Robust Real-Time 3D Respiratory Motion Detection Using Time-of-Flight Cameras

Jochen Penne

Chair of Pattern Recognition, Friedrich-Alexander-University Erlangen-Nuremberg
Martensstr.3 91058 Erlangen, Germany

`Jochen.Penne@informatik.uni-erlangen.de`

Christian Schaller

Chair of Pattern Recognition, Friedrich Alexander University Erlangen-Nuremberg
Martensstr.3, 91058 Erlangen, Germany

`Christian.Schaller@informatik.uni-erlangen.de`

Joachim Hornegger

Chair of Pattern Recognition, Friedrich Alexander University Erlangen.Nuremberg
Martensstr.3 91058 Erlangen, Germany

`Joachim.Hornegger@informatik.uni-erlangen.de`

Torsten Kuwert

Chair of Clinical Nuclear Medicine, Friedrich Alexander University Erlangen-Nuremberg
Krankenhausstraße 12 91054 Erlangen, Germany

`torsten.kuwert@nuklear.imed.uni-erlangen.de`

Abstract

Purpose: *Respiratory motion of the patient during data acquisition causes artefacts in the field of emission or computed tomography. Respiratory gating allows to track and correct these artifacts.* **Material and Methods:** *In this paper, we present a system that uses the fairly new and off-the-shelf Time-of-Flight (ToF) technology to compute a dense estimate of the three-dimensional respiratory motion of a patient. The work is characterized by three key contributions. The first is the employment of ToF sensors. Using ToF sensors it is feasible to acquire a dense 3D surface model of the chest and abdomen of the patient with more than 15 frames per second. The second contribution is an algorithm to derive a surface representation which enables the estimation of the 3D respiratory motion of the patient, which is sufficient to compute 1D breathing signals for scalable specific regions of interest like chest and abdomen. The proposed data-driven algorithm models the chest and abdomen three-dimensionally by fitting distinct planes to different regions of the torso of the patient. The third contribution is the possibility to derive a sub-millimeter accurate 1D respiratory motion signal by observing the displacement of each plane.*

Results: *Our ToF modeling approach enables marker less, real-time, 3D tracking of patient respiratory motion with an accuracy of 0.1 mm.* **Conclusion:** *Thus, our approach provides 1D breathing signals for scalable anatomical regions of interest with sufficient accuracy for artifact reduction in SPECT or x-ray angiography.*

1. Introduction

Physiological motion in emission or computed tomography leads to motion blurring, a reduction of overall image contrast, and a loss of sensitivity [9]. Difficulties in lesion detection, loss of accuracy in functional volume determination and more difficult activity concentration recovery are the consequences of the associated blurring. Widely accepted approaches utilize respiratory gating: Only data acquired during a certain respiration state is used for the computation of 3D reconstructions. Multiple 3D reconstructions lead to a 4D data set. Up to eight respiration states have proven to deliver a good trade-off between temporal resolution and noise in the 3D reconstructions [14].
Furthermore, adaptive radiotherapy of lung cancer requires a breathing signal to enable the computation of 4D-CT mod-

els. Based on these data, radiation delivery can be tracked or gated.

2. State of the Art

Current methods to acquire or compute a breathing signal are:

1. **Spirometry**: measurement of the amount (volume) and/or speed (flow) of air that can be inhaled and exhaled.

2. **Markers** are placed on the skin of the patient. The position of the markers is tracked [11, 3].

3. **Acquisition of stereoscopic images** of torso and computation of 3D torso surface points. Derivation of volumetric information from the surface [13].

4. **Hybrid techniques** [7], which combine two or more of the above mentioned techniques.

It is worth noting that method 3 is the only published method which is contactless and which enables the distinction between abdominal and thoracic breathing at a framerate of 5 Hz. The method is equivalent to spirometry and provides drift-free volume information, which is not the case for spirometry. In the case of spirometry a global decrease of the volume can often be observed, which is caused by the relaxation of the patient due to breathing in a lying position. We propose an alternate method which is also contact-less and therefore more closely related to method 3 than to methods 1 or 2. We suggest a system based on the emerging ToF technology. ToF sensors provide a direct way for acquiring 3D surface information of objects with a single all-solid-state ToF camera by measuring the time of flight of an actively emitted optical reference signal in the infrared spectral range [2]. More recently, applications like obstacle detection [12], gesture recognition [5][6] and automotive passenger classification [4] make us of ToF sensors. Currently, ToF cameras are also on their way to become a component of consumer electronics, i.e. off-the-shelf technology. Therefore, a decrease of production costs for ToF sensors due to mass production can be expected in the near future. As ToF sensors provide data at rates higher than 15 Hz, they are suitable for real-time imaging. Examples of available ToF cameras are shown in Figures 1(a) and 1(b). The 3D data available of a scene observed with a ToF camera is a 3D point cloud. As each 3D point corresponds to a pixel of the sensor matrix, a triangulation of the 3D point can be derived in a straight forward manner. The distance estimation of a point is accomplished by measuring the time of flight of an optical reference signal emitted by the camera and reflected by the scene. Besides the 3D information for each pixel an intensity value corresponding to the reflected

Lateral Resolution [px]	144×176
Depth Resolution [mm]	≥ 1
Framerate [fps]	≥ 15
Number 3D points	25344
Camera Dimensions [mm$\times$mm$\times$mm]	$50 \times 67 \times 43.2$
Field of view [degree$\times$degree]	47.5×39.6

Table 1. Technical specification of the ToF camera SR 3100 used for the experiments. *[M1.4]*

amount of the reference signal is available. These intensity values are normally encoded as grey values and can be used to provide a texture for the 3D surface reconstruction. A schematic overview of the ToF principle and an example of the data available from ToF cameras is given in Figures 1(c) and 1(d).

We investigated the possibilities of extracting a respiration signal from these 3D data and to distinguish abdominal and thoracic breathing based on this information.

To emphasize technical differences between marker-based/marker-less stereo-vision approaches, spirometry and our ToF-based approach we briefly will compare certain techniques in the following:

Marker-less stereo vision vs. ToF: The main difference between both systems is how depth information is achieved. Stereo approaches can only extract depth information at pixels where corresponding features/texture information is available. In contrast, due to their high lateral resolution ToF cameras provide dense depth information at constant lateral resolution (see Table 1) and high framerates. A step required mandatory for a stereo-setup is calibration. Unless correctly calibrated, a stereo system cannot compute true depth estimates from disparity values. For ToF cameras no calibration step in order to acquire 3D information is required at all.

Marker-based stereo vision vs. ToF: Calibration is also required when using markers. Furthermore, 3D information is only available at the positions of the markers, i.e. the number of available 3D points is defined by the number of markers and a dense 3D map cannot be computed. In comparison, ToF cameras provide 3D information for many thousand 3D points (see Table 1) at constant lateral resolution.

Spirometry vs. ToF: Spirometry techniques measure the volume or the change of volume to quantify respiration. Nevertheless, no information about the anatomical deformations of the torso due to respiration can be derived from this volume information. Additionally, no distinction between abdominal and thoracic respiration can be made using the data. In contrast, we will show in our paper that this distinction can be made using ToF based respiratory motion detection.

(a) ToF camera MESA Imaging GmbH [8].

(b) ToF camera PMDTec GmbH [10].
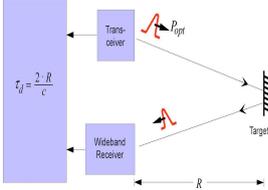
(c) ToF principle ($\mathbf{R}$..distance, $\mathbf{c}$..speed of light, τ_d..travel time of impulse $\mathbf{P}_{opt}$).
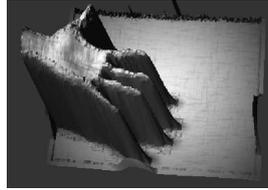
(d) 3D closed surface reconstruction of human hand.

Figure 1. Different ToF cameras (Figure 1(a) and 1(b)), ToF principle (Figure 1(c) and example visualization of available data (Figure 1(d)).

3. Methods

Our method requires the ToF camera to be rigidly mounted in a way which brings the full torso of the patient into the field of view of the camera. This implies a distance of approx. 60 cm–100 cm between ToF camera and patient. Given today's systems for morphological and functional imaging this is no serious constraint. In most systems it is expected to mount the camera in existing caves. Figure 2(a) shows the examples of the acquired 3D surface data. The patient is assumed to be lying on an approximately planar table. The main steps of the method are the following:

1. **Calibration**: An image of the empty patient table is acquired. A best-fitting plane is computed for the surface of the patient table. This plane will be used for segmenting the torso of the patient in subsequent processing steps. We will term this plane table-plane.

2. **Segmentation of the torso**: If the patient is lying on the table, his complete torso is segmented by rejecting all points which are more far away from the ToF camera than the table-plane, i.e. which are behind the table-plane from the viewpoint of the ToF camera.

3. **Defining Regions-of-Interest (ROI)**: Two regions of interest are defined. One for the chest region and one for the abdomen. Of course, the number of regions of interest can be chosen appropriately for the specific application. The definition of regions of interest is necessary when setting up the camera the first time in the

therapy room or possibly when patients of significantly differing size are treated (like kids and adults).

4. **Derivation of multidimensional breathing signal**: The 3D points of each ROI are processed with a mean filter of kernel size five to reduce the influence of noise and subsequently a best-fitting plane in least square sense is computed for the processed 3D points of each ROI. Each plane is restricted to be parallel to the previously computed table plane. In our case, the plane for the chest is termed chest plane, the plane for the abdomen is termed abdomen plane. In Figure 2(b) and 2(c) these planes are visualized: abdomen plane is marked with number 2, chest plane is marked with number 3. The table plane which was used for segmentation is marked with number 1. Each best-fitting plane is not infinite but limited by a bounding polygon which is determined by the silhouette of the segmented torso. The Euclidean distance of each best-fitting plane to the table-plane constitutes one dimension of the breathing signal. Thus, for each of the ROIs (chest and abdomen in our case) a one-dimensional, time-variant breathing signal is derived. The specific signals derived for the chest and abdomen region are displayed in Figures 2(d) and 2(e).

Using planes to derive the respiratory motion of certain ROIs is advantageous to using the initially available 3D points for the following reasons, which are basically motivated by the goal to provide a stable approach in order to meet the high quality requirements of clinical use:

- Noise reduction: Using a plane fitted to a certain number of 3D points significantly lowers the influence of outliers, i.e. severely wrong measured 3D points. This ensures meeting the high quality requirements implied by the clinical application. If each single 3D point would be used to conclude on respiratory motion, an erroneously measured 3D point might indicate a state of maximum exhalation while in fact a state of maximum inhalation is present: If the 3D point is not used alone but in conjunction with adjacent correctly measured 3D points, this effect is efficiently prevented. Please refer to Section 4 to see this proven by the robustness and validity of our results.

- Stability: The indication of respiratory motion by the Euclidean distance between table plane and chest/abdomen plane is justified by the robustness of this approach. Volume estimation is error-prone as it would rely on the original 3D points, which may contain outliers which in consequence lead to a volume which is estimated too small or big. Furthermore, the table plane is computed by fitting a plane through all
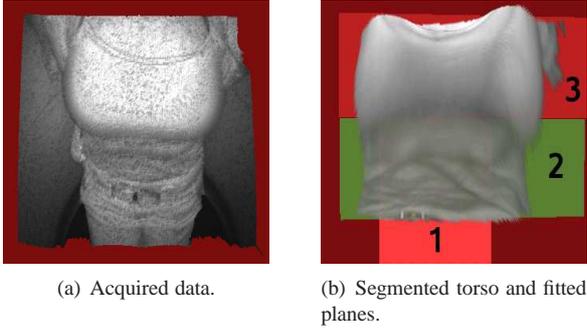
(a) Acquired data.

(b) Segmented torso and fitted planes.
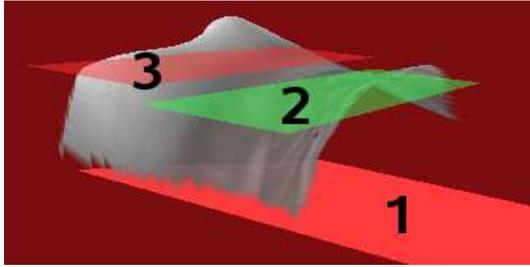
(c) Side-view of segmented torso and fitted planes.
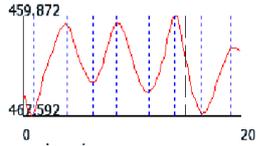
(d) Derived breathing signal for breast region: Distance of thorax plane (**3**) to table plane (**1**).
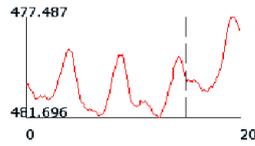
(e) Derived breathing signal for abdomen region: Distance of abdomen plane (**2**) to table plane (**1**).

Figure 2. Figure 2(a) shows the original data: Patient lying on table. Figure 2(b) shows the segmented patient body and the table plane (**1**), the abdomen plane (**2**) and the thorax plane (**3**). Figure 2(c) shows the same data observed virtually from the side. Figure 2(d) and 2(e) show the breathing signals computed over a period of 20 seconds (horizontal axis): The breathing signal is computed as the Euclidean distance (in mm; vertical axis) of thorax plane to table plane (Figure 2(d)) or abdomen plane to table plane (Figure 2(e)).

measured 3D points of the patient table and thus a stable reference plane for the subsequent steps of respiratory motion detection is available. Additionally, restricting the chest and abdomen plane to be parallel to the table plane enables a stable distance computation as the distance of parallel planes is always the same independent from the specific point on each plane which is used for distance computation.

4. Evaluation

Available ToF cameras have a depth resolution of up to 1 mm. This accuracy is only reached under optimal illu-mination conditions. Our approach derives the breathing signal from partial regression planes which are computed from probably noisy 3D points. The usage of best-fitting planes increases the stability. To validate this fact a rigid, non-deformable model of the human torso was subject to the whole processing chain described in section 3. Thus, the breathing signal of a completely non-breathing patient was computed for varying viewing angles of the camera (from perpendicular viewing angle onto the patient table upto $30°$ viewing angle). The still observed breathing motion indicates the achievable z-resolution of the breathing signal. Considering a time-span of 20 seconds the standard deviation of the distance of best-fitting plane for the torso and the table-plane was computed. The computed values were always smaller than 0.1 mm, i.e. the z-resolution of the breathing signal is about a factor ten greater than the resolution of the original ToF camera data and respiratory motion can be detected in the sub-millimeter range. Table 2 shows the detailed results. The results indicate that the best measurement results is obtained when observing the patient directly from above. Observing the patient from a lower point of view, i.e. a smaller viewing angle, decreases the measurement quality.

Angle α	Standard deviation in mm
90°	0.06
75°	0.07
60°	0.09
45°	0.09
30°	0.1

Table 2. Relation between viewing angle α and standard deviation of computed respiratory motion for a perfectly non-breathing model of the human thorax and abdomen.

To validate the correctness of the computed breathing signal we calculated for 13 patients the correlation coefficient of the computed breathing signal with the breathing signal delivered by an ANZAI belt AZ-733V[1]. The patient was advised to breath with the chest when the ANZAI belt was attached to its chest and advised to breath with the abdomen when the ANZAI belt was attached to its abdomen. For the experiments a ToF camera SwissRanger SR3100 was used. The results are displayed in Figure 3. The plot is briefly examined in the following. The horizontal axis corresponds to the 13 candidates. The vertical axis corresponds to the correlation coefficient, where two such values were computed for each patient in the following manner:

1. **Correlation coefficient for abdominal respiration:** The candidate was advised to breath with the abdomen and the ANZAI belt was attached to the abdomen. Thus, a 1D respiratory signal was acquired via the ANZAI belt. Then the proposed method (see Section 3) was applied and from the 3D information of the ab-

dominal region (automatically defined by choosing the hardware setup such that the upper of the field of view of the ToF camera contained the thorax and the lower half the abdomen) a 1D respiratory signal was derived (defined as the Euclidean distance of the best fitting plane of the ROI and the table plane). Then the correlation coefficient of these two 1D signals was computed.

2. **Correlation coefficient for thoracic respiration:** The same setup as above was used, but the candidate was advised to breath with the thorax, the ANZAI belt was attached to the thorax and the ToF camera derived the 1D respiratory signal from the 3D information of the thorax region.
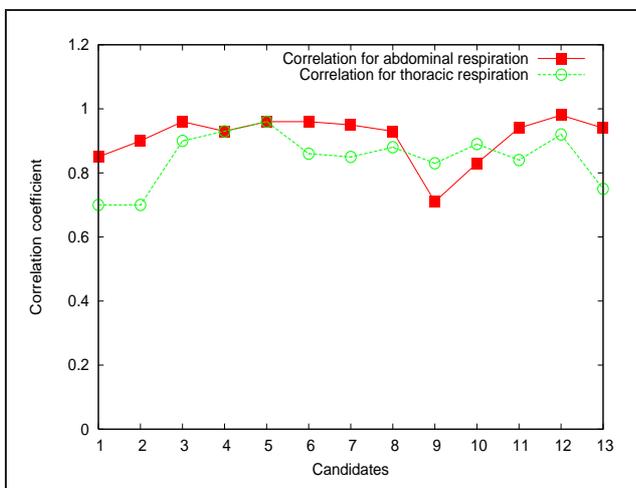
Figure 3. Evaluation results: Correlation coefficients between respiration signal computed from ToF camera data and respiration signal delivered by ANZAI belt. Circle: Correlation coefficient for thoracic respiration; Square: Correlation coefficient for abdominal respiration. See items 1 and 2 in Section 4 for detailed explanation.

5. Conclusion

We presented a non-invasive non-contact method for the computation of respiratory signals based on 3D point cloud surfaces which are acquired with a ToF camera. We have validated that the 3D information of the patient body, which is provided by a ToF camera, can be decomposed into abdominal and thoracic components by defining ROIs. By computing best fitting planes for each ROI, respiration signals for each ROI are derived by computing the displacement of each plane to a plane modeling the patient couch. These signals are significantly correlating with the reference breathing signal acquired with an ANZAI belt. The average correlation coefficient between the ToF derived respiration

signal and the ANZAI belt was 0.85 for thoracic respiration and 0.91 for abdominal respiration. Thus, its clinical relevance is given, but further research will be invested to improve the validity of the ToF based respiratory motion detection. In comparison the achieved correlation values are in the range of the correlation values (0.69-0.87) reported for systems currently in clinical use [13]. It is doubtful that much higher correlation values can be achieved as the signals under investigation measure the same physiological process, i.e. the respiration, but each of the signals is derived from a different sensor (ANZAI uses a pressure sensor attached to the skin; ToF cameras measure the time-of-flight of an optical reference signal). It can be assumed that bad correlation values obtained in our experiments stem from bad respiration signals delivered by the ANZAI belt: The respiration signal is derived from one pressure sensor, i.e. the signal is measured at only one point on the skin of the patient and thus relying strongly on a proper and stable attachment of the pressure sensor. Thus, measurement errors from the ANZAI belt are more likely to occur than errors in our ToF based respiration signal which is derived from the whole 3D surface data of the thorax of the patient. Consequently, the measurement errors of the ANZAI belt will of course decrease the computed correlation value. The computational time for computing the respiratory signals does take approx. 25 ms on a 2.0GHz single core CPU. Thus, the proposed method is real-time capable. By deriving the respiratory signal of a certain anatomical region like chest or abdomen from best-fitting planes and not the original 3D point clouds a depth-resolution of 0.1 mm is achieved.

6. Acknowledgements

References

[1] Anzai Medical Co.,Ltd. www.anzai-med.co.jp, 2007.

[2] B. Büttgen, T. Oggier, M. Lehmann, R. Kaufmann, and F. Lustenberger. CCD/CMOS Lock-In Pixel for Range Imaging: Challenges, Limitations and State-of-the-Art. In *1st Range Imaging Day*, Zürich, Switzerland, June 2005.

[3] P.-C. M. Chi, P. Balter, D. Luo, R. Mohan, and T. Pan. Relation of external surface to internal tumor motion studied with cine CT. *Medical Physics*, 33(9):3116–3123, 2006.

[4] P. Devarakota, M. Castillo-Franco, R. Ginhoux, B. Mirbach, and B. Ottersten. Occupant Classification Using Range Images. *IEEE Transactions On Vehicular Technology*, 56(4):1983–1993, July 2007.

[5] M. Holte, T. Moeslund, and P. Fihl. View Invariant Gesture Recognition using the CSEM SwissRanges SR-2 Camera . In *Proceedings of the Workshop Dynamic 3D Imaging in conjunction with DAGM'07* , pages 53 – 60. DAGM e.V., 2007. Heidelberg, Germany.

[6] E. Kollorz and J. Hornegger. Gesture recognition with a time-of-flight camera . In *Proceedings of the Workshop Dynamic 3D Imaging in conjunction with DAGM'07* , pages 86 – 93. DAGM e.V., 2007. Heidelberg, Germany.

[7] D. A. Low, P. J. Parikh, W. Lu, J. F. Dempsey, S. H. Wahab, J. P. Hubenschmidt, M. M. Nystrom, M. Handoko, and J. D. Bradley. Novel breathing motion model for radiotherapy. *Int J Radiat Oncol Biol Phys*, 63(3):921–929, November 2005.

[8] MESA Imaging AG. www.swissranger.ch, 2007.

[9] S. Nehmeh and Y. Erdi. Respiratory Motion in Positron Emission Tomography/Computed Tomography. *Seminars in Nuclear Medicine*, 38(3):152–167, May 2008.

[10] PMDTec GmbH. www.pmdtec.com, 2007.

[11] E. Rietzel and G. T. Chen. Improving retrospective sorting of 4D computed tomography data. *Medical Physics*, 33(2):377–379, 2006.

[12] T. Schamm, S. Vacek, J. Schröder, M. Zöllner, and R. Dillmann. Obstacle detection with a PMD-camera in autonomous vehicles . In *Proceedings of the Workshop Dynamic 3D Imaging in conjunction with DAGM'07* , pages 70 – 77. DAGM e.V., 2007. Heidelberg, Germany.

[13] S. Tarte, J. McClelland, S. Hughes, J. Blackall, D. Landau, and D. Hawkes. A Non-Contact Method for the Acquisition of Breathing Signals that Enable Distinction Between Abdominal and Thoracic Breathing. *Radiotherapy and Oncology*, 81(1):209, October 2006.

[14] J. Wang, J. Byrne, J. Franquiz, and A. McGoron. Evaluation of amplitude-based sorting algorithm to reduce lung tumor blurring in PET images using 4D NCAT phantom. *Comput. Methods Prog. Biomed.*, 87(2):112–122, August 2007.