

Comparison of video-based methods for respiration rhythm measurement

M. Mateu-Mateus^{a,*}, F. Guede-Fernández^a, V. Ferrer-Mileo^a, M.A. García-González^a, J. Ramos-Castro^a,
M. Fernández-Chimeno^a

^a*Department of Electronic Engineering, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain*

Abstract

The aim of this work is to characterize the differences in the respiratory rhythm obtained through three video based methods by comparing the obtained respiratory signals with the one obtained with the gold standard method in adult population. The analysed methods are an RGB camera, a depth camera and a thermal camera while the gold standard is an inductive thorax plethysmography system (RespiBand system from BioSignals Plux).

21 healthy subjects were measured, performing 4 tests for each subject. The respiratory rhythm and its variability was obtained from the four respiratory signals (3 video methods and gold standard). The signal acquisition was performed with custom and proprietary algorithms. To characterize the respiratory rhythm and its variability obtained with the different video sources and gold standard, the instantaneous frequency, Bland-Altman plots and standard deviation of the error between video methods and the gold standard have been computed.

The depth and RGB camera present high agreement with no statistical differences between them, with errors when comparing with the gold standard in the range of mHz. The thermal camera performs poorly if compared with the two other methods, nevertheless it cannot be discarded directly because some errors produced by the subjects head movement could not be corrected.

From these results we conclude that the depth and RGB camera, and their respective acquisition algorithms, can be used in controlled conditions to measure respiration rhythm and its variability. The thermal camera on the other hand, although it can not be discarded directly, performed poorly if compared with the other two methods. Further studies are needed to confirm that these methods can be used in real life conditions.

Keywords: Respiration Rhythm, Respiration Variability, Video-Based, Camera, Thermal, Depth.

1. Introduction

Breathing is a fundamental physiological function of the body; the analysis of breath signals has proven reliable in the assessment of many respiratory diseases, but it also has been used to detect different states of mind such as: stress, fear... or even to assess drowsiness [1] and sleep patterns [2]. The detection and analysis of such signals has been a topic in the recent years, and many methods to acquire these signals and extract its embedded information have been proposed [2, 3, 4, 5]. Traditional methods, such the ones based on plethysmography, acquire signals in the thorax region and although they are not invasive, they require the sensing hardware to be worn by the user in order to work [4, 6]. Different approaches to develop a method where the breathing signal could be acquired in a contactless manner have been proposed, examples of these methods compress: Doppler radar systems [7], ultrasonic proximity sensors [8] or even thermistors placed under the nose [9].

*Corresponding author

Email address: marc.mateu.mateus@upc.edu (M. Mateu-Mateus)

More recently, new methods based in video analysis have been presented. These systems use video devices such as: RGB video cameras [10, 11], depth cameras [12, 13, 14] or even thermal cameras [15, 16] to acquire the breathing signals. Although all these methods have been studied there are no comparative analysis about its performance versus a gold standard, or even if they are suitable to work in real life environments. One possible scenario were the methods aforementioned could prove a significant change in the present state of the art, would be detecting drowsiness in vehicles [1, 17] in a contact-less manner. To achieve this objective, methods and algorithms that are capable of extracting the variability of the respiratory rate are needed.

The aim of this work is to compare three different systems based on video analysis with custom and proprietary detection algorithms versus a plethysmography system (Plux RespiBandTM used as a gold standard), in order to asses and characterize the error in respiratory rhythm and variability of these systems and the feasibility to be used in real time with real life conditions. A car chair has been used with its homologated seatbelt to emulate a car cockpit, as the final objective is to detect respiration in drivers. The video sources being compared are: an RGB camera, a depth camera and a thermal camera. Each one of the video sources has been acquired simultaneously with the same computer and synchronized among them.

2. Materials and Methods

2.1. Setup

The gold standard used is the RespiBand system from BioSignalsPlux, this system acquires the volumetric changes of the thorax by the means of an inductive band, sampling at 40 Hz with 12-bit ADC resolution and a band pass filter between 0.058 Hz and 0.9 Hz.

The three cameras used to acquire the breathing signals are the ones described in Table 1.

The Logitech camera was configured to record at full HD (1080). According to the manufacturer specifications, this camera allows to capture at full HD 30 frames per second (fps). Nevertheless, preliminary tests showed a systematic drop in frame-rate when acquired at 30 fps with exposure and white balance enabled, which in return produced an irregularly sampled signal. For this reason, this camera was configured to record at 15 fps with the white balance disabled and blocked exposure.

The Kinect sensor includes an RGB camera and a Depth acquisition system. This last one returns a point cloud matrix comprised of the distances between the object in front of the camera, to the camera itself [18]. The Kinect was configured to acquire distance at a short range. Both feeds from the Kinect camera were acquired (RGB and Depth) but due to the low resolution of the RGB feed this one was not analysed.

The Thermal camera provides a grey-scale image that encodes the temperature range (being white the hottest and black the coldest in our case compressed between 28 °C to 38 °C) of the object at which the measurements are taken. In order to acquire this image, an analogue video recorder (connected to the composite video output of the camera) was used due to the impossibility of acquiring the raw frames using the USB interface of the camera.



Figure 1: 1a is a representation of the camera disposition on the setup, 1b shows the actual measurement Setup. (Fig. 1a icons made by Freepik from www.flaticon.com).

The cameras were placed as it is shown in Figure 1a. The features of interest that were targeted with each one of the video sources were the following: for the RGB camera the position of a given object will

be tracked inside the camera’s field of view (FOV), for the Depth camera the depth changes of a certain region within the camera’s FOV will be acquired, finally for the thermal camera the temperature changes of a region inside the camera’s FOV will be acquired.

In order to acquire the respiration signal with the RGB camera, three distinctive patterns were placed on top of the seatbelt for its posterior detection as seen in detail in Figure 2d.

The light source used for all the tests was an LED bulb from the Verbatim manufacturer with reference 52130, warm white color (CCT: 3000 K), 6.5 W, luminous flux of 480 lm, a beam angle of 130° and a distance between the light source and the subject of 70 cm. The room temperature was $24\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$. Prior to any test, a calibration of the RGB camera exposure was performed, as well as a warm-up period for the thermal camera and its posterior thermal bias correction, this steps were performed to avoid burned or over exposed images in both RGB and Thermal cameras.

The three video systems and the gold standard were synchronized on acquisition using the timestamp of the PC.



Figure 2: From left to right: Depth camera with false colour, Thermal camera, RGB Camera and detail of the pattern placed on the seatbelt. In the RGB picture the RespiBand system can be appreciated.

All the extraction algorithms were designed to run in real time, but in order to make the study more reliable and repeatable the video sources were recorded using a custom software based on the OpenCV framework, to save each video feed as separate video file for further analysis.

Table 1: Camera Specifications

Type	Model	Resolution	FPS	FOV (H x V)	Serial Number
RGB	Logitech C920	1920x1080	15	70.42°x43.30°	1611LZ0MXLU8
Depth	Kinect V1.	640x480	10	58.5°x46.6°	00353322342
Thermal	NEC InfReC G100	320x240	25	32°x24°	1041391 (apr/2011)

2.2. Measurement Protocol

Twenty-one healthy subjects with ages between 20 years and 54 years (Mean: 28.6 years, SD: 6.8), 10 females, 11 males, with height between 160 cm and 190 cm (Mean: 170.8 cm, SD: 7.4 cm) and chest perimeter between 74 cm and 110 cm (Mean: 88.4 cm, SD: 10.2 cm) volunteered for the study. Each subject gave their oral informed consent to freely participate in this study, and this study was performed in accordance with the principles of the Declaration of Helsinki [19]. All the measurements performed comply with the regulations of the Universitat Politècnica de Catalunya (UPC).

Prior to the measurements, the subjects were asked to put on the RespiBand system, to seat on the setup and fasten the seatbelt placed at the setup.

The subject was asked to perform four tests. The aim of these were to compare the performance of the studied methods at different breathing frequencies and with different constraints, such as breathing freely or reading out loud. In two of the four tests, the subject was asked to breath at a constant frequency (0.1 Hz for the first test and 0.3 Hz for the second test). To do this, a custom visual aid was developed for the

subject to follow, which consisted on a moving bar with 1/3 of the given period for inhaling and 2/3 for exhaling. In the third test the subject was asked to breathe freely, finally, in the fourth test the subject was asked to read out loud a text. Each subject was asked to remain as still as possible during the whole measurement protocol, all four test had a duration of 3 minutes each with a 30 seconds pause between tests were the subject was allowed to breathe freely. In total, the whole test duration for each subject was 15 minutes.

2.3. Signal Extraction

Prior to the comparison between the respiration signals from the video sources and the gold standard, the raw respiration signal had to be extracted from each video source. For this reason custom and proprietary algorithms were developed in order to acquire the respiration signal from each source using multiple approaches such as: pattern recognition or even performing an active tracking of the nostrils region of the subject [5].

The number of frames extracted from each method to conform the respiratory signal was, accordingly with the duration of each experiment, 180 s multiplied by Fs , where Fs was the sampling frequency for each method. For the RGB camera 2700 frames were extracted, for the Depth camera 1800 frames were extracted and for the Thermal camera 4500 frames were extracted.

2.3.1. RGB Video

To obtain the raw respiratory signal from the RGB video, a proprietary algorithm [5] has been used to detect the patterns placed on top of the seatbelt, as shown in Figure 2d. The pattern placed in the middle was the only one used to extract the raw respiration signal, this pattern was the one that yielded better results in all the performed tests previous to this analysis. After the detection stage, an optical flow tracking, by the means of the Kanade-Lucas (KL) algorithm [20, 21], was performed for each frame using the features extracted from the pattern. Once the new position of the features was extracted for a given frame, the mean of all the (x, y) coordinates were computed, then a ℓ_2 -norm of the averaged coordinates was calculated in order to obtain the fiducial point for the frame.

$$\begin{aligned} x_{avg} &= \frac{1}{N} \sum_{i=1}^N x_i \\ y_{avg} &= \frac{1}{N} \sum_{i=1}^N y_i \\ Raw_i &= \sqrt{x_{avg}^2 + y_{avg}^2} \end{aligned} \quad (1)$$

Where N is the maximum number of features, x_{avg} is the averaged x component, y_{avg} is the averaged y component and Raw_i is the i_{th} value of the computed respiration signal.

These averaged distances are proportional to the chest movements of the subject, hence to the respiration[22]. The temporal series obtained by the concatenation of all the fiducial points and its timestamps results in the raw respiration signal of the subject.

2.3.2. Depth Point cloud

In order to extract the raw respiration signal from the depth point cloud, the maximum and minimum distance between the subject and the camera has been identified. From this region (comprised between the raw values 500 to 900 mm) the average distance between the subject and the camera was computed [5].

This averaged distance is modulated by the movement of the thorax of the subject, hence proportional to the respiration[23]. A temporal series obtained by the concatenation of the averaged depth distances and their timestamps results in the raw respiration signal.

2.3.3. Thermal Video

The extraction of the respiration signal from the thermal video was performed in three different steps: the first one consisted on training a HAAR cascade classifier algorithm [24, 25] to detect the nose region automatically. In order to train the classifier, every 18 seconds one frame was extracted from the constant frequency (0.1 Hz) respiration test, 10 frames were extracted from each subject. After that, the portion of the nose was selected from each frame and cropped in order to train the cascade classifier.

The second step was performed once the classifier was trained. A generic region of the video that contained the nose for each subject and test was manually selected, then the classifier was used to detect the exact position of the nose region. After that, the features (using the "goodFeaturesToTrack" function of the OpenCV library [26]) of the resultant region of interest (ROI) were extracted, these features were later used in combination with the KL algorithm to track the movements of the nose.

Finally, the raw respiration value was extracted by averaging the pixels inside the ROI[9]. The thermal raw respiration signal was then obtained by appending all these values and its timestamps. These averaged pixels represent the average temperature of the nostrils produced by the air flow due to the respiration of the subject.

2.4. Signal Processing

Once the raw respiration signal was extracted from each video source accompanied by the gold standard, all the signals were normalized in order to compare them with each other. All the signal processing steps were performed using the Matlab software.

The normalization process was performed using the following steps.

1. The signals were interpolated at 40 Hz using a cubic spline in order to normalize the sample frequencies. This step was necessary for the video based signals in order to obtain a homogeneous sampled signal and to correct the possible frame loss due to the delays produced by the different communication buses that control the cameras (USB 3.0). This step was also crucial because not all the sources were recorded at the same frame-rate.
2. A bandpass filter between 0.05 Hz and 1 Hz was applied to eliminate undesired components and to remove the different offsets. The employed filter was a bidirectional Butterworth of 2nd order.
3. A moving median filter was applied to the signal to remove the peaks, induced in the previous step, by transitory periods produced by the rapid movements of the subject. The length of the filter was set to three seconds (with a sample length of three times the sample frequency), this length is enough to smooth the signal and shorter than an average breath cycle.
4. Finally, in order to compress each signal to a known range a non-linear function was applied as described in the following equation.

$$S_n[n] = \arctan \frac{S[n]}{\sqrt{\frac{\sum_{i=1}^N (S[i] - \bar{S})^2}{N-1}} * \sqrt{2}} \quad (2)$$

Where $S_n[n]$ is the discrete-time normalized respiratory signal, $S[n]$ is the original signal and \bar{S} is the mean of $S[n]$.

This final step transforms each signal to a normalized space, which is crucial in order to compare all the different methods with the gold standard. Figure 3 shows an example of the difference between the raw signal and the normalized signal.

After the normalization step was performed, and in order to characterize the error between methods, the instantaneous frequency (IF) for each one of the normalized signals was extracted. The reason behind this approach was that this method is used in respiration extraction from ECG [27, 28, 29, 30] and baseline wander removal [31, 32] and could be easily adapted to analyse respiration signals.

The IF has been defined as follows: the first derivative of the instantaneous phase of the respiration signal. This step is possible if it is taken into account that the respiration signal can be approximated as

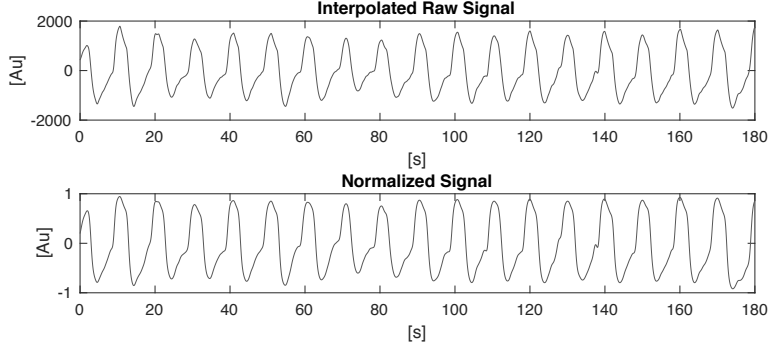


Figure 3: Comparison between raw signal and normalized signal, signals from Bioplux sensor 0.1 Hz test.

a periodic signal, moreover a narrow-band pass-band filter is applied to the signals eliminating the higher frequency harmonics, thus allowing to express these signals by the means of the analytic signal.

The extraction of the IF was performed using the following steps:

The Hilbert transform was applied to the normalized respiration signals.

$$H_s[n] = \text{Hilbert}(S_n[n]) \quad (3)$$

Where $S_n[n]$ is the discrete-time normalized respiratory signal. Then the instantaneous frequency is estimated from the argument of $H_s[n]$:

$$\begin{aligned} \phi[n] &= \arg H_s[n] \\ \phi_n[n] &= \begin{cases} \phi[n], & \text{if } |\phi[n]| < \pi \\ \phi[n] \pm 2\pi, & \text{if } |\phi[n]| \geq \pi \end{cases} \\ \omega[n] &= \phi_n[n] - \phi_n[n-1] \\ IF[n] &= \omega[n] \frac{Fs}{2\pi} \end{aligned} \quad (4)$$

where $\omega[n]$ is the discrete-time instantaneous frequency in radians, Fs is the sampling frequency and $IF[n]$ is the discrete-time instantaneous frequency in Hz.

Finally, a Hodrick-Prescott [33] filter was applied to the IF signals with a smoothing factor of $2 * 10^6$ in order to smooth the resulting envelope of the IF. Afterwards, the respiratory IF signals have been aligned by maximizing the Fisher's intraclass correlation iterating around one respiration period from the origin. This step ensures that all the signals have the same alignment, then the aligned signals have been cropped to the same length.

2.5. Error Characterisation

In order to characterize the error between the studied methods and the gold standard, the standard deviation of the error (SDE) has been used. The SDE was obtained by computing the difference, for each sample, between the IF for a given method and the IF of the gold standard, thus conforming the error series. After the error series was generated the standard deviation was computed for each error series.

$$\begin{aligned}
e[i] &= S[i] - G[i] \\
\bar{e} &= \frac{1}{N} \sum_{i=1}^N e[i] \\
SDE &= \sqrt{\frac{\sum_{i=1}^N (e[i] - \bar{e})^2}{N - 1}}
\end{aligned} \tag{5}$$

Where $S[i]$ is the instantaneous frequency of the studied method, $G[i]$ is the instantaneous frequency of the gold standard and $e[i]$ is the error series.

3. Results

3.1. Signals

Figure 4a illustrates an example of the processed respiration signals obtained by the different methods for a given subject during the breathing test at 0.1 Hz.

Figure 4b on the other hand, contains the computed instantaneous frequency (IF) for the respiration signals shown in Figure 4a for each one of the studied methods. These signals represent the instantaneous frequency changes as a function of time. The mean of these signals is the fundamental frequency at which the person is breathing, in this case, given that the subject was instructed to breath at 0.1 Hz, the mean of the IF is approximately 0.1 Hz.

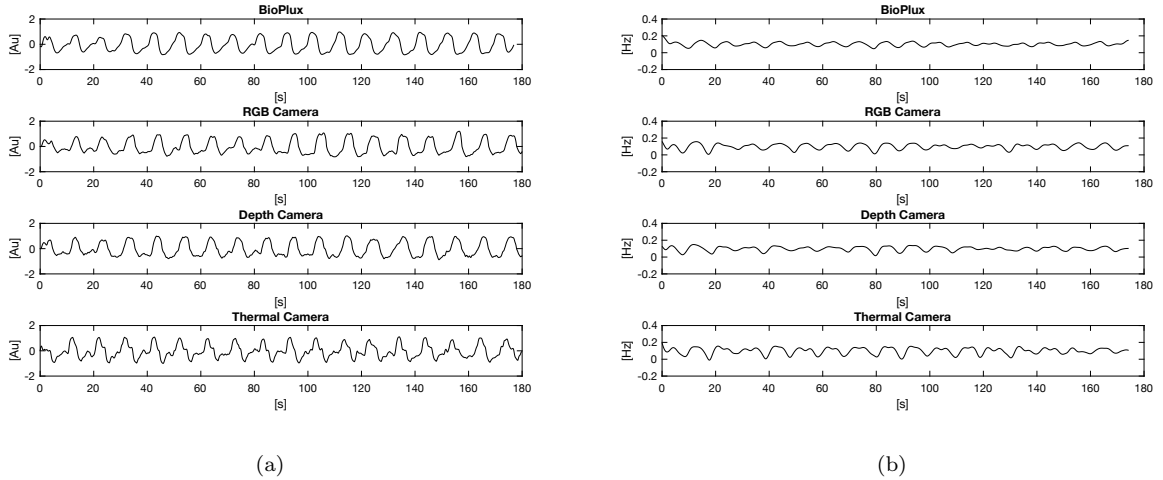


Figure 4: 4a From top to bottom the processed respiration signals for each source: BioPlux, RGB Camera, Depth Camera and Thermal Camera, 4b From top to bottom the instantaneous frequency signals for each source: BioPlux, RGB Camera, Depth Camera and Thermal Camera.

3.2. Performance

In order to clarify the following figures and tables, a new name agreement has been adopted which renames the BioPlux signals as "Plux", RGB Camera signals as "Pattern", the Depth Camera signals as "Depth" and the Thermal Camera signals as "Thermal". Also a new naming convention has been applied for the different breathing frequencies as: "0.1 Hz", "0.3 Hz", "Free" instead of Free breathing and finally Reading.

In order to establish a base frequency, the gold standard has been used as a reference. It can be appreciated in Table 2 that for the 0.1 Hz test the gold standard presents a mean IF of 0.1 Hz and 0.022 Hz of standard deviation and for the 0.3 Hz test it presents a mean of 0.293 Hz and 0.017 Hz of standard deviation, these two tests have a small standard deviation due to the subject being instructed to breath approximately at the target frequency for each test.

On the other hand, for the Free test the mean IF is 0.242 with a standard deviation of 0.053; this increase in the standard deviation can be explained if it is taken into account that every subject has its own natural breathing frequency, which is determined by the unique physiology of the subject. The same can be applied for the Reading test with a mean frequency of 0.234 Hz and 0.09 Hz of standard deviation.

In order to obtain the agreement of the different methods with the gold standard the Bland-Altman plots have been used as well as two different t-test analysis, one to compare the differences of the Bland-Altman plots to the null bias hypothesis and the other one to compare if the differences between methods are statistically independent. All the statistical tests were performed using the R software.

Table 2: Mean \pm Standard deviation for all the tests and methods.

	Plux [Hz]	Depth [Hz]	Pattern [Hz]	Thermal [Hz]
0.1 Hz	0.100 \pm 0.022	0.098 \pm 0.029	0.098 \pm 0.036	0.169 \pm 0.093
0.3 Hz	0.293 \pm 0.017	0.289 \pm 0.018	0.289 \pm 0.018	0.267 \pm 0.072
Free	0.242 \pm 0.053	0.245 \pm 0.065	0.236 \pm 0.059	0.227 \pm 0.094
Reading	0.234 \pm 0.09	0.233 \pm 0.095	0.237 \pm 0.098	0.234 \pm 0.125

For the Bland-Altman representation, some outliers have been removed from each one of the tests with the following criteria: if the difference between the mean IF of the gold standard and the mean IF of the given method exceeded two times the standard deviation of the whole test, then the given subject was tagged as an outlier and removed from the pool. For the 0.1 Hz test four outliers were removed, for the 0.3 Hz test three outliers were removed, for the Free test five outliers were removed and finally for the Reading test three outliers were removed from a total of 21 measurements for each test. This criterion was applied in order to compensate for the measurement errors produced by the consecutive loss of frames during the acquisition stage, and for the excessive movement of certain subjects that produced errors in the estimation of the instantaneous frequency. Another important factor in the removal of these outliers was that they had a negative impact in the statistical analysis of the Bland-Altman differences. Random removal of subjects for each test did not alter significantly the statistical results.

Figure 5 contains 12 Band-Altman plots comparing the mean IF of the gold standard versus the mean IF for each one of the studied methods. The columns represent the different breathing frequencies (breathing at 0.1 Hz, breathing at 0.3 Hz, Free breathing and Reading) and the rows represent the method being compared to the gold standard. In Figure 6 the magnitudes being compared are the standard deviation of the IF of the gold standard versus the standard deviation of the IF for each one of the studied methods, in this case the same representation as in Figure 5 applies.

In Table 3 and Table 4 the mean and standard deviation for each one of the Bland-Altman differences can be seen (Figure 5 and Figure 6 respectively) as well as the significance for each tests.

Given the results in Figure 5, the most relevant are the ones for the Thermal camera, with special focus at the 0.1 Hz and 0.3 Hz test which clearly presents a gain error.

In Table 3 the results for the Bland-Altman differences in Figure 5 and its respective mean and standard deviation are the following: for the Depth and Pattern methods at the 0.1 Hz and 0.3 Hz tests, both present very significant differences ($p < 0.001$) if compared to the null bias. Both bias and standard deviation are very low as well and practically negligible for both methods as the differences can be due to statistical fluctuation.

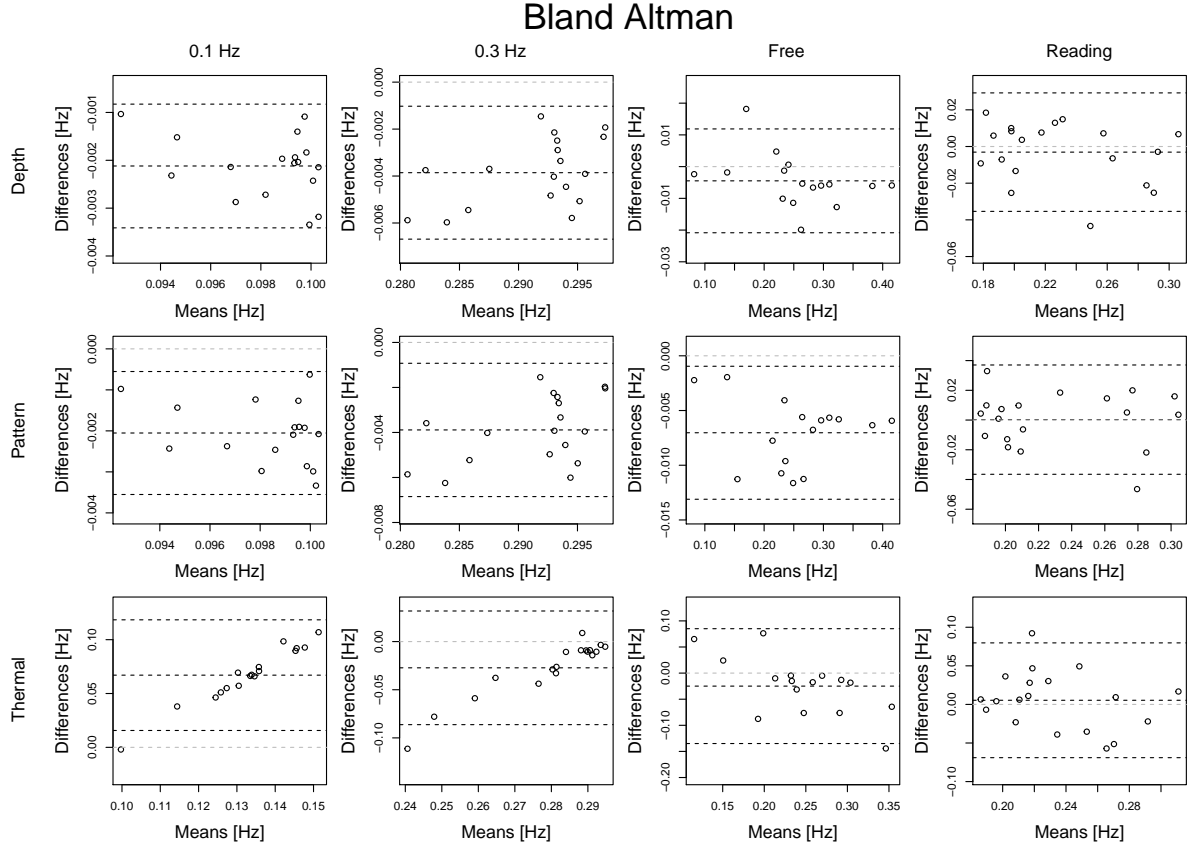


Figure 5: Bland-Altman plots for each one of the methods and frequencies of the mean IF.

The central dashed line for each plot represents the mean of the points, the two upper and lower dashed line represent the confidence level given by the mean plus/minus the 95% of the standard deviation of the points, finally the dashed grey line represents the zero-mean line.

For the Free test, the Depth method presents a $p < 0.05$ for the null bias test and an increased standard deviation if compared to the previous tests, for the Reading test the standard deviation almost doubles the one in the Free test but it can also be appreciated that there are no significant differences in the null bias t-Test.

Regarding the Pattern method, the Free test presents a $p < 0.001$ for the null bias t-test, and an increased bias and standard deviation if compared to the previous tests. For the Reading test it can be appreciated an increased standard deviation up to six times greater than the Free test, but with no significant differences in the null bias t-Test.

The Depth and Pattern methods do not present significant differences if compared between each other.

For the Thermal method, the bias at 0.1 Hz, 0.3 Hz and Free test exceeds all the bias of the other two methods given the same tests. The 0.1 Hz test presents a $p < 0.001$ for the null bias t-test, with a bias and standard deviation of one order of magnitude greater than the other two methods for the same test. The 0.3 Hz test presents a $p < 0.05$ for the null bias t-test, with the same characteristics as the 0.1 Hz test. Both tests present very significant (0.1 Hz test) and significant differences (0.3 Hz test) if compared to the other two methods. Finally, the Free and Reading tests both presents no significant differences in the null bias t-test and neither in the comparison between methods, with both bias and standard deviation of the same order of magnitude as the previous tests.

The Bland-Altman representation of the standard deviation of the studied methods versus the gold

Table 3: Mean \pm SD of the differences in the Bland-Altman plot in Figure 5, for the mean IF.

	Depth [mHz]	Pattern [mHz]	Thermal [mHz]	Significance
0.1 Hz	-2.119 ± 0.660	-2.052 ± 0.765	67.045 ± 26.213	**; ##; $\vee\vee$; <i>NS</i> ; ††; ∞
0.3 Hz	-3.857 ± 1.444	-3.888 ± 1.511	-27.247 ± 30.109	**; ##; \vee ; <i>NS</i> ; ††; \diamond
Free	-4.483 ± 8.348	-7.032 ± 3.100	-24.933 ± 56.106	*; ##; <i>NS</i> ; <i>NS</i> ; <i>NS</i> ; <i>NS</i>
Reading	-3.067 ± 16.482	0.275 ± 18.761	5.366 ± 37.945	<i>NS</i> ; <i>NS</i> ; <i>NS</i> ; <i>NS</i> ; <i>NS</i> ; <i>NS</i>

The results for the null bias t-test are represented as follows, "*" for the Depth, "#" for the Pattern and finally " \vee " for the Thermal Bland-Altman differences. The results for the t-test between methods are represented as follows, "†" for the Depth vs Pattern, "‡" for the Depth vs Thermal and finally " \diamond " for the Pattern vs Thermal. Regarding the significance: "*NS*" for $p > 0.05$, "*" for $p < 0.05$ and "**" for $p < 0.001$.

standard, has special interest for us as it shows the relationship between the frequency range of the gold standard versus the frequency range of the studied methods, this is important as we are not only interested in characterising the studied methods for the mean respiration frequency but also in its variability.

Considering the results in Figure 6, the most relevant are the ones for the Thermal camera, with special focus at the 0.1 Hz and 0.3 Hz test which clearly presents a gain error. For the Free test it can also be appreciated a gain error but not as clearly seen as in the previous tests.

In Table 4 the results for the t-test analysis of the Bland-Altman differences in Figure 6 are the following: for the Depth method, there are two relevant results for the null bias t-test, the 0.1 Hz test which shows a $p < 0.001$ and the Free test which shows a $p < 0.05$, for the other tests all results are non-significant. For the Pattern method, there are two relevant results for the null bias t-test, the one for the 0.1 Hz test that shows a $p < 0.001$ and the one for the Free test which shows a $p < 0.05$, the remaining tests are all non-significant.

Table 4: Mean \pm SD of the differences for the Bland-Altman plot in Figure 6, for the standard deviation of the IF.

	Depth [mHz]	Pattern [mHz]	Thermal [mHz]	Significance
0.1 Hz	4.81 ± 3.413	12.432 ± 6.469	71.308 ± 21.521	**; ##; $\vee\vee$; ††; ‡‡; ∞
0.3 Hz	-0.46 ± 0.936	-0.115 ± 2.329	54.887 ± 30.135	<i>NS</i> ; <i>NS</i> ; $\vee\vee$; <i>NS</i> ; ‡‡; ∞
Free	7.562 ± 9.997	4.612 ± 6.783	50.753 ± 30.282	*; #; $\vee\vee$; <i>NS</i> ; ‡‡; ∞
Reading	3.554 ± 10.488	7.314 ± 15.262	37.079 ± 24.641	<i>NS</i> ; <i>NS</i> ; $\vee\vee$; <i>NS</i> ; ‡‡; ∞

The results for the null bias t-test are represented as follows, "*" for the Depth, "#" for the Pattern and finally " \vee " for the Thermal Bland-Altman differences. The results for the t-test between methods are represented as follows, "†" for the Depth vs Pattern, "‡" for the Depth vs Thermal and finally " \diamond " for the Pattern vs Thermal. Regarding the significance: "*NS*" for $p > 0.05$, "*" for $p < 0.05$ and "**" for $p < 0.001$.

Regarding the results of the t-test between methods only for the 0.1 Hz test there is significant differences between the Depth method and the Pattern method, for the rest of the tests all results are non-significant.

Finally, the Thermal method show both bias and standard deviation one order of magnitude greater than the ones presented for the other two methods. All tests all tests show a $p < 0.001$ for the null bias t-test. And for the comparison between methods, all results yield a $p < 0.001$ between the Thermal method and the other two methods.

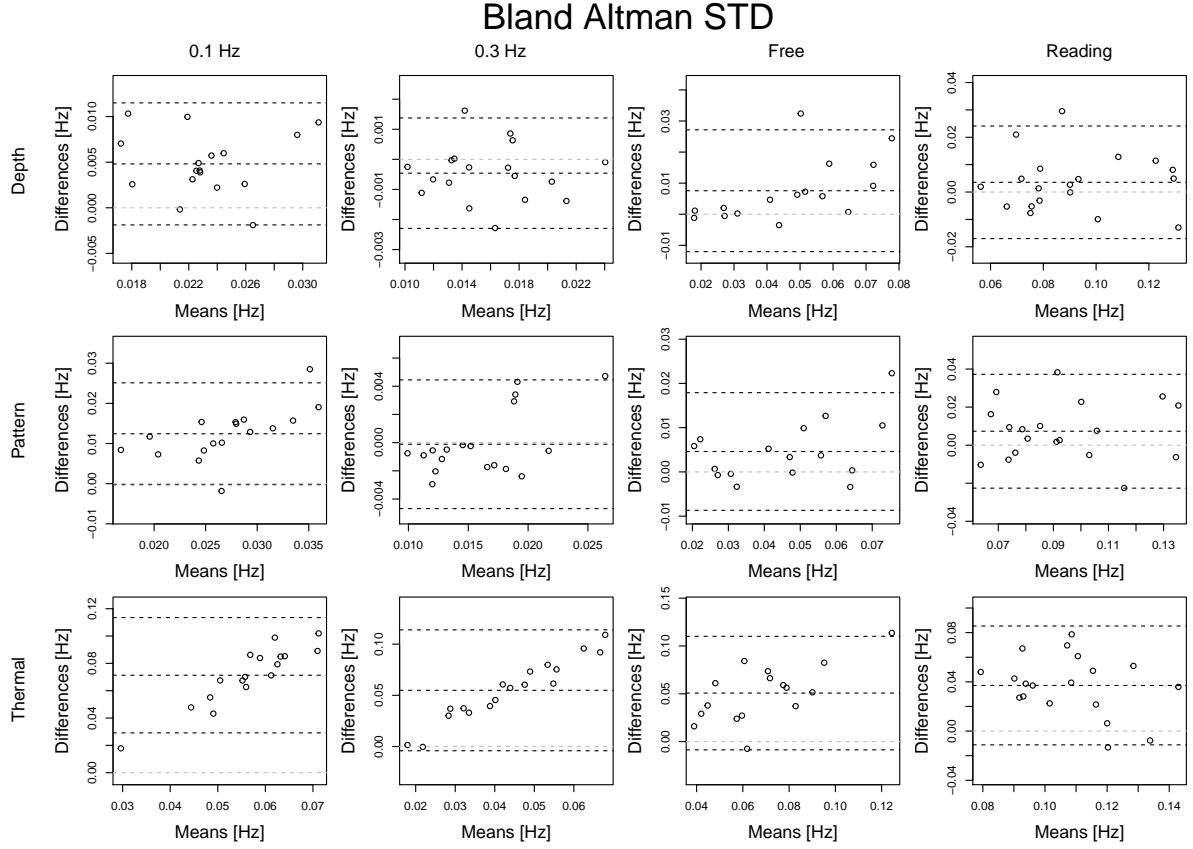


Figure 6: Bland-Altman plots for each one of the methods and frequencies of the standard deviation of the IF.

The central dashed line for each plot represents the mean of the points, the two upper and lower dashed line represent the confidence level given by the mean plus/minus the 95% of the standard deviation of the points, finally the dashed grey line represents the zero-mean line.

3.3. Error Characterisation

Given the SDE for all methods and subjects, a Shapiro-Wilk test [34] was applied for each SDE, grouped by methods, in order to determine if the SDEs could be approximated by a normal distribution. Two of the three error groups (Plux vs Depth and Plux vs Pattern) were not normal.

For this reason, non-parametric statistical tests have been employed. First a Friedman test [35] was applied in order to compare the three methods for all the given frequencies. A Wilcoxon-Nemenyi-McDonald-Thompson post-hoc test [35] was applied in order to confirm the relationship between each one of the error groups.

The Table 5 shows the results of the post-hoc tests as well as the median and interquartile range for each one of the four respiration frequencies and for each one of the three error groups.

4. Discussion and Conclusions

Taking into account the results of the RGB camera and the Depth camera by the Bland-Altman analysis [36] on Figure 5 and Table 3; for the first two tests (0.1 Hz and 0.3 Hz) it can be appreciated that the limits of agreement are very close to the mean difference, this indicates a high agreement between the studied methods and the gold standard. The results shown are practically identical with non-significant differences

Table 5: Median and Interquartile Range [25% – 75%] of the standard deviation of the error for each one of the methods and breathing frequencies.

	Plux Vs Depth [Hz]	Plux vs Pattern [Hz]	Plux vs Thermal [Hz]	Significance
0.1 Hz	0.018 [0.016 - 0.019]	0.024 [0.019 - 0.027]	0.096 [0.084 - 0.102]	<i>NS</i> ; ##; √√
0.3 Hz	0.011 [0.008 - 0.014]	0.013 [0.009 - 0.018]	0.074 [0.049 - 0.092]	<i>NS</i> ; ##; √√
Free	0.038 [0.014 - 0.052]	0.021 [0.015 - 0.034]	0.106 [0.073 - 0.114]	<i>NS</i> ; #; √√
Reading	0.091 [0.073 - 0.117]	0.096 [0.079 - 0.126]	0.14 [0.131 - 0.155]	<i>NS</i> ; ##; √

The given results for the different post-hoc test are represented as follows, "*" for the interaction between Plux vs Depth and Plux vs Pattern, "#" for the interaction between Plux vs Depth and Plux vs Pattern and finally "√" for the interaction between Plux vs Pattern and Plux vs Thermal. Regarding the significance: "*NS*" for $p > 0.05$, "√" for $p < 0.05$ and "√√" for $p < 0.001$.

between each other; this could indicate a high agreement between both methods and the gold standard. For both tests and methods the bias is practically negligible. Although the t-test results indicate significant differences between the mean differences and the null bias ($p < 0.001$), this can be explained if it is taken into account that the standard deviation of the differences is close to zero.

In the analysis of the Free breathing test, in both methods the bias and limits of agreement have increased if compared to the previous tests, this increase could be caused by the non-periodic nature of the respiration signal during the test and by the unique morphology of every subject, which determines their natural breathing frequency. If both methods are compared at the Free test, it can be appreciated that the RGB camera presents more bias than the Depth camera but a lower standard deviation, but if it is taken into account the t-test between each other, which yields non-significant differences, a possible explanation for these differences could be statistical fluctuation. As far as the null bias t-test results go, the same explanation as the 0.1 Hz and 0.3 Hz test can be applied.

Regarding the reading test on both methods, while the standard deviation increases if compared to the previous tests it maintains the same proportion as in the Free test. This increase in the standard deviation can be explained if it is taken into account that for this test the subjects were asked to read out loud a text, therefore producing unique variations to the respiration signals for each subject; this variations translate into a greater standard deviation than the previous tests. Regarding the t-test results for both methods, non-significant differences can be found in both the null bias hypothesis and the difference between methods.

Given the statistical results and its aforementioned interpretation for all the tests of the Depth and Pattern methods, we can observe that the Depth method yields slightly better results than the Pattern method in the Free test but almost identical results in the other tests.

Finally, for the results of the Thermal camera, in the first two tests (0.1 Hz and 0.3 Hz) a gain error can be appreciated, this can be explained as a possible bias between the sampling frequency of the camera and the one used to process all the data. It can also be appreciated for the null bias t-test a p-value inferior to 0.001 for the 0.1 Hz test and a p-value inferior to 0.05 for the 0.3 Hz, this results are a direct consequence of the gain error presented in the Bland-Altman plot. For the Free and Reading its standard deviations are far greater than the ones in the other methods for the same tests but without gain error. Regarding the t-test analysis, both tests yield non-significant differences for the null bias hypothesis, regarding the results of the t-test between methods they show very significant differences when compared with both the Pattern method and the Depth method.

Given the results of the Bland-Altman analysis in Figure 6 and Table 4; for the RGB camera and Depth camera, for the 0.1 Hz test the bias of the RGB camera is almost three times the bias of the Depth camera, this indicates that the RGB camera has lower performance than the Depth camera if compared to the gold standard for this test. Regarding the t-test between this methods, a $p < 0.001$ can be seen which indicate significant differences between this methods for the given test. This could be due to the fundamental frequency of the test being close to the high-cut frequency of the filters used. Regarding the

0.3 Hz test, both methods perform similarly with practically the same bias and with slight differences in the standard deviation but with no significant differences between each other. In the Free and Reading test on the other hand, an increase in the bias and standard deviation for both methods can be appreciated, this can be caused by the unique morphology of each subject as explained above. Referent to the null bias t-test analysis, both methods present a $p < 0.001$ for the 0.1 Hz test, they present a $p < 0.05$ for the free test and non-significant differences for the 0.3 Hz and Reading test. The results for the 0.1 Hz test can be due to the fundamental frequency of the test being close to the high-cut frequency of the filters used.

In this case the Depth method yielded better results than the Pattern method in the controlled tests (0.1 Hz and 0.3 Hz) and slightly better results in the Free and Breathing tests.

Regarding the Thermal method, a gain error at 0.1 Hz, 0.3 Hz and Free test can be seen, these gain errors can be interpreted in the same way as the ones presented in Figure 5. The standard deviation for all four tests is also very high if compared to the other two methods, and this could be interpreted as a low agreement between the studied method and the gold standard. Regarding the null bias t-test analysis, all the tests yield a $p < 0.001$ which also indicated a low agreement between the gold standard and this method.

Referring to the standard deviation of the error (SDE) given the results in Table 5 Depth yields a low median SDE with a narrow inter quantile range (IQR) for all tests, the Plux vs Pattern method also yields a low median SDE and a narrow IQR range, this implies a high agreement between the gold standard and the studied methods. For the Plux vs Thermal method the median of the SDE practically doubles the other two for every test and it also presents a wider IQR range which implies that this method presents a very high variability compared with the gold standard. This means a low agreement between the gold standard and the Thermal method.

For all methods, an improvement has been shown in the median and IQR when the respiration frequency increases. This can be clearly seen in the 0.1 Hz test as the median is greater than the one presented for the 0.3 Hz test for all methods, the same applies to the IQR range. This increase on the median could be due to 0.1 Hz being too close to the high-cut frequency of the filters applied to the processing stage.

As far as the interactions between the SDE of the methods and given the results of the post-hoc tests, non-significant differences between the SDE of the Plux vs Depth and the Plux vs Pattern are shown, this could indicate a likeness between these two methods. For the Plux vs Thermal SDE significant differences between the other two methods can be found ($p < 0.001$), which imply that this method yields different results than the other two.

There were several limitations to the study, the first one was the number of subjects that participated in the study, not all the tests could be included due to errors during the acquisition stage or due to errors during the signal extraction stage (being of special relevance in the case of the thermal camera), only the data of 21 subjects of 23 could be used.

The second limitation of the study was the errors introduced to the thermal signal (occlusion of the nostrils, sudden movements...) due to the subject's movement during the test. Although we tried to correct this issue by applying optical flow tracking to the region of the nostrils, these movements still induced some error on the Thermal signal.

The third limitation was, due to the limitations aforementioned, that the signal from the Thermal camera was not as close to a sinusoid as the other two methods are, thus producing errors to the processed Instantaneous Frequency.

Because all the tests were performed in a very controlled environment, it cannot be assured that the conclusions would be still valid in a light changing or moving environment [1], for this reason we will perform further studies in different environments such as light changing conditions (fast or slow) and in environments that present vibrations which will certainly affect the results of the tracking and depth algorithm.

Regarding the feasibility of using these methods from a technical point of view, the Pattern method is the one with less sophisticated hardware, but on the other hand relies on a pattern placed on the subject to work. The Depth method relies on a more sophisticated hardware (an IR camera with a laser projector) but it does not need any external factor to work. Finally the Thermal method is the one with the most sophisticated hardware and requires of a complete calibration each time a measurement is done.

In conclusion, the Depth and Pattern methods can be used to acquire respiration rhythm and respiratory variability in controlled environments. Both yield very good results compared with our gold standard, with differences between these methods and the gold standard in the order of mHz in both mean and standard deviation.

For the Thermal method, it cannot be discarded directly because of the limitations explained above, but as shown in the results it performed poorly compared with the other two methods. Further studies are needed to confirm that these methods can be used as a reliable tool to acquire respiration in real life conditions; such as light changing environments or where the movements of the subject have a negative impact to the respiration acquisition using these methods.

Acknowledgement

This work was partially supported by the MINECO Project DEP2015-68538-C2-2-R, and with the support of the Universitat Politècnica de Catalunya (UPC).

References

- [1] N. Rodriguez-Ibanez, M. A. Garcia-Gonzalez, M. Fernandez-Chimeno, J. Ramos-Castro, Drowsiness detection by thoracic effort signal analysis in real driving environments, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 6055–6058. doi:10.1109/IEMBS.2011.6091496.
URL <http://ieeexplore.ieee.org/document/6091496/>
- [2] G. Benchetrit, Breathing pattern in humans: diversity and individuality, *Respiration Physiology* 122 (2-3) (2000) 123–129. doi:10.1016/S0034-5687(00)00154-7.
URL <http://linkinghub.elsevier.com/retrieve/pii/S0034568700001547>
- [3] C. Bruser, C. H. Antink, T. Wartzek, M. Walter, S. Leonhardt, Ambient and Unobtrusive Cardiorespiratory Monitoring Techniques, *IEEE Reviews in Biomedical Engineering* 8 (2015) 30–43. doi:10.1109/RBME.2015.2414661.
URL <http://ieeexplore.ieee.org/document/7063243/>
- [4] F. AL-Khalidi, R. Saatchi, D. Burke, H. Elphick, S. Tan, Respiration rate monitoring methods: A review, *Pediatric Pulmonology* 46 (6) (2011) 523–529. doi:10.1002/ppul.21416.
URL <http://doi.wiley.com/10.1002/ppul.21416>
- [5] M. Fernandez Chimeno, J. Ramos Castro, M. A. García Gonzalez, F. Guede Fernandez, M. Mateu Mateus, N. Rodriguez Ibañez, B. Bas Pujols, J. M. Alvarez Gomez, Respiratory Signal Extraction, *WO/2018/121861* (2018).
URL <https://patentscope.wipo.int/search/en/detail.jsf?docId=W02018121861>
- [6] Z. Zhang, J. Zheng, H. Wu, W. Wang, B. Wang, H. Liu, Development of a Respiratory Inductive Plethysmography Module Supporting Multiple Sensors for Wearable Systems, *Sensors* 12 (12) (2012) 13167–13184. doi:10.3390/s121013167.
URL <http://www.mdpi.com/1424-8220/12/10/13167>
- [7] Y. S. Lee, P. N. Pathirana, R. J. Evans, C. L. Steinfort, Noncontact Detection and Analysis of Respiratory Function Using Microwave Doppler Radar, *Journal of Sensors* 2015 (February) (2015) 1–13. doi:10.1155/2015/548136.
URL <http://www.hindawi.com/journals/js/2015/548136/>
- [8] Se Dong Min, Jin Kwon Kim, Hang Sik Shin, Yong Hyeon Yun, Chung Keun Lee, Myoungho Lee, Noncontact Respiration Rate Measurement System Using an Ultrasonic Proximity Sensor, *IEEE Sensors Journal* 10 (11) (2010) 1732–1739. doi:10.1109/JSEN.2010.2044239.
URL <http://ieeexplore.ieee.org/document/5571922/>
- [9] J. Lindemann, R. Leiacker, G. Rettinger, T. Keck, Nasal mucosal temperature during respiration., *Clinical otolaryngology and allied sciences* 27 (3) (2002) 135–9. doi:10.1046/j.1365-2273.2002.00544.x. <http://www.ncbi.nlm.nih.gov/pubmed/12071984>
URL <http://doi.wiley.com/10.1046/j.1365-2273.2002.00544.x>
- [10] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, C. Pugh, Non-contact video-based vital sign monitoring using ambient light and auto-regressive models, *Physiological Measurement* 35 (5) (2014) 807–831. doi:10.1088/0967-3334/35/5/807.
- [11] M. Bartula, T. Tigges, J. Muehlsteff, Camera-based system for contactless monitoring of respiration, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Vol. 2013, IEEE, 2013, pp. 2672–2675. doi:10.1109/EMBC.2013.6610090.
URL <http://ieeexplore.ieee.org/document/6610090/>
- [12] A. Procházka, M. Schätz, O. Vyšata, M. Vališ, Microsoft Kinect Visual and Depth Sensors for Breathing and Heart Rate Analysis, *Sensors* 16 (12) (2016) 996. doi:10.3390/s16070996.
URL <http://www.mdpi.com/1424-8220/16/7/996>
- [13] F. Ernst, P. Saß, Respiratory motion tracking using Microsoft’s Kinect v2 camera, *Current Directions in Biomedical Engineering* 1 (1) (2015) 192–195. doi:10.1515/cdbme-2015-0048.

- [14] S. Longhi, A. Monteriù, A. Freddi, F. Benetazzo, Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results, *Healthcare Technology Letters* 1 (3) (2014) 81–86. doi:10.1049/htl.2014.0063.
- [15] C. B. Pereira, X. Yu, M. Czaplík, R. Rossaint, V. Blazek, S. Leonhardt, Remote monitoring of breathing dynamics using infrared thermography, *Biomedical Optics Express* 6 (11) (2015) 4378. doi:10.1364/BOE.6.004378.
URL <https://www.osapublishing.org/abstract.cfm?URI=boe-6-11-4378>
- [16] J. Fei, I. Pavlidis, Thermistor at a distance: Unobtrusive measurement of breathing, *IEEE Transactions on Biomedical Engineering* 57 (4) (2010) 988–998. doi:10.1109/TBME.2009.2032415.
- [17] N. Rodriguez Ibanez, M. Fernandez Chimeno, J. J. Ramos Castro, M. A. Garcia Gonzalez, E. Montseny Masip, D. Bande Matinez, Method and System for Determining an Individual's State of Attention, US2011028857 (A1) (2014).
- [18] Microsoft Research, Kinect for Windows SDK beta (2011).
URL <https://www.microsoft.com/en-us/research/project/kinect-for-windows-sdk-beta/>
- [19] World Medical Association., World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects., *Bulletin of the World Health Organization* 79 (4) (2001) 373–4.
URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566407/>
- [20] J. K. Suhr, Kanade-Lucas-Tomasi (KLT) Feature Tracker, *Computer Vision (EEE6503)* (2009) 9–18.
- [21] E. A. Bernal, L. K. Mestha, E. Shilla, Non contact monitoring of respiratory function via depth sensing, in: 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI 2014, IEEE, 2014, pp. 101–104. doi:10.1109/BHI.2014.6864314.
URL <http://ieeexplore.ieee.org/document/6864314/>
- [22] A. Chatterjee, A. P. Prathosh, P. Praveena, Real-time respiration rate measurement from thoracoabdominal movement with a consumer grade camera, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBSdoi:10.1109/EMBC.2016.7591289*.
- [23] S. Kumagai, R. Uemura, T. Ishibashi, S. Nakabayashi, N. Arai, T. Kobayashi, J. Kotoku, Markerless Respiratory Motion Tracking Using Single Depth Camera, *Open Journal of Medical Imaging* 06 (01) (2016) 20–31. doi:10.4236/ojmi.2016.61003.
URL <http://www.scirp.org/journal/doi.aspx?DOI=10.4236/ojmi.2016.61003>
- [24] P. Viola, M. J. Jones, Robust Real-Time Face Detection, *International Journal of Computer Vision* 57 (2) (2004) 137–154. doi:10.1023/B:VISI.0000013087.49260.fb.
URL <http://link.springer.com/article/10.1023/B:VISI.0000013087.49260.fb>
- [25] R. Lienhart, A. Kuranov, V. Pisarevsky, Empirical analysis of detection cascades of boosted classifiers for rapid object detection, *Proceedings of the 25th DAGM Pattern Recognition Symposium* (2003) 297–304doi:10.1007/978-3-540-45243-0_39.
- [26] Jianbo Shi, Tomasi, Good features to track, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, IEEE Comput. Soc. Press, 1994, pp. 593–600. doi:10.1109/CVPR.1994.323794.
URL <http://ieeexplore.ieee.org/document/323794/>
- [27] C. Orphanidou, Derivation of respiration rate from ambulatory ECG and PPG using Ensemble Empirical Mode Decomposition: Comparison and fusion, *Computers in Biology and Medicine* 81 (September 2016) (2017) 45–54. doi:10.1016/j.combiomed.2016.12.005.
URL <http://dx.doi.org/10.1016/j.combiomed.2016.12.005>
- [28] R. Balocchi, D. Menicucci, E. Santarcangelo, L. Sebastiani, A. Gemignani, B. Ghelarducci, M. Varanini, Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition, *Chaos, Solitons and Fractals* 20 (1) (2004) 171–177. doi:10.1016/S0960-0779(03)00441-7.
- [29] S. B. Park, Y. S. Noh, S. J. Park, H. R. Yoon, An improved algorithm for respiration signal extraction from electrocardiogram measured by conductive textile electrodes using instantaneous frequency estimation, *Medical and Biological Engineering and Computing* 46 (2) (2008) 147–158. doi:10.1007/s11517-007-0302-y.
- [30] M. Mateu-Mateus, F. Guede-Fernández, M. A. García-González, RR Time Series Comparison Obtained by H7 Polar Sensors or by Photoplethysmography Using Smartphones: Breathing and Devices Influences, in: I. Lacković, D. Vasic (Eds.), 6th European Conference of the International Federation for Medical and Biological Engineering SE - 66, Vol. 45 of IFMBE Proceedings, Springer International Publishing, 2015, pp. 264–267. doi:10.1007/978-3-319-11128-5_66.
URL http://link.springer.com/10.1007/978-3-319-11128-5_66
- [31] D. S. Benitez, P. a. Gaydecki, a. Zaidi, a. P. Fitzpatrick, A new QRS detection algorithm based on the Hilbert transform, *Computing in Cardiology* 27 (2000) 379–382. doi:10.1109/CIC.2000.898536.
- [32] S. Sahoo, P. Biswal, T. Das, S. Sabut, De-noising of ECG Signal and QRS Detection Using Hilbert Transform and Adaptive Thresholding, *Procedia Technology* 25 (Raerest) (2016) 68–75. doi:10.1016/j.protcy.2016.08.082.
URL <http://linkinghub.elsevier.com/retrieve/pii/S2212017316304297>
- [33] R. J. Hodrick, E. C. Prescott, Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking* 29 (1) (1997) 1. doi:10.2307/2953682.
URL <http://www.jstor.org/stable/2953682?origin=crossref>
- [34] J. P. Royston, An Extension of Shapiro and Wilk's W Test for Normality to Large Samples, *Applied Statistics* 31 (2) (1982) 115. doi:10.2307/2347973.
URL <http://www.jstor.org/stable/10.2307/2347973?origin=crossref>
- [35] M. Hollander, D. A. Wolfe, E. Chicken, *The Two-Way Layout*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2015, pp. 289–392. doi:10.1002/9781119196037.ch7.
URL <http://doi.wiley.com/10.1002/9781119196037.ch7>
- [36] D. Giavarina, Understanding Bland Altman analysis, *Biochemia Medica* 25 (2) (2015) 141–151. doi:10.11613/BM.2015.015.