Respiration Rate Estimation with Depth Cameras: An Evaluation of Parameters

Jochen Kempfle*
University of Siegen
Siegen, North Rhine-Westphalia, Germany
jochen.kempfle@uni-siegen.de

Kristof Van Laerhoven
University of Siegen
Siegen, North Rhine-Westphalia, Germany
kvl@eti.uni-siegen.de

ABSTRACT

Depth cameras have been known to be capable of picking up the small changes in distance from users' torsos, to estimate respiration rate. Several studies have shown that under certain conditions, the respiration rate from a non-mobile user facing the camera can be accurately estimated from parts of the depth data. It is however to date not clear, what factors might hinder the application of this technology in any setting, what areas of the torso need to be observed, and how readings are affected for persons at larger distances from the RGB-D camera. In this paper, we present a benchmark dataset that consists of the point cloud data from a depth camera, which monitors 7 volunteers at variable distances, for variable methods to pin-point the person's torso, and at variable breathing rates. Our findings show that the respiration signal's signal-to-noise ratio becomes debilitating as the distance to the person approaches 4 metres, and that bigger windows over the person's chest work particularly well. The sampling rate of the depth camera was also found to impact the signal's quality significantly.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; • **Applied computing** → *Health care information systems*; *Health informatics*;

KEYWORDS

respiration measurement,non-contact measurement, respiratory rate, ToF sensing, Kinect v2

*This is the corresponding author

iWOAR '18, September 20–21, 2018, Berlin, Germany

ACM ISBN 978-1-4503-6487-4/18/09…$15.00
https://doi.org/10.1145/3266157.3266208

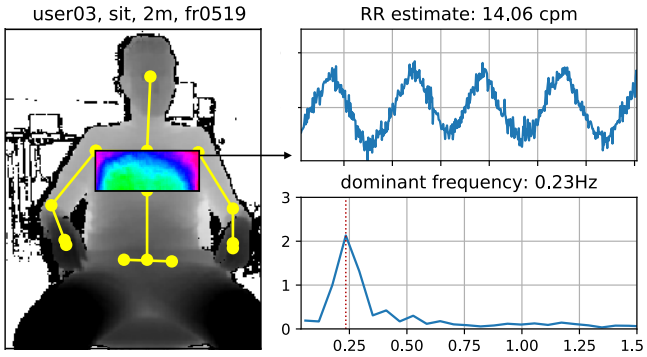

Figure 1: This paper systematically examines the impact of the most prominent parameters to estimate a person's respiration rate from depth imaging: distance to the user, region of interest, respiration rate, and sampling frequency.

ACM Reference Format:
Jochen Kempfle and Kristof Van Laerhoven. 2018. Respiration Rate Estimation with Depth Cameras: An Evaluation of Parameters. In *5th international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18), September 20–21, 2018, Berlin, Germany.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3266157.3266208

1 INTRODUCTION

Monitoring a person's respiratory rate is important in many settings, from medical applications to monitoring in sports, meditation programs, or sleep monitoring. There exist many solutions to estimate a person's respiratory rate, including the use of a respiration belt, a spirometer, PPG-sensors, or body-worn inertial sensing methods. Most of these methods have in common that they in some form require physical contact to the user's body, which may become either uncomfortable or restraining for the person to wear the respective sensors over longer stretches of time. The focus of this paper lies on contact-less respiratory rate detection utilizing a depth camera in particular. This approach has been shown to be feasible, and in contrast to the other methods does not require the user to wear a sensor.

Although several depth sensing approaches have been proposed to capture the respiration rate of a nearby user, the impact of many parameters remains widely unknown.

The assessment and a thorough evaluation of several key parameters are therefore first discussed in the following.

2 RELATED WORK

Non-contact respiration measurement is not limited to depth cameras, but was also successfully carried out using standard color or near-infrared cameras. These range from image subtraction based methods such as the one presented in [22], to optical flow based methods such as [15], [14], and [11], which use Lucas-Kanade [12] or Horn-Schunck [9] methods. In [5], the respiratory rate is measured with both, optical flow computation with the combined local-global method [7] and a depth sensor with surface registration using [4]. It was shown in the latter that the respiration measurement based on the optical flow is potentially better suited for respiratory rate measurement than depth based measurements, which is supported by [10], showing that human breathing mainly occurs along the superior-inferior direction.

During respiration, the change of the volume of certain body parts, namely the chest and abdomen, brings these body parts closer to the depth sensor or moves them away. These changes in distance are visible on the depth sensor and typically are in the range of about 10 mm for normal breathing. When holding still, they directly reflect inhalation and exhalation of the observed subject. The approaches in the literature on capturing the respiratory rate with the help of depth images can roughly be split into two different methods: Based on calculating the distance to, or based on estimating the volume of the relevant parts of the body during respiration. Both methods however are related to each other, as calculating the volume relies on the distance of the single surface points to the sensor. In most cases, a region of interest (short: ROI), usually set to include data from the subject's torso, is defined to focus only on the relevant portion of the usually big amount of depth sensing data.

A simple proof of concept of measuring the respiratory rate with a structured light depth sensor is given in [25]. Here, a solid plane is attached on the chest of the examined body. This plane defines the ROI. Its surface points are captured with the Kinect v1 and are averaged for each received depth frame. The resulting values reflect the average distance of the plane to the Kinect sensor at the different time instants given by the Kinect's sampling rate. The motion caused by respiration and thus the respiratory rate directly is visible on this data. In [6], different parameters are evaluated, including sampling rates of 5 Hz, 7 Hz, and 9 Hz, orientations of 0°or 25°, three light intensities, and different clothing (sweater, jacket, and T-Shirt). Unfortunately the evaluation is specialized on their approach, contains unexpressive numbers, a limited set of parameters, and hardly any information about the evaluation setup. So their evaluation is of little help when transferred to a different project. For respiration measurement, the ROI is set to the chest. With help of the Kinect SDK, it is possible to extract a skeletal model of any observed human from the depth data. The respective shoulder and hip joint positions of the skeleton are used as the corners of the ROI. This skeleton and thus the ROI is updated for each single frame. The depth values inside the ROI are averaged and afterwards a weighted average of four successive such mean values is computed to yield the respiration data over time. In summary, this is the only paper where multiple parameters are considered for evaluation, but it only is shown, that these do not have any impact to their algorithm. In [18], along with the respiratory rate, also the heart rate is measured using the built-in RGB and infrared camera of the Kinect v2 and monitoring the slight color changes of the mouth area caused by the blood pressure change in the vessels during a heart beat. The respiratory rate is measured by caculating the mean of a selected ROI at the torso. Both signals are bandpass filtered with cut-off frequencies set that all frequency components are rejected that are not part of breathing or the heart rate.

In [20] and [8], two papers from nearly the same group and the same year, the respiratory rate is detected using a Kinect v2 as depth sensor. With the frequency and the regularity of the respiration as feature, the state of sleep (wake, REM, NonREM) is classified. To achieve this, the ROI is manually set to the subject's chest and the depth values within the ROI are averaged. In [20], it is mentioned that also the average of the pixel-wise differences of two successive depth images is calculated and used for respiration detection, but this method is not further discussed. [8] describes how to find the features for sleep classification in more detail. It mentions linear interpolation between two successive depth images to fix the Kinect's varying sampling time to a certain value and the wavelet transform as a tool for denoising. Many details for the correct assessment of the respiratory rate are missing or unclear, however, in particular the selection of the ROI or the need to interpolate successive depth frames which is not documented in other work and was not found to have an impact in our experiments.

In [17] and [19], again two papers from nearly the same group and the same year, a time of flight (ToF) camera is used for respiratory rate measurement. They argue that a ToF camera has many advantages to all other depth sensors, like accuracy, no calibration required, real-time capabilities, and so on. For respiration detection, a plane is fitted to the ground surface the observed subject will be lying on. This is the so-called table plane. Two other planes are then fitted to the chest and the abdomen of the subject now lying on the ground surface. Plane fitting according to [17] is achieved in a least squares manner and according to [19] by clustering and averaging the observed surface normals. The euclidean distance between the table plane and the chest or abdomen

plane reflect the subjects respiration. According to the authors, this approach results in a signal with less noise and a better stability than calculating the volume.

In [16] a 10 cm x 20 cm ROI on the center of the subjects thorax is monitored. The mean orientation of the surface plane is computed over 10 successive Kinect v1 frames and then the motion component along the surface normal is observed to extract the respiratory rate.

In [1], [3], and [2], the volume of a ROI in the torso region is computed from the depth readings of the Kinect v1 or the Kinect v2 in the case of [2]. The ROI is defined by the shoulder and hip joint positions of the skeleton that is extracted from the respective Kinect SDK. The depth values inside the ROI are converted to 3D coordinates using the depth sensor's intrinsics. From these points the volume is computed using the Delaunay triangulation with linear interpolation. According to the authors, this volume is proportional to the air volume measured by a spirometer and is called quasi-volume. The observed person is on a bicycle ergometer and its pedaling motion is present in the obtained signal. This motion successfully is filtered out by a FFT bandpass filter.

The distance- or volume-based methods described above are sometimes supported by a respiration model that is computed either previously or on the fly from the incoming data. In most cases, the respiration model is obtained from principal component analysis (PCA) of the respective region of interest of a specified amount of successive depth images, the so-called training set. [23] computes the PCA of the torso and applies the orthomax rotation (more precisely its special case, the varimax rotation) to the obtained PCA model. According to the authors, with the varimax rotation, the obtained model has more relevance to respiration than the model from the standard PCA that exhibits meaningless variations. In contrast to that, as stated in the paper, the principal axes obtained from the orthomax model feature local deformations that are highly correlated to thoracic and abdominal breathing, respectively. In [13] a zoom lens is attached to the Kinect v1 IR projector to increase the size of the dots. These dots then are tracked over 30 s and are stored in a matrix containing the trajectory of each dot. A PCA is applied to this matrix and with the iterative EM algorithm the 16 strongest components are calculated. All bases that fail the Durbin-Watson-test are thrown away. Then, applying FFT on the remaining bases, all bases with less power in the interest region of 0.02 Hz to 1 Hz are discarded. Using the remaining bases, an average trajectory containing less noise is computed. The approach is used for measuring the respiratory rate of sleeping subjects. A special ROI is not required, but at the cost, that only one person can be within the field of view of the Kinect. The approach was tested at different distances and works best at 200 cm. [24] use white markers to define a ROI covering the chest and abdomen. These white markers are visible in the RGB data and thus precisely define the ROI for the depth sensor. As a depth sensor, an Asus Xtion PRO RGB-D camera is used. A PCA of the first 100 frames is computed to obtain a motion model. These frames are preprocessed to reduce noise and to fill holes, as the PCA is very susceptible to it. The first three principal components then are used to reconstruct noise free data without the need of preprocessing. The resulting depth values then are used to generate a surface mesh and to compute its volume. The calculated volume shows strong correlation to spirometer data.

3 OVERVIEW

The depth data is captured from a Kinect v2, a time of flight sensor, connected to a PC by using the Kinect SDK 2.0 built with our custom software. The SDK allows to capture the depth data and, when detected in the depth image, provides skeletal information of up to eight humans, each having 25 joints. The way the Kinect SDK 2.0 detects a human and maps a skeletal model is not reported, but (like in previous versions) most probably is achieved similar to [21]. The depth data, the pixel coordinates of all joints, and a pixel mask that describes which depth pixels belong to which person, along with the respective timestamp, continuously is transferred to hard disk without further processing. Processing the data in real-time was successfully carried out and is possible without restrictions, but storing all data in the first place makes data processing more flexible and repeatable. After recording, the data goes through several processing states, beginning with the preprocessing.

Preprocessing

Salt and pepper noise is a common source of errors when dealing with depth images and usually is caused by a defect pixel. It is characterized as being either totally white or black, which means it either has the maximum possible value or zero. Salt and pepper noise can effectively be eliminated by a median filter. In our implementation it is applied to all pixels showing that kind of noise, primarily to reduce the impact of device specific parameters like pixel defects on our evaluation and to ensure a robust measurement, especially when looking only at few pixels. In depth data also other types of noise are present like quantization noise and white noise. Also surfaces with a steep angle towards the sensor show more noise than surfaces with a flat angle as according to Lambert's cosine law much of the light energy is not reflected back to the sensor. As mentioned in the related work section, there exist many ways of dealing with the different types of noise, be it with band-pass filtering, PCA based approaches, a simple moving average or similar techniques. In our case, only the salt and pepper noise is removed from the single depth images such that further processing is not

affected too much by device specific parameters. All other types of noise consequently remain a part of the signal. This way, denoising will not distort our results and the impact of different parameters on the overall noise level can be investigated better.

Region of Interest

After preprocessing, the joint positions of the left and right shoulders, the hip, and the mid spine are used to define three rectangular regions of interest: the chest (1), the abdomen (2), and the torso (3), as shown in Figure 2. Each ROI furthermore gets divided into ten different sizes. The full size, that may contain parts of the surroundings, step-wise is decreased by subtracting 5% of the initial extent value from each side of the rectangle at each step i. Starting from the initial width w_0 and height h_0, the i^{th} width w_i and height h_i is computed with (1). The smaller ROIs consequently still are centered on their initial position and the smallest size will be 1% of its initial size. The smallest possible ROI size is fixed to at least contain 2x2 pixels to avoid empty or single pixel ROIs.

$$\begin{aligned} w_i &= w_0 - (2 \cdot 0.05 \cdot i) \cdot w_0 \\ h_i &= h_0 - (2 \cdot 0.05 \cdot i) \cdot h_0 \\ 0 \leq i \leq 9, &\quad i \in \mathbb{N} \end{aligned} \tag{1}$$

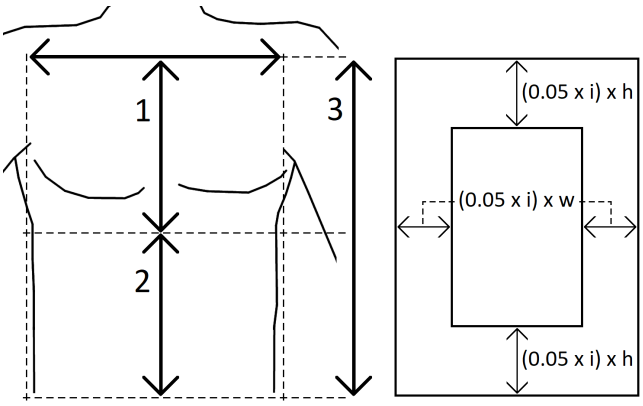

Figure 2: Definition of ROI position and size in our study.

Data Processing

Once the region of interest is defined, further data processing has to concern itself with the extraction of the respiration signal from the respective region. This can be reduced to estimating the ROI's volume or its distance to the sensor. Due to slight movements of the observed person and moving or unstable joint positions, the size and position of the ROI varies over time. These variations can not be neglected and directly will affect the computed volume of the respective ROI. When not fixed to a certain size and position, the evolution of the ROI's volume over time will not represent the respiratory rate, but will be dominated by the ROI's variations. Estimating the volume consequently only makes sense with our approach, when the ROI is fixed to for example that computed in the first frame. This however is equal to estimating the ROI's distance to the sensor as two dimensions are fixed to constant values by the definition of the ROI and only the distance remains as parameter for volume calculation. So, independent of how the single depth values are processed for volume estimation, the resulting value will also be a valid measure for the ROI's distance. In contrast to volume calculation, the computation of a distance measure does not require any further processing like for example multiplying it with the ROI dimensions.

Estimating the ROI's distance with our approach therefore is easier and, as the interesting body region gets tracked over time, is more accurate and flexible. A valid distance measure is to find the mean distance of the ROI to the sensor. For this purpose all depth values within the respective ROI are averaged. This average figuratively represents the distance of an imaginary surface plane fitted to the tracked body part. During breathing this surface plane moves towards or away from the depth sensor. Its evolution over time is the sought respiration signal. One drawback of this method is, that it only is valid when the observed person does not move too much and the ROI does not contain disturbing motions.

4 PARAMETERS

We distinguish several important parameters that may influence the quality of respiratory rate measurement. These include the subject (size, gender,...), its respiration rate, the distance of the subject to the sensor, and size and position of the region of interest.

User

The user, namely the observed person, may play a crucial role in respiratory rate measurement. The person's size, BMI, gender, personal respiration patterns such as thoracic or abdominal breathing, clothing, or age, to name but a few parameters, are likely to impact the respiration estimate. It however is difficult to exactly pinpoint the influence of each single parameter and their correlations to each other. As the respiratory rate will not only be measured on a certain group of people, these parameters are combined. Their combination will be here referred to as the user.

Distance

The distance of the subject to the sensor is an important factor of the overall accuracy of the measurement due to two reasons. First, the number of sampling points seen by the sensor is inversely proportional to the distance squared. Thus, with increasing distance a reduced amount of data samples that actually contribute to the measurement are present in

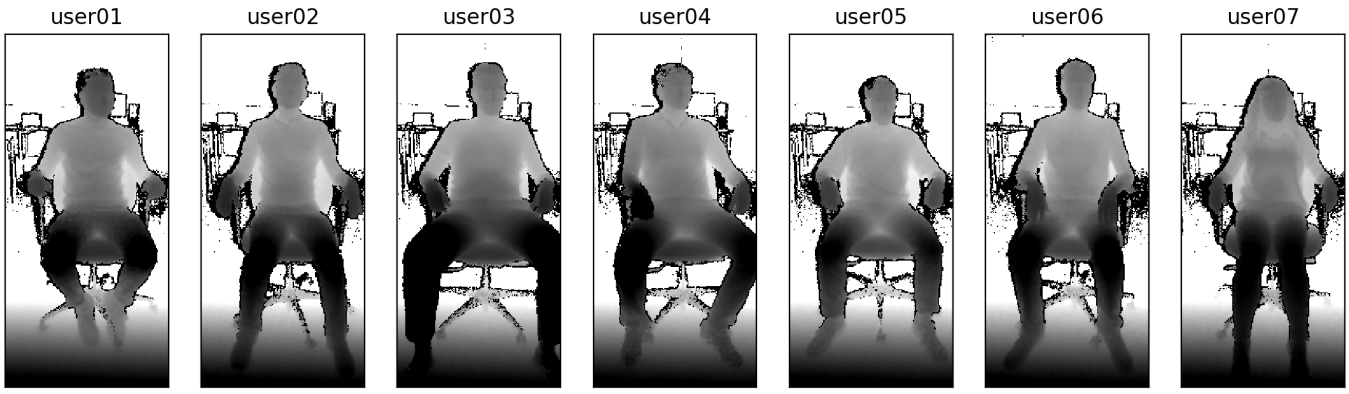

Figure 3: Depth images of the experiments' volunteers, as seen from a 2 meter distance, while sitting and facing the camera. Note that the volunteers were told to avoid occluding their torso but were free to choose their sedentary posture.

the depth data. Secondly, the overall depth resolution of a depth sensor gets worse with increasing distance. The impact of the distance on the depth resolution depends on the sensor. Structured light sensors for example are known to be very susceptible to depth, whereas time of flight sensors seem to be more robust here. Both sensor types however are dependent from the light intensity reflected back from the actively illuminated object which again is inversely proportional to the distance squared. With increasing distance therefore also an increasing noise level is expected.

Respiratory Rate

The respiratory rate primarily is the parameter to be measured. In order to get more variance in the data and to ensure the measurement is valid on a wider range, the respiratory rate is parameterized. This allows to investigate not only different users, but also different respiratory rates per user, as these may influence the measurement. One reason is that people may breath differently on different respiratory rates. For example, a different rate may cause deep or shallow breathing or the user changes from thoracic to abdominal breathing or vice versa. Furthermore it defines the lowest possible sampling rate according to the Nyquist-Shannon sampling theorem. Overall, its influence on the measurement is worth being investigated.

Sampling Rate

The sampling rate is an important parameter when it comes to power efficiency, computation time, or storage size. Usually a higher sampling frequency is preferred, but as the depth stream contains a huge amount of data to be processed, in many cases the processing time limits the sampling rate. Also in case of a long time recording (for example over night) without real-time processing, the data size may set a limit to the sampling rate. In our case, without data compression, one minute of depth data sampled at 30 Hz amounts to nearly one gigabyte of data. Halving the sampling rate also means only half the data. On the other hand, a lower sampling frequency, at least at a certain point, affects the quality of the measured signal. Thus it is important to know how different sampling rates will impact the measurement of the respiratory rate.

Region of Interest

The respiration signal only is present in the depth image pixels that cover the relevant body parts for breathing, namely the observed subject's torso. As the whole depth image usually contains a substantial amount of data and as the relevant image part in most cases is much smaller, the further processing is focused on this area or even smaller parts of it. The observed area is called the region of interest (ROI). The ROI may be as big as the whole depth image or as small as a single pixel, given the respiration signal is contained in it. A smaller ROI is expected to be more susceptible to noise while a bigger ROI requires more processing time and, if chosen too big, will cover too much unnecessary details of the surroundings that may influence the measurement. The size consequently is one of two important ROI parameters to observe. The other parameter is its position. When the region of interest is chosen smaller than the relevant image part, its position will influence the measurement quality, depending on how well the respective body part is suited for measuring the respiration. In the following, the ROI will also be called window. Finding its optimal size and position is one of the aims of this paper and is carried out in the evaluation.

5 EVALUATION

The aim of this section is to evaluate all meaningful or interesting parameter combinations. Throughout the evaluation, the signal-to-noise ratio (SNR) is used as a quality measure of the obtained signal. It is calculated with (2). The higher

the SNR, the higher the signal stands out from the noise and the better it can be extracted from the data.

$$SNR_{dB} = 10 \cdot \log_{10}\left(\frac{P_{Signal}}{P_{Noise}}\right) \quad (2)$$

Study Design

The study was performed with seven participants, six of them male, with all wearing a T-shirt or sleeved shirt. Figure 3 shows the depth images from a distance of 2 meters for all participants while sitting.

The experiment setup is as follows: Each participant has to sit on a chair in front of a Kinect v2 sensor. The sensor is adjusted on the height of 1.40 m and has an angle of 25° towards the floor plane. The height of the chair is 0.5 m. In the Kinect's viewing direction in a regular interval of 1 m four markers are placed on the floor. The first marker thus has the distance of 1 m to the Kinect and the last marker is 4 m away from it. These markers help to define the exact positions where the chair has to be centered on, such that the different measurements all are taken from the same distance. The participant is asked to face the Kinect and to place the arms on the armrests of the chair such that its upper body is fully visible to the depth sensor.

Each participant is asked to breath with a certain frequency that is given via visual feedback from a custom program. The experiment is started when the respective participant has adapted his respiration to the given frequency. During the experiment, they are asked to breath in a natural and comfortable way. The experiment is repeated at each distance marker with two different frequencies each at 0.17 Hz and 0.25 Hz. The duration of the experiment at the slower frequency of 0.17 Hz is at least 2 min and varies depending on the participant feeling comfortable. As the higher frequency is exhausting to the participants and to ensure their well-being, the duration for the faster frequency only is at least 1 min and again varies depending on the participant. Also to ensure well-being and good quality measurements, a short break is done between two successive measurements.

Results

The results of our study are visualized with box plots. All parameters that are not explicitly stated on the diagram are mixed up and yield the variance of the drawn box plots for the given parameter configuration. The only exceptions to that are the ROI sizes and the sampling rate. When not explicitly shown on the diagram, the sampling rate is fixed to 30 Hz and only the largest ROI size is used to not omit valuable data and to not confuse the data with smaller subsets of itself. The ROI size in return is among the first parameters to focus on. All findings only apply to our setup where the

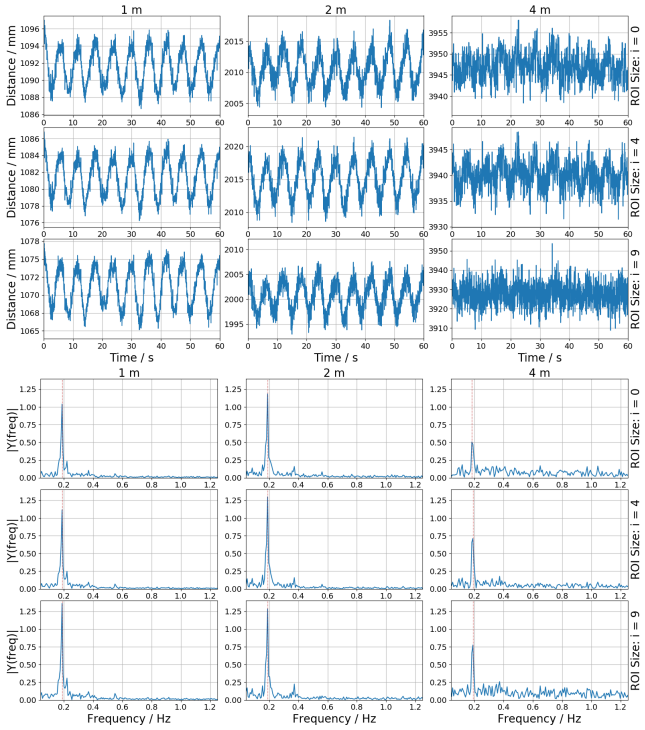

Figure 4: Raw data and FFT data of user 4 at 10 cps, recorded at the abdomen. These examples show that both distance and ROI size matter significantly for the signal quality.

participants were sitting on a chair and not necessarily are true for different setups like for example a standing position.

When looking at the signal (see Figure 4), it can be seen that the respiratory rate measurement with higher distances or with decreasing size of the ROI increasingly gets noisier. Remarkably, in the frequency domain the signal still dominates the upcoming noise even at the highest distance and the smallest ROI size. The SNR at the different distances and with different ROI configurations is pictured in Figure 5. The SNR values indicate, that the signal quality is best at close distances and big ROI sizes. The largest ROI size however does not show the best signal quality. An explanation for this is, that also parts of the surroundings are covered by the ROI. The surroundings do not contribute to the respiration and consequently everything that is captured from it is considered noise. The optimal ROI position is found to be the chest, followed by the torso, and last but not least the abdomen. This optimality order is independent from the distance and, more importantly, it also is independent from the size of the chosen ROI. A more fine grained hint of how to place the region of interest can be taken from Figure 6. Here the difference of the minimum torso extent after exhaling and the maximum torso extent after inhaling is shown with a heat map. The brighter the area, the higher the distance between

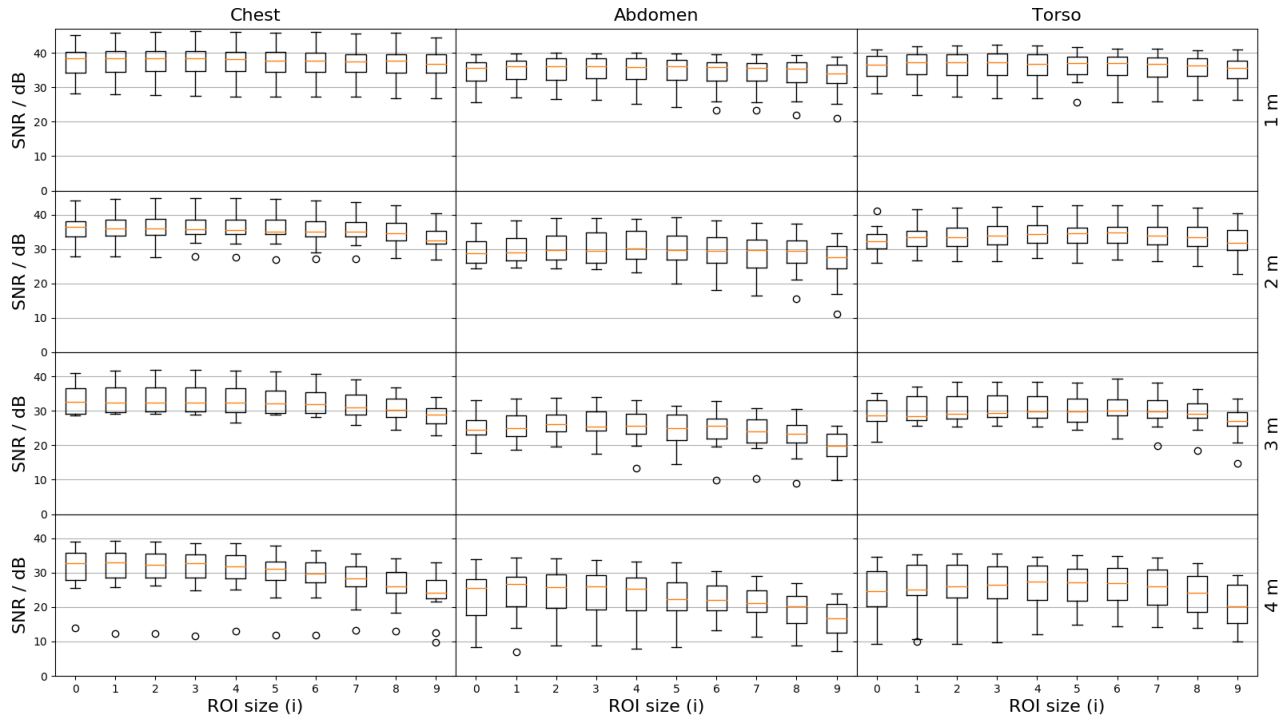

Figure 5: The signal-to-noise ratio (SNR) at different region of interest (ROI) configurations and distances shows that across all data, the trends from Figure 4 hold: The respiration rate signal quality contains more noise when the person is at a larger distance from the depth camera and as the region of interest becomes smaller.

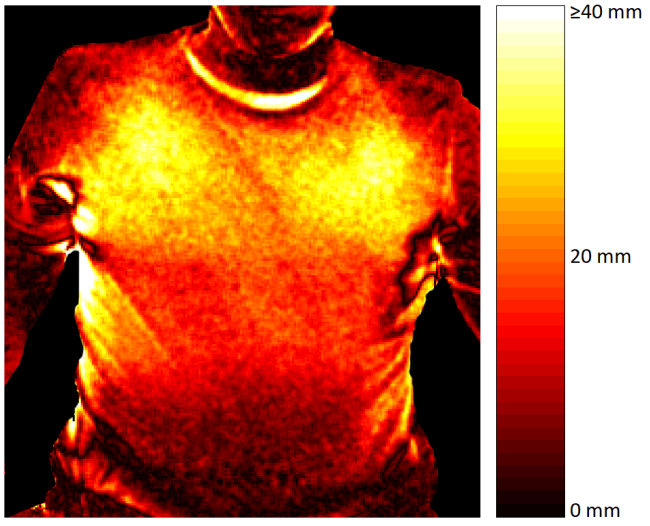

Figure 6: A typical example of a heat map from a person's torso over a single respiration cycle, showing the maximum difference in distance between inhalation and exhalation.

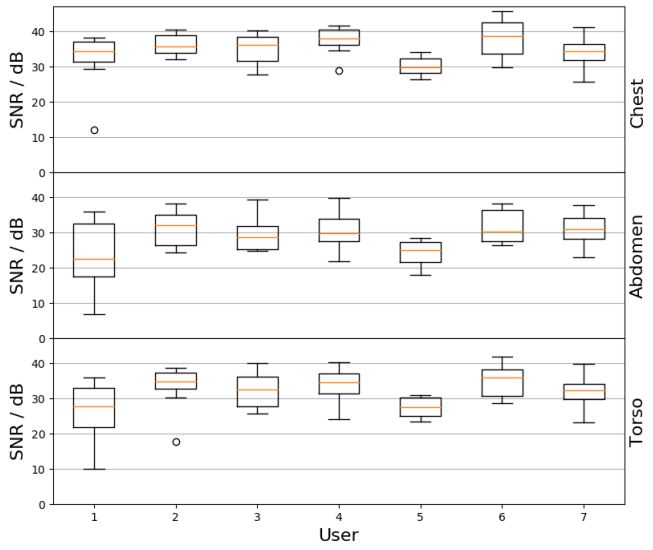

Figure 7: SNR for different users (1 to 7) at the three different region of interest (ROI) positions: chest, abdomen, and torso.

full exhalation and full inspiration. The heat map visually demonstrates that during respiration the chest experiences a greater expansion than the abdomen and thus yields a higher SNR. The shown heat map however just is an example from one single user and one single respiration cycle. It therefore can not be generalized, neither to different respiration cycles of the same user nor to different users. Both cases are likely to produce a different heat map due to a different respiration pattern, for example abdominal breathing.

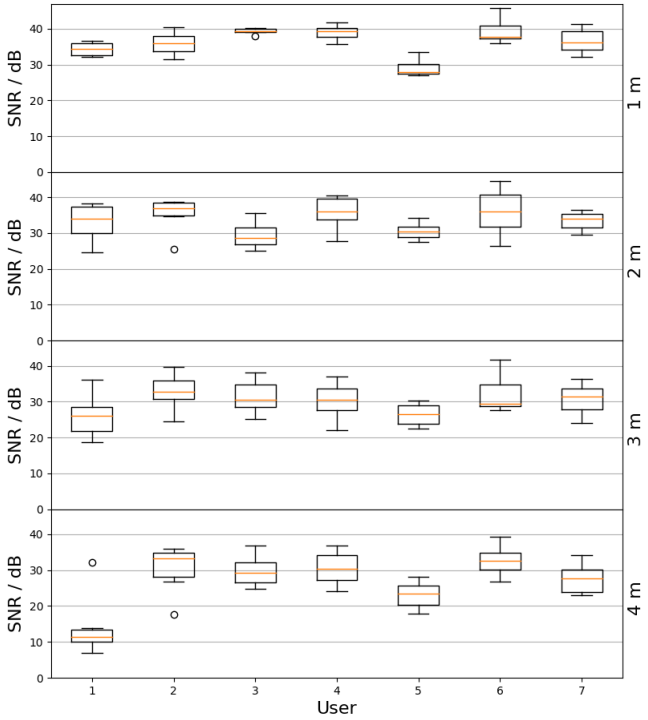

Figure 8: The signal-to-noise ratio (SNR) for different users (1 to 7) at increasing distances (1 meter, top, to 4 meters, bottom). A wide per-person variety can be seen, as well as a more noisy signal for every user as the distance is increased.

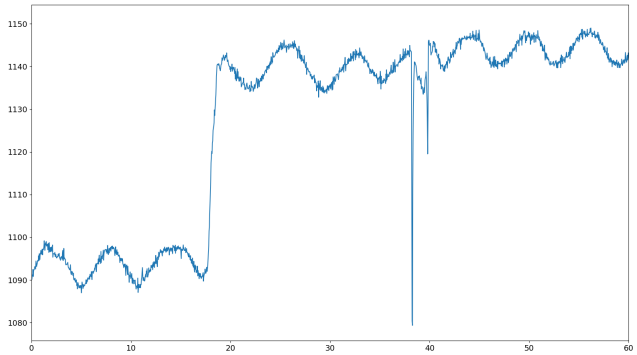

Figure 9: An example of motion artifacts in the respiration signal (Y axis: distance in millimeters, X axis: time in seconds, for user 5) while repositioning at 18 seconds and occluding the torso at 38 seconds.

The single signal-to-noise ratios per user show strong variations across diferent persons, as depicted in Figures 7 and 8. This implies that the many parameters associated with a user do influence the respiration measurement significantly. Although just one female subject participated in the experiments, the test results indicate that the user's gender seems to play a minor role in respiratory rate measurement. User 7 (female) has, independent from the ROI position or her distance to the sensor, similar SNR values compared to the male participants. Another factor that comes into play can be seen when analyzing the poor performance of user 1 in Figure 7: It is caused by the user's body posture (skeleton model) not being correctly recognized by the Kinect SDK at a distance of 4 m. This in turn caused the ROI to flicker and not being aligned correctly. Figure 8 demonstrates how the distance impacts all users performances, especially that of user 1. From a distance of about 4.5 m onwards, a skeleton model no longer can be estimated by the Kinect SDK. Thus it is expected that a unfavorably aligned ROI plays a certain role in higher distances.

User 5 also shows a worse performance than the average, mostly due to motion artifacts during recording. In user 5's data (see Figure 9) several jumps are present in the respiratory signal that are caused by slight movements during the measurement. The participants were asked to sit still during the experiment because we did not want to single out movement as parameter. At least for user 5 it seems that slight movements can not be avoided on longer runs as they are present in nearly each single recording, especially those lasting longer than one minute. The ability to sit still is an unexpected but influential user parameter in this case as due to motion the overall SNR gets worse.

Decreasing the sampling rate yields less data and more time for processing but also negatively influences the SNR. In Figure 10 the impact of different sampling rates is plotted against different respiratory rates and at different ROI positions. Figure 11 furthermore shows how the distance may affect the performance at different sampling rates. The first to observe is that independent of the ROI position, the respiratory rate, and the distance, the signal quality gets worse with decreasing sampling rate. A smaller sampling rate therefore is expected to increase the noise level and should only be considered when due to the other parameter configurations the increased noise level will not exceed a certain limit. An explanation for the higher SNR is, that on higher sampling rates the overall noise power gets distributed over a wider frequency range, lowering the mean noise power level and thus increasing the signal to noise ratio.

As suggested above, the user may switch from thoracic to abdominal breathing when changing the respiratory rate. The correlation of both the respiratory rate and the ROI position is depicted in Figure 10. For this, only the box plots at the same sampling rates have to be compared: It is important to note that the higher respiratory rate has a worse performance than the lower one throughout all other parameter configurations. Apart from this observation, the optimal ROI position for both respiratory rates again is the chest area, followed by the torso, and finally the abdomen. While sitting, increasing the respiratory rate consequently does not change the optimality order of the ROI position.

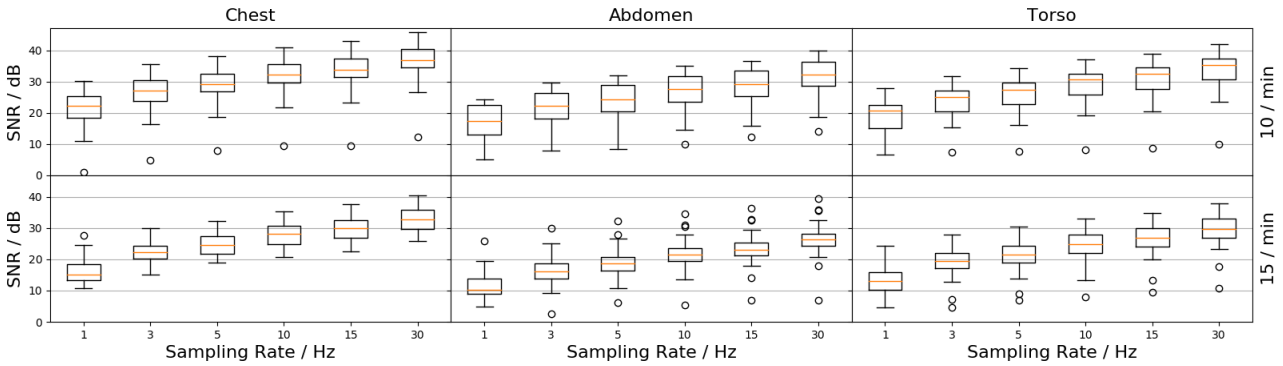

Figure 10: SNR at different sampling rates, respiration rates, and ROI positions.

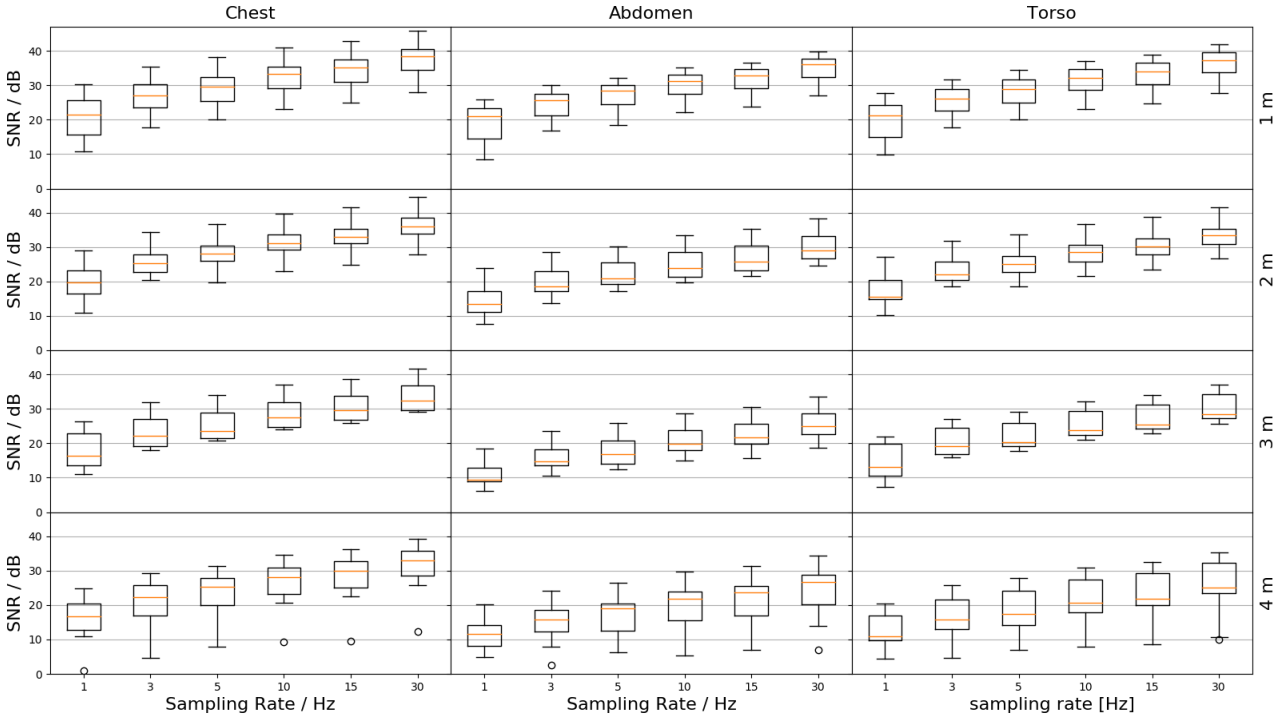

Figure 11: SNR at different sampling rates, distances, and ROI positions.

6 CONCLUSIONS

This paper presents a systematic parametric study for the use of depth imaging to estimate respiration rate. For this purpose, we have recorded the raw depth data from seven study participants for variable breathing rates, and distances between RGB-D camera and the persons. By also varying the size and type of the observed region of the users' torso and the sampling rate, we obtained the following results:

User specific parameters have great influence on the signal quality. Gender seems to play a minor role.

Distance proportionally increases the noise level and limits the respiration signal quality. From a certain distance on, the region of interest can no longer be detected reliably.

Respiratory Rate plays a role as well, as a higher frequency tends to lower the signal quality.

Sampling Rate should be chosen as high as possible. Lowering it immediately decreases the signal-to-noise ratio independently from other parameter settings.

Region of Interest denotes two sub parameters, namely the ROI size and position that have to be treated independently. A smaller size especially in combination with increasing distance has detrimental effects on the signal quality. The ROI's position ideally is placed on a body segment that reveals large respiratory movements, in our case the chest.

While some parameters were deliberately left out from the experiments, the evaluation revealed that many parameters still are hidden or neglected although they have a greater effect on respiratory measurement than expected. One example for this is considering the user as black-box in our case. Apart from the obvious outliers, the variances across the users cannot be explained satisfactorily as it is not known what user-specific parameters may have caused these. The user therefore needs a more sophisticated evaluation in the future and we hope to single out the most influencing user characteristics. Also, we like to expand our current research to further parameters that were left out in this study, like different clothing, different rotations to the sensor, different user postures like standing or lying, deep and shallow breathing, thoracic and abdominal breathing, partial occlusions, setting the ROI larger than required, and different user movements like for example walking, to name some parameters of interest.

The dataset and python scripts supporting this paper's evaluations can be obtained by contacting the first paper author or visiting http://ubicomp.eti.uni-siegen.de.

REFERENCES

[1] Hirooki Aoki, Masaki Miyazaki, Hidetoshi Nakamura, Ryo Furukawa, Ryusuke Sagawa, and Hiroshi Kawasaki. 2012. Non-contact respiration measurement using structured light 3-d sensor. In *SICE Annual Conference (SICE), 2012 Proceedings of*. IEEE, 614–618.

[2] Hirooki Aoki and Hidetoshi Nakamura. 2018. Non-Contact Respiration Measurement during Exercise Tolerance Test by Using Kinect Sensor. *Sports* 6, 1 (2018), 23.

[3] Hirooki Aoki, Hidetoshi Nakamura, Kengo Fumoto, Kunihisa Nakahara, and Masaru Teraoka. 2015. Basic study on non-contact respiration measurement during exercise tolerance test by using kinect sensor. In *System Integration (SII), 2015 IEEE/SICE International Symposium on*. IEEE, 217–222.

[4] Sebastian Bauer, Benjamin Berkels, Joachim Hornegger, and Martin Rumpf. 2011. Joint ToF image denoising and registration with a CT surface in radiation therapy. In *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 98–109.

[5] Sebastian Bauer, Jakob Wasza, and Joachim Hornegger. 2012. Photometric estimation of 3D surface motion fields for respiration management. In *Bildverarbeitung für die Medizin 2012*. Springer, 105–110.

[6] Flavia Benetazzo, Alessandro Freddi, Andrea Monteriù, and Sauro Longhi. 2014. Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results. *Healthcare technology letters* 1, 3 (2014), 81–86.

[7] Andrés Bruhn, Joachim Weickert, and Christoph Schnörr. 2005. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International journal of computer vision* 61, 3 (2005), 211–231.

[8] Fabio Centonze, Martin Schätz, Aleš Procházka, Jiří Kuchyňka, Oldřich Vyšata, Pavel Cejnar, and Martin Vališ. 2015. Feature extraction using MS Kinect and data fusion in analysis of sleep disorders. In *Computational Intelligence for Multimedia Understanding (IWCIM), 2015 International Workshop on*. IEEE, 1–5.

[9] Berthold KP Horn and Brian G Schunck. 1981. Determining optical flow. *Artificial intelligence* 17, 1-3 (1981), 185–203.

[10] Paul J Keall, Gig S Mageras, James M Balter, Richard S Emery, Kenneth M Forster, Steve B Jiang, Jeffrey M Kapatoes, Daniel A Low, Martin J Murphy, Brad R Murray, et al. 2006. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Medical physics* 33, 10 (2006), 3874–3900.

[11] Yung-Ming Kuo, Jiann-Shu Lee, and Pau-Choo Chung. 2010. A visual context-awareness-based sleeping-respiration measurement system. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 255–265.

[12] Bruce D Lucas, Takeo Kanade, et al. 1981. An iterative image registration technique with an application to stereo vision. (1981).

[13] Manuel Martinez and Rainer Stiefelhagen. 2012. Breath rate monitoring during sleep using near-IR imagery and PCA. In *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 3472–3475.

[14] Kazuki Nakajima, Yoshiaki Matsumoto, and Toshiyo Tamura. 2001. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiological Measurement* 22, 3 (2001), N21.

[15] Kazuki Nakajima, Atsushi Osa, and Hidetoshi Miike. 1997. A method for measuring respiration and physical activity in bed by optical flow analysis. In *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, Vol. 5. IEEE, 2054–2057.

[16] Philip J Noonan, Jon Howard, Deborah Tout, Ian Armstrong, Heather A Williams, Tim F Cootes, William A Hallett, and Rainer Hinz. 2012. Accurate markerless respiratory tracking for gated whole body PET using the Microsoft Kinect. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*. IEEE, 3973–3974.

[17] Jochen Penne, Christian Schaller, Joachim Hornegger, and Torsten Kuwert. 2008. Robust real-time 3D respiratory motion detection using time-of-flight cameras. *International Journal of Computer Assisted Radiology and Surgery* 3, 5 (01 Nov 2008), 427–431. https://doi.org/10.1007/s11548-008-0245-2

[18] Aleš Procházka, Martin Schätz, Oldřich Vyšata, and Martin Vališ. 2016. Microsoft kinect visual and depth sensors for breathing and heart rate analysis. *Sensors* 16, 7 (2016), 996.

[19] Christian Schaller, Jochen Penne, and Joachim Hornegger. 2008. Time-of-flight sensor for respiratory motion gating. *Medical Physics* 35, 7Part1 (2008), 3090–3093.

[20] Martin Schätz, Fabio Centonze, Jiří Kuchyňka, Ondřej Ťupa, Oldřich Vyšata, Oana Geman, and Aleš Procházka. 2015. Statistical recognition of breathing by MS Kinect depth sensor. In *Computational Intelligence for Multimedia Understanding (IWCIM), 2015 International Workshop on*. IEEE, 1–4.

[21] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time Human Pose Recognition in Parts from Single Depth Images. *Commun. ACM* 56, 1 (Jan. 2013), 116–124.

[22] K Song Tan, Reza Saatchi, Heather Elphick, and Derek Burke. 2010. Real-time vision based respiration monitoring system. In *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*. IEEE, 770–774.

[23] Jakob Wasza, Sebastian Bauer, Sven Haase, and Joachim Hornegger. 2012. Sparse principal axes statistical surface deformation models for respiration analysis and classification. In *Bildverarbeitung für die Medizin 2012*. Springer, 316–321.

[24] Udaya Wijenayake and Soon-Yong Park. 2017. Real-Time External Respiratory Motion Measuring Technique Using an RGB-D Camera and Principal Component Analysis. *Sensors* 17, 8 (2017), 1840.

[25] Junyi Xia and R Alfredo Siochi. 2012. A real-time respiratory motion monitoring system using KINECT: Proof of concept. *Medical physics* 39, 5 (2012), 2682–2685.