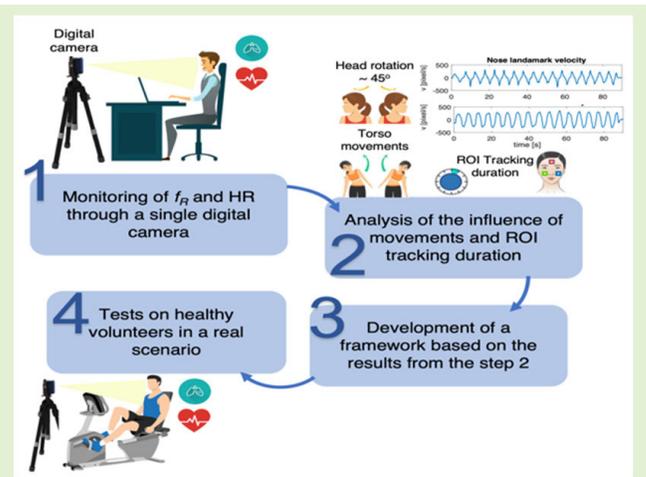


Cardiorespiratory Parameters Monitoring Through a Single Digital Camera in Real Scenarios: ROI Tracking and Motion Influence

Nunzia Molinaro¹, Student Member, IEEE, Federico Zangarelli, Emiliano Schena¹, Senior Member, IEEE, Sergio Silvestri¹, Senior Member, IEEE, and Carlo Massaroni¹, Senior Member, IEEE

Abstract—Monitoring of heart rate (HR) and respiratory rate (f_R) is fundamental to assessing the health status of an individual. To address this scope, technologies that frame the upper body and the face regions of a subject without any physical contact can be used. Motion artifacts can affect the applicability of non-contact methods to the continuous monitoring of these parameters as well as the computational burden. This article focuses on a technique based on images captured with a single digital camera for the continuous estimation of HR and f_R . The main goal is to analyze how the velocity of facial movements and region of interest (ROI) tracking duration influence the performance of the method. Tests were performed on healthy volunteers during motionless trials (i.e., at rest and after exercise), head and torso movements, and during physical exercise. Results demonstrated that a continuous estimation of HR and f_R can be performed with acceptable errors under changing ROI tracking duration and velocity in motionless trials (mean absolute error (MAE) below 5 bpm and 3.42 breaths·min⁻¹ for HR and f_R , respectively), whereas during movements (mimicking head and torso movements, and during exercise) the error increases (MAE up to 5.42 bpm and up to 5.03 breaths·min⁻¹ for HR and f_R , respectively). The proposed investigation can provide a framework for the continuous estimation of HR and f_R during both static and dynamic activities by optimizing the ROI tracking duration under different velocities of facial movements.

Index Terms—Digital camera, heart rate (HR) monitoring, motion assessment, region of interest (ROI) tracking, respiratory rate (f_R) monitoring.



I. INTRODUCTION

VITAL signs monitoring, such as heart rate (HR) and respiratory rate (f_R), is fundamental to assessing the

Manuscript received 11 May 2023; accepted 4 June 2023. Date of publication 23 June 2023; date of current version 31 August 2023. This work was supported by the project titled “Sviluppo di sistemi di misura e modelli per la stima senza contatto di parametri fisiologici”, POR Lazio FSE 2014/202 (CUP code n. F87C21000190009), under Project 23562. The associate editor coordinating the review of this article and approving it for publication was Prof. Xiaofeng Yuan. (Corresponding author: Carlo Massaroni.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Approvals received from Università Campus Bio-Medico di Roma under Application No. ST-UCBM 14/22 OSS, and performed in line with the Declaration of Helsinki.

The authors are with the Unit of Measurements and Biomedical Instrumentation, Departmental Faculty of Engineering, Università Campus Bio-Medico di Roma, 00128 Rome, Italy (e-mail: n.molinaro@unicampus.it; f.zangarelli@alcampus.it; e.schena@unicampus.it; s.silvestri@unicampus.it; c.massaroni@unicampus.it).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSEN.2023.3287555>, provided by the authors.

Digital Object Identifier 10.1109/JSEN.2023.3287555

health status of an individual. These physiological parameters are included in the Early Warning Scores (EWS), which are indicators used to determine the severity of the patient's physiological deterioration in the hospital [1]. HR and f_R are the primary physiological signs to be monitored since their abnormal values can be an index of some important diseases (e.g., cardiac arrest and cardiovascular diseases) [2], [3]. Moreover, the influence of f_R and HR, also known as cardiorespiratory coupling (CRC) plays an important role in the assessment of sleep-related disorders [4].

Commonly, contact-based systems which require the direct contact of the sensor with the skin of the subject are employed to measure HR and f_R [5]. In this context, wearable technologies based on different kinds of sensors have been widely used [6], [7]. However, contact-based technologies present some drawbacks related to discomfort for the subjects, ease of loss of contact, skin irritation in patients with fragile skin, and lastly the need for dedicated instrumentation [8]. Thus, in recent years non-contact technologies (e.g., radar and depth cameras) have gained a lot of interest due to

their ease of use, no discomfort for the subject during the monitoring [9], [10], [11], [12], [13]. Among them, digital cameras (e.g., smartphone's built-in camera, laptop built-in camera) present several advantageous features such as costs, availability, portability, and easiness of use [14], [15], [16]. Commonly, RGB sensors can be employed to detect the remote photoplethysmographic (r-PPG) signal associated with volumetric changes of blood in facial capillaries [17]. Both HR and f_R can be estimated from the r-PPG signal, as the respiratory activity modulates the cardiac activity [18]. When the r-PPG technique is used to retrieve information about the pulsatile activity through a video of the subject's face, different factors must be considered, like lighting source [19], user-camera distance [20], the resolution of the device used to acquire the video [21], the region of interest (ROI) [22], [23], and motion artifacts [24], [25], [26].

Head rotations and whole-body movements must be detected and deleted to improve the accuracy of vitals monitoring and possibly to extend the applicability of video-based systems outside constrained environments (e.g., during sports activities and in real scenarios). In the last years, some studies proposed different techniques to detect movements and to make a robust contactless system, mainly for HR estimation [27], [28], [29], [30]. Only a few studies have investigated the influence of respiratory-unrelated movements on the estimation of f_R [31], [32]. To the best of our knowledge, most of the studies available in literature detect body movements and try to filter, compensate, or remove them by implementing complex signal processing techniques (e.g., adaptive filtering) or through deep learning methods [28], [33], [34]. Another approach is to use the motion information obtained from a body or face tracking to filter or compensate for motion [35]. For example, Guo et al. [36] used a near-infrared time-of-flight camera to detect cardiac and respiratory unrelated movements and compensate for the intensity variation caused by motion artifacts using depth information. Another study applies adaptive noise cancellation due to motion artifacts in combination with a modified HSI—hue, saturation, intensity—model to remove motion artifacts and reduce the effects of irregular intensity induced by head movements [33]. However, to the best of our knowledge, none of the studies available in the literature have attempted analyses to quantify the amplitude of body movements and thus compensate for them, except for the recent work of Wu et al. [30] in which a motion level has been computed to identify the magnitude of simulated movements (i.e., stationary, small, and full motion). Strictly related to the detection of motion artifacts is the tracking of ROIs identified to extract the pulsatile and respiratory patterns [25], [37]. However, an ROI tracking optimization analysis based on the investigation of how many seconds to conduct the ROI tracking has not been performed. Additionally, the reliable estimation of HR and f_R simultaneously is uncommon, mainly because only r-PPG signal does not provide a robust measure of f_R . Moreover, most of the studies focus on the average estimation of HR and f_R [38], [39], which can discourage the application of non-contact technologies when abrupt increase

or decrease in HR and f_R are expected (e.g., sports field) or when more data are needed for short-term recordings.

To tackle these drawbacks, in this article, we focus on the analysis of how facial movements and ROI tracking duration influence the performance of a method based on images captured with a single digital camera for the estimation of HR and f_R . We aim at providing a framework for the continuous estimation of these two parameters during both static and dynamic activities by optimizing ROI tracking and the velocity of facial movements. The contributions of our article are the following:

- 1) A ground-truth independent algorithm is proposed for the continuous estimation of both HR and f_R with an update time of 1 s, even in the case of body motion disturbances.
- 2) A quantitative analysis of cardiac and respiratory unrelated movements based on the velocity of facial movements.
- 3) An analysis of the ROI tracking duration to optimize it under different velocities of facial movements.

The article is structured as follows. Section II describes the submodules of the proposed framework. In Section III, the performed experiments on healthy volunteers during motionless trials (i.e., at rest and after exercise), head and torso movements, and during physical exercise to assess the performance of the proposed framework and the evaluation metrics are described. Section IV reports the obtained results, followed by the conclusion in Section V.

II. METHODS

The architecture of the proposed pipeline for the continuous estimation of f_R and HR under different motions consists of submodules including: 1) video pre-processing; 2) signal extraction; 3) signal analysis; and 4) estimation of f_R and HR. All the steps to estimate f_R and HR are described in Sections II-A–II-C.

A. Video Pre-Processing

A video of the face and the torso regions was recorded through a digital camera. The facial region was detected in the video with the pre-trained cascade classifier from the OpenCV library [40] in a Python environment. Sixty-eight facial landmarks were detected and tracked through the *dlib* detector, which maps the facial points on the subject's face [41].

In this work, three rectangular ROIs on the face—i.e., right cheek, forehead, and left cheek since these facial regions have good vascularization [42]—and one on the torso at the level of the jugular fossa were identified to extract the raw r-PPG and the respiratory signals, respectively (see Fig. 1). Specific landmarks automatically identified on the face were used to define geometric rules to construct the ROIs. The tracking of all the identified ROIs was carried out under changing ROI tracking duration to evaluate its influence on the performance of the proposed framework in the estimation of cardiorespiratory parameters and the computational burden during video processing in terms of CPU running time. Five

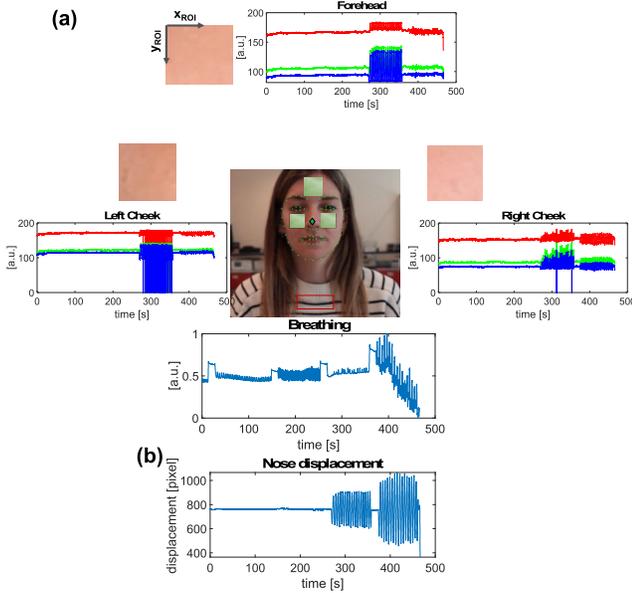


Fig. 1. Examples of (a) raw RGB signal extracted from three ROIs and respiratory signal extracted by OF from ROI torso and (b) raw displacement of nose's landmark along the x-axis.

modalities of video analysis were implemented: 1) each frame; 2) every 30 frames—1 s; 3) every 150 frames—5 s; 4) every 300 frames—10 s; and 5) every 600 frames—20 s. For f_R estimation an additional analysis was performed by using a fixed ROI identified in the first frame of the video [16], [43], [44]. The raw RGB signals extracted from the ROIs on the face were used to retrieve the pulsatile-related signal, whereas the respiratory signal was extracted through a method based on the computation of the optical flow (OF) [44]. Contextually, the displacement of the landmark on the nose was retrieved from the video and then the velocity of this landmark has been calculated. This value is then used to define the threshold on the head movements.

B. Signal Extraction

The raw r-PPG signals were obtained from each ROI on the face by spatially averaging the intensity of the pixels in the three-color channels R, G, and B, corresponding to the red, green, and blue channels according to the following:

$$s = \frac{1}{x_{ROI}} \sum_{x=1}^{x_{ROI}} \sum_{c=R,G,B} I(x, y, c) \quad (1)$$

where x_{ROI} represents the number of pixels in the selected ROI region along the y-axis (see Fig. 1), c is the color channel, and $I(x, y)$ is the intensity component of each channel.

The respiratory signal was extracted through a method based on the computation of OF, which allows the estimation of the displacement between two consecutive images by tracking image features on a pixel-by-pixel basis [44], using the Horn and Schunck (HS) algorithm [45]. A velocity vector for each pixel in the image was obtained and the component along the y-axis was assumed as the most related to movements of the ribcage caused by breathing [44].

C. Signal Analysis

The analysis of raw r-PPG and respiratory signals was carried out in a MATLAB environment through a windowing

analysis. A window length of 20 s was chosen as it represents a good compromise between resolution and noise robustness [16], [46], moving every 1 s. In the same analysis, the f_R and HR traces were obtained. Fig. 2 reports the flowchart illustrating the steps for the respiratory and r-PPG signals analysis when considering the threshold on the head movement, which is computed as the velocity of the nose's landmark.

1) *Respiratory Signal Analysis*: In each identified window, the raw respiratory signals computed with OF were first filtered with a Butterworth bandpass filter between 0.1 and 0.8 Hz (equivalent to the physiological range of f_R 6–48 breaths·min⁻¹). Then, an analysis in the frequency domain was carried out by computing the power spectrum density (PSD) of the signals. f_R values were estimated by considering the maximum frequency at which occurs the highest value of the PSD multiplying by 60, obtaining a value of f_R in bpm in each window. These steps were carried out per each modality of video analysis and when considering the threshold on the velocity.

2) *r-PPG Signal Analysis*: In each window, all the raw RGB signals were filtered through a Butterworth bandpass filter between 0.7 and 3 Hz (equivalent to the physiological range of HR 42–180 bpm). We implemented five well-established r-PPG algorithms that can be used when performing tests with digital cameras [47]: 1) green channel (GC); 2) independent component analysis (ICA); 3) principal component analysis (PCA); 4) chrominance-based signal processing method (CHROM); and 5) plane orthogonal to skin (POS) [48]. This analysis was carried out for all the raw r-PPG signals extracted from the three identified ROIs with ROI tracking in each frame of the video. After identifying the best algorithm to retrieve the r-PPG signal, this algorithm was implemented for HR estimation per each implemented modality of analysis (i.e., each frame, every 1, 5, 10, and 20 s) and when considering the threshold on the movement's velocity of the head, according to the flowchart in Fig. 2. HR estimation was performed through an analysis in the frequency domain by computing the PSD of the signals. After the normalization of the PSD against its maximum peak, HR was calculated by considering the average of the frequencies at which occur the peaks of the PSD above a threshold value of 0.8, weighted against the values of those peaks, according to the following:

$$HR = \frac{\sum_{i=1}^N f_i \cdot P_i}{N} \quad (2)$$

where N is the number of peaks in the PSD above 0.8, P_i is the value of the PSD in the range (0.8, 1], and f_i represents the value of the frequency at which occur P_i . Then, the estimated values were multiplied by 60 to obtain a value of HR in bpm. The HR values obtained per each identified ROI (i.e., left cheek, right cheek, and forehead) were numerically averaged. The mean value of HR was used for further statistical analysis [16].

3) *Motion Threshold*: To evaluate the performance of the proposed framework in the estimation of cardiorespiratory parameters under different head and torso movements, motion thresholds were identified by considering the velocity of the

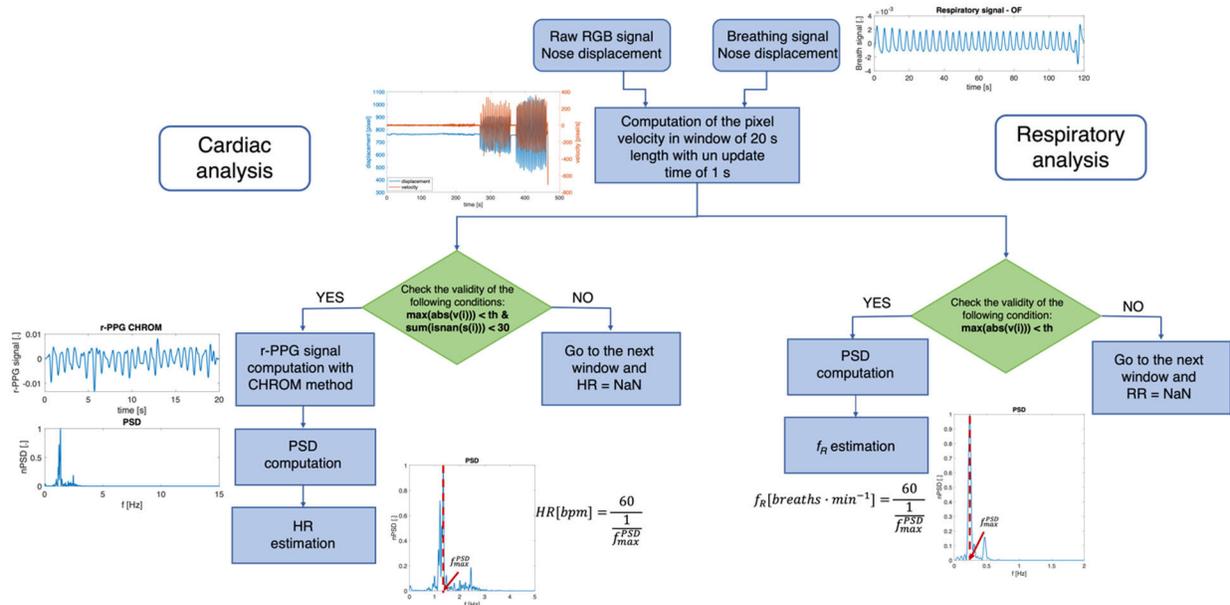


Fig. 2. Schematic of the flowchart illustrating the steps for respiratory and r-PPG signals when considering the threshold on the head movement. The inputs are the raw RGB signals, the breathing signal, and the displacement of the nose's landmark. The windowing analysis is carried out to extract the values of HR and f_R .

nose's landmark (see Section II-C). As reported in Fig. 2, the nose displacement was obtained in all the frames of the video and the velocity (v) was computed as $v = s/t$, where s is the nose displacement in pixels and t is the time. An analysis under gradually changing velocity thresholds from 5 to 400 pixels/s with a step of 5 pixel/s was carried out (the relative results are not reported). We considered just seven threshold velocities to summarize the results: 50, 60, 100, 150, 200, 300, and 400 pixels/s, equivalent to approximately 2, 2.4, 4, 6, 8, 12, and 16 cm/s, respectively. These values were obtained considering the region covered by the face in the whole image, which is about 10% in all the videos recorded. The maximum velocity value was calculated in each window and compared to the threshold to execute the next steps as shortly reported in Fig. 2.

III. EXPERIMENTAL TRIALS

To test the performances of the proposed framework in the continuous estimation of f_R and HR simulating real conditions with motion artifacts (e.g., head rotations and torso movements) and during physical activity, we carried out experiments on healthy volunteers. The study was carried out according to the Declaration of Helsinki, and in compliance with the Ethical Approvals received from Università Campus Bio-Medico di Roma (ST-UCBM 14/22 OSS).

A. Dataset in Laboratory Environment

This article aims at exploring the use of digital cameras in the presence of motions, focusing on the analysis of how facial movements and ROI tracking duration influence the performance of the proposed framework in the continuous estimation of cardiorespiratory parameters. Thus, we first create a dataset using a digital camera setup in a laboratory environment consisting of the following:

- 1) *Camera*: The CANON EOS 1100D camera (Canon, USA) was used to record two videos of the face and the

torso of the subjects in a laboratory environment, with a resolution of 1280×720 pixels and an acquisition frequency of 30 frame per seconds (fps). The camera was placed at about 0.5 m from the subject. The face of the subject covers $\sim 10\%$ of the whole image with a size of 257×361 pixels.

- 2) *Reference Instrument*: The multiparametric wearable device, the Zephyr BioHarness v3 by Medtronic was used to record the electrocardiographic (ECG) and the respiratory waveforms. The device consists of a thoracic belt and an electronic module. It acquires the user's respiratory pattern by detecting the volumetric changes in the thorax through a strain gauge and the ECG waveform via dry electrodes. The reference respiratory and ECG signals were collected at 25 and 250 Hz, respectively.
- 3) *Subjects*: A total of 20 healthy subjects (ten males and ten females, aged between 22 and 31, Fitzpatrick Skin Phototype between II and III) were enrolled for the recordings. Each volunteer was called to perform two video collections. During the first video, the subject was asked to perform: 1) ~ 120 s of quiet breathing (QB); 2) ~ 90 s of tachypnea; 3) ~ 90 s of small head rotations left and right along the craniocaudal axis during QB; and 4) ~ 90 s of small left and right body oscillations. The total video recording lasts around 8 min. During the second test, the subjects were required to perform ~ 110 s of QB after physical exercise (i.e., running at high knees) for approximately 2 min of video recording. An apnea stage of about 15 s was performed between one task and the following to then discriminate them. The experimental trials were guided via a graphical interface running on a tablet located in front of the subject. In addition, subjects were required to not wear glasses and females to not have make-up on their faces.

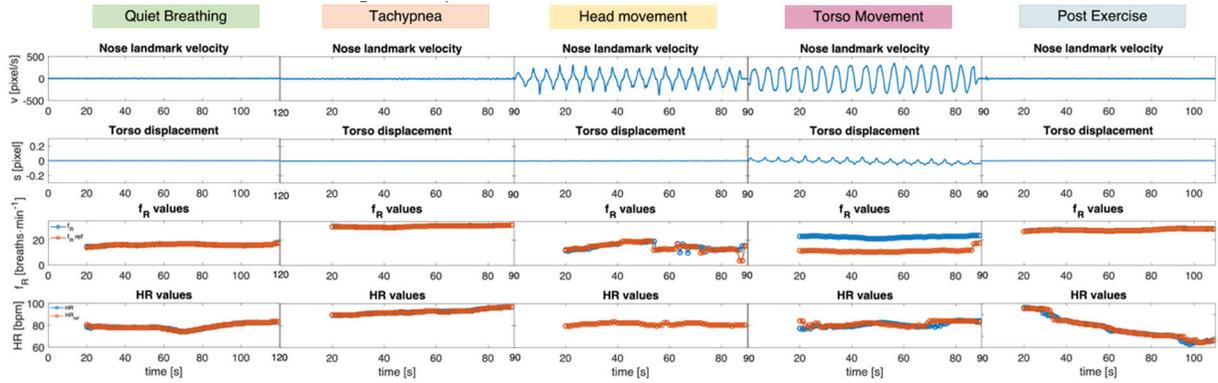


Fig. 3. Example of temporal trends of estimated f_R and HR values against the reference in the five trials according to the nose's landmark velocity (v) and the displacement of the torso along the x -axis (s) for one subject.

B. Dataset in Gym Environment

A dataset using the CANON EOS 1100D camera in a gym environment was made to investigate the use of digital cameras for the continuous estimation of cardiorespiratory parameters during physical activity. The digital camera (Canon, USA) was used to record one video of the face and the torso of the subjects during physical activity (resolution of 1280×720 pixels, at 30 fps). It was placed on a tripod at about 0.5 m from the subject and the zoom was adjusted so that the region covered by the face in the whole image is of about 10%. The multiparametric wearable device Zephyr BioHarness v3 was used as a reference system. A total of nine healthy volunteers (six females and three males, aged between 25 and 33, Fitzpatrick Skin Phototype between II and III) were enrolled for the experiments. During the tests, each subject was required to sit on a cycle ergometer (Electronic bike RHC-100, Air Machine, Cesena, Italy) and to perform the following protocol: 1) initial apnea stage of ~ 15 s; 2) ~ 60 s of QB at rest; and 3) ~ 300 s of cycling at a power of about 100 W. Each video recording lasts around 6 min.

C. Data Analysis

The collected videos both in the laboratory and gym environment were post-processed in MATLAB environment according to the steps reported in Fig. 2. All the r-PPG and the breathing signals were synchronized with the ECG and the respiratory waveforms from the reference system, starting from the first minimum point after the apnea stage. An analysis in the frequency domain was carried out to estimate the values of f_R and HR both for reference and video signals as reported in Section II-C. HR values were estimated from signals retrieved with the CHROM algorithm. Sometimes, some estimated values of f_R and HR are inconsistent with the other estimated values. Thus, these values (i.e., outliers) were removed using the Hampel filter, which identifies and replaces the outliers through a moving median window [49]. Fig. 3 reports the temporal trends of the estimated f_R and HR values against the reference after the outliers' removal in the five trials performed in a laboratory environment for one subject. The first line reports the nose velocities (v) computed in each

trial, whereas the second line shows the torso's displacements along the x -axis (s) extracted through OF in each trial.

D. Evaluation Metrics

To evaluate the performance of f_R and HR extraction, the mean absolute error (MAE) and the mean absolute percentage error (MAPE) were computed. In the first analysis of cardiac signals, MAE was used for evaluating the performance of each r-PPG implemented algorithm and identifying the most promising. In addition, a Bland-Altman analysis was carried out to investigate the agreement between the implemented approaches and the reference values both for f_R and HR estimation and to quantify the discrepancies.

IV. RESULTS

This section presents the results obtained in the continuous estimation of f_R and HR. We first report the results obtained in a laboratory environment for f_R and HR both when considering the threshold on the head movement and the influence of different modalities of video analysis and when neglecting them. Then, the results obtained in the gym environment are reported.

A. Dataset in Laboratory Environment

1) *Respiratory Rate Estimation*: When the threshold on the velocity was not considered and the modality of video analysis with fixed ROI was used, MAE values below 1 breaths·min⁻¹ were obtained for the motionless trials (i.e., QB, tachypnea, head movement, and post-exercise), whereas an MAE of 2.56 and 5.71 breaths·min⁻¹ were obtained during head and torso movements, respectively.

Fig. 4 shows the Bland-Altman plots for the f_R values estimated through OF in the five trials using the fixed ROI. The dashed line represents the mean of difference (MOD), and the red lines represent the upper and low limits of agreement (LOA). A good agreement was achieved during QB, tachypnea, and post-exercise, with a MOD \pm LOAs of 0.04 ± 2.16 breaths·min⁻¹ (mean error of $\sim 11\%$), 0.24 ± 6.11 breaths·min⁻¹ ($\sim 15\%$) and 0.05 ± 4.83 breaths·min⁻¹ ($\sim 19\%$), respectively. The values of LOAs increase when f_R was estimated during head and torso movements (± 8.92 and ± 14.67 breaths·min⁻¹, with a MOD of 0.34 and 2.50 breaths·min⁻¹, respectively).

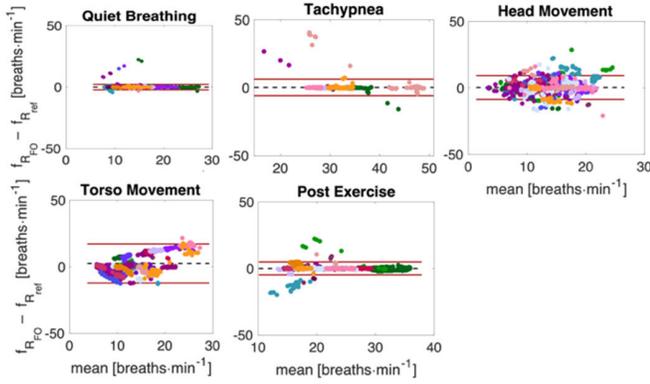


Fig. 4. Bland–Altman plots comparing f_R estimated through OF in the five trials using the fixed ROI and without considering the threshold on the nose’s velocity. Each color corresponds to a single subject.

Table I shows MAE and MAPE values obtained under changing velocity thresholds and modality of video analysis for all five trials. In motionless trials, MAE below $3.5 \text{ breaths}\cdot\text{min}^{-1}$ was obtained when performing ROI tracking between Mode 1 s and Mode 20 s, even if MAPE values reach high values. In Mode 1 an increase of MAE was achieved. During head and torso movements, the values of MAE are generally slightly higher. However, MAE $< 5 \text{ breaths}\cdot\text{min}^{-1}$ with the higher percentage of monitored windows (i.e., 88% during head movements and 63% during torso movements) were achieved at 400 pixel/s for all the implemented modalities of video analysis (see Table S1). Comparing the different modalities of video analysis, Mode fixed ROI results to be the best in motionless trials, although when performing head and torso movements Mode 10 s is most suitable with MAE below $4.09 \text{ breaths}\cdot\text{min}^{-1}$ at all the considered velocity thresholds (see Table S2 for Bland–Altman analysis).

2) *HR Estimation*: The performances of each implemented r-PPG algorithm in the estimation of HR in the different trials were investigated by analyzing the values of MAE. According to the achieved values reported in Table S3, the CHROM algorithm results to be the best method to retrieve the valuable r-PPG signal for the estimation of HR in all the performed trials (i.e., QB, tachypnea, head movement, torso movement, and post-exercise) with an MAE $< 5 \text{ bpm}$. As a result, the CHROM algorithm was used for further analyses, including the estimation of HR values considering the threshold on the velocity of the pixel of the nose’s landmark and the analysis of the ROI tracking on the subject’s face.

To evaluate the performances of the proposed framework under changing velocity thresholds and modality of video analysis, we computed the MAE and the MAPE for all five trials (Table II). MAE values below 4.26 bpm and MAPE $< 3.89\%$ were achieved in motionless trials (i.e., QB, tachypnea, and post-exercise) for all the tested modality of analysis and for all the velocity thresholds. As for f_R estimation, the different number of windows respecting the thresholds has been used to calculate the MAE. The percentage of windows used for HR in each condition are reported in Table S4 for all the implemented modalities of video analysis. When considering Mode 1, MAE of 3.17 bpm was obtained

TABLE I
MAE AND MAPE COMPUTED PER EACH TRIAL UNDER CHANGING VELOCITY THRESHOLDS AND MODALITY OF ANALYSIS FOR f_R ESTIMATION. IN BOLD ARE HIGHLIGHTED THE BEST RESULTS

		MAE [breaths · min ⁻¹]; MAPE [%]				
Trial	Velocity [pixel/s]	Modality of analysis				
		Mode 1	Mode 1s	Mode 5s	Mode 10s	Mode 20s
QB	50	5.22;	2.07;	1.32;	0.61;	0.41;
		32.80	14.96	9.78	5.88	4.27
	60	5.24;	2.07;	1.30;	0.61;	0.41;
		32.96	14.96	10.02	5.88	4.27
	100	5.22;	2.07;	1.32;	0.61;	0.41;
		32.80	14.96	9.78	5.88	4.27
	150	5.22;	2.07;	1.32;	0.61;	0.41;
		32.80	14.96	9.78	5.88	4.27
	200	5.22;	2.07;	1.32;	0.61;	0.41;
		32.80	14.96	9.78	5.88	4.27
300	5.22;	2.07;	1.32;	0.61;	0.41;	
	32.80	14.96	9.78	5.88	4.27	
400	5.22;	2.07;	1.32;	0.61;	0.41;	
	32.80	14.96	9.78	5.88	4.27	
Tachypnea	50	16.73;	3.44;	3.46;	1.24;	1.39;
		53.13	11.16	10.17	3.74	5.49
	60	16.33;	3.40;	3.42;	1.23;	1.38;
		51.84	11.04	10.06	3.70	5.43
	100	16.54;	3.40;	3.42;	1.23;	1.38;
		52.53	11.04	10.06	3.70	5.43
	150	16.54;	3.40;	3.42;	1.23;	1.38;
		52.53	11.04	10.06	3.70	5.43
	200	16.54;	3.40;	3.42;	1.23;	1.38;
		52.53	11.04	10.06	3.70	5.43
300	16.54;	3.40;	3.42;	1.23;	1.38;	
	52.53	11.04	10.06	3.70	5.43	
400	16.54;	3.40;	3.42;	1.23;	1.38;	
	52.53	11.04	10.06	3.70	5.43	
Head Mov	50	5.84;	0.36;	0.36;	0.37;	0.38;
		56.10	6.45	6.07	6.18	6.34
	60	4.76;	0.62;	0.62;	0.64;	0.63;
		45.09	8.27	7.45	7.65	7.46
	100	5.00;	2.33;	2.08;	2.06;	2.11;
		40.92	17.46	15.05	16.53	15.83
	150	4.21;	3.42;	2.31;	2.98;	3.21;
		36.07	30.81	19.47	25.97	26.90
	200	4.46;	3.89;	2.65;	3.08;	3.29;
		41.04	36.01	23.59	28.10	28.98
300	4.65;	3.88;	3.28;	2.85;	2.94;	
	40.24	34.81	26.43	24.77	25.03	
400	5.03;	3.80;	3.17;	2.65;	2.78;	
	44.48	34.11	26.82	24.57	24.94	
Torso Mov	50	-;-	-;-	-;-	-;-	-;-
		2.93;	1.46;	8.31;	2.84;	6.06;
	60	21.93	12.96	62.33	22.26	46.25
		3.04;	2.76;	5.52;	3.39;	4.70;
	100	27.16	27.75	46.55	31.62	40.88
		3.04;	2.76;	5.52;	3.39;	4.70;
	150	27.16	27.75	46.55	31.62	40.88
		2.22;	3.74;	5.33;	4.09;	5.01;
	200	20.28	30.47	41.21	32.46	39.71
		2.71;	3.36;	3.97;	3.55;	4.18;
300	22.98	25.49	29.60	26.56	30.97	
	3.00;	4.00;	4.42;	4.06;	4.53;	
400	27.39	31.87	34.72	31.98	35.92	
	7.29;	2.98;	2.32;	1.79;	1.90;	
50	32.28	13.39	10.24	8.18	8.16	
	7.28;	2.98;	2.43;	1.79;	6.06;	
60	32.26	13.38	10.56	8.17	8.15	
	7.28;	2.96;	2.33;	1.80;	1.88;	
100	32.26	13.29	10.27	8.19	8.08	
	7.28;	2.95;	2.33;	1.79;	1.87;	
150	32.24	13.27	10.25	8.17	8.07	
	7.28;	2.95;	2.33;	1.79;	1.87;	
200	32.24	13.27	10.25	8.17	8.07	
	7.28;	2.95;	2.33;	1.79;	1.87;	
300	32.24	13.27	10.25	8.17	8.07	
	7.28;	2.95;	2.33;	1.79;	1.87;	
400	32.24	13.27	10.25	8.17	8.07	
	7.28;	2.95;	2.33;	1.79;	1.87;	

at 100 pixel/s during head movement, but only 6% of all the windows were used for the estimation.

TABLE II

MAE AND MAPE COMPUTED PER EACH TRIAL UNDER CHANGING VELOCITY THRESHOLDS AND MODALITY OF ANALYSIS FOR HR ESTIMATION. IN BOLD ARE HIGHLIGHTED THE BEST RESULTS

		MAE [bpm]; MAPE [%]					
Trial	Velocity [pixel/s]	Modality of analysis					
		Mode 1	Mode 1s	Mode 5s	Mode 10s	Mode 20s	
QB	50	0.94 ; 1.17	1.06 ; 1.31	1.08 ; 1.32	1.11 ; 1.38	1.10 ; 1.33	
	60	0.94; 1.17	1.06; 1.31	1.08; 1.32	1.11; 1.38	1.10; 1.33	
	100	0.94; 1.17	1.06; 1.31	1.08; 1.32	1.11; 1.38	1.10; 1.33	
	150	0.94; 1.17	1.06; 1.31	1.08; 1.32	1.11; 1.38	1.10; 1.33	
	200	0.94; 1.17	1.06; 1.31	1.08; 1.32	1.11; 1.38	1.10; 1.33	
	300	0.94; 1.17	1.06; 1.31	1.08; 1.32	1.11; 1.38	1.10; 1.33	
	400	0.94; 1.17	1.06; 1.31	1.08; 1.32	1.11; 1.38	1.10; 1.33	
	Tachypnea	50	1.33; 1.57	1.98 ; 2.18	1.74 ; 1.95	1.44 ; 1.65	1.37 ; 1.59
		60	1.33; 1.56	2.01; 2.22	1.76; 1.97	1.44; 1.66	1.38; 1.60
		100	1.33; 1.56	2.02; 2.23	1.76; 1.98	1.44; 1.67	1.38; 1.61
150		1.32 ; 1.56	2.02; 2.23	1.76; 1.98	1.44; 1.67	1.38; 1.61	
200		1.32; 1.56	2.02; 2.23	1.76; 1.98	1.44; 1.67	1.38; 1.61	
300		1.32; 1.56	2.02; 2.23	1.76; 1.98	1.44; 1.67	1.38; 1.61	
400		1.32; 1.56	2.02; 2.23	1.76; 1.98	1.44; 1.67	1.38; 1.61	
Head Mov		50	3.25; 3.48	10.20; 10.79	6.84; 8.21	4.64 ; 5.00	6.31; 6.72
		60	3.47; 4.09	8.59 ; 9.45	5.50 ; 5.72	4.48; 4.77	5.27 ; 5.52
		100	3.17 ; 3.71	8.73; 9.80	7.17; 8.74	5.40; 6.11	5.89; 6.47
	150	3.91; 4.45	9.59; 10.95	7.43; 8.80	7.48; 8.41	6.83; 7.62	
	200	3.36; 3.92	9.25; 11.31	7.55; 9.52	7.42; 8.87	6.56; 7.67	
	300	4.73; 5.77	10.65; 13.41	10.211 3.00	9.75; 12.07	9.83; 12.21	
	400	5.42; 6.65	11.19; 14.65	11.041 4.20	10.40 13.24	10.351 3.03	
	Torso Mov	50	-;-	-;-	-;-	-;-	-;-
		60	0.58 ; 0.84	6.00 ; 8.78	5.63 ; 8.27	7.63 ; 11.20	5.18 ; 7.69
		100	1.89; 2.09	13.83; 15.66	16.321 8.42	18.002 0.45	15.96 18.01
150		1.89; 2.09	13.83; 15.66	16.32 18.42	18.002 0.45	15.961 8.01	
200		1.47; 1.70	14.89; 18.38	17.472 1.14	15.632 1.27	18.362 1.73	
300		1.26; 1.57	12.55; 16.03	15.491 9.29	18.121 9.06	15.251 8.79	
400		1.22; 1.56	11.25; 14.35	15.251 9.63	15.511 9.74	14.591 8.49	
Post Ex		50	2.36 ; 2.39	4.17 ; 3.81	3.73 ; 3.43	3.87 ; 3.49	3.72 ; 3.45
		60	2.36; 2.39	4.17; 3.81	3.74; 3.44	3.88; 3.50	3.72; 3.46
		100	2.41; 2.44	4.24; 3.88	3.79; 3.49	3.94; 3.55	3.79; 3.52
	150	2.43; 2.45	4.26; 3.89	3.80; 3.49	3.96; 3.56	3.80; 3.53	
	200	2.43; 2.45	4.26; 3.89	3.80; 3.49	3.96; 3.56	3.80; 3.53	
	300	2.43; 2.45	4.26; 3.89	3.80; 3.49	3.96; 3.56	3.80; 3.53	
	400	2.43; 2.45	4.26; 3.89	3.80; 3.49	3.96; 3.56	3.80; 3.53	

Whereas, during torso movement a MAE of 1.22 bpm was achieved at 400 pixel/s for Mode 1 and 62% of the total windows were used to estimate HR.

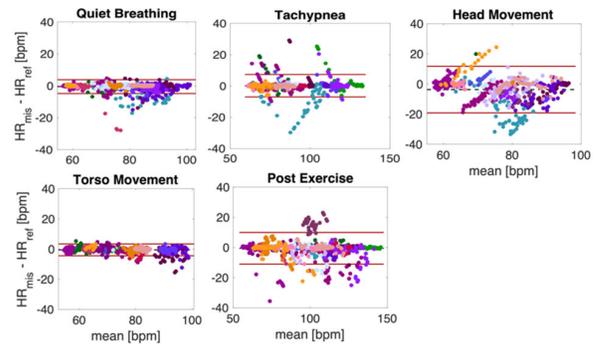


Fig. 5. Bland–Altman plots comparing HR estimated when using Modality 1 of video analysis in the five trials with a velocity threshold of 400 pixel/s. Each color corresponds to a single subject.

Comparing the different modalities of video analysis (i.e., when increasing ROI tracking duration), there is a slight increase of MAE values for all the velocity thresholds in motionless trials (e.g., during QB, MAE values range from 0.94 to 1.10 bpm at 50 pixel/s). Whereas, during head and torso movements, MAE highly increases under changing modality of video analysis (e.g., during torso movement, MAE ranges from 1.22 to 14.59 bpm at 400 pixel/s). The best modality of video analysis results is Mode 1 both for motionless trials and for trials with head and torso movements. However, when analyzing the videos with the Mode 1 the running time required to perform the ROI tracking is 65 min. Increasing the ROI tracking duration, the running time decreases up to 20 min (i.e., Mode 5 s).

Fig. 5 shows the Bland–Altman plots for HR values estimated in the five trials when using Mode 1 for the video analysis with a velocity threshold of 400 pixel/s. Good agreement was achieved during QB, tachypnea, and torso movement, with a MOD ± LOAs of -0.48 ± 4.33 bpm (mean error of $\sim 6\%$), 0.16 ± 7.17 bpm ($\sim 8\%$), and -0.51 ± 3.92 bpm ($\sim 5\%$), respectively. Higher values of LOAs were achieved during head movement and post-exercise (± 15.45 bpm – $\sim 20\%$ and ± 10.49 bpm – $\sim 10\%$, respectively). In accordance with the values of MAE, changing the modality of video analysis to track the identified ROIs, the values of LOAs increase (see Table S5).

B. Dataset in Gym Environment

Based on the promising results obtained in a laboratory environment, experiments were carried out in a gym during physical activity. Assuming that no considerable motions can occur during physical activity on a cycle ergometer, here we report the obtained results in the estimation of f_R with fixed ROI modality and HR with Mode 1 under changing velocity thresholds. MAE values were computed per each velocity threshold during QB and physical activity for the two parameters as reported in Table III.

Regarding f_R estimation, an MAE below 3 breaths·min⁻¹ was achieved when the threshold is 50 pixel/s, whereas for HR estimation, MAE below 4 bpm was obtained at 50 pixel/s. However, during physical activity, 42% and 38% of the windows were used for the computation of f_R and HR, respectively. Increasing the velocity threshold up to 400 pixel/s the number of windows used for the estimation of the vital signs

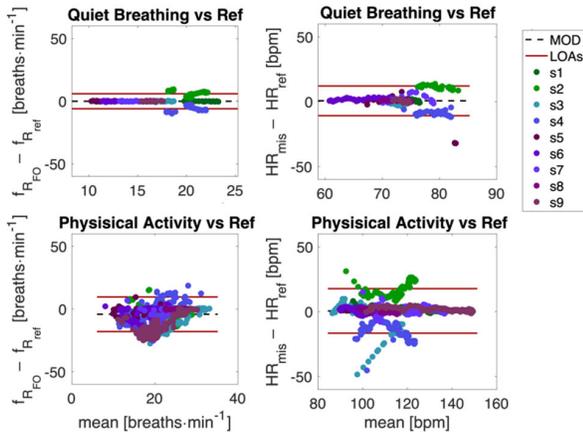


Fig. 6. Bland–Altman plots comparing the estimated values of f_R and HR with a velocity threshold of 400 pixel/s for the two trials.

TABLE III

MAE AND MAPE COMPUTED FOR f_R AND HR ESTIMATION UNDER CHANGING VELOCITY THRESHOLDS PER EACH TRIAL AND WINDOWS NUMBER USED FOR f_R AND HR ESTIMATION IN PERCENTAGE. IN BOLD ARE HIGHLIGHTED THE BEST RESULTS

Trial	Velocity [pixel/s]	MAE; MAPE [%]		Win. Number [%]	
		f_R [breaths·min ⁻¹]	HR [bpm]	f_R	HR
QB	50	1.43;	3.50;	100	100
		7.38	4.55		
	60	1.41;	3.50;	100	100
		7.38	4.55		
	100	1.43;	3.50;	100	100
		7.38	4.55		
	150	1.43;	3.50;	100	100
		7.38	4.55		
200	1.43;	3.50;	100	100	
	7.38	4.55			
300	1.43;	3.50;	100	100	
	7.38	4.55			
400	1.43;	3.50;	100	100	
	7.38	4.55			
Physical Activity	50	2.48;	3.20;	42	38
		18.80	2.93		
	60	3.13;	4.07;	48	45
		14.49	3.67		
	100	4.17;	6.10;	62	60
		17.72	5.41		
	150	4.82;	5.32;	82	80
		20.25	4.72		
200	4.65;	4.98;	91	89	
	19.51	4.43			
300	4.60;	4.87;	98	97	
	19.04	4.30			
400	4.57;	4.84;	98	97	
	18.98	4.27			

increase (i.e., 98% and 97% for f_R and HR, respectively), as well as the MAE.

A Bland–Altman analysis was performed considering the values of f_R and HR estimated with a threshold velocity of 400 pixel/s (Fig. 6). For f_R estimation, a good agreement was achieved during QB at rest before the physical activity with values of LOAs of ± 5.99 breaths·min⁻¹. During the physical activity, a MOD \pm LOAs of -4.10 ± 13.80 breaths·min⁻¹ was obtained. Regarding the HR estimation, values of LOAs of ± 11.45 and ± 17.16 bpm were obtained during QB at rest

before the physical activity and during physical activity, with approximately 15% of error in both cases considering the whole HR ranges.

V. DISCUSSIONS

In this article, we focus on the analysis of how facial movements and ROI tracking duration influence the performance of a method based on images captured with a single digital camera for the estimation of HR and f_R . The proposed framework allows the continuous estimation of these two parameters both in simulated and real scenarios (i.e., during physical exercise in a gym).

A. Dataset in Laboratory Environment

The obtained results from the dataset in a laboratory environment show that the velocity threshold of 50 pixel/s allows achieving low values of MAE both for f_R and HR estimation in all the motionless trials and a very continuous monitoring can be performed since all the windows were used to estimate the two vital signs.

For head and torso movements, it is difficult to find the right compromise between the velocity threshold to be set, the windows number used to perform f_R and HR estimation, and the low values of MAE. Setting a velocity threshold at 400 pixel/s allows continuous monitoring of both f_R and HR during head and torso movements using a high number of windows for the estimation of the vital signs.

Focusing on f_R estimation, we obtained comparable results in motionless trials both when considering the velocity thresholds and when neglecting them. An MAE < 2 breaths·min⁻¹ was achieved and values of LOAs comparable with those obtained in [32] (LOAs of ± 1.04 breaths·min⁻¹) were obtained in motionless trials when considering the Mode 10 s for the video analysis. During head and torso movements, the obtained results in terms of MAE and LOAs change under changing velocity thresholds and with the windows number used to estimate f_R . The best results in terms of low values of MAE and LOAs were achieved at 50 pixel/s. The obtained MOD \pm LOAs of 0.25 ± 2.18 breaths·min⁻¹ at 50 pixel/s (Mode 10 s) during head movement is comparable with that reported in [32] (MOD \pm LOAs of 0.18 ± 2.45 breaths·min⁻¹) during tests in non-stationary scenario.

Regarding HR estimation, we obtained comparable MAE in all the tested modalities of video analysis both when considering the velocity thresholds and when neglecting them. MAE is between 0.94 and 4.30 bpm for the motionless trials and per each modality of video analysis when the velocity threshold was not considered. These values are comparable with those obtained in [29] (MAE of 6.13 bpm). When head and torso movements were performed, MAE increased ranging from 4.26 and 10.88 bpm for head movement and ranging from 3.74 and 15.22 bpm during torso movement without no threshold on the velocity (see Table S6). Comparing the different modalities of video analysis under changing velocity thresholds Mode 10 s and Mode 1 result to be the most suitable for f_R and HR estimation in all five trials, respectively. However, when performing the HR monitoring in the absence of motions or after a physical exercise, the video

can be analyzed with all the tested modes of analysis with $MAE < 4.26$ bpm (see Table II), which is in accordance with the threshold defined in the ANSI/AAMI EC13:2002 standard. This result suggests that the ROI tracking can be performed every 20 s obtaining a continuous estimation of HR per each velocity threshold in motionless trials with low errors and reducing the computational burden. However, in the case of HR monitoring in the presence of head and/or torso movements, an ROI tracking every frame should be executed, as also reported in other studies [37], [50]. The results obtained under changing velocity thresholds during head and torso movements in terms of MAE and LOAs are comparable with other studies, in which an average value of HR was estimated during lateral movements or head movements and/or natural movements, such as head rotation, blinking, and speaking (LOAs of ± 3.07 bpm in [51], MAE of 5.45 bpm in [31], LOAs of ± 2.45 bpm in [32]). However, unlike the reported studies, our work provides continuous monitoring of HR even in dynamic trials (LOAs of ± 3.92 bpm during torso movements and LOAs of ± 15.45 bpm during head movements) when setting a velocity threshold of 400 pixel/s.

B. Dataset in Gym Environment

To the best of our knowledge, there are few studies concerning the estimation of f_R during physical activity, thus it is difficult to have a comparison with other research. However, regarding HR estimation there are various research articles that explore the potentiality of estimating HR during physical activity [52], [53], [54]. Comparing our results in the estimation of HR in terms of MAE, our results are in line with those obtained in [54], in which an MAE of ~ 2 bpm was achieved in the estimation of average HR during 3 min of physical activity. However, to the best of our knowledge, none of the studies available in the literature have done a quantitative analysis of cardiac and respiratory unrelated movements based on the velocity of the pixel of the nose's landmark during a physical exercise to obtain continuous monitoring of cardiorespiratory parameters.

VI. CONCLUSION

In this article, promising results are obtained for the monitoring of cardiorespiratory parameters in both motionless and dynamic conditions and during physical activity. A single digital camera was used to record videos and a novel approach that considers the velocity of the performed movements is presented, trying to provide a framework for the continuous estimation of f_R and HR during both static and dynamic activities by optimizing the ROI tracking duration under different velocity of facial movements. Results demonstrated that a continuous estimation of f_R and HR can be performed by setting a velocity threshold of 400 pixel/s both in the laboratory and gym environment, with MAE below 3.42 breaths·min⁻¹ for f_R when performing ROI tracking between Mode 1 s and Mode 20 s and MAE < 5 bpm for HR estimation in all the tested modality of video analysis per all the motionless trials. In dynamic trials (i.e., head and torso movements) a continuous estimation of f_R can be performed with a MAE < 5.03 breaths·min⁻¹ in all tested modalities of video analysis,

whereas during cycling on a cycle ergometer f_R can be estimated continuously with a MAE below 4.57 breaths·min⁻¹ when using the fixed ROI modality under the assumption that there are no considerable motions. Continuous estimation of HR can be performed with an MAE below 6.10 bpm when using Mode 1 in all the dynamic trials, with a run processing time of about 65 min. However, the other modalities of video analysis can be used to reduce the run processing time to 20 min, but higher values of MAE are achieved. One limitation of the study is related to the non-inclusion of subjects with different skin colors in the experimental trials, which will be further investigated. In addition, future investigations will be devoted to the improvement of the proposed motion-robust framework, performing tests on a larger population of different ages including the elderly and children, and in other scenarios (e.g., during sports activities, in clinical scenarios, or in home-monitoring).

ACKNOWLEDGMENT

This work was carried out in the framework of the project titled "Sviluppo di sistemi di misura e modelli per la stima senza contatto di parametri fisiologici," POR Lazio FSE 2014/2020 (CUP code n. F87C21000190009, Project ID 23562).

REFERENCES

- [1] S. Gerry et al., "Early warning scores for detecting deterioration in adult hospital patients: Systematic review and critical appraisal of methodology," *BMJ*, vol. 369, no. 3, pp. 1–16, 2020, doi: [10.1136/bmj.m1501](https://doi.org/10.1136/bmj.m1501).
- [2] K. Fox et al., "Resting heart rate in cardiovascular disease," *J. Amer. College Cardiol.*, vol. 50, no. 9, pp. 823–830, 2007.
- [3] R. R. Mølgaard, P. Larsen, and S. J. Håkonsen, "Effectiveness of respiratory rates in determining clinical deterioration: A systematic review protocol," *JBI Database Systematic Rev. Implement. Rep.*, vol. 14, no. 7, pp. 19–27, Jul. 2016.
- [4] W. Schüller, N. Spicher, and T. M. Deserno, "Cardiopulmonary coupling analysis using smart wearables and mobile computing," *Current Directions Biomed. Eng.*, vol. 7, no. 2, pp. 291–294, Oct. 2021, doi: [10.1515/cdbme-2021-2074](https://doi.org/10.1515/cdbme-2021-2074).
- [5] M. Harford, J. Catherall, S. Gerry, J. Young, and P. Watkinson, "Availability and performance of image-based, non-contact methods of monitoring heart rate, blood pressure, respiratory rate, and oxygen saturation: A systematic review," *Physiol. Meas.*, vol. 40, no. 6, Jun. 2019, Art. no. 06TR01, doi: [10.1088/1361-6579/ab1f1d](https://doi.org/10.1088/1361-6579/ab1f1d).
- [6] C. Massaroni, M. Zaltieri, D. Lo Presti, A. Nicolò, D. Tosi, and E. Schena, "Fiber Bragg grating sensors for cardiorespiratory monitoring: A review," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14069–14080, Jul. 2021.
- [7] N. Molinaro et al., "Wearable textile based on silver plated knitted sensor for respiratory rate monitoring," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2865–2868, doi: [10.1109/EMBC.2018.8512958](https://doi.org/10.1109/EMBC.2018.8512958).
- [8] M. Paul et al., "Non-contact sensing of neonatal pulse rate using camera-based imaging: A clinical feasibility study," *Physiol. Meas.*, vol. 41, no. 2, Mar. 2020, Art. no. 024001, doi: [10.1088/1361-6579/ab755c](https://doi.org/10.1088/1361-6579/ab755c).
- [9] N. Molinaro et al., "Contactless vital signs monitoring from videos recorded with digital cameras: An overview," *Frontiers Physiol.*, vol. 13, p. 160, Feb. 2022.
- [10] M. Varanini, P. C. Berardi, F. Conforti, M. Micalizzi, D. Neglia, and A. Macerata, "Cardiac and respiratory monitoring through non-invasive and contactless radar technique," in *Proc. Comput. Cardiology*, Sep. 2008, pp. 149–152, doi: [10.1109/CIC.2008.4748999](https://doi.org/10.1109/CIC.2008.4748999).
- [11] S. Franceschini, M. Ambrosiano, and F. Baselice, "MUHD: A multi-channel ultrasound prototype for remote heartbeat detection," in *Proc. 13th Int. Joint Conf. Biomed. Eng. Syst. Technol. (BIODEVICES)*, 2020, pp. 57–63, doi: [10.5220/0008982700570063](https://doi.org/10.5220/0008982700570063).
- [12] S. Suzuki et al., "A novel autonomic activation measurement method for stress monitoring: Non-contact measurement of heart rate variability using a compact microwave radar," *Med. Biol. Eng. Comput.*, vol. 46, no. 7, pp. 709–714, Jul. 2008, doi: [10.1007/s11517-007-0298-3](https://doi.org/10.1007/s11517-007-0298-3).

- [13] M. Ambrosanio, S. Franceschini, G. Grassini, and F. Baseliace, "A multi-channel ultrasound system for non-contact heart rate monitoring," *IEEE Sensors J.*, vol. 20, no. 4, pp. 2064–2074, Feb. 2020, doi: [10.1109/JSEN.2019.2949435](https://doi.org/10.1109/JSEN.2019.2949435).
- [14] D. Qiao, F. Zulkernine, R. Masroor, R. Rasool, and N. Jaffar, "Measuring heart rate and heart rate variability with smartphone camera," in *Proc. 22nd IEEE Int. Conf. Mobile Data Manage. (MDM)*, Jun. 2021, pp. 248–249.
- [15] C. Massaroni, E. Schena, S. Silvestri, F. Taffoni, and M. Merone, "Measurement system based on RGB camera signal for contactless breathing pattern and respiratory rate monitoring," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2018, pp. 1–6, doi: [10.1109/MeMeA.2018.8438692](https://doi.org/10.1109/MeMeA.2018.8438692).
- [16] N. Molinaro, E. Schena, S. Silvestri, and C. Massaroni, "Multi-ROI spectral approach for the continuous remote cardio-respiratory monitoring from mobile device built-in cameras," *Sensors*, vol. 22, no. 7, p. 2539, Mar. 2022.
- [17] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, p. 21434, 2008, doi: [10.1364/oe.16.021434](https://doi.org/10.1364/oe.16.021434).
- [18] L. Bernardi, C. Porta, A. Gabutti, L. Spicuzza, and P. Sleight, "Modulatory effects of respiration," *Auton. Neurosci.*, vol. 90, nos. 1–2, pp. 47–56, 2001, doi: [10.1016/S1566-0702\(01\)00267-3](https://doi.org/10.1016/S1566-0702(01)00267-3).
- [19] R.-N. Yin et al., "Heart rate estimation based on face video under unstable illumination," *Appl. Intell.*, vol. 51, no. 8, pp. 5388–5404, 2021.
- [20] B. Han, K. Ivanov, Y. Yan, and L. Wang, "Exploration of the optimal skin-camera distance for facial photoplethysmographic imaging measurement using cameras of different types," in *Proc. 5th EAI Int. Conf. Wireless Mobile Commun. Healthcare-Transforming Healthcare Through Innov. Mobile Wireless Technol. (MOBIHEALTH)*, 2015, pp. 186–189, doi: [10.4108/eai.14-10-2015.2261689](https://doi.org/10.4108/eai.14-10-2015.2261689).
- [21] R. Song, S. Zhang, J. Cheng, C. Li, and X. Chen, "New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method," *Comput. Biol. Med.*, vol. 116, Jan. 2020, Art. no. 103535, doi: [10.1016/j.combiomed.2019.103535](https://doi.org/10.1016/j.combiomed.2019.103535).
- [22] D. Dăcu, M. Cidota, S. Lukosch, and L. Rothkrantz, "Noncontact automatic heart rate analysis in visible spectrum by specific face regions," in *Proc. 14th Int. Conf. Comput. Syst. Technol.*, Jun. 2013, pp. 120–127.
- [23] D.-Y. Kim, K. Lee, and C.-B. Sohn, "Assessment of ROI selection for facial video-based rPPG," *Sensors*, vol. 21, no. 23, p. 7923, Nov. 2021, doi: [10.3390/s21237923](https://doi.org/10.3390/s21237923).
- [24] M. A. Hassan et al., "Heart rate estimation using facial video: A review," *Biomed. Signal Process. Control*, vol. 38, pp. 346–360, Sep. 2017, doi: [10.1016/j.bspc.2017.07.004](https://doi.org/10.1016/j.bspc.2017.07.004).
- [25] D. J. McDuff, J. R. Estep, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 6398–6404, doi: [10.1109/EMBC.2015.7319857](https://doi.org/10.1109/EMBC.2015.7319857).
- [26] 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, doi: [10.1109/TBME.2014.2356291](https://doi.org/10.1109/TBME.2014.2356291).
- [27] H. Ghanadian, M. Ghodratioghar, and H. Al Osman, "A machine learning method to improve non-contact heart rate monitoring using an RGB camera," *IEEE Access*, vol. 6, pp. 57085–57094, 2018, doi: [10.1109/ACCESS.2018.2872756](https://doi.org/10.1109/ACCESS.2018.2872756).
- [28] W. Wang, A. C. den Brinker, and G. de Haan, "Discriminative signatures for remote-PPG," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 5, pp. 1462–1473, May 2020, doi: [10.1109/TBME.2019.2938564](https://doi.org/10.1109/TBME.2019.2938564).
- [29] M. Artemyev, M. Churikova, M. Grinenko, and O. Perepelkina, "Robust algorithm for remote photoplethysmography in realistic conditions," *Digit. Signal Process.*, vol. 104, Sep. 2020, Art. no. 102737, doi: [10.1016/j.dsp.2020.102737](https://doi.org/10.1016/j.dsp.2020.102737).
- [30] B. Wu, B. Wu, S. Cheng, Y. Sun, and M. Chung, "Motion-robust atrial fibrillation detection based on remote-photoplethysmography," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 6, pp. 2705–2716, Jun. 2023, doi: [10.1109/jbhi.2022.3172705](https://doi.org/10.1109/jbhi.2022.3172705).
- [31] K. Lin, D. Chen, and W. Tsai, "Image-based motion-tolerant remote respiratory rate evaluation," *IEEE Sensors J.*, vol. 16, no. 9, pp. 3263–3271, May 2016, doi: [10.1109/JSEN.2016.2526627](https://doi.org/10.1109/JSEN.2016.2526627).
- [32] A. Al-Naji and J. Chahl, "Simultaneous tracking of cardiorespiratory signals for multiple persons using a machine vision system with noise artifact removal," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–10, 2017, doi: [10.1109/JTEHM.2017.2757485](https://doi.org/10.1109/JTEHM.2017.2757485).
- [33] D. Cho, J. Kim, K. J. Lee, and S. Kim, "Reduction of motion artifacts from remote photoplethysmography using adaptive noise cancellation and modified HSI model," *IEEE Access*, vol. 9, pp. 122655–122667, 2021, doi: [10.1109/ACCESS.2021.3106046](https://doi.org/10.1109/ACCESS.2021.3106046).
- [34] A. Ni, A. Azarang, and N. Kehtarnavaz, "A review of deep learning-based contactless heart rate measurement methods," *Sensors*, vol. 21, no. 11, p. 3719, May 2021, doi: [10.3390/s21113719](https://doi.org/10.3390/s21113719).
- [35] A. Chung et al., "High-resolution motion-compensated imaging photoplethysmography for remote heart rate monitoring," *Proc. SPIE*, vol. 9316, Mar. 2015, Art. no. 93160A, doi: [10.1117/12.2078397](https://doi.org/10.1117/12.2078397).
- [36] K. Guo, T. Zhai, E. Pashollari, C. J. Varlamos, A. Ahmed, and M. N. Islam, "Contactless vital sign monitoring system for heart and respiratory rate measurements with motion compensation using a near-infrared time-of-flight camera," *Appl. Sci.*, vol. 11, no. 22, p. 10913, Nov. 2021, doi: [10.3390/app112210913](https://doi.org/10.3390/app112210913).
- [37] S. Zauneder, A. Trumpp, D. Wedekind, and H. Malberg, "Cardiovascular assessment by imaging photoplethysmography—a review," *Biomedizinische Technik*, vol. 63, no. 5, pp. 529–535, Oct. 2018, doi: [10.1515/bmt-2017-0119](https://doi.org/10.1515/bmt-2017-0119).
- [38] S. Maji, C. Massaroni, E. Schena, and S. Silvestri, "Contactless heart rate monitoring using a standard RGB camera," in *Proc. IEEE Int. Workshop Metro. Ind. 4.0 IoT*, Jun. 2020, pp. 729–733, doi: [10.1109/MetroInd4.0IoT48571.2020.9138202](https://doi.org/10.1109/MetroInd4.0IoT48571.2020.9138202).
- [39] A. Chatterjee, A. P. Prathosh, and P. Praveena, "Real-time respiration rate measurement from thoracoabdominal movement with a consumer grade camera," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 2708–2711, doi: [10.1109/EMBC.2016.7591289](https://doi.org/10.1109/EMBC.2016.7591289).
- [40] Accessed: Feb. 10, 2023. [Online]. Available: <https://opencv.org/>
- [41] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1867–1874.
- [42] T. A. von von Arx, K. Tamura, O. Yukiya, and S. Lozanoff, "The face—A vascular perspective. A literature review," *Swiss Dental J*, vol. 128, no. 5, pp. 382–392, 2018.
- [43] C. Massaroni, D. lo Presti, D. Formica, S. Silvestri, and E. Schena, "Non-contact monitoring of breathing pattern and respiratory rate via RGB signal measurement," *Sensors*, vol. 19, no. 12, pp. 1–15, 2019, doi: [10.3390/s19122758](https://doi.org/10.3390/s19122758).
- [44] C. Romano, E. Schena, S. Silvestri, and C. Massaroni, "Non-contact respiratory monitoring using an RGB camera for real-world applications," *Sensors*, vol. 21, no. 15, p. 5126, Jul. 2021, doi: [10.3390/s21155126](https://doi.org/10.3390/s21155126).
- [45] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.
- [46] R. Stricker, S. Müller, and H. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *Proc. 23rd IEEE Int. Symp. Robot Human Interact. Commun.*, Aug. 2014, pp. 1056–1062, doi: [10.1109/ROMAN.2014.6926392](https://doi.org/10.1109/ROMAN.2014.6926392).
- [47] G. Boccignone, D. Conte, V. Cuculo, A. D'Amelio, G. Grossi, and R. Lanzarotti, "An open framework for remote-PPG methods and their assessment," *IEEE Access*, vol. 8, pp. 216083–216103, 2020, doi: [10.1109/ACCESS.2020.3040936](https://doi.org/10.1109/ACCESS.2020.3040936).
- [48] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017, doi: [10.1109/TBME.2016.2609282](https://doi.org/10.1109/TBME.2016.2609282).
- [49] R. K. Pearson, Y. Neuvo, J. Astola, and M. Gabbouj, "Generalized Hampel filters," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, p. 87, Dec. 2016, doi: [10.1186/s13634-016-0383-6](https://doi.org/10.1186/s13634-016-0383-6).
- [50] D. J. McDuff, E. B. Blackford, and J. R. Estep, "Fusing partial camera signals for noncontact pulse rate variability measurement," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1725–1739, Aug. 2018, doi: [10.1109/TBME.2017.2771518](https://doi.org/10.1109/TBME.2017.2771518).
- [51] M. M. Shoushan, B. A. Reyes, A. M. Rodriguez, and J. W. Chong, "Non-contact HR monitoring via smartphone and webcam during different respiratory maneuvers and body movements," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 602–612, Feb. 2021, doi: [10.1109/JBHI.2020.2998399](https://doi.org/10.1109/JBHI.2020.2998399).
- [52] Q. Zhu, C. Wong, C. Fu, and M. Wu, "Fitness heart rate measurement using face videos," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2000–2004.
- [53] 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 Informat.*, vol. 18, no. 4, pp. 1153–1160, Jul. 2014, doi: [10.1109/JBHI.2013.2291900](https://doi.org/10.1109/JBHI.2013.2291900).
- [54] Y. Chou, B. Ye, H. Chen, and Y. Lin, "A real-time and non-contact pulse rate measurement system on fitness equipment," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022, doi: [10.1109/TIM.2021.3136173](https://doi.org/10.1109/TIM.2021.3136173).