Inhaltsverzeichnis

Sitzung ABC

The SPHERE Project: Sleep Monitoring using Computer Vision

Manuel Martinez and Rainer Stiefelhagen

Karlsuhe Institute of Technology, Institute for Anthropomatics and Robotics,
Computer Vision for Human Computer Interaction (cv:hci) Lab,
Vincenz-Priessnitz-Str. 3, 76131 Karlsruhe, Germany
manuel.martinez@kit.edu rainer.stiefelhagen@kit.edu

Abstract The aim of the SPHERE project is to develop a compact device that monitors sleep using computer vision. Our system attaches to the ceiling above the bed and integrates infrared and depth cameras, alongside other auxiliary sensors. We installed systems in a nursery home and a sleep laboratory, allowing us to evaluate algorithms that analyze respiration, sleep position and agitation. Compared to other sleep monitoring modalities, computer vision is non-intrusive, and provides a holistic understanding of the bed environment, enabling better alarm systems and cleaner sleep summaries.

Keywords sleep monitoring, computer vision, respiration.

1 Introduction

Lack of sufficient quality sleep can lead to mental and physical health problems, diminished awareness status and reduced quality of life. This is aggravated on elderly patients, as the ability to sleep deteriorates with age.

There are effective ways to treat most sleep disorders, but they need to be diagnosed first. To perform such diagnoses, patients spend the night at the hospital while being monitored by up to a hundred contact sensors. The sensor report (named polysomnogram) is then reviewed manually. This protocol has significant human and material costs, resulting in waiting lists of several months long.

In nursery homes it is not feasible to throughly monitor the sleep quality of every resident, as a result, most sleep disorders stay undiagnosed. Night watch nurses already need to deal with residents falling out of the bed, having panic attacks, etc. To assist in their task, modern nursery homes have all sorts of intelligent sensors (infrared, or pressure sensors usually) to detect if a bed is occupied or empty, or if a person has fallen out of it. The aim is to minimize the duration between an accident and the arrival of the assistance. The sensors are very sensitive to avoid missing actual dangerous events, therefore they often trigger false alarms, and thus are a source of alarm fatigue, which lowers their usefulness.

We created the SPHERE project with the aim to help to improve the diagnose rate of sleep disorders by monitoring sleep quality indicators, and to generate more reliable alarms for accidents. By using Computer Vision we can explore the bed and its surroundings, allowing us to develop holistic algorithms that provide better understanding of the situation.

We have developed a Medical Recording Device (MRD) [1] with several cameras alongside other auxiliary sensors that works autonomously. We collaborate with a nursery home and the sleep laboratory from the ThoraxKlinikum Heidelberg (THX), who is interested in portable monitoring systems.

Our collaborations are a great advantage over alternative studies which rely on simulated patients and environments. Our experiments and results reveal a huge performance gap between simulated and real scenarios, the latter clearly being more challenging. We leverage on this advantage by learning from the data we collected, and design our algorithms accordingly.

In this paper we describe in detail the problem of continuously monitoring respiration rate on real patients, and then we show state-of-the-art algorithms for sleep position and agitation quantification, which we use to provide nighly sleep quality summaries

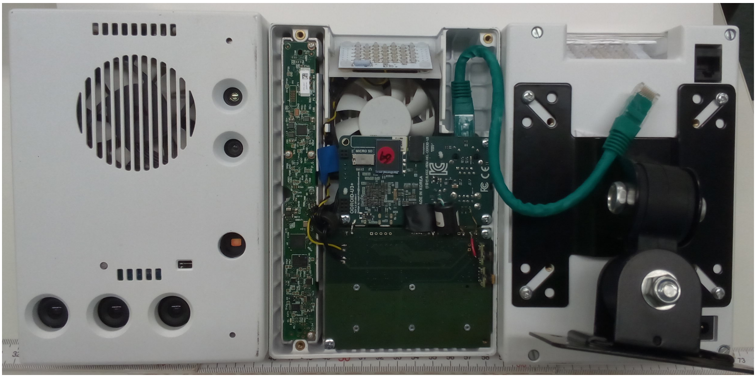

Figure 1.1: Left to right: front view of our Medical Recording Device, view of the internals, and rear view (including VESA mount). The device has five PCBs of which three have been custom designed for the project.

2 Medical Recording Device

Our Medical Recording Device (MRD) [1] is compact (120x180x55mm) and attaches to the ceiling above the bed via a standard VESA mount. Integrates depth [2, 3], stereo and mono cameras; stereo microphone, temperature, pressure, humidity and light sensors; a 4-core ARM CPU with storage, ethernet, WiFi, and bluetooth.

As we need to record in the darkness, we use infrared sensible image sensors with active infrared illumination. The infrared light is projected to the ceiling which reflects over the patient, providing uniformly illuminated images.

The device requires a fan for cooling, but its speed is dynamically controlled and the device is virtually silent by night. It is CE compliant, in order to be installed in hospitals and nursery homes.

We use the hardware compression engine to compress the grayscale images in h264 format, while the quadcores is used to compress the depth maps from the camera using a custom lossless codec.

Our management software recovers automatically from any major flaw, our 7 installed units have accumulated more than 5,000 hours of continuous use without malfunction.

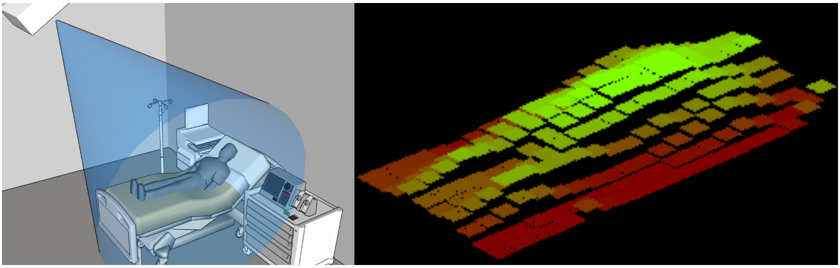

Figure 1.2: Left: our Medical Recording Device (MRD) on the ceiling of an Intensive Care Unit (ICU). Right: the Bed Aligned Map (BAM), a low resolution, height-based descriptor aligned to the bed. Obtained from a depth camera, its privacy conscious and robust to light and orientation variations.

3 Bed Aligned Maps

We capture spatial information from a depth camera and use the bed position, which is automatically estimated, to align the point cloud. The bed matress is divided in equal sized cells, and the mean cell height above the mattress is stored, as seen in Fig. 1.2. We call the resulting low dimensional descriptor Bed Aligned Map (BAM) [4].

Resolution is an important trade-off for BAM. Lower resolution provides better depth estimates, minor storage requirements, and better privacy protection. However, too low of a resolution may discard important spatial information. Unless stated otherwise, we use 10cmx10cm cell BAMS, which translates to a descriptor size between 8x20 for the most narrow bed in our database, to 13x20 for the widest.

BAMs are scale, orientation, light and alignment independent, while occlusions are generally filtered out by a raytracing algorithm. This not only makes our algorithms robust to the common ailments of Computer Vision, but also reduces the amount of data we need to collect to train our machine learning tools, as the differences induced by varying scenarios are reduced.

Furthermore, storing BAMs has practical advantages over storing RAW image data: it reduces the storage requirement, which are substantial when performing long term sleep monitoring, and BAMs are ethically friendly, as the subjects are not recognizable.

4 Respiration Analysis

Against the general impression, respiration is a very complex signal to retrieve in real conditions. The respiration control system is semi-autonomic: the muscles involved can be voluntarily controlled, but the autonomous system will take care as soon the voluntary control stops. This is important as our respiration pattern changes when we speak, or move our body or become agitated. There exist multitude of events that alter our breathing, some are very common (*e.g.* snoring, coughing), and some are less common but important nevertheless, like obstructive apnoea. In obstructive apnoea the upper airways are blocked and the diaphragm moves the air from the lungs to the stomach and back, resulting in chest motion but no air exchange.

If an instant breath rate measurement is required in an hospital, it is usually taken by a nurse. The patient will be *told not to move or talk* for a while, and the nurse will count the number of chest excursions during a set period of time. In case of coughing or agitation, the nurse will *repeat the test*.

On polysomnograms, respiration is monitored using no less than 5 sensors: a **thermistor** placed under the nose measures the temperature differential, a **barometer** placed 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, finally, a **video stream** records the full session.

Measuring the respiration rate can be as simple as counting chest excursions, but only on very simple scenarios. In SPHERE we want to evaluate how well can we *continuously* estimate breath rate using a depth camera to measure chest excursions in an unconstrained scenario.

4.1 Methodology

We use a dataset obtained from the ThoraxKlinik Heidelberg containing 99 recorded polysomnogram sessions (81 different patients). We took 40 samples for each night, generating a total of 3960 samples, each being 30 seconds long. Several samples contain challenging situations: empty beds, patients sitting, changing sleep positions, having apnoeas, etc. We use the thermistor signal as a reference for evaluation purposes.

From our MRD, we use a grayscale camera (752x480@10Hz) and a

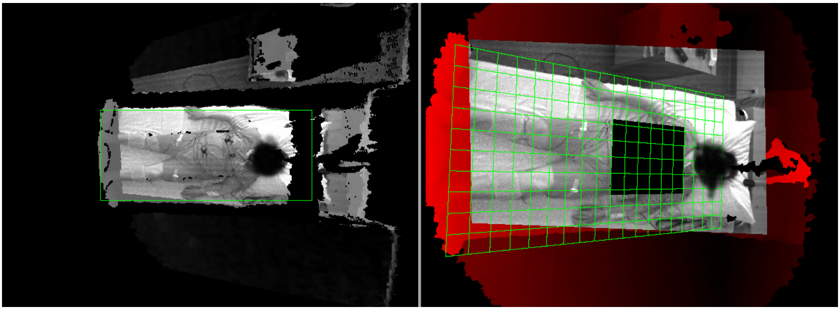

Figure 1.3: Left: Bed Aligned Map (BAM) generated depth map with infrared image overimposed. Right: Raw disparity image with infrared camera superimposed. In green are displayed the BAM grid used for alignment. The black square is the Region of Interest selected (the same for all patients). Face is hidden to preserve privacy.

depth camera (PS1080 based, 640x480@30Hz). Three different bed sizes are used in the study.

To obtain the breathing rate we calculate the Power Spectral Density (PSD) with 8x interpolation. On 30 second windows, this gives us a resolution of 0.25 Breaths Per Minute (BPM). The breathing rate reported is given by the position of the largest peak.

Our evaluation is performed using an acceptance curve: on the **x** axis we have our acceptance threshold, and on the **y** axis we plot the percentage of samples that provide estimates within the threshold distance to our reference (the thermistor signal).

4.2 Breathing rate recognition from images

In a previous work [5], we showed analytically that the signal-to-noise ratio for the breathing signal is inversely proportional to the 4th power of the distance when captured by cameras. Most studies place the depth camera at distances between 70cm and 1m [6] to the chest, at those distances, no signal processing is needed to obtain a clean signal. However, as we need to attach the camera to the ceiling to capture the whole environment without obstructions, our distance to the chest is around 4 meters. At 4 meters, the breathing signal we record is 256 times (24dB)

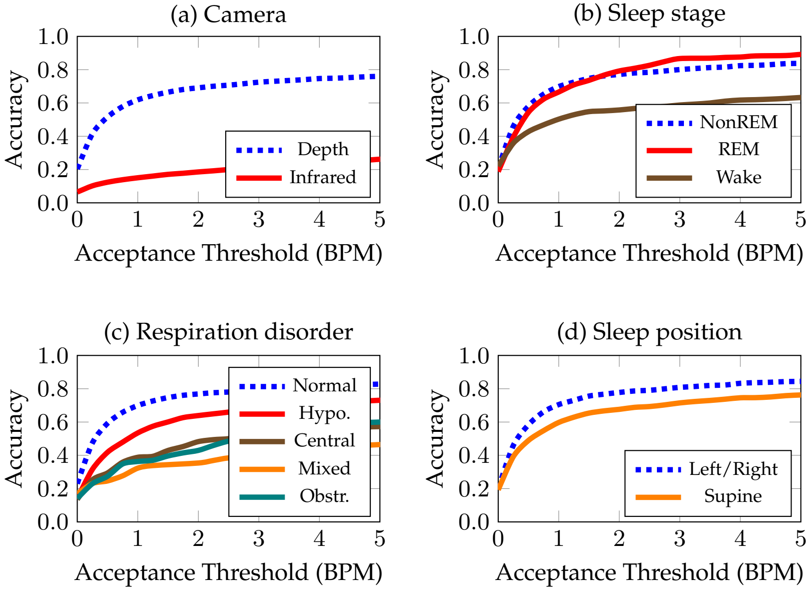

Figure 1.4: The acceptance curves for our breath rate recognition algorithm show the ratio of samples providing an estimate within the accepted threshold. We evaluate the algorithm under different sensor modalities (a), sleep stages (b), respiration disorders (c) and sleep positions (d).

weaker than at 1 meter, therefore it is crucial to perform signal processing to recover the breathing signal from the background noise.

Our previous approach used PCA combined with Durbin-Watson filters to fuse trajectories [5]. This approach aggressively discards noisy samples to create a very clean signal estimate, however at 4 meters all samples are noisy, and the approach fails to produce an estimate at all.

We use a simpler fusion strategy. First, we create a trajectory for each image pixel. Second, we filter out the pixels using a Region-of-Interest (RoI). Third, we discard trajectories with significant discontinuities. Fourth, we obtain the PSD of each trajectory, and discard pixels whose power outside our interest band (3-30 BPM) is larger than inside. Last, we aggregate the remaining PSD to create a single PSD estimation.

In our evaluation, we found that depth cameras perform significantly better than infrared cameras (see Fig. 1.4.a).

4.3 Effects of sleep stage, respiration disorders, sleep position

Due to the semi-autonomous nature of breathing, it is easier to recognize breathing rate if the patient is sleeping (see Fig. 1.4.b). We appreciate a strong impact of sleep disorders in our estimations (Fig. 1.4.b). Sleep position has a surprising impact on our estimates: it is more difficult to estimate breathing rate if the patient is in supine position (Fig. 1.4.d). Our experience with the dataset suggests that our method is as reliable in supine position than in left or right positions, however there is a larger incidence of sleep disorders when sleeping in supine position, inducing a bias in the measurement.

Our findings confirm that measuring breath rate is reliable if the subject is relaxed and breaths normally, but a chest movement estimator is not sufficient by itself to diagnose respiration disorders.

4.4 Developing a confidence metric

As currently exposed, our system simply reports the location of the maximum peak of the PSD, therefore it provides a breathing estimate in all cases (even if there is no patient in the bed).

We use a simple metric to rate the confidence of our measurements. We consider the ideal measurement to be a PSD consisting of a single, powerful peak, while the worst measurement would provide an almost uniform PSD with no discernible peak. The power of the signal is not a good measure, as the breathing signal may be very weak, and a distractor signal might be very powerful, therefore we normalize the PSD before rating it. Then we compare the normalized PSD to a uniform PSD using the Earth Mover's Distance [7], which is the natural metric to use when comparing histograms. A low distance would imply that our measurement is similar to the uniform PSD, and thus, not very reliable. Conversely, a large distance implies higher reliability.

By applying a simple threshold on such reliability metric, our estimates using the depth camera coincide with the thermistor with a correlation 0.998 (p-value < 0.0001), having a confidence interval of 0.383 Breaths per Minute for a 95% confidence level.

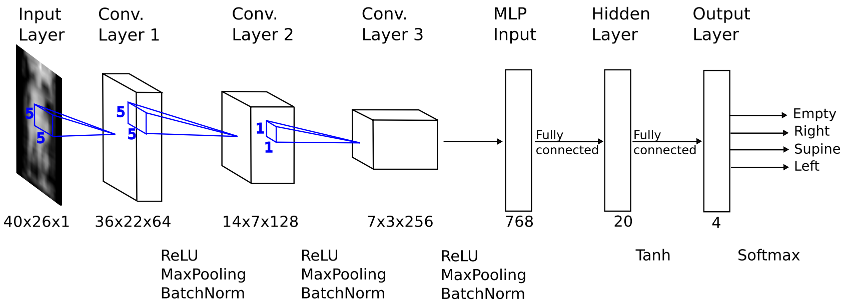

Figure 1.5: CNN architecture used for sleep position classification. It follows a conventional architecture of three convolutional and two fully connected layers.

	Empty	Right	Supine	Left
Empty	98.61	0.00	0.69	0.69
Right	1.36	93.49	6.24	0.44
Supine	0.00	1.01	96.98	2.00
Left	0.86	0.45	13.24	85.45

(a) Chest Sensor

	Empty	Right	Supine	Left
Empty	98.40	0.00	0.00	1.60
Right	0.40	93.20	4.40	2.00
Supine	0.40	1.60	94.00	4.00
Left	0.00	1.20	4.80	94.00

(b) BAM + CNN

Table 1.1: Confusion matrix for the the gravity-based chest sensor worn in the sleep laboratory (left) and our approach based on BAMs and CNNs (right).

5 Sleep Position

Sleep Position monitoring is crucial in nursery homes. Medication and illnesses may prevent patients from changing sleep position themselves, and this causes pressure ulcers. If patients do not change sleep position themselves, nurses must move them. But keeping track of the sleep position of all patients is complicated.

We use the deep convolutional neural network pictured in Fig. 1.5 to classify sleep position from a single BAM into one of the following four classes: "Empty bed", "Left", "Supine", or "Right". Evaluating on the 81 patinents from our sleep laboratory dataset, our algorithm achieves an average accuracy of 93.0% with a Matthews Correlation Coefficient (MCC) of 0.86. Therefore we outperform the gravity sensor used in the sleep laboratory, which is directly attached to the patient's chest and uses an accelerometer to localize the gravity vector, and has an accuracy of only 91.9% with a MCC of 0.84 on the same dataset (see Table. 1.1).

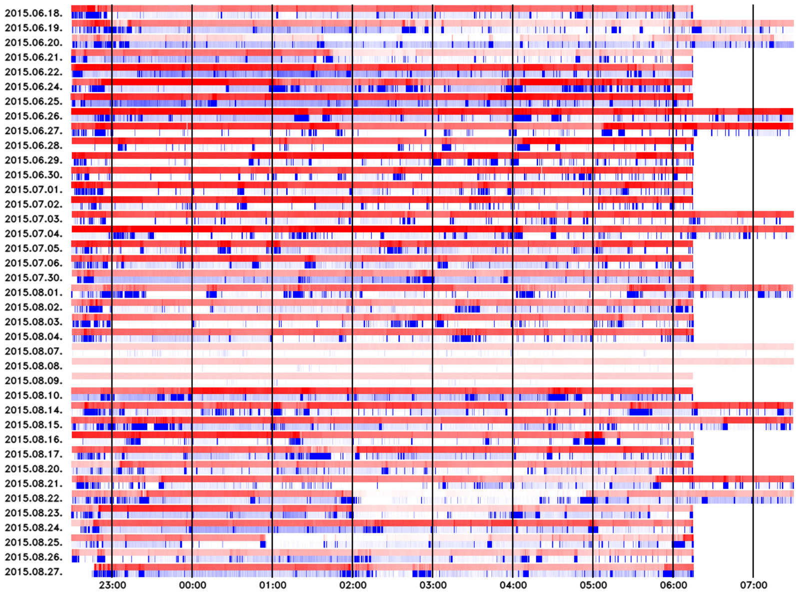

Figure 1.6: 39 night sleep summary of a nursery home resident. The red bar indicates bed occupancy, showing how the patient spent long hours outside the bed towards the end of the study. The blue bar indicates agitated periods. Best viewed in color.

6 Long term sleep summaries

One of the goals of the SPHERE project is to help assessing the long term sleep quality of nursery home residents. Towards this goal, we generate nightly summaries of the patient sleep using two objective metrics based on BAM: bed occupancy, and agitation.

We define bed occupancy as the volume occupied above the bed mattress. It is simply calculated by adding together all BAM cells. This indicator registers exactly when the patient goes to the bed and wakes up, and helps to quantize the amount of sleep.

We use a custom designed agitation metric to complement bed occupancy. Agitation is a strong health indicator, however there is no objective gold standard to measure it. We suggest to use the absolute variation of BAMs within one second as an agitation measure, which has already shown compelling results [1,4].

Both metrics together can summarize a large amount of information in a compact view (see Fig. 1.6).

7 Conclusions

We have presented the SPHERE project, whose aim is to develop a sleep monitoring system using computer vision. We developed a Medical Recording Device (MRD). It is compact but integrates a wide variety of sensors, including depth and infrarred cameras and it is CE certified. Seven units currently installed in real locations have accumulated more than 5,000 hours without incident.

While the MRD is the hardware backbone of the project, the software backbone is the Bed Aligned Map (BAM), a compact image descriptor based on depth that provides alignment and is robust to common image aliments (light, position, rotation, scale). Using BAMs obtained from the MRD, we have shown algorithms that estimate breath rate, sleep position, agitation and bed occupancy.

More importantly, SPHERE has been designed to be evaluated in real scenarios instead of simulated ones. This has revealed how tasks that were considered simple are actually very challenging when performed in unconstrained scenarios (*e.g.* breath rate estimation).

We hope that SPHERE represents a big step forward towards the development of automated and non-intrusive sleep monitoring devices that can be deployed in nursery homes and assisted living installations.

Acknowledgements: This work is supported by the German Federal Ministry of Education and Research (BMBF) within the SPHERE project. We want to thank our partners at Thoraxklinik Heidelberg, Evangelische Heimstiftung and Videmo for their invaluable aid.

References

1. M. Martinez and R. Stiefelhagen, "Automated Multi-Camera System for Long Term Behavioral Monitoring in Intensive Care Units," in *Machine Vision Applications (MVA)*, 2013.

2. ——, “Kinect Unleashed: Getting Control over High Resolution Depth Maps,” in *MVA*, 2013, pp. 247–250.
3. ——, “Kinect Unbiased,” in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5791–5795.
4. M. Martinez, B. Schauerte, and R. Stiefelhagen, ““BAM!” Depth-Based Body Analysis in Critical Care,” in *Computer Analysis and Image Patterns (CAIP)*, 2013, pp. 465–472.
5. M. Martinez and R. Stiefelhagen, “Breath Rate Monitoring During Sleep using Near-IR Imagery and PCA,” in *International Conference on Pattern Recognition (ICPR)*, 2012.
6. 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.
7. Y. Rubner, C. Tomasi, and L. J. Guibas, “The earth mover’s distance as a metric for image retrieval,” *International Journal of Computer Vision (IJCV)*, vol. 40, no. 2, pp. 99–121, 2000.