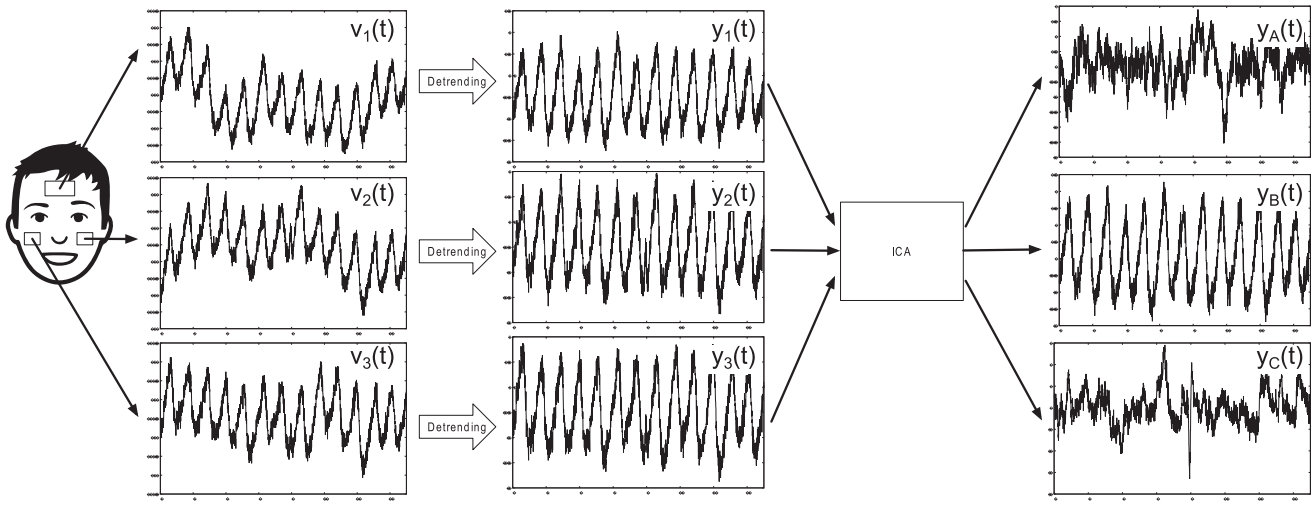

Fig. 3. Procedure based on ICA.

IV. Results

Data analysis was based on absolute errors observed by comparing iPPG with ECG. HR and HRV descriptors estimated from video (both by M1 using band-pass filter only and M2 based on ICA pre-processing) were compared to corresponding parameters obtained from ECG. Mean, standard deviation, minimum, and maximum values of the data obtained by ECG and video signal are reported in next sections for the various descriptors. For each of them, we give the mean absolute error (MAE), with standard deviation, minimum, and maximum. The median of signed errors (M_e) is also reported as a robust indicator of systematic error. In addition, as correlation coefficients are commonly used by many researchers, Pearson coefficient r_p and Spearman's rho ρ_s are also reported for completeness.

A. Analyzing Video at the Native Frame Rate

Mean, standard deviation, minimum, and maximum values of the data obtained by ECG and video signal analysis (both obtained from M1 and M2) are reported in Table I. The observed MAEs and correlations obtained from each ROI (ROI1, ROI2, and ROI3) are reported in Table II. With respect to NN values, both video signals and ICA perform well against ECG showing low values of MAEs, with the smallest ones provided by $x_1(t)$ (3.812 ms) and $y_S(t)$ (3.822 ms), and strong correlations ($x_1(t)$, $x_2(t)$, $x_3(t)$, and $y_S(t)$) with $r_p \geq 0.999$ and $\rho_s \geq 0.995$. Similarly, for SDNN, LF, HF, and LF/HF, the smallest MAEs values and strongest correlations are observed in $x_1(t)$ and $y_S(t)$; instead, the video signals from the cheeks ROIs ($x_2(t)$ and $x_3(t)$) are more prone to artifacts, so that the extracted parameters exhibits a weaker correlation with ECG data than signal from the forehead ROI. For $x_1(t)$ the MAEs observed values are 5.643 ms for SDNN, 3.303% for LF, 6.556% for HF, and 0.657 for LF/HF; for $y_S(t)$ we have 5.853 ms for SDNN, 2.775% for LF, 4.372% for HF, and 0.482 for LF/HF. For $x_1(t)$ the observed correlations are $r_p = 0.960$ ($\rho_s = 0.903$) for SDNN, $r_p = 0.878$ ($\rho_s = 0.813$) for LF, $r_p = 0.786$ ($\rho_s = 0.779$) for HF, and $r_p = 0.842$ ($\rho_s = 0.779$) for LF/HF. For $y_S(t)$ the observed correlations are $r_p = 0.961$ ($\rho_s = 0.928$) for SDNN, $r_p = 0.916$ ($\rho_s = 0.895$) for LF, $r_p = 0.885$ ($\rho_s = 0.884$) for HF, and $r_p = 0.931$ ($\rho_s = 0.876$) for LF/HF. All the parameters obtained from $y_S(t)$ (M2) were strongly correlated with the reference ones, the correlation being larger than those observed for M1 in all cases. Scatter plots of parameters computed from $x_1(t)$ and $y_S(t)$ and the corresponding ECG ones are summarized in Fig. 4 (time domain) and Fig. 5 (frequency domain).

TABLE I
Mean, Standard Deviation, Minimum, and Maximum Values of the Data Obtained by ECG and Video Signal Analysis at the Native Frame Rate

Parameter	Metrics	ECG	$x_1(t)$ ROI1	$x_2(t)$ ROI2	$x_3(t)$ ROI3	$y_S(t)$ ICA
NN (ms)	Mean	891.281	894.241	890.732	893.661	894.521
	SD	141.412	142.372	143.170	142.012	142.575
	Min	666.470	664.189	648.958	669.968	664.090
	Max	1127.100	1133.250	1133.390	1127.130	1132.080
SDNN (ms)	Mean	55.663	58.885	71.285	69.498	59.667
	SD	20.302	19.902	23.646	19.637	19.684
	Min	26.189	32.928	41.714	37.094	30.944
	Max	113.834	122.372	123.666	124.734	117.419
LF (%)	Mean	29.153	29.112	30.142	30.039	29.404
	SD	9.035	8.901	8.309	8.008	8.730
	Min	11.780	13.894	12.267	12.626	14.068
	Max	43.315	41.532	48.140	45.490	41.574
HF (%)	Mean	25.404	26.911	33.091	32.916	27.692
	SD	12.742	12.899	12.711	12.755	12.286
	Min	4.042	9.182	13.560	11.980	8.447
	Max	56.621	61.521	67.840	68.999	58.625
LF/HF	Mean	1.752	1.409	1.099	1.114	1.377
	SD	1.814	0.930	0.609	0.672	0.970
	Min	0.275	0.235	0.181	0.183	0.259
	Max	9.767	4.427	2.476	3.349	4.910

B. Varying the Frame Rates

In Table III and Table IV we analyzed the parameters estimated from down-sampled video signals. In particular, we analyzed data obtained by M1 from the forehead signal $x_1(t)$ (the best performing at 133 fps) and those obtained from

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

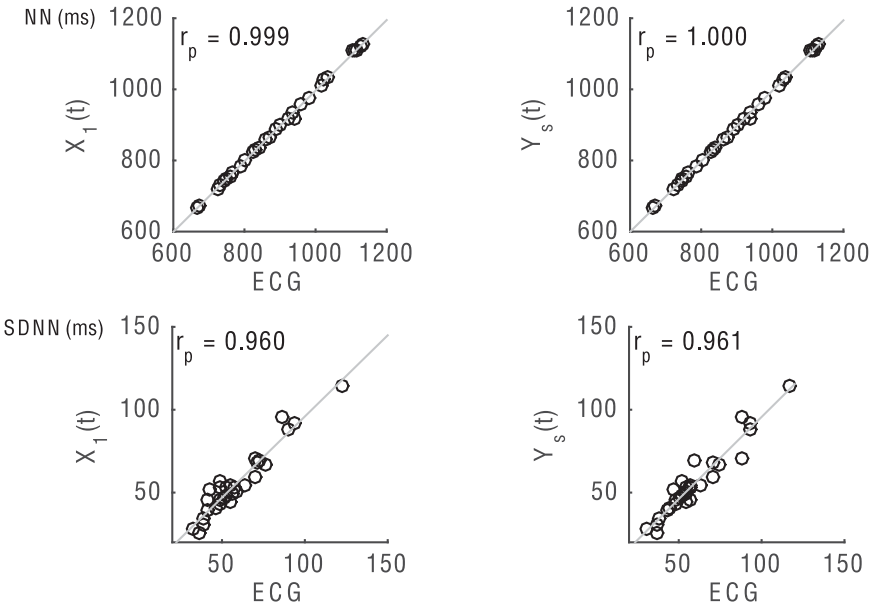

Fig. 4. Scatter plots comparing NN and SDNN values between the video signal and ECG signal. The video signal data are obtained from $x_1(t)$ applying M1, on the left side, and from $y_S(t)$ applying M2, on the right side.

TABLE II
ABSOLUTE ERRORS (MEAN, STANDARD DEVIATION, MINIMUM, AND MAXIMUM), MEDIAN OF SIGNED ERRORS (M_e), PEARSON COEFFICIENTS, AND SPEARMAN'S rho OBTAINED BY USING M1 AND M2, RESPECTIVELY

Parameter	Metrics	$x_1(t)$ ROI1	$x_2(t)$ ROI2	$x_3(t)$ ROI3	$y_S(t)$ ICA
NN	MAE (ms)	3.812	6.124	4.309	3.822
	SD_{AE} (ms)	4.174	6.253	4.441	4.076
	AE_{min} (ms)	0.332	0.075	0.114	1.210
	AE_{max} (ms)	24.311	24.217	23.912	24.424
	M_e (ms)	2.193	1.512	2.156	3.157
	r_p	0.999	0.998	0.999	1.000
	ρ_s	0.998	0.995	0.996	0.998
SDNN	MAE (ms)	5.643	16.653	14.750	5.853
	SD_{AE} (ms)	3.197	11.851	11.722	4.076
	AE_{min} (ms)	0.028	0.251	1.593	1.500
	AE_{max} (ms)	10.604	53.427	47.773	17.355
	M_e (ms)	4.202	15.164	12.231	3.867
	r_p	0.960	0.830	0.794	0.961
	ρ_s	0.903	0.757	0.723	0.928
LF	MAE (%)	3.303	4.666	5.459	2.775
	SD_{AE} (%)	2.810	4.417	5.421	2.341
	AE_{min} (%)	0.068	0.181	0.110	0.262
	AE_{max} (%)	11.261	15.715	21.345	8.361
	M_e (%)	0.783	0.205	0.579	0.758
	r_p	0.878	0.730	0.596	0.916
	ρ_s	0.813	0.685	0.637	0.895
HF	MAE (%)	6.556	11.716	9.605	4.372
	SD_{AE} (%)	5.324	8.460	8.978	4.684
	AE_{min} (%)	0.353	0.076	0.359	0.592
	AE_{max} (%)	20.238	36.965	33.823	18.450
	M_e (%)	3.324	8.046	5.270	1.436
	r_p	0.786	0.529	0.638	0.885
	ρ_s	0.779	0.519	0.643	0.884
LF/HF	MAE	0.657	0.925	0.806	0.482
	SD_{AE}	0.994	1.444	1.213	0.927
	AE_{min}	0.007	0.003	0.001	0.002
	AE_{max}	5.339	7.978	6.418	4.857
	M_e	-0.187	-0.356	-0.242	-0.046
	r_p	0.842	0.512	0.827	0.931
	ρ_s	0.779	0.487	0.584	0.876

$y_S(t)$ by M2. The MAE values related to NN extracted by the down-sampled video signals for both M1 and M2 do not exhibit appreciable differences (being all slightly larger than 3 ms) at different frame rates. High correlation coefficients are observed in all cases ($r_p \geq 0.999$ and $\rho_s \geq 0.998$). SDNN behavior shows increasing MAEs when frame rate decreases for both M1 and M2; in parallel, correlation coefficients decrease at the slowest frame rate. Also in frequency domain, MAE values increase as frame rate decrease, this being more pronounced for HF and LF/HF. Generally, MAE values and correlation coefficients obtained for M2 are better than those obtained for M1. Furthermore, in frequency domain, the contribution of the ICA is outstanding when the sampling rate is less than 100 fps.

C. Varying the Acquisition Time

We analyzed the parameters obtained from the video signals (both M1 and M2) having different durations. As in the case of subsection IV-B we considered $x_1(t)$ and $y_S(t)$ only. In Table V we reported the basic statistics for parameters measured from ECG, $x_1(t)$, and $y_S(t)$, respectively. In Table VI MAEs and correlation coefficients are given. Both M1 and M2 show similar error pattern for NN: MAE increases when the recording time decreases, difference being slightly more evident for durations less than or equal to 2 min. High correlation coefficients were observed at any duration for $x_1(t)$ ($r_p \geq 0.998$ and $\rho_s \geq 0.995$) and $y_S(t)$ ($r_p \geq 0.999$ and $\rho_s \geq 0.994$). The MAEs of remaining parameters tend to increase as the recording duration decreases, HF showing the more pronounced dependence. As shown in Table VI, dependence on acquisition time was slightly more pronounced for frequency domain descriptors. In any case, only for the shortest observation window (1 min) we observed a MAE approximately doubled with respect to other durations.

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

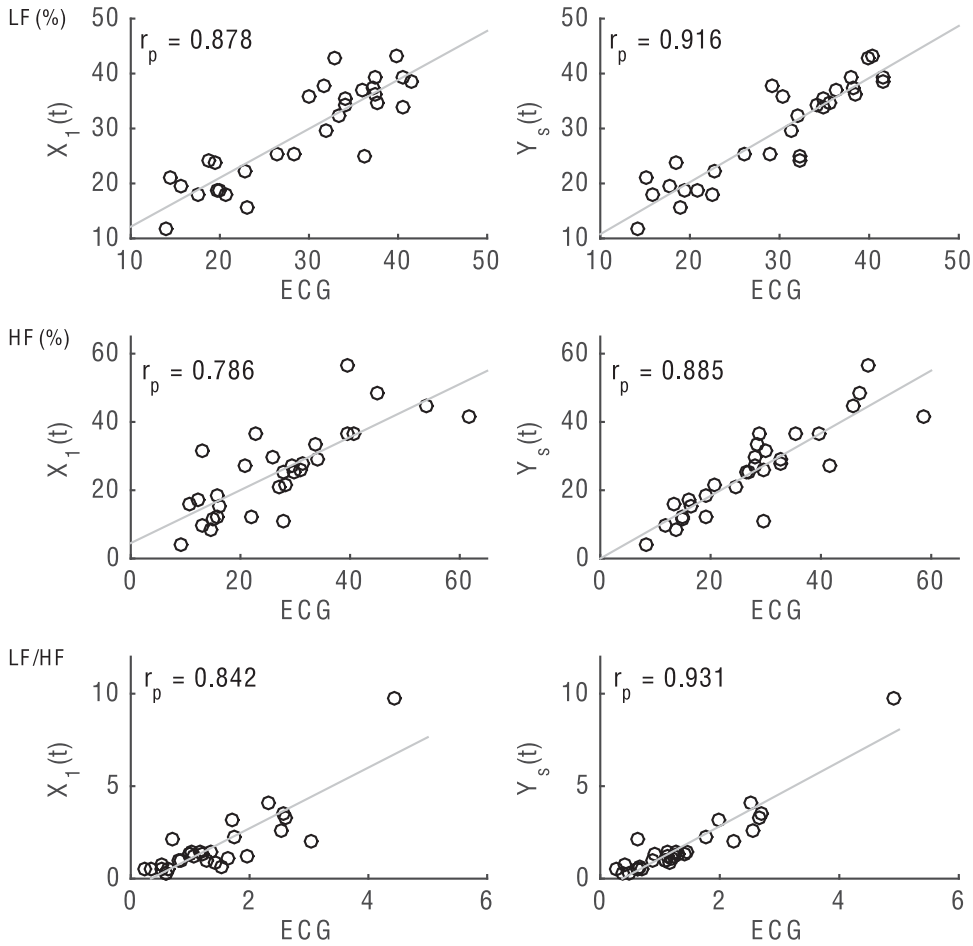

Fig. 5. Scatter plots comparing LF, HF, and LF/HF values between the video signal and ECG signal. The video signal data are obtained from $x_1(t)$ applying M1, on the left side, and from $y_S(t)$ applying M2, on the right side.

TABLE III
MEAN, STANDARD DEVIATON, MINIMUM, AND MAXIMUM VALUES OF THE DATA OBTAINED BY VIDEO SIGNAL ANALYSIS VARYING THE FRAME RATES

Parameter	Metrics	$x_1(t)$ (ROI1) video rate (fps)			$y_S(t)$ (ICA) video rate (fps)		
		133	66	33	133	66	33
NN (ms)	Mean	894.241	892.987	892.381	894.521	893.122	893.041
	SD	142.372	142.134	142.009	142.575	142.130	142.373
	Min	664.189	662.083	661.517	664.090	661.946	661.843
	Max	1133.250	1131.990	1131.960	1132.080	1130.350	1130.680
SDNN (ms)	Mean	58.885	59.891	62.022	59.667	61.780	63.273
	SD	19.902	19.867	20.663	19.684	20.638	20.001
	Min	32.928	32.952	34.147	30.944	31.485	32.262
	Max	122.372	122.391	123.756	117.419	117.636	116.960
LF (%)	Mean	29.112	29.111	29.394	29.404	29.448	28.875
	SD	8.901	8.637	8.245	8.730	8.202	8.154
	Min	13.894	14.090	14.271	14.068	14.897	14.440
	Max	41.532	42.007	41.195	41.574	41.300	41.185
HF (%)	Mean	26.911	27.280	28.275	27.692	28.368	29.335
	SD	12.899	13.142	11.962	12.286	11.977	12.055
	Min	9.182	9.930	11.363	8.447	8.948	10.637
	Max	61.521	60.534	60.194	58.625	57.935	57.330
LF/HF	Mean	1.409	1.386	1.302	1.377	1.321	1.240
	SD	0.930	0.909	0.824	0.970	0.898	0.807
	Min	0.235	0.240	0.245	0.259	0.261	0.268
	Max	4.427	4.025	3.528	4.910	4.616	3.872

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

TABLE IV

ABSOLUTE ERRORS (MEAN, STANDARD DEVIATION, MINIMUM, AND MAXIMUM), MEDIAN OF SIGNED ERRORS (M_e), PEARSON COEFFICIENTS, AND SPEARMAN'S rho OBTAINED AT THREE DIFFERENT VIDEO RATES FOR $x_1(t)$ AND $y_S(t)$ USING M1 AND M2, RESPECTIVELY

Parameter	Metrics	$x_1(t)$ (ROI1) video rate (fps) 133	66	33	$y_S(t)$ (ICA) video rate (fps) 133	66	33
NN	MAE (ms)	3.812	3.337	3.526	3.822	3.136	3.251
	SD_{AE} (ms)	4.104	3.993	4.263	4.008	3.840	3.929
	AE_{min} (ms)	0.332	0.284	0.248	1.210	0.637	0.057
	AE_{max} (ms)	24.311	22.713	22.643	24.424	22.872	22.512
	M_e (ms)	2.193	1.991	1.812	3.157	2.227	1.868
	r_p	0.999	0.999	0.999	1.000	0.999	0.999
	ρ_s	0.998	0.998	0.998	0.998	0.998	0.998
SDNN	MAE (ms)	5.643	6.518	8.668	5.835	7.118	8.497
	SD_{AE} (ms)	3.143	3.212	4.8273	3.502	3.810	4.732
	AE_{min} (ms)	0.028	0.300	1.210	1.500	3.036	1.939
	AE_{max} (ms)	10.604	12.596	23.027	17.355	19.968	20.645
	M_e (ms)	4.202	5.500	8.137	3.867	6.137	7.786
	r_p	0.960	0.955	0.929	0.961	0.966	0.953
	ρ_s	0.903	0.887	0.874	0.928	0.937	0.911
LF	MAE (%)	3.303	3.179	3.365	2.775	3.311	3.094
	SD_{AE} (%)	2.850	3.278	4.827	2.302	2.293	4.732
	AE_{min} (%)	0.068	0.015	0.002	0.262	0.190	0.077
	AE_{max} (%)	11.261	12.971	13.862	8.361	7.996	9.861
	M_e (%)	0.783	0.629	0.525	0.758	0.774	0.534
	r_p	0.878	0.863	0.855	0.916	0.892	0.889
	ρ_s	0.813	0.827	0.828	0.895	0.872	0.863
HF	MAE (%)	6.556	6.580	7.083	4.372	4.851	5.571
	SD_{AE} (%)	5.234	5.571	5.423	4.606	4.906	4.898
	AE_{min} (%)	0.353	0.454	1.074	0.592	0.093	0.092
	AE_{max} (%)	20.238	19.251	21.200	18.450	21.294	22.976
	M_e (%)	3.324	3.157	4.304	1.436	2.937	3.944
	r_p	0.786	0.782	0.760	0.885	0.870	0.868
	ρ_s	0.779	0.766	0.774	0.884	0.871	0.850
LF/HF	MAE	0.657	0.657	0.708	0.482	0.539	0.620
	SD_{AE}	0.978	1.037	1.126	0.912	0.975	1.084
	AE_{min}	0.007	0.074	0.049	0.002	0.004	0.003
	AE_{max}	5.339	5.742	6.239	4.857	5.151	5.895
	M_e	-0.187	-0.196	-0.229	-0.047	-0.086	-0.197
	r_p	0.842	0.817	0.786	0.931	0.922	0.887
	ρ_s	0.782	0.739	0.739	0.876	0.844	0.828

TABLE V

MEAN, STANDARD DEVIATON, MINIMUM, AND MAXIMUM VALUES OF THE DATA OBTAINED BY ECG AND VIDEO SIGNAL ANALYSIS VARYING THE ACQUISITION TIME FROM M1

Parameter	Metrics	5 min ECG	5 min $x_1(t)$ ROI1	5 min $y_S(t)$ ICA	4 min ECG	4 min $x_1(t)$ ROI1	4 min $y_S(t)$ ICA	3 min ECG	3 min $x_1(t)$ ROI1	3 min $y_S(t)$ ICA	2 min ECG	2 min $x_1(t)$ ROI1	2 min $y_S(t)$ ICA	1 min ECG	1 min $x_1(t)$ ROI1	1 min $y_S(t)$ ICA
NN (ms)	Mean	891.281	894.241	894.521	893.874	896.17	896.454	899.108	901.651	902.224	903.163	905.632	906.013	912.381	913.224	913.700
	SD	141.412	142.372	142.575	144.211	144.084	144.471	144.510	144.735	145.262	148.181	148.043	148.426	148.816	149.370	149.751
	Min	666.470	664.189	664.090	664.637	661.864	661.906	664.689	660.676	660.504	655.403	648.219	647.808	669.886	655.705	656.141
	Max	1127.100	1133.250	1132.080	1141.610	1139.580	1138.990	1149.250	1147.850	1147.000	1175.780	1167.650	1168.820	1183.200	1176.390	1176.090
SDNN (ms)	Mean	55.663	58.885	59.667	55.346	58.700	59.277	54.915	58.361	57.903	53.622	56.799	55.790	51.433	55.447	53.842
	SD	20.302	19.902	19.684	21.705	20.560	20.281	23.200	21.709	21.244	20.471	19.372	19.067	18.416	19.115	17.760
	Min	26.189	32.928	30.944	23.985	33.88	31.365	24.901	34.952	31.948	25.845	30.653	28.958	24.750	30.757	29.153
	Max	113.834	122.372	117.419	118.823	127.048	123.851	126.919	133.067	131.597	99.561	107.222	105.608	99.203	103.733	99.853
LF (%)	Mean	29.153	29.112	29.404	26.318	28.559	28.627	27.524	30.209	30.190	28.801	31.144	31.240	30.809	35.339	35.576
	SD	9.035	8.901	8.730	8.703	8.632	8.935	10.566	10.846	11.182	10.809	10.413	10.210	13.886	15.492	15.552
	Min	11.78	13.894	14.068	11.560	14.021	13.932	12.515	13.954	14.527	11.702	13.170	13.739	14.347	12.143	14.838
	Max	43.315	41.532	41.574	42.623	46.050	45.967	49.951	58.773	58.833	55.247	57.848	57.797	64.490	77.275	75.240
HF (%)	Mean	25.404	26.911	27.692	32.142	27.467	27.943	34.318	28.936	29.006	38.412	33.405	32.786	43.135	35.284	34.604
	SD	12.742	12.899	12.286	12.499	12.406	11.951	11.968	12.411	11.255	16.414	16.736	15.651	16.441	16.475	15.041
	Min	4.042	9.182	8.447	5.540	9.380	8.602	6.055	9.362	9.125	3.708	6.496	7.173	5.617	5.769	7.162
	Max	56.621	61.521	58.625	59.097	56.189	52.518	54.561	60.605	51.887	75.813	71.767	70.853	76.578	73.960	72.378
LF/HF	Mean	1.752	1.409	1.377	1.124	1.333	1.323	1.055	1.356	1.311	1.150	1.411	1.304	1.001	1.620	1.505
	SD	1.814	0.930	0.970	1.189	0.866	0.941	0.979	1.008	0.891	1.368	1.273	0.990	1.147	2.278	1.805
	Min	0.275	0.235	0.259	0.229	0.263	0.297	0.267	0.262	0.300	0.178	0.282	0.301	0.214	0.179	0.239
	Max	9.767	4.427	4.910	6.683	4.145	4.720	5.167	4.742	3.645	7.094	5.382	4.439	6.256	12.296	9.779

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
ABSOLUTE ERRORS (MEAN, STANDARD DEVIATION, MINIMUM, AND MAXIMUM), MEDIAN OF SIGNED ERRORS (M_e), PEARSON COEFFICIENTS, AND SPEARMAN'S rho FOR FIVE DIFFERENT DURATIONS OF VIDEOS FOR $x_1(t)$ AND $y_S(t)$ USING M1 AND M2, RESPECTIVELY

Parameter	Metrics	5 min		4 min		3 min		2 min		1 min	
		$x_1(t)$ ROI1	$y_S(t)$ ICA	$x_1(t)$ ROI1	$y_S(t)$ ICA	$x_1(t)$ ROI1	$y_S(t)$ ICA	$x_1(t)$ ROI1	$y_S(t)$ ICA	$x_1(t)$ ROI1	$y_S(t)$ ICA
NN	MAE (ms)	3.812	3.822	3.742	3.656	3.872	4.204	4.689	4.936	5.737	5.543
	SD_{AE} (ms)	4.174	4.076	4.254	4.130	4.551	4.520	5.178	4.708	6.012	5.559
	AE_{Min} (ms)	0.332	1.210	0.028	0.190	0.063	0.535	0.000	0.221	0.150	0.150
	AE_{Max} (ms)	24.311	24.424	24.467	24.496	26.318	26.857	27.621	27.317	29.123	28.371
	M_e (ms)	2.193	3.157	2.173	2.405	2.525	2.947	2.459	3.582	2.174	2.130
	r_p	0.999	1.000	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.999
	ρ_s	0.998	0.998	0.998	0.998	0.999	1.000	0.997	0.996	0.995	0.994
SDNN	MAE (ms)	5.643	5.853	5.951	5.331	6.563	5.855	6.568	6.466	7.324	6.898
	SD_{AE} (ms)	3.197	3.562	3.204	3.057	3.429	3.746	4.478	5.005	4.751	5.191
	AE_{Min} (ms)	0.028	1.499	0.848	0.083	0.340	0.329	0.864	0.061	0.026	0.360
	AE_{Max} (ms)	10.604	17.355	12.313	12.301	12.866	13.929	16.380	19.619	17.744	21.812
	M_e (ms)	4.202	3.867	4.526	4.265	4.553	4.122	3.812	3.634	4.280	3.810
	r_p	0.960	0.961	0.962	0.976	0.958	0.963	0.933	0.921	0.913	0.893
	ρ_s	0.903	0.928	0.884	0.956	0.895	0.944	0.920	0.911	0.883	0.860
LF	MAE (%)	3.303	2.775	3.577	3.516	4.185	3.915	3.983	4.005	6.204	7.059
	SD_{AE} (%)	2.898	2.341	2.970	2.727	3.491	3.056	3.305	2.863	6.183	6.027
	AE_{Min} (%)	0.068	0.262	0.290	0.024	0.032	0.005	0.037	0.236	0.097	0.381
	AE_{Max} (%)	11.261	8.361	14.605	10.131	13.424	13.275	11.414	13.922	20.921	22.453
	M_e (%)	0.783	0.758	2.397	2.529	2.537	2.626	1.893	2.600	2.302	3.540
	r_p	0.878	0.916	0.888	0.906	0.901	0.926	0.904	0.917	0.874	0.857
	ρ_s	0.813	0.895	0.883	0.890	0.887	0.939	0.864	0.873	0.833	0.730
HF	MAE (%)	6.557	4.372	6.172	5.398	7.290	6.620	8.000	8.351	11.159	11.359
	SD_{AE} (%)	5.324	4.684	6.054	4.027	6.312	4.876	7.971	6.399	9.273	8.832
	AE_{Min} (%)	0.353	0.592	0.180	0.825	0.398	0.012	0.099	0.243	0.480	0.431
	AE_{Max} (%)	20.238	18.450	23.394	16.177	23.250	17.098	33.554	28.548	30.003	34.608
	M_e (%)	3.324	1.436	-3.002	-4.042	-3.046	-4.660	-2.459	-6.187	-4.262	-8.664
	r_p	0.786	0.885	0.828	0.907	0.782	0.854	0.811	0.845	0.721	0.729
	ρ_s	0.779	0.884	0.810	0.889	0.758	0.814	0.690	0.713	0.646	0.705
LF/HF	MAE	0.657	0.482	0.416	0.361	0.481	0.409	0.543	0.428	0.832	0.821
	SD_{AE}	0.994	0.927	0.543	0.413	0.634	0.446	0.874	0.547	1.871	1.484
	AE_{Min}	0.007	0.002	0.008	0.003	0.018	0.020	0.003	0.016	0.004	0.019
	AE_{Max}	5.340	4.857	2.537	1.963	3.160	1.576	3.541	2.656	9.832	7.315
	M_e	-0.187	-0.046	0.130	0.208	0.173	0.183	0.112	0.175	0.157	0.249
	r_p	0.842	0.931	0.843	0.909	0.723	0.830	0.716	0.881	0.513	0.468
	ρ_s	0.779	0.876	0.826	0.915	0.863	0.902	0.833	0.879	0.752	0.765

V. DISCUSSION

This study used a monochrome video camera to estimate the HR and HRV from the plethysmographic signal obtained from human face recordings. Videos from 30 subjects were acquired in rest conditions allowing limited natural movements. An artificial white light was used and the camera input was band-pass filtered so as to match the light absorption peaks of hemoglobin and enhance the BVP contribution to the image gray level. We adopted direct BVP estimation via temporal band-pass filtering (method M1) and BVP peak detection applied to single ROI signals. In addition, ICA pre-processing to extract BVP signal from multiple ROIs (method M2) was implemented. Standard HR and HRV measurements derived from videos were compared with the corresponding descriptors obtained from ECG signals recorded simultaneously to videos.

We analyzed both mean absolute errors between video and ECG in three different regions of the face (forehead and each cheek). For completeness, Pearson and Spearman correlation coefficients were computed. In all cases, the ROI on the forehead provided smallest errors and highest correlations. This seems consistent with a reduced impact of subject motion. Fusion of the three ROIs signals via ICA pre-processing led to substantially reduced errors and improved correlation in all frequency domain descriptors, whereas time domain descriptors exhibit a behavior similar to the ones from ROI on the forehead.

At the original video rate of 133 fps method M1 produced acceptable results for the forehead ROI (MAEs were 3.812 ms for NN, 5.643 ms for SDNN, 3.303% for LF, 6.556% for HF, and 0.657 for LF/HF).

Anyway, M2 provided data more consistent with ECG estimation as made evident by frequency domain parameters (MAEs were 3.822 ms for NN, 5.853 ms for SDNN, 2.775% for LF, 4.372% for HF, and 0.482 for LF/HF).

When the video signal is down-sampled from the native 133 fps to 66 fps and 33 fps estimates of HRV descriptors worsen, whereas HR does not look significantly affected. This finding is inline with the expected need of an accurate detection of BVP pulses for HRV analysis, a requirement that can be significantly relaxed for HR assessment. In particular, the LF/HF parameter seems to be the most affected. It gets worse with the decrease of the sampling rate. In frequency domain, the

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

contribution of the ICA pre-processing is significant at low sampling rates.

As to the impact of video duration, HR, as expected, does not exhibit a particular sensitivity and both M1 and M2 produce constant and similar results. In the case of HRV, method M1 looks less robust than method M2 and discrepancies arises between iPPG and ECG for very short recordings (1 min). In general, MAEs increase when sampling rate and acquisition time decrease.

To summarize, both M1 and M2 can be expected to allow a very confident estimation of the HR. As concerns HRV descriptors, ICA pre-processing enables a more reliable estimation in all conditions.

VI. Conclusion

The proposed ICA-based integration of multiple simultaneously-acquired BVP signals seems to permit a reliable measurement of HR and HRV. Good results have been obtained from signals recorded with varying frame rates and temporal durations. The contribution of ICA is quite significant especially at lower frame rates. In this study the subjects were sitting still in front of the camera at a distance of about one meter and the chair has a headrest to contain head motion and making the video recording comfortable for the volunteer. Our next goal is to apply our method to monitor diseased people as well as healthy subjects under stimulation of the autonomic nervous system. To this end we are currently developing a robust tracking procedure so as to build a system able to measure vital signs from video in normal environments of daily life and during activities such as driving, playing a video game, or brushing teeth.

References

[1] M. T. Cooney, E. Vartiainen, T. Laakitainen, A. Juolevi, A. Dudina, and I. M. Graham, "Elevated resting heart rate is an independent risk factor for cardiovascular disease in healthy men and women," *Amer. Heart J.*, vol. 159, pp. 612–619.e3, 2010.

[2] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, 2006.

[3] E. Watanabe, K. Kiyono, Y. Yamamoto, and J. Hayano, *Heart Rate Variability and Cardiac Diseases*. Tokyo, Japan: Springer, 2017, pp. 163–178.

[4] P. C. Dinas, Y. Koutedakis, and A. D. Flouris, "Effects of active and passive tobacco cigarette smoking on heart rate variability," *Int. J. Cardiology*, vol. 163, pp. 109–115, 2013.

[5] B. S. Tegegne, T. Man, A. M. van Roon, H. Riese, and H. Snieder, "Determinants of heart rate variability in the general population: The lifelines cohort study," *Heart Rhythm*, vol. 15, no. 10, pp. 1552–1558, 2018.

[6] L. Capdevila, J. Moreno, J. Movellan, E. Parrado, and J. Ramos-Castro, "HRV based health&sport markers using video from the face," in *Proc. 34th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 5646–5649.

[7] B. G. Lee, S. J. Jung, and W. Y. Chung, "Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection," *IET Commun.*, vol. 5, no. 17, pp. 2461–2469, Nov. 2011.

[8] D. S. Quintana and J. A. J. Heathers, "Considerations in the assessment of heart rate variability in biobehavioral research," *Frontiers Psychol.*, vol. 5, pp. 1–10, 2014.

[9] I. R. Tayibnapis, D. Y. Koo, M. K. Choi, and S. Kwon, "A novel driver fatigue monitoring using optical imaging of face on safe driving system," *Proc. Int. Conf. Control, Electron., Renewable Energy Commun.*, 2016, pp. 115–120.

[10] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans. Affect. Comput.*, vol. 8, no. 1, pp. 15–28, Jan.–Mar. 2017.

[11] A. Mesleh, D. Skopin, S. Baglikov, and A. Quteishat, "Heart rate extraction from vowel speech signals," *J. Comput. Sci. Technol.*, vol. 27, no. 6, pp. 1243–1251, 2012.

[12] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, pp. 1418–1426, Aug. 2007.

[13] O. Boric-Lubecke, V. M. Lubecke, I. Mostafanezhad, B. K. Park, W. Massagram, and B. Jokanovic, "Doppler radar architectures and signal processing for heart rate extraction," *Mikrotalasna Revija*, vol. 15, no. 2, pp. 12–17, 2009.

[14] D. Obeid, S. Sadek, G. Zaharia, and G. El Zein, "Multitunable microwave system for touchless heartbeat detection and heart rate variability extraction," *Microw. Opt. Technol. Lett.*, vol. 52, no. 1, pp. 192–198, 2010.

[15] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Med. Eng. Phys.*, vol. 29, no. 8, pp. 853–857, 2007.

[16] M. Z. Poh, D. J. Mcduff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.

[17] F. Bousefsaf, C. Maaoui, and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 568–574, 2013.

[18] T. Pursche, J. Krajewski, and R. Moeller, "Video-based heart rate measurement from human faces," *Proc. IEEE Int. Conf. Consum. Electron.*, 2012, pp. 544–545.

[19] C. Brüser, C. H. Antink, T. Wartzek, M. Walter, and S. Leonhardt, "Ambient and unobtrusive cardiorespiratory monitoring techniques," *IEEE Rev. Med. Eng.*, vol. 8, pp. 30–43, 2015.

[20] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 2168–2194, Jul. 2014.

[21] L. Wei, Y. Tian, Y. Wang, and T. Ebrahimi, "Automatic webcam-based human heart rate measurements using Laplacian eigenmap," in *Proc. Asian Conf. Comput. Vision*, 2013, pp. 281–292.

[22] L. Feng, L. M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, May 2015.

[23] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam—A non-contact method for evaluating cardiac activity," in *Proc. Federated Conf. Comput. Sci.*, 2011, pp. 405–410.

[24] W. Wang, S. Stuijk, and G. De Haan, "Exploiting spatial redundancy of image sensor for motion robust rPPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 415–425, Feb. 2015.

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

[26] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological Meas.*, vol. 35, no. 5, pp. 807–31, 2014.

[27] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, pp. 2878–2886, Oct. 2013.

[28] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote measurements of heart and respiration rates for telemedicine," *PloS ONE*, vol. 8, 2013, Art. no. e71384.

[29] W. Cui, L. E. Ostrander, and B. Y. Lee, "In vivo reflectance of blood and tissue as a function of light wavelength," *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 632–639, Jun. 1990.

[30] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2013, pp. 3430–3437.

[31] Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.

[32] L. Iozzia, L. Cerina, and L. Mainardi, "Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA," *Physiological Meas.*, vol. 37, pp. 1934–1944, 2016.

[33] W.-H. Lin, D. Wu, C. Li, H. Zhang, and Y.-T. Zhang, "Comparison of heart rate variability from PPG with that from ECG," *IFMBE Proc. Int. Conf. Health Inform.*, 2014, pp. 213–215.

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.

[34] N. Pinheiro, R. Couceiro, J. Henriques, J. Muehlsteff, I. Quintal, L. Goncalves, and P. Carvalho, "Can PPG be used for HRV analysis?," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 2945–2949.

[35] SEMEOTICONS Project. https://www.semeoticons.eu. [Online]. Available: https://www.semeoticons.eu

[36] W. Wang, A. C. Den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote-PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 99, pp. 1479–1491, Jul. 2017.

[37] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official J. Eur. Union*, vol. L119, pp. 1–88, 2016. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[38] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, Mar. 1985.

[39] J. Vila, F. Palacios, J. Presedo, M. Fernandez-Delgado, P. Felix, and S. Barro, "Time-frequency analysis of heart-rate variability," *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 5, pp. 119–126, Sep./Oct. 1997.

[40] G. D. Clifford and L. Tarassenko, "Quantifying errors in spectral estimates of HRV due to beat replacement and resampling," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 630–638, Apr. 2005.

[41] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, no. 3, pp. 626–34, May 1999.

[42] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Boston, MA, USA: Springer, 1994.

Authorized licensed use limited to: Tsinghua University. Downloaded on February 26,2021 at 09:49:10 UTC from IEEE Xplore. Restrictions apply.