Video-based respiration monitoring with automatic region of interest detection

Video-based Respiration Monitoring with Automatic Region of Interest Detection

Rik Janssen[1], **Wenjin Wang**[1], **Andreia Moço**[1], **and Gerard de Haan**[1,2]

[1] Eindhoven University of Technology, PO Box 513, 5600MB, Eindhoven, NL
[2] Philips Group Innovation, Research, High Tech Campus 36, 5656AE, Eindhoven, NL

E-mail: rik.janssen@philips.com, w.wang@tue.nl, a.moco@tue.nl, g.de.haan@philips.com

Abstract. Vital signs monitoring is ubiquitous in clinical environments and emerging in home-based healthcare applications. Still, since current monitoring methods require uncomfortable sensors, respiration rate remains the least measured vital sign. In this paper, we propose a video-based respiration monitoring method that automatically detects respiratory Region of Interest (RoI) and signal using a camera. Based on the observation that respiration induced chest/abdomen motion is an independent motion system in a video, our basic idea is to exploit the intrinsic properties of respiration to find the respiratory RoI and extract the respiratory signal via motion factorization. We created a benchmark dataset containing 148 video sequences obtained on adults under challenging conditions and also neonates in the neonatal intensive care unit (NICU). The measurements obtained by the proposed video respiration monitoring (VRM) method are not significantly different from the reference methods (guided breathing or contact-based ECG; p-value=0.6), and explain more than 99% of the variance of the reference values with low limits of agreement (-2.67 to 2.81 bpm). VRM seems to provide a valid solution to ECG in confined motion scenarios, though precision may be reduced for neonates. More studies are needed to validate VRM under challenging recording conditions, including upper-body motion types.

Keywords: Biomedical monitoring, remote sensing, respiration, object detection

1. Introduction

Respiratory rate is an important early indicator for the deterioration of a person's health. Conditions such as cardiopulmonary arrest [1], sudden infant death syndrome [2] and other diseases that lead to an increase or decrease in the arterial partial pressure of carbon dioxide ($PaCO_2$) [3] can be detected by monitoring a person's respiratory rate. Conventional methods in respiration monitoring require contact sensors to be attached to the human body, such as an airflow sensor, electrodes, or a strain gauge. However,

these contact-based methods are inconvenient and uncomfortable to use, and cannot be applied to all patients, i.e., patients with burned skin and neonates with sensitive skin. Therefore, there is a preference to monitor respiration without contact, especially for sleep monitoring, ambient assisted living and long term respiration monitoring. To this end, non-contact respiration monitoring methods have been proposed using radar, thermal sensors, or optical sensors [4]. However, the respiratory signal detected by the Doppler radar approach is easily corrupted by other motion noises, while thermal sensors require a visible face for detecting temperature changes in the nose/mouth areas. In addition, they are professional medical devices that are not affordable for home-based use. Therefore in this paper, we focus on a camera-based approach to extract the respiratory signal from chest and abdominal motion using a consumer-level camera.

Previously, several approaches have been proposed to monitor respiration with a camera based on the respiratory-induced movements of chest and abdomen [5,8,9,14,15]. Tan *et al.* [5] used frame differencing to detect respiratory motion between two adjacent frames. However, it highly depends on the type of clothing and is not robust to non-respiratory motion or illumination changes. In the work of Wiesner *et al.* [15], three colored fiducials placed on the patients abdomen are automatically detected and tracked to extract respiration. Beside the requirement of the fiducials, robustness to non-respiratory motion is limited. Bartula *et al.* [8] proposed to manually select the respiratory Region of Interest (RoI) at initialization. This, however, requires manual interaction, which we rather prevent as it fails whenever the patient significantly moves after initialization. Alternatively, Li *et al.* [9] takes the entire video frame for tracking motion features. The motion features are then decomposed and the signal with the highest variance is initialized as the respiratory signal. Another approach is presented by Lukáč *et al.* [14], where, instead of tracking feature points, motion trajectories are calculated for each subregion of the image using optical flow. Consequently, the respiratory signal is selected based on the signal-to-noise ratios (SNR) of the subregions. Although accurate for stationary conditions, the proposed algorithm is not able to detect the respiratory signal when there are non-respiratory motions present. Furthermore, other camera-based respiration monitoring methods based on the respiratory-induced color differences of the skin have been presented [13, 16, 17].

However, automatic localization of valid RoI in videos for reliable respiratory signal extraction remains an open problem. If it can be accomplished, motion robustness improves in comparison with manual selection, a cornerstone advantage for long-term monitoring applications. Indeed, the RoI (e.g., location, size or shape) can be regularly and automatically adjusted in different environments, or even reinitialized after temporary motion-related failure. Additionally, manual selection of the respiratory RoI during initialization becomes obsolete, thus improving ease-of-use of this technology.

Inspired by the prior works [9,14], we aim to improve the robustness of motion-based respiration estimation and enable the automatic RoI detection. Similar to Li *et al.* [9] and Lukáč *et al.* [14], we extract the respiratory signal by using pixel-based motion vectors as features (e.g., optical flows) and motion factorization (e.g., singular value

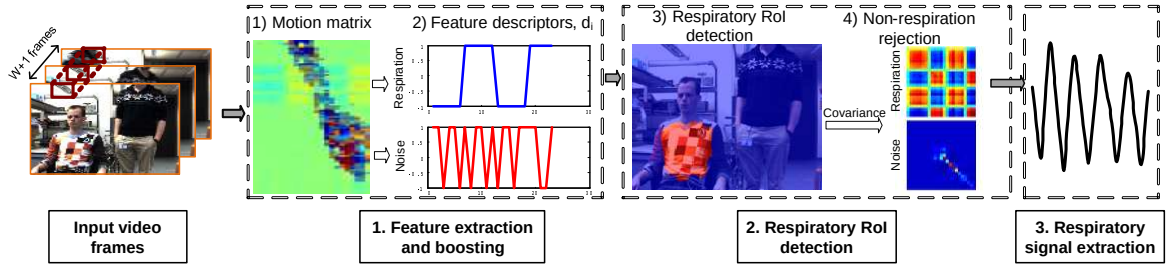

Figure 1. The flowchart of the proposed video-based respiration monitoring system. Given a video sequence of a subject, the respiratory RoI is automatically detected for respiratory signal extraction.

decomposition (SVD) or principal component analysis (PCA)). The contributions of our work are: (1) we propose a robust feature representation for respiratory signal based on motion features, which exploits the intrinsic properties of respiration; and (2) we enable the automatic respiratory RoI detection to enhance the monitoring performance. The proposed method is thoroughly evaluated by a significant number of challenging videos, and demonstrates competitive performance against guided breathing and contact-based ECG method.

2. Video-based respiration monitoring method

To monitor respiration, we exploit two basic observations: (1) respiration induces subtle trunk motions on chest or abdomen, which can be detected by dense optical flow; (2) the respiratory motion is physically uncorrelated from the remaining motion sources, which means that motion factorization can extract the intended respiratory signal. Therefore, we propose the following steps: (1) feature extraction and boosting, (2) respiratory RoI detection, and (3) respiratory signal extraction. Fig. 1 shows an overview of the proposed algorithm, which will be explained in detail in following subsections.

2.1. Feature extraction and boosting

2.1.1. Motion matrix To detect respiration-induced chest or abdominal motion, we employ the dense optical flow algorithm proposed by Brox *et al.* [10] to estimate pixel motion vectors. Since respiration-induced motion is mainly in the vertical direction [8], we only use the vertical flows. Care is taken to ensure that the window size includes at least one complete inhale/exhale cycle, meaning that the window size, W, is made larger than $60 \cdot f_{sampling}/f_{resp,min}$, where $f_{resp,min}$ is the minimum respiratory rate; $f_{sampling}$ is the camera sampling rate. As such, the rows of the constructed motion matrix M (size $N \times W$, where N is number of pixels in a video frame) containing motion derivatives that represent the velocity of a pixel's trajectory in the vertical direction.

2.1.2. Mid-level respiratory descriptors The overall motion matrix is then factorized into separate motion trajectories, which is a strategy that is previously described by Hou

et al. [14]. A drawback, however, is that pixel-based motion derivatives are inherently sensitive to subtle changes and often deviate from one another, i.e., they are noisy to be clustered into a factorized basis. We propose to overcome this limitation by generating robust mid-level feature representations from pixel motion vectors. To this end, we partition M into spatio-temporal regions, m_i, where i is the region index, i.e., m_i stores W consecutive squared blocks from the input flow sequence. For each m_i, we can generate eigenvectors that satisfy the general condition:

$$m_i \cdot D_i = \lambda \cdot D_i \quad \text{s.t.} \quad \det(m_i - \lambda_i \cdot I) = 0, \tag{1}$$

where $\det(\cdot)$ denotes the matrix determinant; I is the identity matrix; and D_i and λ_i correspond to eigenvectors and eigenvalues, respectively. By convention, the first eigenvector with the largest eigenvalue dominates the feature space. And it is often the case that the relevant respiratory signal is at the strongest first component in the segmented spatio-temporal tube containing respiration. Hence, we proceed by reducing m_i to a feature vector $f_i = \lambda_1 \cdot D_{i,1}$, i.e., a robust least square estimation of pixel-based motion vectors. Lastly, we sought to promote the subtle respiratory motion and suppress large motions by quantizing f_i into two levels:

$$d_i(k) = \begin{cases} 1 & \text{, if } f_i(k) \geq 0 \\ -1 & \text{, otherwise} \end{cases}, \tag{2}$$

which binarizes the feature descriptor $f_i(k)$. Since f_i is derived from the derivative of the respiratory motion-signal, we use the threshold 0 to separate the inhaling motion from the exhaling motion. The benefit of binarization is to magnify and equalize the subtle motion changes (e.g., respiratory motion), while suppressing large motion distortions.

2.1.3. Reference descriptor formation Since d_i is not specified for respiratory motion yet, we continue to exploit respiratory properties that would allow us to form a reference descriptor to query the respiratory trajectories in the motion matrix. To this end, we develop/assign a score for each descriptor that would meet the following criteria:

• **Boost respiratory range** In line with [9,14] , we exploit the prior of respiration frequency to restrict outliers. Since the human respiratory rate is typically in the range of $[12, 44]$ breaths per minute (bpm) [11], a boosting term is included to preserve descriptors within this frequency band and penalize the rest. Thus a χ^2 kernel is selected and tuned to satisfy the band-pass condition in the respiratory range.

• **Penalize noise** Since noise is generally found at the high-frequency components of the spectrum, we count the rising-edge transitions within each descriptor. If the descriptor exceeds the upper limit that corresponds to the maximum respiratory rate, the excess, ϵ_i, is used to penalize the score of d_i.

• **Enforce temporal consistency** Similar to [9], we also take the temporal consistency of estimation into account. Here we sought to improve motion robustness by also considering the correlation between current and previous descriptors, d_i^t and d_i^{t-1}, respectively. Temporally coherent d_i leads to a higher correlation value.

We translate above items into a compound score, S_i, to grade d_i. It is expressed as:

$$S_i = \overbrace{\frac{R_i^{\beta} \cdot e^{-\frac{R_i}{\eta}}}{\max(r^{\beta} \cdot e^{-\frac{r}{\eta}})}}^{\text{Respiration boosting}} \cdot \overbrace{e^{-\alpha\epsilon}}^{\text{Noise penalty}} \cdot \overbrace{\sum_{k=1}^{W-1} d_i^{t-1}(k+1) \cdot d_i^t(k)}^{\text{Temporal consistency}} \tag{3}$$

where R_i is the respiratory rate detected in subregion i and $r \in [0, 100]$ breaths per minute (bpm). The additional parameters, α, β and η, are tuned at the beginning of the algorithm to configure the error tolerance and regulate the frequency response of the χ^2 kernel.

The total amount of scores is then condensed into an histogram representation. After normalization, histograms allow separation of noise bins from respiratory bins, as only the latter are awarded scores close to unity (see Fig. 2). Accordingly, we refine the respiration descriptor by selecting the most frequent bin with a higher score (e.g., > 0.5). We shall refer to this final descriptor as $\hat{d}$.

2.2. Respiratory RoI detection

In this section, we describe a cascade of coarse-to-fine processing stages that allow us to obtain a final respiratory RoI. Subscripts 1 to 4 shall be used to refer to sequential RoIs obtained in the proposed pipeline.

As a starting approach, the *RoI* is obtained as the self-similarity between the pixel-flows stored in the rows of M and the previously obtained respiratory signal, $\hat{d}_i$, here is regarded as a reference function. As a metric, we apply the normalized inner product, resulting in:

$$RoI_1 = \frac{\overline{M} \cdot \hat{d}}{\max(\overline{M} \cdot \hat{d})}, \tag{4}$$

where $\overline{M}$ is the normalized motion matrix. RoI_1 was then vectorized and reshaped, resulting in RoI_2:

$$RoI_2 = \begin{cases} 1 & \text{, if } RoI_1 \geq THR \\ 0 & \text{, otherwise} \end{cases}, \tag{5}$$

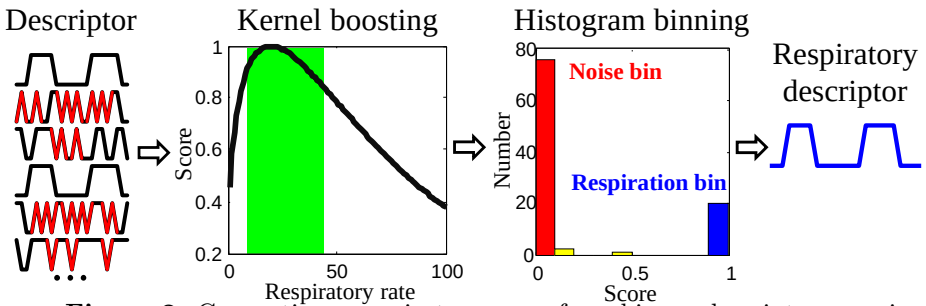

Figure 2. Computing a respiratory score from binary descriptors requires terms that translate kernel boosting, noise penalty and temporal consistency check.

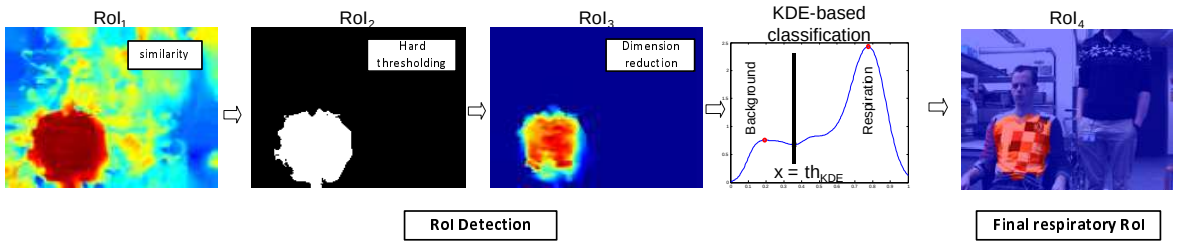

Figure 3. The pipeline for coarse-to-fine respiratory RoI detection.

where the threshold $THR = 0.7$ is provisionally defined but will be refined later. Recognizing that the respiration is independent from other motion sources, we proceed by separating motion trajectories by orthogonalizing them. So we apply the SVD on the subset of s rows of M, hereafter denoted as M_s, for which RoI_2 is one (i.e., valid respiratory regions):

$$\overset{N_s \times W}{M_s} = \overset{N_s \times N_s}{U} \cdot \overset{N_s \times W}{\Sigma} \cdot \overset{W \times W}{V^\top}, \tag{6}$$

The assumption of respiration being the dominant component is generally valid for seated or lying subjects in healthcare monitoring, which is also the use-case focused in this work. Thus we perform the low-rank approximation on M_s as:

$$\overset{N_s \times W}{M_s} \approx \overset{N_s \times 1}{RoI_{3,s}} \cdot \sigma_1 \cdot \overset{1 \times W}{d^\top}, \tag{7}$$

where RoI_3 stands for the overall vectorized respiratory region and σ_1 is its associated singular value. At this stage, one might rightly suspect that RoI_3 is suboptimal due to hard-coded threshold of Eq. 5. Since the chest/abdomen is a spatio-temporally coherent region that does not dramatically change the appearance or location, we apply Kernel Density Estimation (KDE) to obtain a smoothed histogram distribution of the non-zero elements of RoI_3. Accordingly:

$$KDE(x) = \frac{1}{n \cdot B} \sum_{i=1}^{n} \frac{e^{-\frac{1}{2} \cdot \left(\frac{RoI_{i,3} - x}{B}\right)^2}}{\sqrt{2\pi}} \text{ with } B = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}}, \tag{8}$$

where n is the number of non-zero entries in RoI_3; $\hat{\sigma}$ is the standard deviation of the non-zero entries in RoI_3; and x denotes the values where the kernel density is estimated.

An adaptive threshold, th_{KDE}, is derived from $KDE(x)$ by detecting the two largest peaks in $KDE(x)$ and selecting the x-value that corresponds to the minimum value between those peaks as th_{KDE}, i.e., the valley. The location of the peak is selected as th_{KDE} when only one peak exists. Afterwards, th_{KDE} is used to improve the respiratory mask RoI_4 by separating "background" from respiration, where Eq. 5 is replaced by:

$$RoI_4 = \begin{cases} 1 & \text{, if } RoI_3 \geq th_{KDE} \\ 0 & \text{, otherwise} \end{cases}. \tag{9}$$

Fig. 3 exemplifies a possible density plot for the non-zero entries of a RoI_4. When there is a clear-cut separation between "background" from respiration, the separating line, $x = th_{KDE}$, is placed at the valley between peak distributions. Otherwise, th_{KDE} is set at the middle of the single peak of the KDE.

2.3. Respiratory signal extraction

From the respiratory RoI detected in previous steps, we are now able to extract the respiratory signal. A lingering issue, however, is that RoI detectors for respiration-monitoring applications are required to favor high precision over recall. As such, it is worthwhile to perform an additional consistency check to prevent non-respiratory motions from being falsely reconstructed. In this regard, we design a new score to penalize non-respiratory motions and prune corrupted motion matrices. It consisted of three parts:

• **Standard deviation of covariance** We observe that respiration is more temporally consistent than noise, implying that its patterns can appear multiple times within a sliding window. However, respiration is not always periodic. To overcome this issue, we propose a soft metric, S_c, that stands for the standard deviation of the *covariance* of M. S_c is used to measure the frequency of respiration patterns in M. Formally:

$$S_c = \mathrm{std}\left(\mathrm{vec}\left(\frac{(M - \mathrm{E}(M))^\top \cdot (M - \mathrm{E}(M))}{N}\right)\right), \tag{10}$$

where the operators $\mathrm{vec}(\cdot)$, $\mathrm{std}(\cdot)$ and $\mathrm{E}(\cdot)$ denote vectorization, standard deviation and expectation, respectively. In essence, the covariance matrix of M reflects the self-similarity of features; if M only contains repeatable patterns like respiratory waveforms, S_c will be small. Conversely, if non-respiratory distortions pollute M, the subtle respiratory patterns will be overshadowed in the covariance matrix and S_c will increase rapidly. For illustration purposes, Fig. 1 shows covariance matrices for respiratory regions and noise. It is visible that the respiratory covariance matrix has repeated patterns (represented as red and blue rectangular subregions in non-diagonal entries) that result from mutual-correlation between trajectories. In contrast, the noise covariance matrix contains most of its energy at diagonal entries that result from self-correlation.

• **Temporal variability of the dominant singular value** The singular value σ of motions in *RoI* is used as an indicator for sudden changes; i.e., a sudden decrease indicates the breath holding, while the opposite means that abrupt non-respiratory motions corrupt the overall *RoI*. Accordingly, a score measuring the temporal changes of σ is given by:

$$S_\sigma = \begin{cases} 1 & , \text{if } \sigma > \sigma^t \\ e^{-\frac{\sigma^t - \sigma}{\omega_2}} & , \text{otherwise} \end{cases}, \tag{11}$$

where t denotes t-th frame; ω_2 weights the temporal changes between σ^t and σ. σ is recursively updated as follows:

$$\sigma = \alpha \cdot \sigma^t + (1 - \alpha) \cdot \sigma^{t-1} \,\text{with}\, \alpha = 0.5\; e^{-\frac{(\sigma^t - \sigma^{t-1})^2}{\omega_1}}, \tag{12}$$

where ω_1 denotes the tolerance to temporal changes in σ.

• **Temporal consistency of RoI** In addition, the temporal consistency of RoI_3 and its binary mask RoI_4 are also measured/maintained, which is derived by the inner product between current and previous measurements as:

$$S_m = \sum_{x=1}^{N} \left(\frac{RoI_3^t(x)}{\|RoI_3^t\|} \cdot \frac{RoI_3(x)}{\|RoI_3\|} \right) \cdot \sum_{y=1}^{M} \left(\frac{RoI_4^t(y)}{\|RoI_4^t\|} \cdot \frac{RoI_4(y)}{\|RoI_4\|} \right), \tag{13}$$

where x and y denote the pixel location within RoI; $\|\cdot\|$ denotes the L2-norm. Similarly, RoI_3 and RoI_4 are also recursively updated as:

$$\begin{cases} RoI_3 & = \gamma RoI_3^t + (1-\gamma) RoI_3^{t-1} \\ RoI_4 & = \gamma RoI_4^t + (1-\gamma) RoI_3^{t-1} \end{cases}, \text{with}\, \gamma = \frac{1}{W}, \tag{14}$$

where the updating speed, γ, is inversely proportional to the length of the sliding window. It is now possible to obtain an improved binary mask, ω, that rejects non-respiratory motions by combining S_c, S_σ and S_m. ω is obtained as follows:

$$\omega = \begin{cases} 1 & \text{if } \psi + (1-\psi) e^{-\phi \cdot S_\sigma^2} < 0.9 \\ 0 & \text{otherwise} \end{cases}, \tag{15}$$

where the threshold 0.9 is empirically defined; $\psi = S_c \cdot S_m$ and ϕ denotes the tolerance of S_σ. Consequently, we are able to compute the intended respiratory signal by numerical integration of valid flow vectors. To this end, we decompose the motions within RoI_4. This results in the derivative of the respiratory motion d_y and it's energy σ_y. Afterwards, we cumulate sum the derivatives of d_y to generate the respiratory signal as:

$$resp(n) = \sum_{k=1}^{n} d_y(k), \tag{16}$$

which is then normalized as:

$$resp = \begin{cases} \sigma_y \cdot \left(\frac{resp - \text{avg}(resp)}{\text{std}(resp)} \right) & \text{, if } \omega = 1 \\ 0 & \text{, othwise} \end{cases}, \tag{17}$$

where avg$(\cdot)$ denotes the mean value. Subsequently, the respiratory strides from sequential sliding windows are multiplied with a Hanning window and overlap-added into a long-term signal. From the extracted respiratory signal, we estimate its instantaneous respiratory rate using a simple peak detector.

3. Materials and methods

We assessed the performance of our proposed Video Respiration Monitoring (VRM) system based on a benchmark dataset comprising 148 video recordings, under different conditions, from 4 healthy adults (3 males and 1 female) and 2 neonates. The study was approved by the Internal Committee Biomedical Experiments of Philips Research, and the informed consent has been obtained for each adult subject. In addition, the medical ethical research committee at Maxima Medical Center (MMC) approved the

study‡ and informed parental consents were obtained prior to data acquisition. The following subsections specify the benchmark dataset, evaluation metric and parameter setting.

3.1. Benchmark dataset

In adults, the reference signals (e.g., ground-truth) are guided breathing patterns, i.e., subjects were instructed to mimic a sinusoidal breathing pattern that was displayed in a front screen during recordings. We compared VRM against a contact-based method with respect to the reference. Thus the respiratory signals were measured by the thoracic impedance plethysmography method (Philips Intellivue MP50., The Netherlands), which we shall denote as ECG. To investigate performance, we created recordings for three challenge categories: (1) breathing patterns ("slow"-10 bpm, "normal"-16 bpm, "fast"-20 bpm, "very fast"-40 bpm, "deep breath"-8bmp and "held breaths"-0 bpm); (2) type of non-respiratory motions ("body motion", "foreground occlusion" and "background motion"); and (3) lighting conditions ("bright", "dark" and "varying"). Each category was recorded under two different camera distances ("1m" and "2m") and clothing styles ("textured" and "textureless"). In total, 128 videos are recorded in guided breathing scenario. The videos were recorded by a regular CCD-RGB camera (IDS, model UI-2230SE-C, Germany) in global-shutter mode and stored in an uncompressed data format (size 768×576 pixels, 8 bit depth). The average duration of the videos was around 4.3 min. Note that all subjects seated in a chair in all recordings (even when performing body motions), which are not vigorous body motions presented for example in pedestrian or sports field.

In neonates, ECG was considered as the reference method. Since the ECG signals recorded from (sleeping) neonates in this scenario are rather stable, it is important to know whether the non-contact method can replace the contact-based method in real clinical settings. The videos were recorded in a neonatal intensive care unit (NICU; MMC, Veldhoven, The Netherlands). We collected 20 videos from different scenes and viewpoints: (1) zoomed top-view of head and a part of the chest, (2) wide range top-view, (3) zoomed side-view of head and a part of the chest, and (4) wide range side-view. In 16 scenes, the neonate's chest was covered by a blanket.

Note that none of the existing studies in the field of video-based respiration monitoring attempted to make such a large dataset comprising a diverse range of respiration rates and challenging measurement variables, namely motion patterns, distance to camera, camera view angles, lighting conditions and clothing styles.

‡ The medical ethical research committee at Maxima Medical Center has reviewed the research proposal and considered that the rules laid down in the Medical Research involving Human Subjects Act (also known by its Dutch abbreviation WMO), do not apply to this research proposal.

3.2. Statistical analysis

The starting purpose of the statistical analysis was to determine if measurement errors of VRM (reference minus alternative method) varied as a function of "motion types", "lighting conditions", "clothing patterns", and "camera distance". In addition, since multiple recordings were obtained from each subject, the significance of a categorical variable subject index (1-4: adults, 5-6: newborns) was tested. The second purpose was to assess agreement of VRM with respect to the guided breathing or reference ECG.

The first analysis was performed to determine whether the effects of the independent variables should be examined with parametric or nonparametric statistics. First, the BrownForsythe test [18] was used to investigate the homogeneity of the sample error variances. Next, multiple regression analysis was performed to determine if VRM errors were independent of the subjects from which they were obtained. To this end, linear regression analysis was performed using "subject index", "clothing patterns", "motion types", "camera distance" and "lighting conditions" as predictor variables, and VRM error as the dependent variable. If it shows that subject index is not significantly related to the dependent variables after adjustment for the other predictors, it will be appropriate to combine video recordings from all subjects.

The effects of "clothing patterns" and "camera distance" were inspected, separately, using an independent sample t-tests. The effects of "lighting conditions" and "motion types" on VRM measurement error were tested, separately, using one-way ANOVA and the Kruskal-Wallis test. Post-hoc pairwise comparisons were done to investigate which motion pattern was statistically different. Supplementary, the Pearson correlation values were obtained between VRM and the reference, under each motion category.

Since multiple challenges were included in a single recording (e.g., a video in motion category contains different motion types), we sliced the recordings into short video intervals to investigate the challenges independently, which leads to 259 video segments. We assessed the agreement between methods using a pooled 259 video segments obtained under different scenarios. Thus, each measurement corresponds to the average respiratory rate measured in each video sequence. All values were expressed as mean $\pm$ standard deviation (STD) or 95% confidence intervals (CI) as measures of central tendency and variability. Mean values for respiratory rate were compared between the reference and VRM. We used paired sample t-tests to compare the methods. A correlation plot between reference and VRM is presented across video segments, including the coefficient of determination, r^2, and the standard error of the estimate (SEE). Bland–Altman analysis [19] was performed to test for magnitude bias in respiratory rate differences. Here, the 95% limits of agreement were determined by $[-1.96$ STD, 1.96 STD$]$. Statistical analysis was performed using IBM SPSS Statistics version 22.0, 2015 (SPSS Inc., IBM, Chicago, Illinois, USA) and MATLAB R2011a (The Mathworks, MA, USA). The statistical significance was set at $p < 0.05$.

To assess the effect of individual measurement conditions, the challenges of guided breathing videos were analyzed into detail. We used the accuracy to denote the quality

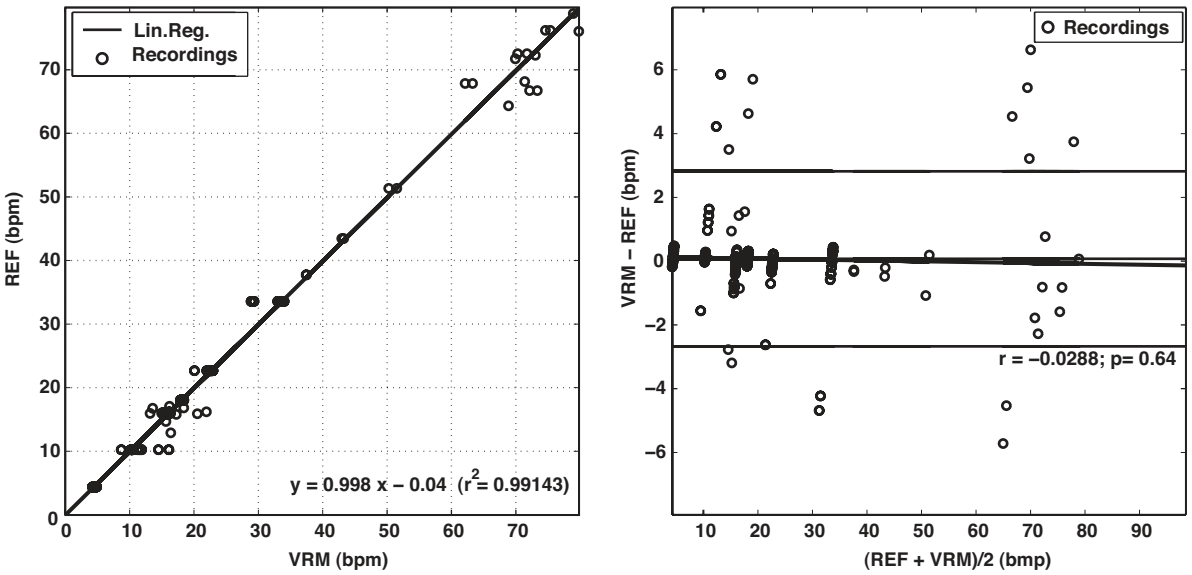

Figure 4. The Pearson correlation and Bland-Altman plots of VRM and ECG in the complete benchmark dataset.

of estimated respiratory rate, which is defined as the percentage of correctly detected breaths as $P = \frac{C_{breaths}}{T_{breaths}}$, where $T_{breaths}$ is the total number of breaths and $C_{breaths}$ is the number of correctly detected breaths. Note that P was calculated for each simulated sub-challenge in a category.

3.3. Parameter settings

We tuned parameters for adult and neonate populations based on videos from one adult and one neonate, respectively. For all videos, the sliding window size, W, was 23 and the RoI verification parameters, ω_1, ω_2 and ϕ, were set to 300, 50 and 200, respectively. However, the remaining parameters need to be tuned to the expected respiratory range of the intended population. Specifically, the regular respiratory rate in adults ranges from 12 to 44 bpm [11], whereas neonates have a much higher respiratory rate in the order of 40 to 70 bpm. A first implication of respiratory range variability is the choice for the sampling rate; for a tradeoff between computation speed and temporal resolution, the camera frame rate was set to $R_{cam} = 4$ fps for adults and to 10 fps for capturing the fast breaths of neonates. Also, the adaptive bandpass filter parameters of S_i were programmed to fit the respective ranges of intended populations (α, β and η were set to 1.33, 0.39 and 50 for adults, and to 0, 3 and 25 for neonates, respectively). VRM was implemented in Matlab R2014a (the Mathworks, Inc.) and run on an Intel Core i5 platform (3.20 GHz, 4 GB RAM).

4. Results and discussion

4.1. Measurement conditions

The Brown-Forsythe test statistics were not significant for "subject index" ($p = 0.69$), "lighting conditions" ($p = 0.99$), "clothing patterns" ($p = 0.29$) and "camera distance" ($p = 0.41$), but was significant for "motion types" ($p = 0.02$), which indicated that the sample error variances were not equal for this variable. Given this finding, samples t-tests were selected for "camera distance" and "clothing patterns", and regression analysis was applied for "camera distance", whereas the nonparametric Kruskal-Wallis test was used to analyze VRM measurement errors in "motion types".

Regression analysis showed that no categorical variable was significantly associated with measurement error, i.e., "subject index" ($p = 0.43$), "lighting conditions" ($p = 0.91$), "clothing patterns" ($p = 0.25$), "camera distance" ($p = 0.41$), "motion types" ($p = 0.08$). These results suggested that we can confidently combine all the recordings into a complete dataset for all further analysis. This dataset consisted of video recordings of adults and neonates with a median respiratory rate of 16.04 bpm (range in $4.38-78.86$ bpm).

The next analysis used independent samples t-tests or ANOVA to test the significance of "lighting conditions" ($p = 0.99$), "clothing patterns" ($p = 0.29$), and "camera distance" ($p = 0.41$) respectively. None of these variables was found to be a significant predictor of VRM measurement errors. The Kruskal-Wallis test was used to evaluate the influence of motion types on the VRM measurement error. We found significant differences between different motion types ($p = 0.02$). However, subsequent post-hoc pairwise comparisons of motion types indicated that in the "motion types" category, only the differences between upper-body motion and occlusion achieved significance ($p = 0.02$), whereas the remaining motions (including stationary) did not differ significantly from one another ($p > 0.08$).

Overall, the results for analyzing the effect of recording conditions on VRM measurement errors (in adults and neonates) can be summarized as follows: (1) individual variability has a negligible relationship on VRM errors; (2) "camera distance", "lighting conditions", "clothing patterns", and "motion types" are not significantly related to VRM errors.

4.2. Agreement analysis

In our study, the average respiratory rate in the overall benchmark dataset is 20.7 ± 15.1 bpm for ECG and 20.8 ± 15.1 bpm for VRM. The dataset is homogenous, as indicated by Shapiro-Wilk test ($p < 0.001$), and the average differences between our method and the reference are not significant (CI 95%, $[-0.098, 0.244]$ bpm; $p = 0.6$). The between-individual measures of respiratory rate were highly correlated ($r^2 > 0.99$, $p < 0.001$, SEE=1.40 bpm) and closely comparable to the respective measurements from guided breathing or ECG (see Fig. 4(a); for comparison, the SEE of ECG method is 1 bpm, as

described in the datasheet of Philips IntelliVue Patient Monitor). Furthermore, Bland-Altman analysis revealed neither significant magnitude bias nor trend ($r^2 = 0.0008$, p = 0.6) in prediction of respiratory rate. Also, the 95% CI was low ([-2.67, 2.81] bpm) and competitive with ECG (CI 95%, $[-2, 2]$ bpm) (see Fig 4(b)).

Overall, these are encouraging results for the prospective application of VRM in healthcare monitoring scenarios (e.g., subjects in seated posture or sleeping). However, future studies are needed to confirm the validity of our algorithm in larger populations, with varying heath conditions that may challenge the periodicity assumption of our algorithm, as well as a wide age range. In addition, one may question the interference that guided breathing may have over natural breathing patterns, thus recommending more ecological methods to provide a reference respiratory signal, such as thermal imaging. Lastly, one should be critical about that fact that Bland-Altman analysis provided higher limits of agreement and outliers for recordings above 50 bpm; i.e., videos obtained from neonates. We hypothesize that this is due to suboptimal temporal resolution of respiratory waveforms for higher respiratory rates, for a fixed sampling rate of 10 Hz. This investigation considered seated adult subjects and newborns lying in NICU settings. Future research is valuable to clarify if the performance of VRM holds in different lying postures and additional motion types.

4.3. Discussion of individual challenges

In order to investigate the performance of VRM in different circumstances (e.g., a particular challenge like upper-body rotation), we show the comparison of average accuracy between ECG and VRM in each sub-challenge of (1) the guided breathing scenario in Fig. 5, and (2) the neonatal monitoring scenario in Fig. 6. The examples of respiratory RoI detected in different challenges are shown in Fig. 7.

•**Varying breathing patterns** Fig. 5 shows that both the ECG and VRM achieve almost the same high accuracy (all above 95%) in normal, fast, very fast, deep and held breaths. It implies that both methods can accurately capture different breathing patterns. However, we notice a modest quality drop (around 3% less) of VRM in slow breath (10 bpm). This is due to the optical flow errors at the last part of the exhaling motion, where a short pause appears before the next inhaling motion. A longer pause introduces more optical flow errors, and results in a lower score for motion descriptors. Note that for the breath holding challenge, the detected RoI in VRM is released when the subject holds breath for more than 3 seconds. When the subject starts to breathe again, the RoI is quickly recovered.

•**Varying motion patterns** In Fig. 5, we find that VRM clearly outperforms ECG in challenges containing subject motions, such as upper-body motion, rotation, translation and body shaking, approximately 20%-30% improvement. This is because that (1) the sensors closely attached to body are seriously affected by body-motion, and (2) ECG cannot recognize non-respiratory motions. In contrast, VRM can detect and reject non-respiratory motions when estimating the respiratory signal, which is thus

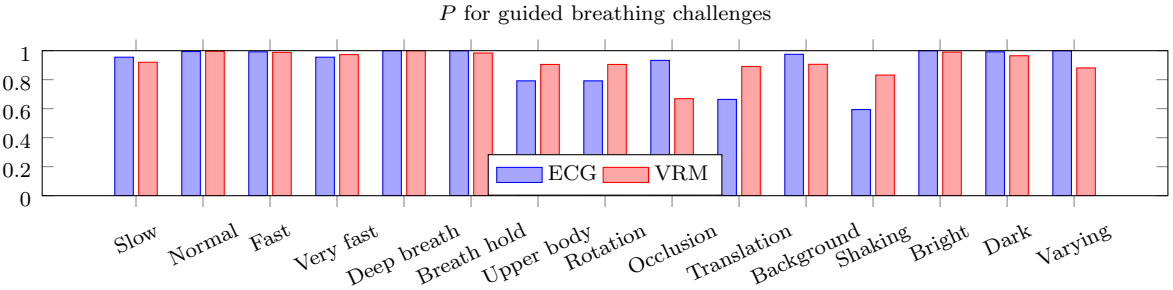

Figure 5. The comparison of average respiration accuracies obtained by ECG and VRM under different sub-challenges in guided breathing scenario.

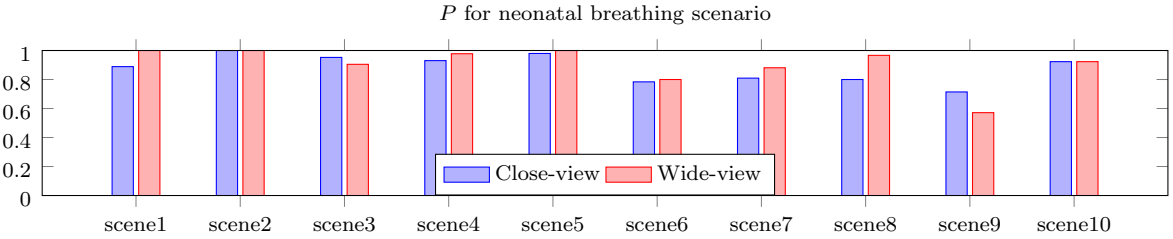

Figure 6. The average respiration accuracy obtained by VRM in neonate monitoring scenarios, where videos recorded in NICU contain two views: the camera view is set to (1) a close distance to the head or small part of the chest, and (2) a far distance for capturing the entire body of neonate.

more robust to body motions. However, since ECG does not use a remote camera, it will not be influenced by the challenges of foreground occlusion and background motion. The sudden occlusion between subject and camera can seriously pollute the RoI and signal of VRM. Besides, VMR can better deal with background motion than foreground occlusion.

•**Varying lighting conditions** Fig. 5 shows that VRM obtains similar accuracy as ECG in bright and dark lighting conditions, but performs worse in varying lighting condition. The challenge of varying lighting condition is simulated by creating shadows on subject and background. These shadows appear/disappear rapidly with a frequency within the regular respiratory band. It does not cause a problem to VRM when the chest is a larger area compared to the shadows, since the respiratory descriptor still dominates the feature space. But if the distance-to-camera is very large (e.g., RoI is very small), VRM will suffer from performance degradation.

•**Varying distance and clothing** Furthermore, we investigate the overall results of ECG and VRM obtained in the complete dataset in terms of "camera distance" and "clothing patterns". In fact, these two challenge categories have no impact on ECG. This is mainly due to the improvement of VRM in motion category. For VRM, the close distance between subject and camera is an advantage, because larger chest RoI can be monitored, while the clothing style (e.g., textured or textureless) is not critical for VRM.

•**Neonatal breathing scenario** To assess the effect of camera angle on the performance of VRM in NICU settings, we computed the accuracy of VRM in scenes of neonates, obtained under different viewing angles (scenes 1-3: top-view camera; 4-8:

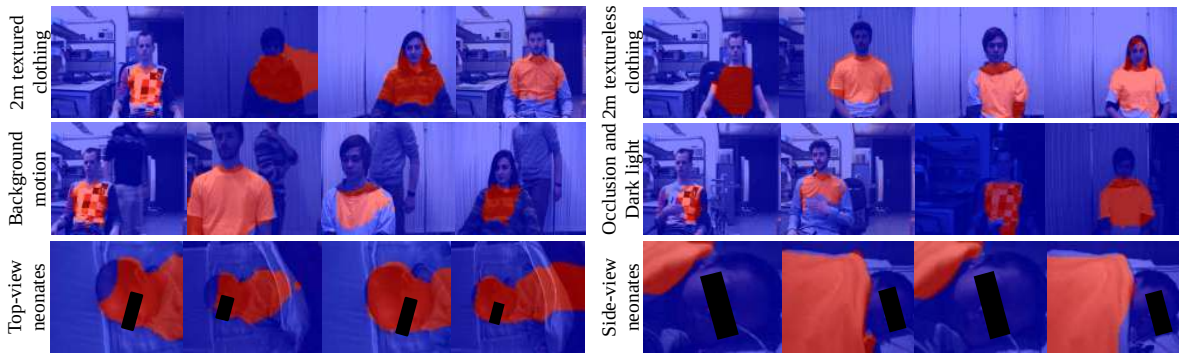

Figure 7. Some snapshots of the respiratory RoI (red) detected by proposed VRM.

side-view camera; 9-10: top-view camera). The results suggest that VRM generally performs better in wide-view than in close-view (Fig. 6). The fact that the extent of foreground occlusion also vary between scenes (1-3: neonate partially covered by blanket; 4-8: neonate fully covered; 9-10: no blanket) does not affect this conclusion. In overall, the average accuracy of VRM in close-view and wide-view videos are respectively 88.65% and 92.55%. This is because that in wide-view videos, the respiratory RoI is larger than that in close-view videos, i.e., respiratory descriptors dominate the feature space. However in scene 9, the performance of VRM in wide-view video is worse than that in close-view video. This is due to the occurrence of occasional large non-respiratory motions in the wide-view video. Fig. 7 shows some snapshots of the respiratory RoI detected by VRM in neonates monitoring.

5. Conclusion

In this paper, we present a robust video-based respiration monitoring system with automatic RoI detection, comprising three main steps: (1) feature extraction and boosting, where a novel respiratory descriptor is created; (2) respiratory RoI detection and verification, where the RoI containing respiratory motion is detected and non-respiratory motions are rejected; and (3) respiratory rate extraction. The proposed method has been thoroughly evaluated using 148 challenging videos containing seated adults performing guided breathing and neonates. It performs similarly to thoracic impedance plethysmography (ECG) in challenges of various breathing patterns, and shows improved robustness to body-motion but degraded performance in varying lighting conditions. These findings suggest that, regardless of measurement conditions, including lighting settings and distance to camera, our proposed monitoring system is suitable for applications of confined motion, though precision may be reduced for neonate monitoring applications.

6. Acknowledgments

The authors would like to thank Dr. Ihor Kirenko at Philips Research and Mark van Gastel at Eindhoven University of Technology for their support in paper revision, and also the volunteers from Eindhoven University of Technology for their efforts in creating the benchmark dataset.

References

[1] Fieselmann J F, Hendryx M S, Helms C M and Wakefield D S 1993 Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients *Journal of General Internal Medicine* **8** 354–360

[2] Steinschneider A 1972 Prolonged apnea and the sudden infant death syndrome: clinical and laboratory observations *Pediatrics* **50** 646–654

[3] Cretikos M, Bellomo R, Hillman K, Chen J, Finfer S and Flabouris A 2008 Respiratory rate: the neglected vital sign *Med J Aust.* **188** 657–659

[4] AL-Khalidi F, Saatchi R, Burke D, Elphick H and Tan S 2011 Respiration rate monitoring methods: A review *Pediatric Pulmonology* **46** 523–529

[5] Tan K S, Saatchi R, Elphick H and Burke D 2010 Real-time vision based respiration monitoring system *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on* (IEEE) pp 770–774

[6] Alkali A, Saatchi R, Elphick H and Burke D 2014 Eyes' corners detection in infrared images for real-time noncontact respiration rate monitoring *Computer Applications and Information Systems (WCCAIS), 2014 World Congress on* pp 1–5

[7] Yaying L, Yao J and Tan Y 2010 Respiratory rate estimation via simultaneously tracking and segmentation *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on* pp 170–177

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

[9] Li M, Yadollahi A and Taati B 2014 A non-contact vision-based system for respiratory rate estimation *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* pp 2119–2122

[10] Brox T, Bruhn A, Papenberg N and Weickert J 2004 High accuracy optical flow estimation based on a theory for warping *Computer Vision - ECCV 2004* (*Lecture Notes in Computer Science* vol 3024) (Springer Berlin Heidelberg) pp 25–36 ISBN 978-3-540-21981-1

[11] Lindh W, Pooler M, Tamparo C and Dahl B 2010 *Delmar's Clinical Medical Assisting* (Cengage) chap Vital Signs and Measurements, pp 267–269

[12] Balakrishnan G, Durand F and Guttag J 2013 Detecting pulse from head motions in video *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on* (IEEE) pp 3430–3437

[13] Tarassenko L, Villarroel M, Guazzi A, Jorge J, Clifton D and Pugh C 2014 Non-contact video-based vital sign monitoring using ambient light and auto-regressive models *Physiological measurement* **35** 807

[14] Lukac T, Pucik J and Chrenko L 2014 Contactless recognition of respiration phases using web camera *Radioelektronika (RADIOELEKTRONIKA), 2014 24th International Conference* (IEEE) pp 1–4

[15] Wiesner S and Yaniv Z 2007 Monitoring patient respiration using a single optical camera *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE* (IEEE) pp 2740–2743

[16] Poh M Z, McDuff D J and Picard R W 2011 Advancements in noncontact, multiparameter physiological measurements using a webcam *Biomedical Engineering, IEEE Transactions on* **58** pp 7–11

[17] Zhao F, Li M, Qian Y and Tsien J Z 2013 Remote measurements of heart and respiration rates for telemedicine *PloS one* **8** e71384

[18] Brown, Morton B. and Forsythe, Alan B. 1974 Robust Tests for the Equality of Variances *Journal of the American Statistical Association* vol 69 **346** pp 364-367

[19] Bland, J. M. and Altman, D. G. 1986 Statistical methods for assessing agreement between two methods of clinical measurement *Lancet* vol 1 **8476** pp 307-10