Breathing Rate Monitoring during Sleep from a Depth Camera under Real-life Conditions

Manuel Martinez Rainer Stiefelhagen
Karlsruhe Institute of Technology
{manuel.martinez, rainer.stiefelhagen}@kit.edu

Abstract

Computer vision is a non-invasive way to supervise patients in the bed. We introduce a novel algorithm that monitors breathing rate from a depth camera placed above the bed. While visually registering breathing rate has raised significant interest in the health care field, most published approaches are evaluated only on constrained or simulated settings. Conversely, we evaluate our method in a real dataset consisting of 3,239 segments collected from 67 sleep laboratory patients. Our method introduces three novel contributions: a dynamic Region of Interest (RoI) which is aligned to the bed, a confidence metric based on patient agitation, and the Early Fourier Fusion strategy. Overall, our camera based method is accurate on 85.9% of the segments. This performance is similar to the obtained from a chest sensor (88.7%). Most importantly, we report the performance impact related to different sleep conditions, like apnea, position and staging.

1. Introduction

We present a novel algorithm that monitors breathing rate continuously during sleep using a depth camera. Our monitoring device, which requires no calibration, is attached to the ceiling above the bed (see Fig. 1), and is completely autonomous, non-invasive and easy to install and use.

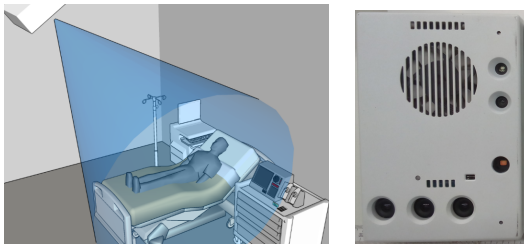


Figure 1: We monitor the breathing rate of a patient from a depth camera attached to the ceiling (left). The device used in this experiment (right) contains, among other sensors, a depth camera and an infrared camera, allowing us to record at night.

Registering the breathing rate from a depth camera in ideal conditions is simple and has been done multiple times. However, the main application of this technology is to monitor elderly people who often have significant medical conditions which alter their sleep and breathing patterns, *e.g.*, insomnia, and apnea. We aim to develop technology that improves the quality of life of elderly people for both nursing homes and ageing-at-home scenarios.

As our setting is far from ideal, we need to deal with its associated challenges, namely a very large distance between patient and camera, and the different sleep conditions that the patient may present. To this end, we evaluate our algorithm on 67 patients referred to a sleep laboratory with various degrees of sleep apnea. To the best of our knowledge, this is the first system that has been evaluated in

This work is supported by the German Federal Ministry of Education and Research (BMBF) within the SPHERE project.

real-life conditions on apnea-prone patients.

Respiration in real-life conditions is a complex signal to retrieve. Respiration is a semi-autonomous mechanism where the muscles involved are generally controlled unconsciously, but we can also control them at will. In fact, we interrupt our breathing motion often to speak, cry, cough, or move our body. Furthermore, there are several conditions that interrupt or alter our breathing, such as apnea.

The semi-autonomous nature of breathing becomes a significant problem if we monitor breathing rate continuously. In medical settings, there are two main ways to register breathing rate: instant measurements, or continuous measurements. Instant measurements are usually taken by a nurse, where the patient is told not to move or talk while the nurse counts the number of chest excursions. In case of coughing, talking or agitation, the nurse can simply repeat the measurement.

Conversely, continuous monitoring is a more complex task. In sedated patients (*e.g.*, Intensive Care Units) an airflow measuring mask is used, but to diagnose respiration conditions during sleep, doctors perform a multi-sensor analysis named polysomnogram.

We aim to provide continuous breathing rate monitoring in domestic and elderly care environments, where polysomnograms are not cost-effective and excessively intrusive.

Our system improves over the state of the art by using a dynamic region of interest aligned to the bed, a novel confidence test based on agitation, and an Early Fourier Fusion approach that allows us to combine out-of-phase breathing signals in a constructive way and thus recover a strong signal from a noisy sensor.

We evaluate our algorithm on 67 real patients and use the polysomnogram as a reference. This way we can measure how much different sleep conditions impact our accuracy. We found out that breathing rate can be accurately recovered from any sleep position and sleep stage, but apnea events degrade our estimates significantly. Compared to a chest band, our algorithm achieves similar quantitative and qualitative results.

2. Previous work

Due to the large cost and invasiveness of the polysomnogram, many alternative sensing methods have been proposed like microwave radar [5], audio radar [11], motion sensors [6], laser interferometry [15], and cameras [1, 2, 3, 8, 10, 12, 16, 17].

Computer vision systems are becoming popular because they are cost effective and can be used for more than one monitoring task, however, are affected by illumination and point-of-view changes and require more advanced signal processing than approaches that measure chest movement directly. To simplify the problem and obtain a more robust solution, cameras with different modalities have been used. Thermal cameras can visualize the temperature of the exhaled air, however are expensive and localizing the region of interest is difficult [17]. Infrared cameras are inexpensive and can capture images during the night albeit they loose the color information. Typical algorithms used are temporal differences [16] and optical flow [10].

3D approaches are based on stereo, structured light [1], and time-of-flight [4] cameras. Stereo cameras depend on textured patterns on the scene, which are often not available in hospital environments. Structured light and time-of-flight cameras are illumination independent and require no texture, hence are preferred [1, 2, 3, 8].

Most relevant to our work, Nakajim et Al. [10] used optical flow on infrared cameras for sleep analysis, Poh et Al. [12] noted how accurate filtering can detect a fine vital signal (*i.e*. heath rate) from color images, Centonze et Al. [4] used Kinect v2 with a manual a Region of Interest to capture a precise respiratory signal, this signal was used to analyze breathing rate in different sleep stages, and Wang et Al. [16] used infrared cameras on 15 simulated patients to detect apnea episodes using an expert system, heavily parametrized.

Few approaches provide a fully automated analysis system, from the sensor to the actual analysis result. Most require heavy manual calibration and parameter tuning, and several require manual steps like selection the Region of Interest, or visually evaluating the results.

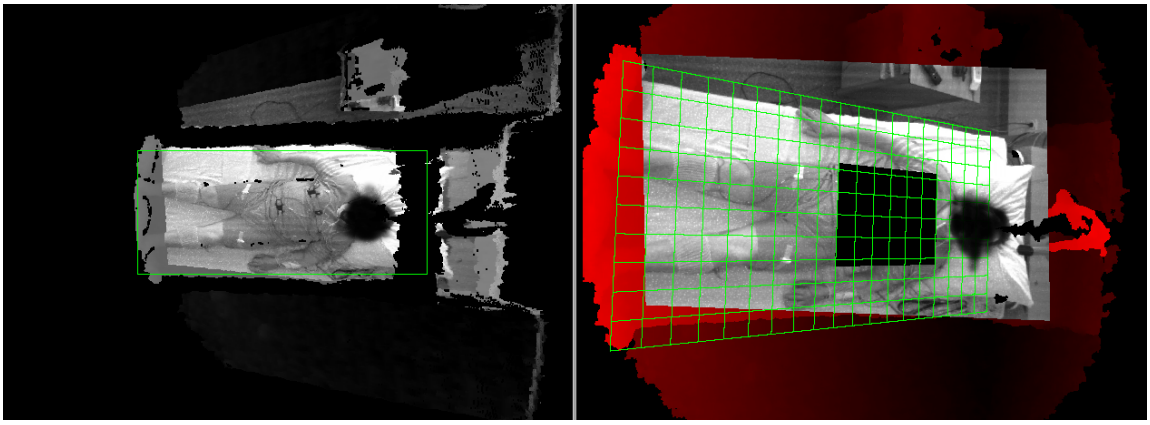


Figure 2: Left: artificial top-down view generated from a Bed Aligned Map (BAM). The polysomnogram sensors are visibly attached to the patient, however most patients slept under a blanket. Right: raw depth map with the infrared camera superimposed. The black square is our bed aligned Region of Interest, which is the same for all patients. Face obscured to preserve privacy.

3. Methodology

3.1. Dataset

We captured a dataset from a sleep laboratory. Patients were simultaneously recorded using our recording system (Fig. 1) and a polysomnogram (Fig. 2) which we use as a reference. The dataset is annotated by the sleep laboratory doctors with labels for sleep position, stage, and apnea events.

Our monitoring system contains several cameras alongside other sensors. In this study we use the infrared camera (752x480@10Hz) and the depth camera (PS1080, 640x480@30Hz).

We recorded 94 sleep analysis sessions of 67 patients from three rooms. Three bed sizes are used, and the camera to chest distance is between 4 and 5 meters. From the polysomnogram, we use signals from the four contact sensors that are related to respiration: a **thermistor** placed under the nose measures the temperature differential, a **barometer** placed also under the nose measures the pressure differential, a **chest band** measures the extension of the thorax, an **abdomen band** measures the extension of the abdomen.

We took 40 samples each night (a total of 3,760) containing many challenging situations: empty beds, patients sitting, changing sleep positions, having apneas, etc. We discarded samples if at least one polysomnogram sensor was disconnected, reducing the total number to 3,239. We use a window length of 30 seconds in order to obtain results that are comparable to the ones captured manually by doctors and nurses.

All patients were informed of the procedure and but we did not place any limitations on their actions or routines. Therefore, patients used at will: blankets of various thicknesses, a variable amount and size of pillows, and several read books, newspapers and magazines during the recording. Those variables had no discernible impact on the results.

There is no uniquely defined ground truth for the number of breaths of a sequence, precisely because the definition of what constitutes a single breath is fuzzy in several corner cases (*e.g.*, interrupted breaths, minimal expansion, coughing, etc.). Based on the recommendations of our colleagues from the sleep laboratory, we use the estimate obtained from the thermistor placed under the nose as a reference.

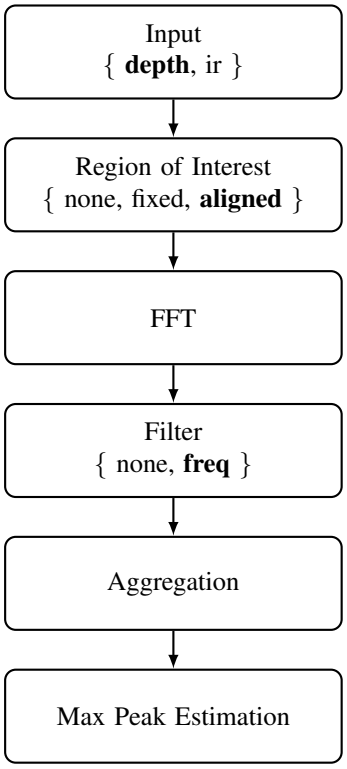


Figure 3: Workflow, default settings are in **bold**

3.2. Breathing rate estimation

Our algorithm (Fig. 3) takes as input a sequence of images, obtains a Region of Interest (RoI), and for each pixel within the RoI creates a trajectory in the time domain, which contains all the pixel values over the sequence.

The Power Spectral Density (PSD) of every trajectory is calculated using fast Fourier transform, and a filter is optionally applied to each PSD.

The PSDs from all trajectories are aggregated together, and the location of its peak is found using quadratic interpolation.

Note that the information about the frequency is on the location of the peak, not its magnitude, hence the units and magnitudes actually used for the trajectories are irrelevant.

3.2.1 Input sources

We evaluate both depth and infrared cameras as input sources. In both cases the images were compressed using a lossless algorithm, as we found out that even the slightest image degradation had a strong negative performance impact.

The depth camera provides distance readings, and thus is preferred to monitor respiration related movements. However, the resolution and noise of depth cameras degrade easily with the distance, and at the 4-5 meter range used in our setting the PS1080 provides very noisy data. Only by considering the sequence as a whole, the main spectral component can be recovered.

Trajectories from the infrared camera represent the intensity variation of a pixel over the sequence. Although its performance depends strongly on the environment, light conditions, and available texture, it is intrinsically less affected by the distance to the chest, and has a significantly better effective resolution (for the effective resolution of PS1080 cameras, see [9]). Furthermore, infrared cameras are less expensive than depth cameras, so it is worth to evaluate its performance.

3.2.2 Dynamic Region-of-Interest

We boost the signal-to-noise ratio by dropping out non-relevant pixels using a Region-of-Interest (RoI).

We suggest to use a dynamic RoI that is anchored to the bed position. This is necessary for unattended monitoring, as our camera is fixed to the ceiling, but beds may change position often (*e.g.*, in hospitals and nursery homes).

The bed is automatically located once per night using the Bed Aligned Map algorithm [7] and the region of interest is centered on the chest area, as seen in Fig. 2.

We also evaluate a fixed RoI. The fixed RoI is equivalent to the dynamic RoI, only that the bed position is not updated for each recording, and thus it is not dynamically aligned to the bed.

3.2.3 Early Fourier Fusion

To generate a strong signal we must fuse as many trajectories as possible. There are two main trends on how to perform the fusion. We can simply aggregate all temporal trajectories, hoping that the noise gets canceled while the main signal raises above the noise level. Or we can do some intelligent fusion using either Principal Component Analysis [14, 8] or Independent Component Analysis [13].

In both cases, it is assumed that the signal we capture is in phase on all trajectories, however this is not the case for breathing. As we breath, we move our environment, parts of the bed clothing rise, while other parts sink. Those parts would have the same frequency, but different phase.

This is the main motivation behind our Early Fourier Fusion technique. Instead of fusing temporal trajectories and estimate the PSD of the fused trajectory, we first estimate the PSD of each individual trajectory before fusing them. As the PSD effectively eliminates the phase of the signal, it ensures that there is no destructive inference between two useful signals.

3.2.4 Frequency filter

Before aggregation, we apply a very conservative filter to each PSD: we keep only the trajectories whose mean power/BPM is larger inside the region between 6 BPM and 60 BPM than outside. This effectively limits our range of detected frequencies from 6 BPM to 60 BPM, which is an acceptable range to detect using a 30 second window.

3.3. Confidence test

As our system performs continuous monitoring instead of independent measurements, we need to deal with cases where the bed is empty, or the patient is coughing, talking or changing sleep positions, etc. A nurse would not measure breathing rate in those situations where the outcome would not be considered reliable. Thus an automated monitoring system should not report the breathing rate in those situations. However, this raises the challenge of detecting unreliable situations.

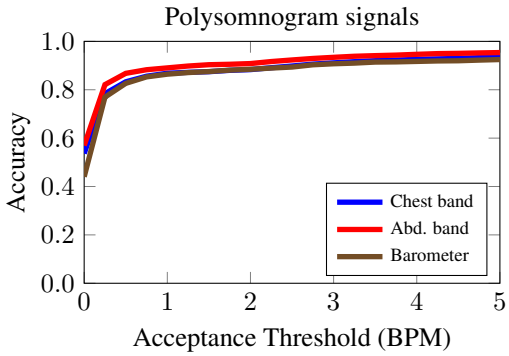


Figure 4: Breathing rate predicted from other polysomnogram signals is accurate to ± 1 Breaths per Minute (BPM) only in 87% of the sequences.

We suggest a confidence test based on agitation. If we detect minimal levels of agitation, we expect the bed to be empty and thus we discard the segment. If we detect excessively high levels of agitation, we expect that a singular event happened (change in sleep position, coughing, etc.), and we discard the estimation.

We use an agitation metric based on BAMs [7]. It is defined as the volume between the maximum and the minimum convex hull of the bed within a second. We average the agitation over the sequence, and we set the low and high thresholds at $0.03m^3$ and $0.6m^3$ respectively.

4. Evaluation

We use acceptance curves for evaluation. The **x** axis is the acceptance threshold, while the **y** axis is the percentage of samples that provide an estimate within the acceptance threshold to our reference.

4.0.1 Baseline

As a baseline, we evaluate how well can we predict our reference breathing rate, obtained from the nose thermistor, using the other polysomnogram signals as a source (see Fig. 4).

We can see that an agreement better than 1 breath per minute (BPM) is achieved in only 87% of the sequences.

4.1. Algorithm

4.1.1 Sensor input

Our results show how the depth camera provides significant better performance than infrared images, as expected (see Fig. 5.a). In the hospital setting, where our dataset was captured, the bed clothing has very little texture where changes in the infrared image can be tracked. Therefore we use the depth camera as default.

4.1.2 Region of Interest

Our experiments show that using a RoI is critical to obtain good results (see Fig. 5.b). In our dataset there is little variation between bed positions, hence the fixed RoI, whose position is fixed relative to the camera field of view, provides a large performance boost. However, the dynamic RoI, whose position is relative to the bed localization, is even better.

These results suggest that there is still room for improvement when selecting the right RoI size. The RoI size needs to balance two properties. A large RoI will capture breathing in segments where the patient is lying on the edges of the bed, or in less common positions. But a smaller RoI would provide better signal-to-noise ratio, and thus improve the accuracy if the patient is breathing faintly. As a consequence, we expect that adapting dynamic also the RoI size would improve the results.

4.1.3 Merging Style

Given the distance between the camera and the patient, Principal Component Analysis (PCA) based algorithms do not provide acceptable results, in particular, if a denoise step like Durbin-Watson is used (see Fig. 5.c).

The Durbin-Watson statistical test is meant to remove trajectories with low autocorrelation, however the minimum resolution of the depth camera at such distances is so large that the signal appears randomly dithered, and thus shows very little autocorrelation. The same applies for the correlation between different trajectories, which is minimal, hence the failure of the PCAs.

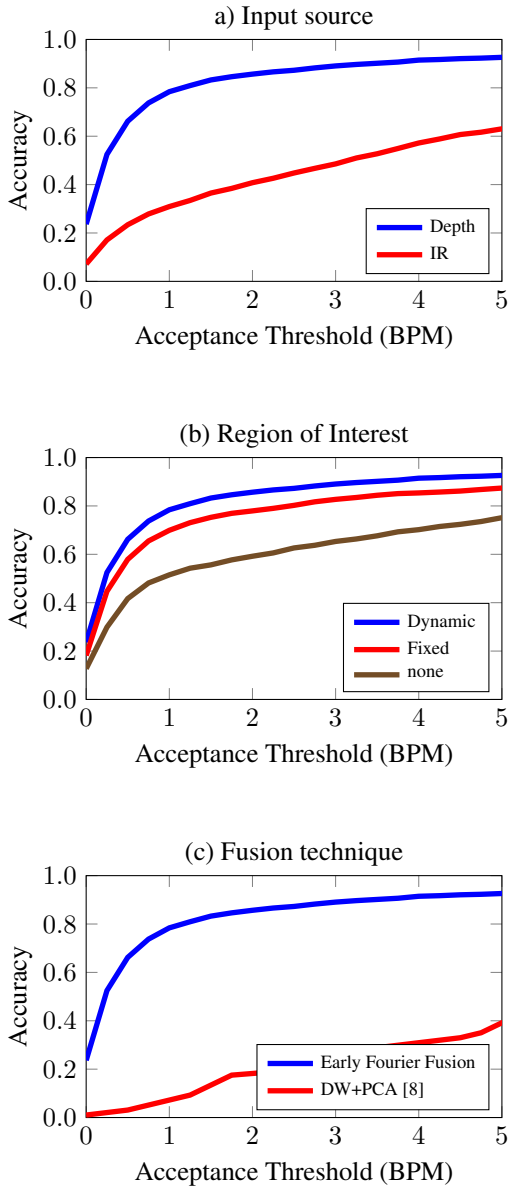


Figure 5: Accuracy under different algorithm configurations: (a) the depth camera performs significantly better than an infrared camera as a source, (b) the dynamic RoI provides a performance edge over the fixed RoI, (c) our Early Fourier Fusion outperforms a traditional fusing algorithm (Durbin-Watson+PCA).

4.2. Sleep conditions analysis

4.2.1 Sleep position

Although the breathing motion of the chest is not directly observable if the patient is lying in a lateral position, the motion is transferred to the surroundings of the patient (bed clothing, arms, pillow) where it can be observed. Hence, the sleep position of the patient has no significant impact on the accuracy (see Fig. 6.a). Albeit the supine position shows the worst performance, this happens because most apnea events occur while the patient is in supine position.

4.2.2 Sleep stage

Results for sleep stage accuracy exhibit our expected behavior as the system performs better if the patient is relaxed (see Fig. 6.b). At the same time, patients have difficulties to sleep deep when having apneas, thus deep sleep is correlated with sequences with low rate of apneas.

4.2.3 Apnea events

Apnea events have a large accuracy impact (see Fig. 6.c). Hypoapneas are defined as a 30% reduction of air flow for more than 10 seconds, therefore breathing is usually shallow, but nonetheless existent. Central apneas represent a breathing pause larger than 10 seconds, which then resumes uneventfully. As 10 seconds are a significant part of the 30 second segment we use, the signal degradation is larger.

Obstructive apneas happen when the upper respiratory tract is obstructed: the diaphragm tries to expel the air from the lungs, but only manages to send it to the stomach, where it is sent again to the lungs. This stage is known as paradoxal breathing, and is manifested in the polysomnogram as a phase shift between chest and abdomen breathing signals. This stage is interrupted when the oxygen saturation in the blood drops enough to arouse the patient, who makes an conscious effort to unlock the airways. The breathing rate is non regular, and thus do not express a large peak on the spectra.

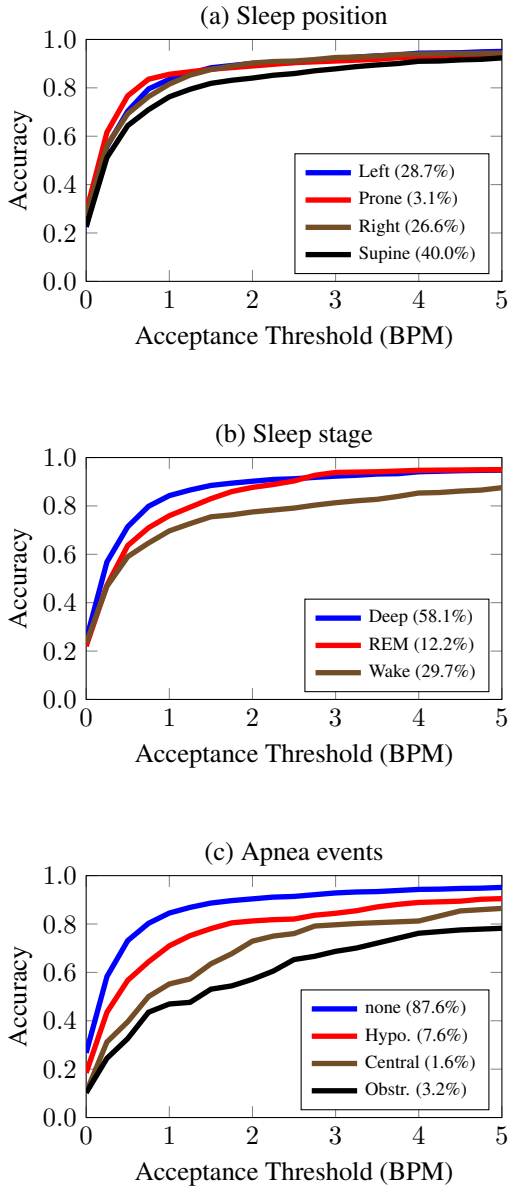
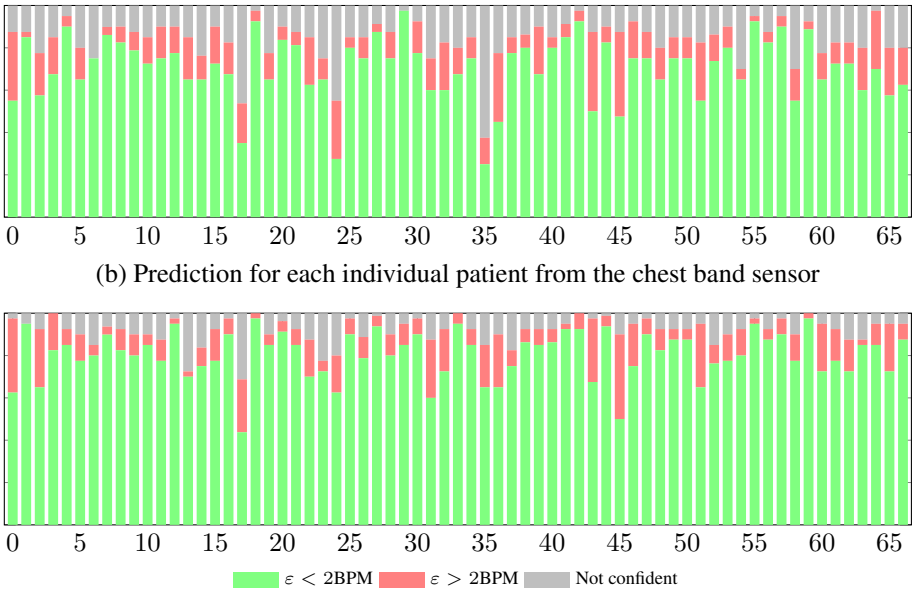


Figure 6: Accuracy under different sleep conditions: (a) sleep position has no significant performance impact, (b) the algorithm performs better if the patient is sleeping, (c) the performance of the system is degraded during apnea events, in particular during obstructive apnea. The number enclosed in parentheses is the incidence rate of each class.

(a) Prediction for each individual patient from depth camera



(b) Prediction for each individual patient from the chest band sensor

Figure 7: We evaluate our system's performance on each patient. Top: results from our algorithm (accuracy 85.9%). Bottom: Results from the chest band used in the polysomnogram (accuracy 88.7%). Note the behavior similarity between the depth camera and the chest band sensor.

4.3. Per patient evaluation

We evaluate our system predictions for each patient, and compare to the breathing rate predictions from the chest band. In our representation, each sequence is color coded. Green represents that the sequence produced a confident estimation, and it was within 2 BPM of the reference, the thermistor. Red means that the sequence produced a confident estimation, and it was not within 2 BPM of the reference. Gray indicates that the sequence did not pass the confidence test (see Fig. 7).

This way we can evaluate if the results are consistent among all patients. We found out that, for most patients, our system behaves like the chest band. However we have two outliers: patient 24 and 35 were significantly better recognized using the chest band than from the depth camera. Two patients out of 67 are not significant enough to extract conclusions, therefore further research is required.

5. Conclusions

We have presented an algorithm that monitors breathing rate during sleep from a depth camera in a very challenging setting: a sleep laboratory.

Our Early Fourier Fusion algorithm fuses the information from different parts of the image at the Power Spectral Density level, allowing us to combine signals that are not in phase.

Most importantly, we evaluated our approach on 67 patients from a sleep laboratory, showing that breathing rate can be detected accurately under different sleep positions and stages, however respiratory conditions like apnea can reduce the performance of the system significantly.

Our system achieves similar results of that of a contact chest sensor. This indicates that computer vision is precise enough for the task, and to improve our results we would need to improve the recognition algorithm instead.

References

[1] H. Aoki, Y. Takemura, K. Mimura, and M. Nakajima. Development of non-restrictive sensing system for sleeping person using fiber grating vision sensor. In *Micromechatronics and Human Science*, 2001.

[2] F. Benetazzo, A. Freddi, A. Monteriù, and S. Longhi. Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results. *Healthcare technology letters*, 1(3):81, 2014.

[3] N. Burba, M. Bolas, D. M. Krum, and E. A. Suma. Unobtrusive measurement of subtle nonverbal behaviors with the microsoft kinect. In *International Workshop on Ambient Information Technologies*, 2012.

[4] F. Centonze, M. Schatz, A. Prochazka, J. Kuchynka, O. Vysata, P. Cejnar, and M. Valis. Feature extraction using MS Kinect and data fusion in analysis of sleep disorders. In *International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, 2015.

[5] J. Geisheimer and I. Greneker, E.F. A non-contact lie detector using radar vital signs monitor (RVSM) technology. *Aerospace and Electronic Systems Magazine*, 16(8):10 –14, 2001.

[6] S. Gradl, H. Leutheuser, P. Kugler, T. Biermann, S. Kreil, J. Kornhuber, M. Bergner, and B. Eskofier. Somnography using unobtrusive motion sensors and Android-based mobile phones. In *International Conference of Engineering in Medicine and Biology Society (EMBC)*, 2013.

[7] M. Martinez, B. Schauerte, and R. Stiefelhagen. BAM! depth-based body analysis in critical care. In *Computer Analysis of Images and Patterns (CAIP)*, 2013.

[8] M. Martinez and R. Stiefelhagen. Breath Rate Monitoring During Sleep using Near-IR Imagery and PCA. In *International Conference on Pattern Recognition (ICPR)*, 2012.

[9] M. Martinez and R. Stiefelhagen. Kinect unleashed: Getting control over high resolution depth maps. In *International Conference on Machine Vision Applications (MVA)*, 2013.

[10] K. Nakajim, Y. Matsumoto, and T. Tamura. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiological Measurements*, 22(3), 2001.

[11] R. Nandakumar, S. Gollakota, and N. Watson. Contactless sleep apnea detection on smartphones. In *International Conference on Mobile Systems, Applications, and Services*, 2015.

[12] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Noncontact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 2010.

[13] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011.

[14] B. Prathyusha, T. S. Rao, and D. Asha. Extraction of respiratory rate from ppg signals using pca and emd. *International Journal of Research in Engineering and Technology*, 1(2):164–184, 2012.

[15] L. Scalise, I. Ercoli, P. Marchionni, and E. Tomasini. Measurement of respiration rate in preterm infants by laser Doppler vibrometry. In *Medical Measurements and Applications Proceedings*, 2011.

[16] C.-W. Wang, A. Hunter, N. Gravill, and S. Matusiewicz. Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea. *IEEE Transactions on Biomedical Engineering*, 61(2):396–404, 2014.

[17] Z. Zhu, J. Fei, and I. Pavlidis. Tracking human breath in infrared imaging. In *Bioinformatics and Bioengineering*, 2005.