Research Article

Contactless Monitoring of Breathing Patterns and Respiratory Rate at the Pit of the Neck: A Single Camera Approach

Carlo Massaroni [ID],[1] **Daniel Simões Lopes** [ID],[2] **Daniela Lo Presti** [ID],[1] **Emiliano Schena** [ID],[1] **and Sergio Silvestri** [ID][1]

[1]*Unit of Measurements and Biomedical Instrumentation, Department of Engineering, Università Campus Bio-Medico di Roma, via Álvaro del Portillo 21, 00128 Rome, Italy*
[2]*Unit of INESC-ID Lisboa, IST Taguspark, Avenida Professor Cavaco Silva, 2744-016 Porto Salvo, Portugal*

Correspondence should be addressed to Carlo Massaroni; c.massaroni@unicampus.it

Vital signs monitoring is pivotal not only in clinical settings but also in home environments. Remote monitoring devices, systems, and services are emerging as tracking vital signs must be performed on a daily basis. Different types of sensors can be used to monitor breathing patterns and respiratory rate. However, the latter remains the least measured vital sign in several scenarios due to the intrusiveness of most adopted sensors. In this paper, we propose an inexpensive, off-the-shelf, and contactless measuring system for respiration signals taking as region of interest the pit of the neck. The system analyses video recorded by a single RGB camera and extracts the respiratory pattern from intensity variations of reflected light at the level of the collar bones and above the sternum. Breath-by-breath respiratory rate is then estimated from the processed breathing pattern. In addition, the effect of image resolution on monitoring breathing patterns and respiratory rate has been investigated. The proposed system was tested on twelve healthy volunteers (males and females) during quiet breathing at different sensor resolution (i.e., HD 720, PAL, WVGA, VGA, SVGA, and NTSC). Signals collected with the proposed system have been compared against a reference signal in both the frequency domain and time domain. By using the HD 720 resolution, frequency domain analysis showed perfect agreement between average breathing frequency values gathered by the proposed measuring system and reference instrument. An average mean absolute error (MAE) of 0.55 breaths/min was assessed in breath-by-breath monitoring in the time domain, while Bland-Altman showed a bias of -0.03 ± 1.78 breaths/min. Even in the case of lower camera resolution setting (i.e., NTSC), the system demonstrated good performances (MAE of 1.53 breaths/min, bias of -0.06 ± 2.08 breaths/min) for contactless monitoring of both breathing pattern and breath-by-breath respiratory rate over time.

1. Introduction

Accurate measurement and monitoring of physiological parameters, such as body temperature, heart rate, respiratory patterns, and, above all, the respiration rate, play a crucial role in a wide range of applications in healthcare and sport activities [1, 2].

Temporal changes of physiological parameters can indicate relevant variations of the physiological status of the subject. Among the wide range of parameters which can be measured in clinical settings, the respiratory rate is the most crucial vital sign to detect early changes in the health status of critically ill patients. For instance, respiratory rate is typically collected at regular interval by operators (i.e., every 8–10 hours) in the clinical setting, while it is often neglected in home-monitored people (i.e., telemonitoring and telerehabilitation). However, the respiratory rate has been demonstrated to be a significant and sensitive clinical predictor for serious adverse events; its value increases during exacerbation of COPD [3], it can be used to determine hospitalization, and it offers the opportunity for early intervention. Moreover, the respiratory rate has been found to be a more discriminatory parameter between stable and unstable patients than heart rate [1, 4].

Conventional techniques for measuring respiration parameters require sensors in contact with the subject. Measuring techniques based on the monitoring of several parameters sampled from inspiratory and/or expiratory flow (e.g., temperature, RH, CO_2, and flow) are widely used. [4]. Sensors may also be attached directly on the torso or integrated in clothes fibers to collect respiratory-related chest or abdominal movements [5]. Such monitoring systems may cause undesirable skin irritation and discomfort, especially when long-term monitoring is required or during sleep. Moreover, it has been shown that such kind of contact-based measurement techniques may influence the underlying physiological parameters being measured [6]. Therefore, contactless monitoring systems are welcomed to surpass issues related to placing sensors on patients.

For this reason, solutions—even commercial ones—based on the analysis of the sound recorded surrounding the person and on the monitoring of temperature map changes adopted to thermal cameras and depth map changes due to breathing have been designed and tested. However, they suffer from high cost, needs of specialized people, and sometimes of low signal-to-noise ratio. Optical motion capture systems have gained greater interest in the field of respiratory monitoring in both research and clinical scenarios [12]. Other approaches resort on markers as color cues to track breathing motion [7]. Recent advancements in video technology and machine vision software have allowed RGB cameras to become exciting solutions as they provide low-cost and easy-to-use noncontact approaches for measuring and monitoring physiological signals.

Different types of cameras have been used to measure physiological parameters, including heart rate and respiratory rate, either by adopting specific sensor camera technology, principle of work, or signals processing procedures. Two main methods have been used, based on remote photoplethysmography and body motion estimation.

Several attempts have been proposed to extract respiratory features from video frames recording breathing-related movements of thorax [8–10], thoracoabdominal area [8, 11], face area [12]–[14], and area at the edge of the shoulder [15]. Even though some studies consider region of interest (ROI) which include the neck region [14], none specifically considers the pit of the neck that is a large, visible dip in between the neck and the two collarbones that may be easily identifiable from the video.

Different approaches have been also used to postprocess the pixel data to extract signal related to the respiration from such videos by the subtraction of two continuous images [8, 11], analysis of pixel intensity changes based upon independent component analysis [12, 13], analysis of average contributions of red, green, and blue channel of the video [14, 16, 17], and analysis of optical flow [9]. Even though breathing patterns and respiratory rates have been faithfully estimated using high-quality cameras [14, 16], several other approaches that rely on off-the-shelf webcams also are able to achieve the same level of monitoring accuracy [7, 8, 12, 13]. Even so, only few numbers of studies used notebook built-in webcams—usually used for video chat and video conferences—for contactless physiological monitoring [12].

Then, a large amount of studies does not declare the details of the camera adopted and a there is a lot of variability in terms of camera resolutions used in such studies (i.e., from 640×480 [7] up to 2560×1920 [17]). According to sensor resolutions and the postprocessing methods adopted, the ROI used for extracting the signal varies study by study. Although several studies acquired video data with subjects at different distances from the camera [8, 11], none performed a comparative study with different camera resolutions.

Despite the large number of studies adopting video cameras for respiratory monitoring purposes, there is a lack of results about validity and accuracy of such methods in the practice, since the majority of these studies present proof of concepts or preliminary tests and no error metrics are reported [8, 11, 14, 16]. On available quantitative results, frequency-domain analysis is generally performed to extract the frequency content of the signal collected with the video method, to estimate the average breathing rate. However, time-domain analysis techniques may be useful to investigate breath-by-breath respiratory values among the time and to analyze additional features of respiration, otherwise unfeasible with a frequency-domain analysis.

In this paper, we present a single-camera video-based respiratory monitoring system based on the selection of the pit of the neck area. The aim of the present study is threefold: (*i*) the development of the measuring system capable of noncontact monitoring of respiratory pattern by using RGB video signal acquired from a single built-in high-definition (HD) webcam; (*ii*) the experimental test of this monitoring system in extracting average and breath-by-breath breathing rate values using both frequency-domain and time-domain analyses; and (*iii*) the evaluation of the influence of the sensor setting (i.e., resolution of the video sensor) on the accuracy of both average and breath-by-breath breathing rate values.

The respiratory pattern is estimated by analyzing the intensity of reflected light at the level of the pit of the neck. Experimental trials are presented to test the measuring system in the real practice, that is, in the monitoring of breathing pattern of twelve healthy subjects at self-paced breathing rate. Lastly, an analysis of performances of the proposed measuring system at different camera resolution is presented and discussed.

2. Measuring System: CCD Sensor and Video Processing Algorithm

The proposed measuring system is composed by a hardware module for data recording and preprocessing and a software for respiratory pattern extraction.

2.1. Hardware for Video Data Recording and Preprocessing. A video recorded with a CCD camera is considered as a series of f RGB frames, that is, polychromatic images. Each image is split in red (R), green (G), and blue (B) channels. Each element of the image matrix is a pixel with (x, y) coordinates that vary along the reference coordinate system placed at the bottom left corner of the image. Each pixel value represents a color light intensity. Zero-valued pixels correspond
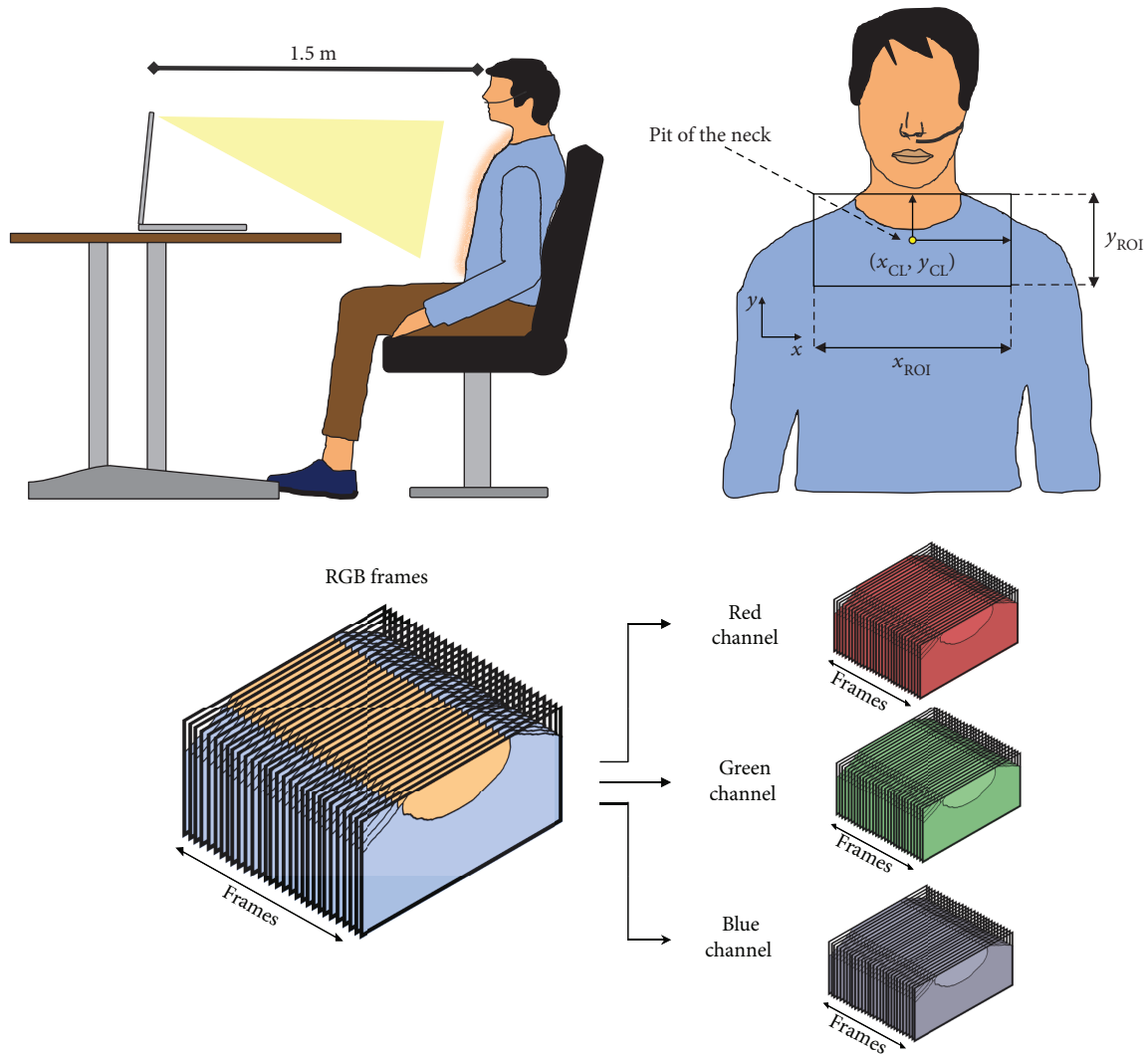
FIGURE 1: Schematic representation of the setup used to record the video and the pressure at the level of the nose. The volunteer is seated in the position shown in the figure at a distance between the camera and chest wall of around 1.5 m. The red highlighted area on the chest is the area recorded by the video. The area used to extract the breathing is the ROI of size $x_{ROI} \times y_{ROI}$, which is a rectangular region determined by the selection of pixel (x_{CL}, y_{CL}) at the pit of the neck. From the RGB video, all the frames composing it are split in the red, green, and blue components to be postprocessed. Details are provided in Section 2.2.

to a black color, whereas the maximum value renders white. The numerical values of each pixel depend on the number of bytes used to represent a given channel. When considering commercial 8-bit/channel cameras (24-bit for RGB colors), the maximum value is 2^8 (i.e., 256 colors including zero).

In this work, we used the built-in CCD RGB webcam (iSight camera) from a laptop (MacBook Pro, Apple Inc.). Images were recorded at 24-bit RGB with three channels, 8 bits per channel. An *ad hoc* interface was developed in MATLAB to manage video recording and provide useful event information to the subject (i.e., "hold breath," "start breathing," and "data collection completed") during the data collection. The duration of data collection can also be defined through the interface. Each subject was asked to perform a sequence of actions which properly informed to the subject via the graphical user interface and also timed by the experimentalist.

The video is collected at a set frame rate of 30 Hz, which is enough to discretize the breathing movements that commonly occur up to 60 breaths per minute, equal to 1 Hz.

The proposed system needs to collect an RGB video of a person seated in front of the camera (see Figure 1). The algorithm developed for the analysis was intended to recognize the respiratory signal from intensity value variations encountered in the recorded video. So, after video recording, users are asked to select a pixel at the level of pit of the neck (x_{CL}, y_{CL}) in the first frame of the video. The pit of the neck is the anatomical point near the suprasternal notch (fossa jugularis sternalis), also known as the jugular notch. It is a large, visible dip in between the neck and the two collarbones that may be easily identifiable at the superior border of the manubrium of the sternum, between the clavicular notches.

The script automatically delineates ROI that consists of a rectangular region with dimensions $x_{ROI} \times y_{ROI}$ where
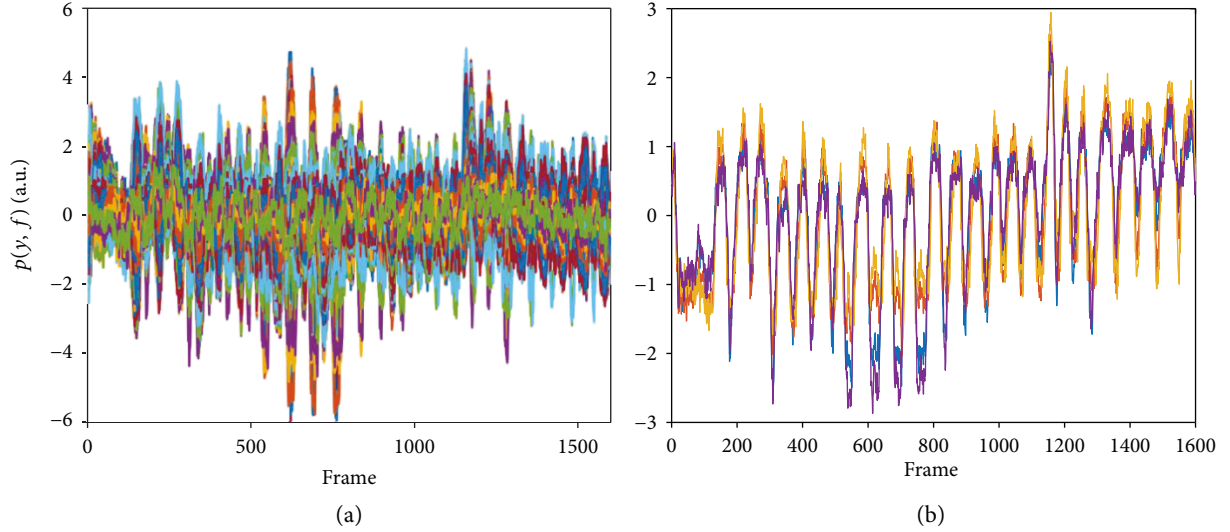
(a)

(b)

FIGURE 2: (a) Example of a trend of $p(y, f)$ collected from one volunteer during ~50 seconds of breathing (~1500 frames). (b) Trend of the 5% of the lines selected by the analysis of higher standard deviation.

$x_{\mathrm{ROI}} = [x_{\mathrm{CL}} - 1/15 \cdot 100 \cdot x, x_{\mathrm{CL}} + 1/15 \cdot 100 \cdot x]$ and $y_{\mathrm{ROI}} = [y_{\mathrm{CL}} - 1/15 \cdot 100 \cdot y, y_{\mathrm{CL}} + 1/15 \cdot 100 \cdot y]$. Details of the ROI selected starting from $(x_{\mathrm{CL}}, y_{\mathrm{CL}})$ are shown in Figure 1.

2.2. Extraction of Respiratory Pattern. To extract the respiratory pattern from the video, firstly, the selected ROI is split in the red, green, and blue channels (Figure 1). At each frame f, the intensity components of each channel $I(x, y, c, f)$ are obtained, where c is the color channel (i.e., red (R), green (G), and blue (B)). The intensity components are averaged for each line y of the ROI according to the following equation:

$$p(y, f) = \frac{1}{x_{\mathrm{ROI}}} \sum_{x=1}^{x_{\mathrm{ROI}}} \left(\sum_{c=R,G,B} I(x, y, c, f) \right), \quad y \in [1, y_{\mathrm{ROI}}]. \quad (1)$$

Therefore, p is a function of y and f. Each $p(y, f)$ is then detrended, that is, the mean is removed from the signal (Figure 2(a)).

To extract the respiratory pattern from the number of trends $p(y, f)$, the standard deviation of each $p(y, f)$ line is computed. Then, the 5% of the $p(y, f)$ signals with the higher standard deviations are selected (Figure 2(b)). The $p(f)$ pattern is then obtained computing the mean value considering the selected lines at each frame.

Figure 2(a) presents a typical trend where the variation of the intensity signal concerning the baseline is characterized by low amplitude and high frequencies in the absence of chest wall movements. When the RGB camera sensor is facing the chest wall, then respiratory content can be extracted using filtering operations. Thus, adequate cut-off frequencies and bandwidth need to be defined. It is crucial to accurately design the filter parameters to obtain proper performance of the measuring system. A band-pass configuration was chosen for controlling the whole bandwidth. The general configuration of the method requires one to fix the low cut-off frequency f_{low} around 0.05 Hz, to avoid the slow signal variations unrelated to respiratory movements. By filtering the

signal content up to a high cut-off frequency f_{high} equal to 2 Hz, the changes generated by the respiratory movements recorded to the CCD sensor can be adequately isolated and relayed to the subsequent elaboration stages.

For the filter, an infinite impulse response (IIR) filter was designed: a 3rd-order Butterworth digital filter was employed. The transfer function is expressed in terms of b and a coefficients as in the following equation:

$$H_b(z) = G_b \frac{\sum_{n=0}^{\mathrm{order}} b_n z^{-n}}{\sum_{n=0}^{\mathrm{order}} a_n z^{-n}}, \quad (2)$$

where G_b is the filter stage gain and order is the filter order. Hence, filtered output in the z-domain ($Y(z)$) can be expressed as a function of input signal ($X(z)$) which is $p(f)$ as the following equation:

$$Y(z) = H_b(z)X(z). \quad (3)$$

At that point, the signal $p(f)$ is normalized allowing us to obtain $\hat{p}(f)$, by following the equation in (3):

$$\hat{p}(f) = \frac{p(f) - \mu(p(f))}{\sigma(p(f))}, \quad (4)$$

where $\mu(p(f))$ and $\sigma(p(f))$ are the mean and standard deviation of signal $p(f)$, respectively.

Normalized signal $\hat{p}(f)$ is used for extracting respiratory pattern and temporal information since $\hat{p}(f)$ would be proportional to the changes in the intensity component and thus to the underlying respiratory signal of interest (Figure 3).

2.3. Respiratory Rate Calculation. The breathing rate can be extracted from $\hat{p}(f)$ either in frequency or time domain.

In the frequency domain, the breathing rate can be identified via power spectral density (PSD) estimate. The PSD estimation aims to assess the spectral density of a signal from a sequence of time samples of the same signal (finite
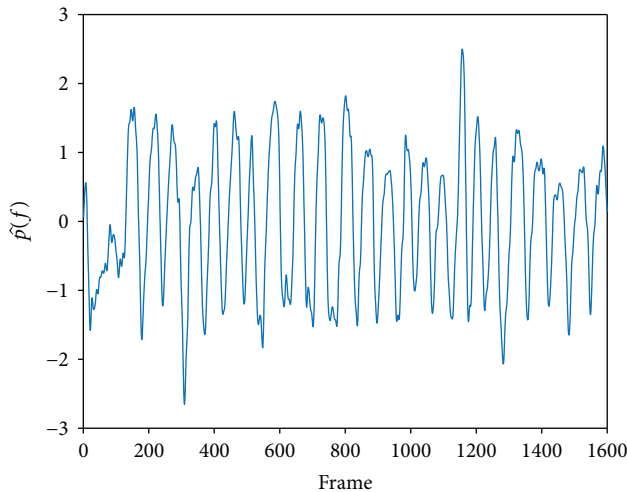
FIGURE 3: Example of the normalized $\hat{p}(f)$ signal obtained after summing the most significative $p(f)$ signals of each frame.

set of data). PSD is useful in signal detection, classification, and tracking for detecting any periodicities in the data, by observing peaks at the frequencies corresponding to these periodicities [18, 19].

The main approaches for frequency analysis consist of parametric methods (such as AR, ARMA) and nonparametric methods (window methods). Here, we focus on a nonparametric method.

Let $y(t)$ be a deterministic discrete-time signal. Assuming that $\sum_{t=-\infty}^{\infty} |y(t)|^2 < \infty$, then the discrete-time Fourier transform of the data sequence is

$$Y(\omega) = \sum_{t=-\infty}^{\infty} y(t)e^{-i\omega t}, \quad \omega = [-\pi, \pi]. \tag{5}$$

Let $S(\omega) = |Y(\omega)|^2$ be the energy spectral density; then we got

$$\sum_{t=-\infty}^{\infty} |y(t)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega)d\omega, \tag{6}$$

where $S(\omega)$ is the distribution of energy as a function of frequency. The power spectrum of a zero-mean stationary stochastic process $y(t)$ can be calculated as the Fourier transform of its covariance function $r(k)$. Hence, PSD can be defined as

$$\Phi(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-i\omega k}, \tag{7}$$

where $\Phi(\omega)$ represents the distribution of signal power over frequency [18, 19].

When using this method, the most pronounced maximum frequency peaks of the spectrum identify the periodicity of the signal. Each spectrum obtained with PSD describes how the power of the $\hat{p}(f)$ signal is distributed with frequency. In other words, the power of the $\hat{p}(f)$ signal in a given frequency band can be calculated by integrating over the frequency values of the band. Consequently, PSD can

be used to evaluate both (*i*) the variability of the pattern among the time in all the frequency band and (*ii*) the average value of the respiratory rate.

Contrary to frequency domain analysis, the time-domain specific points on the signal must be identified. Different approaches can be used based on the detection of maximum and minimum points, as well as on zero-crossing point individuation. We used a method based on both these approaches split into two steps. In the first step, the algorithm identifies the zero-crossing points on the video signal. It allows determining the onset of each respiratory cycle, characterized by a positive going zero-crossing value as

$$i : \begin{cases} x_i \geq 0, \\ x_i < 0, \end{cases} \tag{8}$$

where x_i is the value x of the signal for frame (or time) index i corresponding to the onset of a respiratory cycle. In the second step, the algorithm provides the individuation of local minimum points on the signal and their indices between respiratory cycle onsets determined in the first step as

$$\forall x \in [a, b], \quad f(x) \leq f(X), \tag{9}$$

where a and b are the time indexes of the signal corresponding to the onset of two consecutive respiratory cycles of video signal and X is the local signal minimum of a respiratory cycle. The duration of each ith breath (T_{Ri}) is then calculated as the time elapsed between two consecutive minima points. Consequently, the ith breath-by-breath breathing rate f_{Ri} is calculated as $60/T_{Ri}$.

3. Tests and Experimental Trial

3.1. Participants and Reference Data. Our dataset consists of recordings of 12 participants (six males, six females, mean ages 25 ± 3 years old, mean height of 163 ± 8 cm, mean weight 58 ± 9 kg). All the participants provided their informed consent. Each participant was invited to sit on a chair in front of the RGB camera at distance of about 1.5 m (see Figure 1). The experiments were carried out indoor and with a stable amount of light delivered by neon lights and one window as sources of illumination. During the experiments, each video was recorded for ~170 s. Participants were asked to keep still and seated, breathe spontaneously, and face the webcam.

At the same time, the pressure drop (ΔP), which occurs during exhalation/inhalation phases of respiration, was collected by a differential pressure sensor [20] (i.e., Sensirion SDP610, pressure range up to $\pm$125 Pa, Figure 4(a)). To not obstruct the data collection with the webcam, the ΔP was recorded at the level of the nose. The ΔP was sampled at 100 Hz, and the data were sent to a remote laptop via a USB connection and archived via MATLAB. All the steps carried out on signals are summarized in Figure 5.

Then, we carried out a temporal standard cumulative trapezoidal numerical integration of the ΔP signal (i.e., integrated ΔP) to provide a smooth signal for further analysis and to emphasize the maximum and minimum peaks on
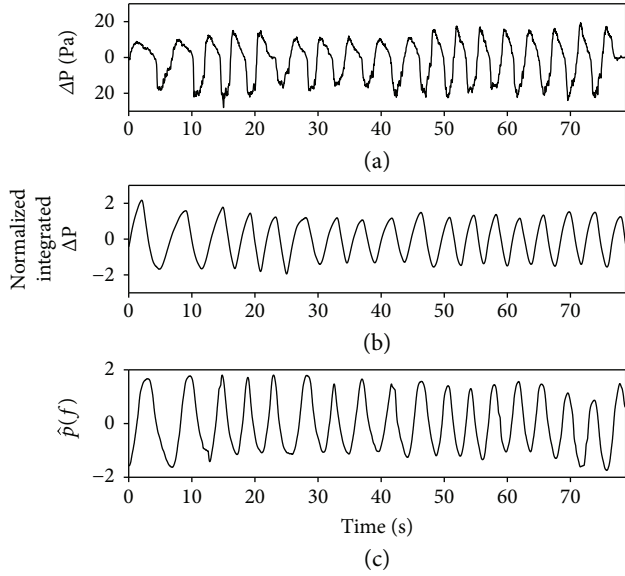
Figure 4: Example of trends relative to (a) pressure drop ΔP recorded by the differential pressure sensor, (b) integrated and normalized ΔP signal used to extract reference pattern and respiratory rate values, and (c) $\widehat{p}(f)$ signal obtained by the video processing algorithm using the data collected from the CCD sensor of the camera.

the signal (see Figure 4). This approach has been used in previous preliminary studies for extracting temporal respiratory features ([20–22]) from the pressure signals.

Afterwards, such integrated ΔP has been filtered using a bandpass Butterworth digital filter in the frequency range 0.05–2 Hz and normalized following the formula in (4). This normalized and integrated ΔP (Figure 4(b)) is the signal used for extracting reference respiratory pattern and respiratory rate values.

An example—obtained from one volunteer—of the ΔP trend collected by the pressure sensor, the normalized and integrated ΔP signal, and the $\widehat{p}(f)$ signal extracted by the video processing algorithm is reported in Figure 4.

3.2. Respiratory Pattern and Respiratory Rate Comparisons.
Signals obtained from the measuring systems have been compared to the reference signals in terms of similarity of curves and respiratory rate values. The similarity of the frequency content of signals and average respiratory rate values have been investigated from the normalized PSD.

The similarity between signals has been evaluated by overlapping the two normalized PSD, considering the one of the reference instrument as the reference PSD. From frequency dominant peak, the average respiratory rate value can be extracted. From average values of breathing rate, the accuracy (expressed in %) of the proposed method can be calculated as

$$\text{accuracy} = 100 - 100 \cdot \frac{\left| f_R - \widehat{f}_R \right|}{f_R}, \tag{10}$$

where f_R and $\widehat{f}_R$ are the breathing rate measured using the reference signal and proposed method, respectively.

Additionally, the breath-by-breath respiratory rate values have been compared between instruments by extracting such values with the time-domain analysis. To compare the values gathered by the reference instrument and computed by the video-based method, we calculate the mean absolute error (MAE) of breaths per minute as

$$\text{MAE} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \widehat{f}_R(i) - f_R(i) \right|, \tag{11}$$

where n is the number of breaths estimated for each subject in the trial, $\widehat{f}_R(i)$ is breaths per minute, and $f_R(i)$ is reference breaths per minute using reference signal data. The standard error (SE) of the mean is then calculated as

$$\text{SE} = \frac{\text{SD}}{\sqrt{n}}, \tag{12}$$

where SD is the standard deviation of the absolute difference between estimations and reference data $(\widehat{f}_R(i) - f_R(i))$.

Additionally, the strength of associations between the breath-by-breath values collected with the proposed method and those collected by the reference instrument were evaluated with the Spearman correlation coefficient. Then, the slope β of the simple linear regression ($y = \alpha + \beta x$, with y-intercept α) computed on such values has been calculated fixing $\alpha = 0$. Then, the Bland-Altman analysis was used to evaluate the differences between the two methods: mean of the differences (MOD) and the limits of agreements (LOAs) were used to determine the accuracy and the dispersions of the breath-by-breath respiratory rate differences [23].

3.3. Influence of Sensor Size Resolution.
To investigate the influence of camera sensor resolution on the accuracy of the proposed measuring system, we postelaborated the videos to decrease each frame resolution. This postprocessing was carried out in MATLAB. Bicubic interpolation was used for interpolating data points on a two-dimensional regular grid (sensor matrix). With this method, the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood.

We decided to investigate the performances of 6 camera sensor resolutions (including the resolution of the original video, HD 720) since they can be considered the most used resolution of commercial in-built webcam as HD 720, PAL, WVGA, VGA, NTSC, and SVGA, characterized by three different aspect ratios (i.e., 4:3, 5:3, and 16:9). Attributes such as sensor's size and number of x and y used in the ROI selection are reported in Table 1.

Since the ROI size is linked with the maximum size of x and y, the ROI size ($x_{\text{ROI}} \times y_{\text{ROI}}$) depends on the resolution of the CCD sensor. As a consequence, the number of lines used to compute the respiratory pattern from the video changes with the resolution (see Table 1). The same data analysis was carried out on signals according to Section 3.2, by considering each postelaborated decreased-resolution signal as a separate signal. Furthermore, same indicators for the
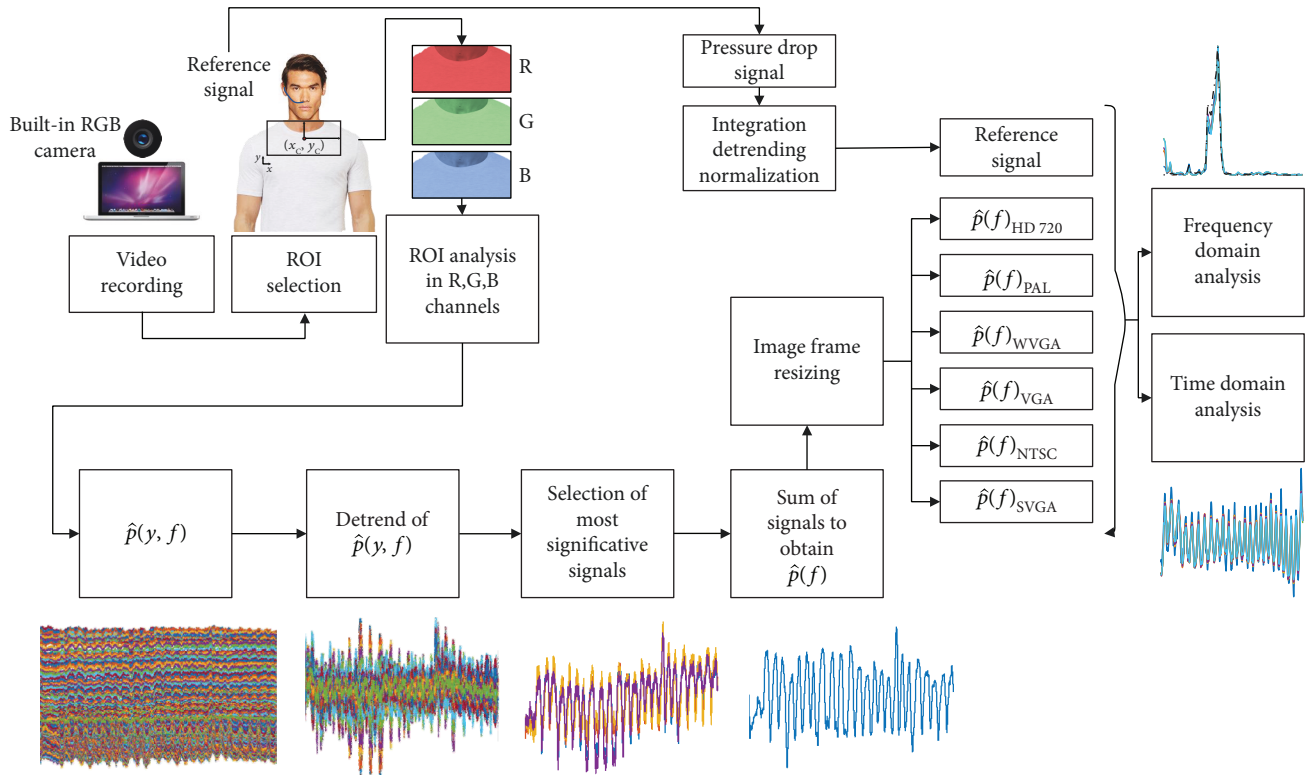
FIGURE 5: Flowchart presenting all the steps carried out to extract the respiratory pattern and then the respiratory rate from in both the frequency domain and time domain.

TABLE 1: Size of ROI in function of CCD camera setting: x_{ROI}, y_{ROI}, and the number of pixels considered for the analysis in the ROI are reported.

CCD camera setting		Characteristics of the proposed method		
Aspect ratio	Resolution (sensor setting)	x_{ROI} [px]	y_{ROI} [px]	Number of pixels in the ROI [px]
	1280 × 720 (HD 720)	384	216	82944
16 : 9	1024 × 576 (PAL)	306	172	52632
	854 × 480 (WVGA)	256	144	36864
5 : 3	800 × 480 (VGA)	240	120	28800
4 : 3	640 × 480 (NTSC)	192	144	27648
	800 × 600 (SVGA)	240	180	43200

respiratory pattern and respiratory rate comparisons against reference signal were used.

4. Results

The results obtained from the proposed measuring system are compared to the reference ones. The analysis is carried out on both frequency and time domains, separately.

4.1. Breathing Rate Estimation in the Frequency Domain. In the frequency domain, we computed the normalized PSD obtained in each trial. Average breathing rate is calculated indirectly by taking the maximum peak of the normalized PSD plot. The values for each volunteer are reported in Table 2.

The similarities between normalized PSD obtained with the reference signal and the $\widehat{p}(f)_{\text{HD720}}$ are shown in Figure 6. Since the volunteers were called to perform self-pace breathing, some of them present a pattern with high variations of the respiratory rate during data collection. In these cases, it is normal to obtain a PSD with dominant peaks at different frequencies (see Figure 6). With the proposed method we got an average accuracy on average f_R estimation of 100%.

4.2. Breathing Rate Estimation in the Time Domain. The analysis in the time domain provides additional information compared to the analysis in the frequency domain with normalized PSD, for example, the breath-by-breath respiratory rate values. An average MAE value of 0.55 breaths/min was found, while the maximum value was 1.23 breaths/min. To specify the uncertainty around the estimate of the mean measurement, we use SE since it provides a confidence interval. Thus, we calculated the 95% confidence interval as $1.96 \times \text{SE}$. The computed confidence interval was always better than 0.45 breaths/min.

By considering all the breaths collected in the 12 trials ($n = 205$), the average accuracy of the proposed method based on the analysis of $\widehat{p}(f)_{\text{HD720}}$ signal is of 97% in the breath-by-breath analysis. Linear regression analysis demonstrated a Spearman correlation coefficient of 0.97, with a slope $\beta = 1.001$. The Bland–Altman analysis reveals

TABLE 2: Dominant peak of the normalized PSD from the reference signal and the video signal ($\widehat{p}(f)_{HD720}$) and relative estimated respiratory rate value.

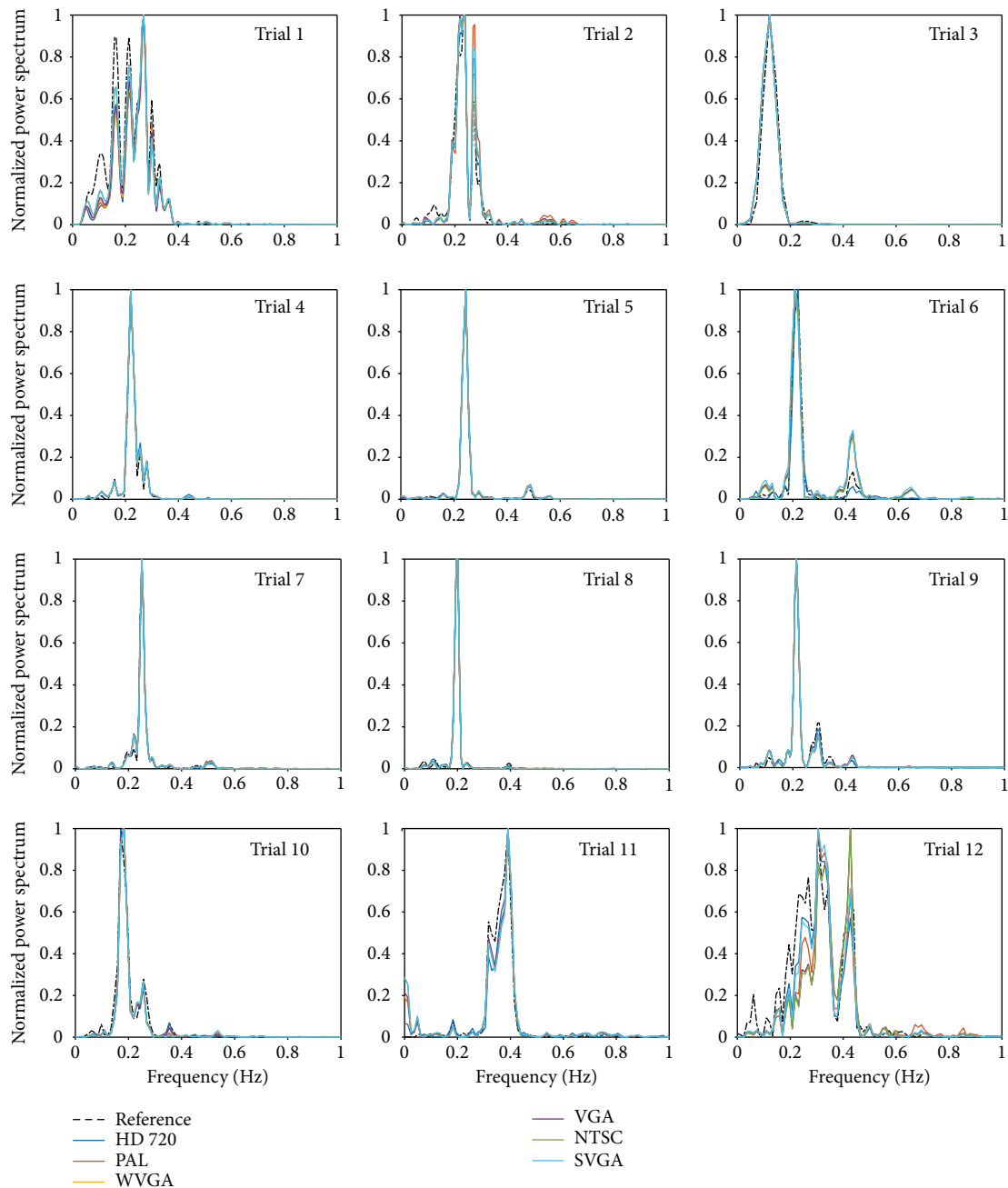|		Trial											
		1	2	3	4	5	6	7	8	9	10	11	12
Reference signal — $\widehat{p}(f)_{HD720}$	Dominant peak PSD (Hz)	0.27	0.23	0.12	0.22	0.24	0.22	0.25	0.20	0.21	0.17	0.31	0.39
		—	—	—	—	—	—	—	—	—	—	—	—
		0.27	0.23	0.12	0.22	0.24	0.22	0.25	0.20	0.21	0.17	0.31	0.39
	Estimated f_R (breaths/min)	16.11	13.92	7.32	13.18	13.18	13.20	15.01	12.09	12.82	10.25	18.31	23.44
		—	—	—	—	—	—	—	—	—	—	—	—
		16.11	13.92	7.32	13.18	13.18	13.20	13.20	12.09	12.82	10.25	18.31	23.44

FIGURE 6: Power spectrum density estimates for each camera sensor-resolution signal and reference instrument, for each volunteer.

TABLE 3: Breathing frequency extracted with the 5 sensor settings and from reference with the PSD analysis.

Trial	Respiratory rate (breaths/min) Dominant peak frequency (Hz)					
	Reference	PAL	WVGA	VGA	NTSC	SVGA
1	16.11	16.11	16.11	16.11	16.11	16.11
	0.27	0.27	0.27	0.27	0.27	0.27
2	13.92	14.28	13.18	13.18	13.92	14.28
	0.23	0.24	0.22	0.22	0.23	0.24
3	7.32	7.32	7.32	7.32	7.32	7.32
	0.12	0.12	0.12	0.12	0.12	0.12
4	13.18	13.18	13.18	13.18	13.18	13.18
	0.22	0.22	0.22	0.22	0.22	0.22
5	14.65	14.65	14.65	14.65	14.65	14.65
	0.24	0.24	0.24	0.24	0.24	0.24
6	13.18	12.45	12.45	12.45	12.45	12.45
	0.22	0.21	0.21	0.21	0.21	0.21
7	15.01	15.01	15.01	15.01	15.01	15.01
	0.25	0.25	0.25	0.25	0.25	0.25
8	12.09	11.72	11.72	11.72	12.09	11.72
	0.20	0.20	0.20	0.20	0.20	0.20
9	12.82	12.82	12.82	12.82	12.82	12.82
	0.21	0.21	0.21	0.21	0.21	0.21
10	10.25	10.99	10.99	10.99	10.99	10.99
	0.17	0.18	0.18	0.18	0.18	0.18
11	18.31	18.31	25.63	25.63	25.63	18.31
	0.31	0.31	0.43	0.43	0.43	0.31
12	23.44	23.44	23.44	23.44	23.44	23.44
	0.39	0.39	0.39	0.39	0.39	0.39

a slight overestimation of breathing rate with the proposed method (MOD of −0.03 breaths/min) and small LOAs amplitude (±1.78 breaths/min).

4.3. Influence of Sensor Size Resolution. Table 3 reports the dominant peak frequency of the normalized PSD obtained by each trial at the five different resolution investigated (i.e., PAL, WVGA, VGA, NTSC, SVGA). A qualitative analysis of differences between PSDs obtained from the reference signal and the proposed technique with five-different camera settings can be performed through Figure 6 (in Figure 6 there are also the HD 720 camera setting normalized PSDs).

For each subject, the normalized PSDs are very similar regarding shape and dominant peaks using different resolutions. In all cases, except trials 1 and 12, there is one dominant peak which is distinctly sharper from the surrounding peaks. In trial 1 and 12, the presence of several peaks highlighted several changes in breathing rate during the data collection. An example of these differences in the time domain and frequency domain is reported in Figure 7. The average accuracies for average respiratory rate estimation were 98.7%, 94.9%, 94.9%, 95.6%, 98.7% for PAL, WVGA, VGA, NTSC, SVGA, respectively.

The results obtained in the time domain are reported in Figure 8, in terms of MAE and SE. Despite the fact that MAE and SE tend to increase when the resolution decreases (i.e., NTSC), the values are similar to those obtained with higher resolutions. The maximum MAE values were not higher than 1.45 breaths/min. Additionally, Table 4 reports the Bland-Altman analysis results (MOD ± 1.96 × SD), as well as the slopes β of the linear regression curves and the values of the correlation coefficient.

Figure 9(a) shows Bland-Altman and linear regression plots obtained using the original sensor resolution (HD 720), while Figure 9(b) shows the plots obtained using the NTSC resolution (with broader LOAs and worst value of correlation coefficient).

5. Discussion

Within the wide spectrum of physiological measurements that are useful for clinical assessment, respiratory rate plays a crucial role. Especially, in some conditions it must be monitored continuously, for instance when patients are in clinical setting (i.e., intensive care unit) or both needs the monitoring of physiological data at home (i.e., tele monitoring, tele rehabilitation).

The use of unobtrusive solutions is widespread in respiratory monitoring. Optical technologies can allow nonintrusive and low-cost monitoring of respiratory patterns. Different solutions have been proposed based upon photo-reflective markers and frame subtraction. Although relative new techniques based on the analysis of video collected by digital camera have been demonstrated to be promising in the respiratory monitoring, most of them monitoring only the average respiration rate.

In this paper, we present a single-camera video-based respiratory monitoring system based on the selection of a small skin area near the base of the neck. The proposed method for extracting the respiratory pattern and the corresponding respiratory rate consists of three steps: (*i*) trunk wall motion data collection, (*ii*) ROI selection and intensity change analysis to extract video-based respiratory signal, and (*iii*) analysis in the frequency and time domain to investigate the frequency content of the signal and to extract the breath-by-breath respiratory rate.

Since the proposed method can work with very different built-in RGB cameras (webcams) available in most laptops, we have investigated the influence of sensor resolution (from HD 720 to NTSC) on the respiratory pattern and respiratory rate values extracted from video signal. The method has been tested on 12 participants wearing t-shirt or sweaters during data collection in an unstructured environment. Postprocessed pressure drop signal collected at the nose was used as reference signal in this work. Computed error measurements are at par with those reported in the literature [12, 13].

Results show excellent performances of the method with the use of HD resolution (HD 720) with an accuracy of the method equal to 100% in the estimation of average breathing rate from the frequency-domain analysis. Additionally, PSD spectra demonstrated the similarity of all the breathing pattern collected at the different resolutions when compared to
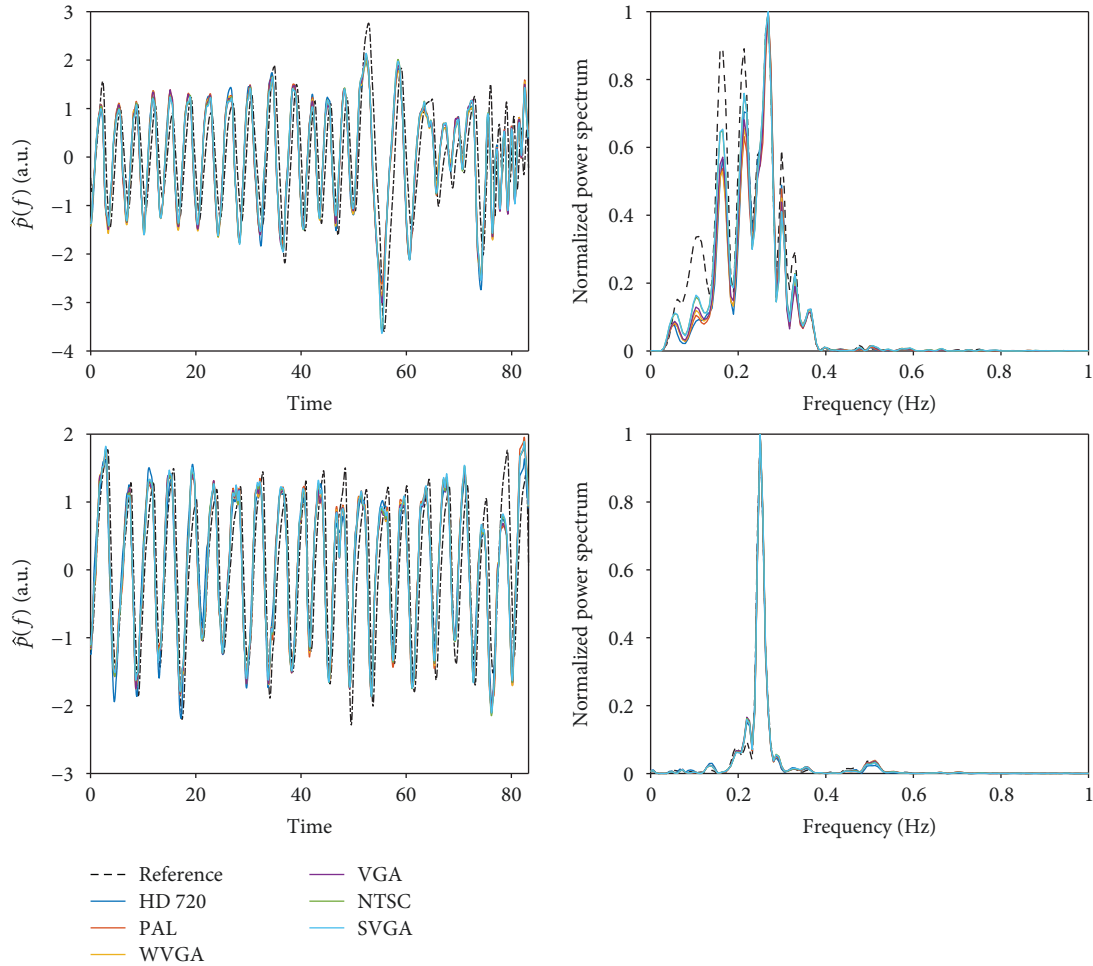
FIGURE 7: Signals obtained by the video recordings at different resolution during 80 s of data collection. In the first subplot, an irregular breathing pattern during self-pace breathing trials, and in the second one, a regular pattern. Differences in frequency domain can be appreciated from the normalized power spectra.

the reference signal frequency content. It results in a lower value of 94.9% of accuracy in the estimate of the average respiratory rate from spectra. Despite the excellent results obtained in the frequency domain, further developments may be devoted to test parametric methods to estimate the PSD, for example, AR methods since the periodicity of the respiratory signal [24].

In the calculation of breath-by-breath respiratory rate, the use of HD 720 camera setting shows the better results in terms of MAE (average value of 0.55 breaths/min) and SE. Additionally, in this case, the method shows a bias of -0.03 ± 1.78 breaths/min in the calculation of breath-by-breath respiratory rate when compared to the reference values. With lower resolution (NTSC), the dispersions of the data are slightly higher (LOAs are wider, ± 2.08 breaths/min), while the MOD value is comparable. These biases are comparable to those obtained [25] on average respiratory rate. It is slightly worse than to those of [26] who used spirometer as reference and pseudo-Wigner-Ville distribution time-frequency representation (with 0.7324 bpm of resolution) for the signal analysis during standing position (-0.02 ± 0.83 breaths/min).

By analyzing more than 200 breaths (from 12 volunteers), sensor resolution seems to influence the accuracy of the proposed method. NTSC resolution (the ROI area is one third of the HD 720 area) shows the worst results, with an accuracy of 95.6% in the estimation of average breathing rate, and a MAE error of 1.45 breaths/min. In the estimation of breath-by-breath parameter, the correlation coefficient is 0.95 with a bias of -0.06 ± 2.08 breaths/min. These values can be compared with respiratory rate bias obtained from wearable sensors like using Doppler radar (via fast Fourier transform) with the use of transmitter and receiver antennas when compared to a respiration strap [27] and to the use of radar which returns using harmonically related filters [28]. Relationship coefficients and bias assessed with the proposed method are in line with those found with wearable systems based on respiratory inductive plethysmographic sensors [29] or based on optical fibers [30], which require expensive systems and contact with the patient for more accurate monitoring. So the most important findings are (i) the proposed measuring system is able to detect small chest wall movements caused by the respiration by calculating the pixel color differences between consecutive frames in order to extract
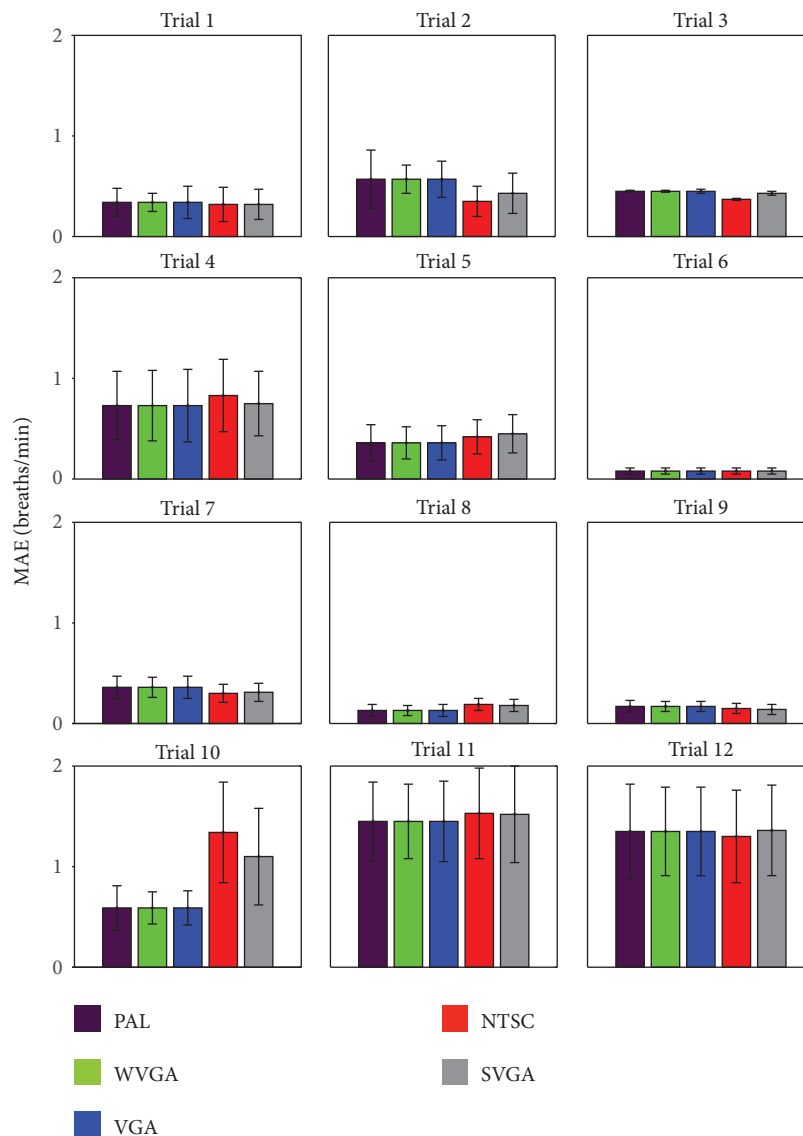
FIGURE 8: Barplot showing mean absolute error (MAE) as coloured bars, and 1.96 × standard error (SE) values as error bar. Each color and bar is related to the 5 investigated sensor settings.

TABLE 4: Results of the Bland-Altman analysis (MOD ± LOAs), the slope (β) of the linear regression curve, and correlation coefficient.

Sensor setting	MOD ± LOAs (breaths/min)	β	Correlation coefficient
PAL	−0.04 ± 1.91	1.001	0.97
WVGA	−0.04 ± 1.76	1.001	0.97
VGA	−0.04 ± 1.84	1.001	0.97
NTSC	−0.06 ± 2.08	1.002	0.95
SVGA	−0.06 ± 2.05	1.002	0.96

respiratory pattern even with low resolution of cameras (i.e., NTSC); (ii) the errors calculated by comparing the average respiratory rate and the breath-by-breath analysis between instruments are acceptable for using the proposed measuring system for accurately monitoring the subject with commercial single camera, even at lower sensor resolutions, in the absence of breathing unrelated movements. Further developments of the proposed system will be the use of an additional technique based on the pixel flow analysis to detect unrelated breathing movements and to reduce their influence in the calculation of breathing pattern. Additionally, we are working on the automatic detection of the chest wall to automatically detect the ROI used for pattern calculation. Further test will be carried out investigating the performances of the proposed method in different scenarios (i.e., clinical setting room and intensive care unit), at different respiratory pattern (i.e., spontaneous breathing, high rate, deep breaths, apnea, and Cheyne-Stokes), and different postures. These upgrades will be useful to test the proposed measuring system for telemonitoring purposes, and in general, for the monitoring of subjects at a distance. Since the encouraging performances in the breath-by-breath monitoring of
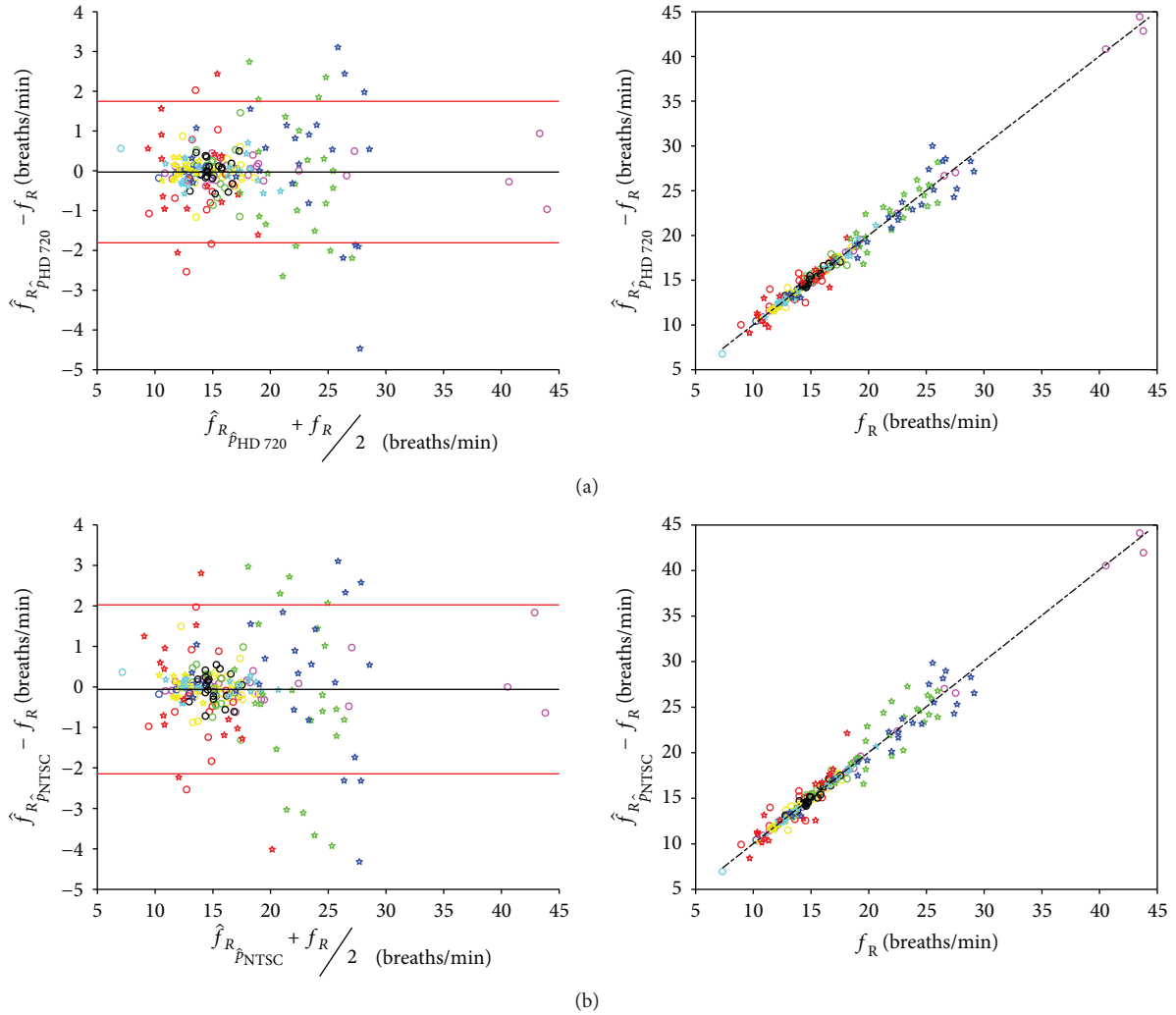
(a)

(b)

FIGURE 9: (a) Bland-Altman graph and correlation analysis using the f_R values obtained by the video signal collected at HD 720 resolution ($\widehat{f}_{R_{\widehat{p}_{HD720}}}$). (b) Bland-Altman graph and correlation analysis using the f_R values obtained by the video signal collected at NTSC resolution ($\widehat{f}_{R_{\widehat{p}_{NTSC}}}$). In the Bland-Altman, the black line is the MOD while the red lines are the upper and lower LOAs, respectively. In the correlation analysis, the dashed line is the linear regression curve. Each volunteer is reported with a different color and/or symbol.

respiratory rate, the present system will be profitably used in the monitoring of subjects in both clinical and home setting as well as spot physiological check by using a commercial insight webcam, commonly used for video call. Other future works include validation with ultrasound images [7] and studying the effect of varying lighting conditions, varying distance and clothing [9].

Data Availability

Reference raw data and videos are available from the corresponding author upon request

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: the neglected vital sign,"

The Medical Journal of Australia, vol. 188, no. 11, pp. 657–659, 2008.

[2] A. Nicolò, C. Massaroni, and L. Passfield, "Respiratory frequency during exercise: the neglected physiological measure," *Frontiers in Physiology*, vol. 8, p. 922, 2017.

[3] A. M. Yañez, D. Guerrero, R. P. de Alejo et al., "Monitoring breathing rate at home allows early identification of COPD exacerbations," *Chest*, vol. 142, no. 6, pp. 1524–1529, 2012.

[4] M. Folke, L. Cernerud, M. Ekström, and B. Hök, "Critical review of non-invasive respiratory monitoring in medical care," *Medical & Biological Engineering & Computing*, vol. 41, no. 4, pp. 377–383, 2003.

[5] A. Aliverti, V. Brusasco, P. T. Macklem, and A. Pedotti, Eds., *Mechanics of Breathing*, Springer Milan, Milano, 2002.

[6] R. Gilbert, J. H. Auchincloss Jr, J. Brodsky, and W. Boden, "Changes in tidal volume, frequency, and ventilation induced by their measurement," *Journal of Applied Physiology*, vol. 33, no. 2, pp. 252–254, 1972.

[7] S. Wiesner and Z. Yaniv, "Monitoring patient respiration using a single optical camera," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2007, pp. 2740–2743, Lyon, France, August 2007.

[8] K. S. Tan, R. Saatchi, H. Elphick, and D. Burke, "Real-time vision based respiration monitoring system," in *2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010)*, pp. 770–774, Newcastle upon Tyne, UK, July 2010.

[9] R. Janssen, W. Wang, A. Moço, and G. De Haan, "Video-based respiration monitoring with automatic region of interest detection," *Physiological Measurement*, vol. 37, no. 1, pp. 100–114, 2015.

[10] K. Y. Lin, D. Y. Chen, and W. J. Tsai, "Image-based motion-tolerant remote respiratory rate evaluation," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 3263–3271, 2016.

[11] Y. W. Bai, W. T. Li, and Y. W. Chen, "Design and implementation of an embedded monitor system for detection of a patient's breath by double webcams," in *2010 IEEE International Workshop on Medical Measurements and Applications*, pp. 171–176, Ottawa, ON, Canada, May 2010.

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

[13] X. Jin, L. Dong, Y. Zhao et al., "Real-time detection of respiration rate with non-contact mode based on low-end imaging equipment," in *Applications of Digital Image Processing XXXVI*, vol. 8856, p. 88561U, San Diego, CA, USA, September 2013.

[14] M. Villarroel, S. Davis, P. Watkinson et al., "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthcare Technology Letters*, vol. 1, no. 3, pp. 87–91, 2014.

[15] D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu, and N. Tao, "Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2760–2767, 2014.

[16] 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 Measurement*, vol. 35, no. 5, pp. 807–831, 2014.

[17] Y. Nam, Y. Kong, B. Reyes, N. Reljin, and K. H. Chon, "Monitoring of heart and breathing rates using dual cameras on a smartphone," *PLoS One*, vol. 11, no. 3, article e0151013, 2016.

[18] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[19] S. L. Marple Jr, "Digital spectral analysis with applications," *The Journal of the Acoustical Society of America*, vol. 86, p. 492, 1987.

[20] G. Rosengarten, B. Ahmed, and C. Marselli, "A novel method to measure respiratory airflow rate based on nasal pressure," *IFAC Proceedings Volumes*, vol. 36, no. 15, pp. 415–419, 2003.

[21] D. L. Presti, C. Massaroni, D. Formica et al., "Smart textile based on 12 fiber Bragg gratings array for vital signs monitoring," *IEEE Sensors Journal*, vol. 17, no. 18, pp. 6037–6043, 2017.

[22] F. Taffoni, D. Rivera, A. La Camera, A. Nicolò, J. R. Velasco, and C. Massaroni, "A wearable system for real-time continuous monitoring of physical activity," *Journal of Healthcare Engineering*, vol. 2018, 16 pages, 2018.

[23] D. G. Altman and J. M. Bland, "Measurement in medicine : the analysis of method comparison studies," *The Statistician*, vol. 32, no. 3, pp. 307–317, 1983.

[24] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, Pearson/Prentice Hall, 2005.

[25] A. Al-Naji and J. Chahl, "Simultaneous tracking of cardiorespiratory signals for multiple persons using a machine vision system with noise artifact removal," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, pp. 1–10, 2017.

[26] B. A. Reyes, N. Reljin, Y. Kong, Y. Nam, and K. H. Chon, "Tidal volume and instantaneous respiration rate estimation using a volumetric surrogate signal acquired via a smartphone camera," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 764–777, 2017.

[27] Y. S. Lee, P. N. Pathirana, R. J. Evans, and C. L. Steinfort, "Non-contact detection and analysis of respiratory function using microwave Doppler radar," *Journal of Sensors*, vol. 2015, Article ID 548136, 13 pages, 2015.

[28] M. Mabrouk, S. Rajan, M. Bolic, M. Forouzanfar, H. R. Dajani, and I. Batkin, "Human breathing rate estimation from radar returns using harmonically related filters," *Journal of Sensors*, vol. 2016, Article ID 9891852, 7 pages, 2016.

[29] P. Grossman, M. Spoerle, and F. H. Wilhelm, "Reliability of respiratory tidal volume estimation by means of ambulatory inductive plethysmography," *Biomedical Sciences Instrumentation*, vol. 42, pp. 193–198, 2006.

[30] C. Massaroni, C. Venanzi, A. P. Silvatti et al., "Smart textile for respiratory monitoring and thoraco-abdominal motion pattern evaluation," *Journal of Biophotonics*, vol. 11, no. 5, article e201700263, 2018.