Robust Remote Measurement of Respiratory Rate Using Infrared Thermography

Xinchi YU¹, Carina B. PEREIRA¹

¹Chair for Medical Information Technology, Helmholtz-Institute for Biomedical Engineering, RWTH Aachen University, Pauwelsstr. 20, 52074 Aachen, Germany
yu@hia.rwth-aachen.de, pereira@hia.rwth-aachen.de

Abstract. The demand for contactless and unobtrusive techniques for vital signs monitoring has increased during the last years. Current approaches for contactless respiration monitoring mostly detect respiration in the area around the nostrils. This implies that the nostrils must be in the camera’s field of view, which however is not always the case. In this paper, we present a novel approach for robust remote respiratory rate monitoring using infrared thermography. Instead of identifying and tracking the nostrils in thermal video sequences, we use a grid-based approach to detect respiration-induced motion and use sensor fusion to obtain a robust estimate for the respiratory rate. Hence, our method does not require the nostrils to be in the camera’s field of view and thus widens the range of possible applications. We evaluated our algorithm in a small preliminary study and obtained excellent results. During normal breathing a mean RMSE of 0.36 min⁻¹ and mean correlation of 0.98 were achieved, while during simulation of different respiration patterns such as Cheyne-Stokes respiration and tachypnea the mean RMSE was around 3.98 min⁻¹ and mean correlation was around 0.92.

In summary, our algorithm is a promising approach to an easily applicable contactless and unobtrusive respiratory rate monitoring. Furthermore, our method is completely passive and robust against variations in ambient light due to the principle of infrared thermography.

Keywords
Infrared thermography, remote respiratory rate monitoring, signal processing, sensor fusion

1. Introduction

Respiration is an important metabolic process in order to supply the human body with oxygen and remove products of metabolism from blood circulation. Furthermore, irregularities in respiration can be an initial indication for various illnesses. For example, Cheyne-Stokes respiration, which is characterized by (a) an increase in depth of inspiration followed by (b) a decrease in depth of inspiration and finally followed by (c) a phase of apnea, can be attributed to cardiac insufficiency or cerebral damage [1]. Therefore, respiration and respiratory rate are crucial vital signs that need to be constantly monitored on patients. According to the state of the art, technology for respiration monitoring such as impedance pneumography require the use of adhesive electrodes, which can trigger skin disorders or allergic reactions. Moreover, it is impossible to attach adhesive electrodes to patients with burns or neonates due to their extremely sensitive skin. Hence, the need for more widely and more comfortably applicable respiration monitoring methods is obvious. Approaches for contact-free respiration monitoring in current literature include camera-based systems in the visible light [2], near-infrared [3] and far-infrared spectrum [4]. In this paper we present a new method for respiratory rate monitoring based on infrared thermography. The main advantages of the usage of infrared thermography / far infrared over visible light or near infrared is the absolute passivity of the measurement principle (i.e. no active illumination required) and robustness against variances in ambient light conditions. Current proposals for respiratory rate monitoring by infrared thermography such as [5] as well as our previous work [6] mainly use the thermal information around the nostrils. However, this information may be unavailable due to the patient breathing through the mouth or the nose being outside the camera’s field of view. Our new algorithm aims to increase robustness by the introduction of a black box approach which works directly without the need to detect and track anatomic landmarks or specific feature points. For that we divided the image into a regular grid and extracted the respiration waveform from each grid cell based on the respiration-induced movement of the human body (e.g. shoulder [7]). The signals were automatically analysed with regard to their quality and the respiratory signal as well as the respiratory rate were computed from the grid cell with the best signal quality.

Section 2 describes our methods in detail, while section 3 details our experimental setup. The results of our experiments are presented in section 4. Finally, the results are discussed in section 5 and a brief outlook is given.
2. Materials and Methods

2.1. Image processing and acquisition of respiration waveform

Since the detection of breathing-induced motion (e.g. around the shoulders and head) in thermal video sequences is crucial to our algorithm, we first need to separate the subject’s body from the background in each frame of the thermal video. For that we used the fact that the human body is significantly warmer than the background. Hence, we could use Otsu’s method [8] to convert each video frame into a binary image consisting only of the background and the subject’s body (see fig. 1). After the conversion a X by Y pixels sized regular grid is laid over the image and the respiration waveform for each cell is calculated according to

\[ s_c(t) = \frac{1}{X \cdot Y} \sum_{x=1}^{X} \sum_{y=1}^{Y} I_{c,bin}(x, y, t) \]  

(1)

where \( s_i(t) \) is the respiration waveform of cell \( c \) at a given time \( t \) and \( I_{c,bin} \) is the area of the binary image which corresponds to cell \( c \).

![Fig. 1. Left: Original infrared thermogram. Right: Binary version of thermogram.](image)

2.2. Calculation of respiratory rate and Signal Quality Index (SQI)

In order to calculate the respiratory rate in each grid cell \( c \), an analysis window \( s_{c,w} \), consisting of the most recent 15 seconds of the respiration waveform, was used for analysis

\[ s_{c,w}(t) = \sum_{i=t-15}^{t} s_c(t) \cdot \delta(t - i) \]  

(2)

The analysis window was mean centered

\[ s_{c,w}(t) = s_{c,w}(t) - \bar{s}_{c,w}(t) \]  

(3)

and multiplied with a hamming window \( h_w(t) \) before being transformed into the frequency domain:

\[ S_c(f) \quad \rightarrow \quad s_{c,w}(t) \cdot h_w(t) \]  

(4)

Based on the normalized spectrum \( S_{c,norm}(f) \)

\[ S_{c,norm}(f) = \frac{S_c(f)}{\max(S_c(f))} \]  

(5)

the respiratory rate \( r_{resp,c} \) [min\(^{-1}\)] in each cell \( c \) is defined according to

\[ r_{resp,c} = 60 \cdot f(S_{c,norm}(f) = \max(S_{c,norm}(f))) \]  

s.t. \( 0.1 \, \text{Hz} \leq f \leq 3 \, \text{Hz} \) \hspace{1cm} (6)

Iterated over time, this algorithm corresponds to a short time fourier transform (STFT). The window size of 15 seconds results in a good trade-off between observation period and stationarity, while the hamming window reduces the influence of side maxima. The frequency boundaries correspond to respiratory rates between 6 and 180 breaths per minute. While the upper limit covers all possible physiological respiratory rates, the lower limit is restricted by the window length. A dedicated algorithm was implemented for apnea detection.

In order to assess the signal quality, the fourier spectrum was normalized. Since spectral components with a frequency above 3 Hz (henceforth denoted as HF-band) can be attributed to noise, this frequency band in particular is of interest for the analysis of signal quality. Depending on the frequency distribution of the normalized spectrum \( S_{c,norm}(f) \) in the HF-band, a normalized signal quality index with values in the range of \([0, 1]\) is computed. While a SQI of 1 indicates a very good signal quality, the SQI’s value decreases with increasing noise until a value of zero indicates a very bad (noisy) respiration waveform.

Finally, the overall respiratory rate is taken from the cell with the highest signal quality index.

3. Experimental Setup

The infrared thermal video sequences were acquired using a long wave infrared (LWIR) camera (InfraTec VariCam HD Head 820S/30mm, InfraTec GmbH, Dresden, Germany). The camera has an uncooled infrared microbolometer focal plane array with a spatial resolution of 1024 x 768 pixels and a spectral range of 7.5 - 44 \( \mu \text{m} \). In addition, the thermal sensitivity at 30\(^\circ\)C is better than 0.05 K and the images were recorded at 30 frames per second (fps). A preliminary study was carried out with the two authors as healthy subjects. The subjects were sitting on a chair facing the camera, which was placed on a tripod 2m away in front of them. Two video sequences were recorded from

![Fig. 2. Schematic visualization of simulated respiration patterns during the second sequence. Eupnea: segments A, C, F. I. Tachypnea: segments B, H. Kussmaul respiration: segment D. Apnea: segment E. Cheyne-Stokes respiration: segment G.](image)
each subject: In the first sequence (9 minutes), the subjects were sitting still and initially only breathing through the nose (3 min). In the next 3 minutes the subjects only breathed through the mouth and in the last 3 minutes the subjects breathed through the nose and mouth simultaneously. In order to investigate the performance of our method during different types of respiration, the subjects simulated different physiologic and pathologic breathing patterns (eu- pnea, tachypnea, apnea, Kussmaul respiration and Cheyne-Stokes respiration) during the second sequence (10 minutes, see fig. 2). Ground truth was obtained with a piezoplethysmograph (Weinmann Somnolab 2, Weinmann GmbH, Hamburg, Germany), which recorded the thoracic effort.

4. Results

The algorithms described above were implemented and evaluated in Matlab R2015a (The MathWorks Inc., Natick, MA, USA). Figure 3 shows an example for grids cells where respiration was detected. The areas with respiration are the cells around the shoulder and head as expected and referenced in literature [7] [2].

Table 1 shows the results of a quantitative analysis of our algorithms for both subjects in our preliminary study. We used the root mean squared error (RMSE)

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(s(i) - ref(i))^2}{n}} \tag{7}
\]
as a metric for the precision between our result \(s\) and the ground truth \(ref\). Furthermore, we also evaluated the correlation

\[
\rho = \frac{\sum_{i=1}^{n}(s(i) - \bar{s})(ref(i) - \bar{ref})}{\sqrt{\sum_{i=1}^{n}(s(i) - \bar{s})^2 \sum_{i=1}^{n}(ref(i) - \bar{ref})^2}} \tag{8}
\]
between our result \(s\) and the ground truth \(ref\). The results

<table>
<thead>
<tr>
<th>Subject</th>
<th>1st sequence</th>
<th>2nd sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Correlation</td>
</tr>
<tr>
<td>1</td>
<td>0.49</td>
<td>0.97</td>
</tr>
<tr>
<td>2</td>
<td>0.22</td>
<td>0.99</td>
</tr>
<tr>
<td>mean</td>
<td>0.36</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Tab. 1. Results for both the first and second sequence

for the first sequence, where the subjects were breathing normally through the nose and/or mouth, are very good showing a low RMSE and very high correlation. Furthermore, the absolute difference between our results and the ground truth was less than 1 breath per minute in 98.69% of the recording time for the first subject and 100% for the second subject. Figure 4 shows a Bland-Altman plot for both subjects. The low bias of 0.04 min\(^{-1}\) and the small limits of agreement (-0.97 - 1.05 min\(^{-1}\)) stress the good results based on the RMSE and correlation. Only very few outliers are existent.

![Fig. 3. Example of grid cells with detected respiration. The respiration signal in the highlighted (red) cell is visualized at the top part of fig. 5.](image)

![Fig. 4. Joint Bland-Altman plot of both subjects for the first sequence](image)

The results for the second sequence, where the subjects simulated different physiologic and pathologic respiration patterns, are still very good, although the RMSE is significantly higher and the correlation slightly lower. The high RMSE can be attributed to the sudden changes in respiratory rate, which occurred during the transitions between the different respiration patterns. Therefore, a small error in the time synchronization between the ground truth signal and our video signal resulted in a large absolute deviation. This issue is illustrated in fig. 4. Figure 4 also shows the excellent overall agreement between our algorithm and the ground truth. This visual impression is supported by the fact that the absolute error between our algorithm and the ground truth was less than 2 breaths per minute in 99.06% of the recording time for the first subject and 100% for the second subject. The transition-segments between the different respiration patterns have been excluded for this analysis, because the deviations are caused by an imperfect synchronization with ground truth and therefore cannot be attributed to our algorithms. The Bland-Altman plot of both subjects for the
second sequence is shown in fig. 6. Compared to the first sequence, the bias is larger (0.71 min$^{-1}$), as well as the limits of agreement (-7.47 - 8.91). However, this was expected due to the larger range of respiratory rates. Moreover, areas with large deviations from ground truth are highlighted. While the areas marked in grey depict the transition between respiration and apnea, the area highlighted in yellow represents the transition from tachypnea to eupnea at the end of the sequence (also see fig. 4). The Bland-Altman plot in combination with fig. 4 shows that the rather large RMSE is only caused by errors in time synchronization between the video signal and ground truth. In conclusion, the performance of our method even during various respiratory rates and respiration patterns is excellent.

5. Discussion and Outlook

In this paper, we presented a novel approach for unobtrusive respiratory rate monitoring. Our approach serves the increasing demand and need for new contactless and unobtrusive vital sign monitoring solutions. In contrast to other proposed methods, our algorithm does not rely on the identification and tracking of (anatomic) feature points such as the nose. Instead, respiration is detected across the whole thermal image using a grid. With this measure, our approach is more widely applicable, e.g. if the nose is outside the camera’s field of view. Moreover, our method is fully automatic and does not require any user interaction. Since our algorithm was only evaluated in a preliminary study with two subjects, we are planning a larger clinical study to obtain more data for a more detailed validation. Although all algorithms are causal and thus real-time capable, our current implementation’s performance is rather slow. Additional efficiency improvements in the code are needed to ensure actual real-time capability. Finally, we are investigating the influence of background noise on the algorithms and means to further reduce requirements on image quality and preprocessing steps.

Acknowledgements

Research described in the paper was supervised by Prof. S. Leonhardt and Prof. V. Blazek, RWTH Aachen University, Germany. Carina Barbosa Pereira wishes to acknowledge FCT (Foundation for Science and Technology in Portugal) for her PhD grant SFRH / BD / 84357 / 2012.
References


About Authors...

Xinchi YU was born in Berlin, Germany. He studied Electrical Engineering at the RWTH Aachen University and received the M.Sc. degree in 2015. He is now a Ph.D. student at the Philips Chair for Medical Information Technology, RWTH Aachen University, Aachen, Germany. His main research fields are geriatrics and unobtrusive vital signs monitoring.

Carina PEREIRA was born in Braga, Portugal. She received the M.Sc. degree in Biomedical Engineering from the University of Minho, Braga, Portugal, in 2011. Currently, she is working toward the doctor degree at the Philips Chair for Medical Information Technology, RWTH Aachen University, Aachen, Germany. Her research interests include infrared thermography, imaging and biosignal processing as well as feedback control in medicine.