Time-of-flight sensor for respiratory motion gating

Christian Schaller,[a] Jochen Penne, and Joachim Hornegger
Chair for Pattern Recognition, Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany

(Received 11 January 2008; revised 16 April 2008; accepted for publication 10 May 2008; published 13 June 2008)

In this technical note we present a system that uses time-of-flight (ToF) technology to acquire a real-time multidimensional respiratory signal from a 3D surface reconstruction of the patient's chest and abdomen without the use of markers. Using ToF sensors it is feasible to acquire a 3D model in real time with a single sensor. An advantage of ToF sensors is that their high lateral resolution makes it possible to define multiple regions of interest to compute an anatomy-adaptive multidimensional respiratory signal. We evaluated the new approach by comparing a ToF based respiratory signal with the signal acquired by a commercially available external respiratory gating system and achieved an average correlation coefficient of 0.88. © *2008 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2938521]

Key words: markerless noncontact respiratory tracking, time-of-flight, gating, 4D reconstruction

Time-of-flight (ToF) sensors provide a direct way for acquiring 3D surface information of objects.[1] More recently, applications like gesture recognition[2] or automotive passenger classification[4] are using ToF sensors and ToF is on its way to become a component of consumer electronics. As ToF sensors provide data at rates higher than 15 Hz, they are suitable for real-time 3D imaging and can also be used as an additional imaging modality in medicine. Though until now, 3D endoscopy is the only medical application where ToF sensors are supposed to be deployed.[5] We are convinced that ToF technology can contribute to enhance applications within medicine and suggest a system to perform multidimensional tracking, where our focus is on respiratory motion.

ToF camera systems can be used to actively illuminate a patient with an incoherent light signal. The signal is modulated by a cosine-shaped signal of frequency f. Usually, the emitted light is part of the nonvisible area of the spectrum in the near infrared spectral range (780 nm). Traveling with the constant speed of light in the surrounding medium, the light signal is reflected by the surface of the patient. By estimating the phase-shift ϕ (in rad) between both, the emitted and reflected light signal, the distance d can be computed as follows:

$$d = \frac{c}{2f} \cdot \frac{\phi}{2\pi}. \quad (1)$$

Based on the periodicity of the cosine-shaped modulation signal, this equation is only valid for distances smaller than $c/2f$. Currently available ToF cameras operate, e.g., at a modulation frequency of about 20 MHz. Thus, the upper limit for observable distances of available ToF camera systems is approx. 7.5 m, which is sufficient for respiratory tracking. In addition to depth values, ToF cameras provide intensity values, representing the amount of light sent back from a specific point. Detailed information about the working principle of ToF cameras can be found in Xu *et al.*[1]

There have been various approaches to account respiratory tracking in the past. Most of them are based upon tracking a discrete point set with a stereo camera system. Manufactures like BrainLAB and Varian are using infrared cameras in order to track marker on the patients chest. VisionRT provides a markerless solution to account respiratory motion using an active stereo camera system. There are also systems utilizing pressure sensors (e.g., AZ-733V, ANZAI Medical Co.) to account respiratory motion.

Today most of these systems are limited to focus on a single point or a limited area of the upper part of the body. This results in an acquisition of a 1D respiratory signal. However, there is only one system (GateRT/GateCT, VisionRT) which is competitive to our proposed approach.[3] The overall process of human respiration is a complex process and is dependent on many different factors. The human body is able to utilize both, thorax muscles as well as the diaphragm for breathing. This yields to an infinite number of possible combinations of these two systems to keep the respiratory cycle alive. Therefore a multi-dimensional respiratory signal accounting different breathing patterns could potentially improve solutions for problems dealing with respiratory motion, like motion artifacts or tumor motion due to respiration.

To emphasize technical differences between both Stereo and Structured light techniques versus ToF sensors we briefly will discuss these issues in the following:

STEREO VS TOF: The main difference between both systems is how depth information is achieved. Most stereo methods extract depth values at pixels where either corresponding features or texture information is available. To obtain dense depth information interpolation between these pixels is required. In contrast, due to their high lateral resolution ToF cameras have innately very dense depth information at constant resolution and high frame rates without the need of interpolation.

Both systems are able to operate in real time, but the accuracy for stereo systems usually drops when trying to achieve real time. Accuracy in stereo systems depends on the setup, where typical parameters are e.g., base line, focal length, etc. ToF cameras are monocular all-solid-state cameras, hence no hardware setup depended parameters have influence on the accuracy. However, the main camera specific parameters for accuracy are the amount of emitted light and the quality of the modulation frequency.

An inevitable step for stereo setup is calibration. Unless calibrated, a stereo system will not be able to convert disparity values to true depth estimates. Furthermore, many correspondence algorithms assume known epipolar geometry which is typically derived from calibration. Last regular recalibration and very robust stereo mounting is an issue for long-term use of stereo systems. Again, due to the fact that ToF cameras are all-solid-state systems, no calibration steps are necessary to acquire depth information using ToF cameras.

STRUCTURED LIGHT VS TOF: Real-time operation is often difficult to achieve in structured light techniques, especially in setups that require scanning the light through the scene. Furthermore, structured light obtains direct measurements only at scene points illuminated by the light pattern. Depth values at nonilluminated points also have to be derived via interpolation or surface fitting like for stereo based systems.

Again the accuracy depends on the setup and is dependent on parameters like base line, focus, light pattern, etc.

For structured light systems it is also mandatory to calibrate the system in order to map the observed light pattern to values. Besides calibrating the camera itself, the relative geometry of the light rays with respect to the camera has to be known. The projection unit has to be mounted very accurately and is also highly application dependent and therefore is linked with high costs.

KEY BENEFITS OF TOF CAMERAS: We think that there are several advantages to using a ToF camera system over stereo/structured light systems for respiratory motion gating. First ToF cameras are off-the-shelf technology. The automotive and consumer electronics industry are very interested in this emerging technology and a lot of research is going on within this communities. Currently a typical ToF camera is available for about 5000 Euros (approximately 7500 USD) (see Fig. 1). As some applications for ToF cameras are already established there will be a significant decrease in manufacturing costs for such cameras. Second, no calibration steps at all are necessary to operate a ToF based system. The utilization of a ToF camera for respiratory motion gating is a Plug and Play application. ToF cameras are highly portable and can easily be used at different sites without any complex setup procedures. Third, ToF cameras are very compact and small. Due to their size and the all-solid-state design ToF cameras can easily be integrated into existing systems. Fourth, several other applications are possible using ToF cameras, like patient positioning, truncation artifact correction, table removal algorithms, etc., which are intensively investigated in the CT community today.

FIG. 1. Example of a ToF camera model SR-3000 from MESA Imaging AG.

The proposed system is able to detect independent respiratory signals for different anatomical regions of the body simultaneously and in real time. This enables the system to deal with changing breathing patterns, e.g., the patient switches from breast breathing to abdominal breathing during an image acquisition or radiotherapy treatment. Hence it is also possible to determine the instantaneous type of breathing. This information can be used to adapt reconstruction algorithms or treatment protocols. Using a ToF camera [e.g., SR-3000, MESA Imaging AG (see Fig. 1)], no additional markers or complex calibration steps are necessary to acquire respiratory signals. Potential applications for the proposed method are manifold and can be found within 4D-CT-[6] and 4D-MRT[7] reconstruction, PET-[9] and SPECT-motion compensation[10] and gated radiotherapy.[11,12]

In the following we will give a brief overview how a ToF sensor enables multi-dimensional respiratory tracking. Indices in the following equations are considered to be integer values. The set $\mathbf{P}$ includes the $K \times L$ 3D points of interest acquired by a ToF camera and the set of corresponding intensity values assigned to each 3D point is denoted $\mathbf{A}$, where $i \in \{0,1,\ldots,K-1\}$, $j \in \{0,1,\ldots,L-1\}$,

$$\mathbf{P} = \{\mathbf{p}_{i,j} | \mathbf{p}_{i,j} \in \mathbb{R}^3\}, \quad \mathbf{A} = \{a_{i,j} | a_{i,j} \in \mathbb{R}\}. \tag{2}$$

Furthermore, we assume, the ToF camera is rigidly mounted above the patient table. Basically, a position perpendicular to the patient table is considered to be best, but is not mandatory for the following algorithm.

The basic idea of our method is to fit distinct planes through all 3D points belonging to a certain region of interest. Therefore, we compute a set $\mathbf{N}$ of unit normals using $\mathbf{h}_{i,j} = \mathbf{p}_{i+1,j} - \mathbf{p}_{i-1,j}$ and $\mathbf{v}_{i,j} = \mathbf{p}_{i,j+1} - \mathbf{p}_{i,j-1}$, where $i \in \{1,2,\ldots,K-2\}$, $j \in \{1,2,\ldots,L-2\}$,

$$\mathbf{N} = \left\{ \mathbf{n}_{i,j} | \mathbf{n}_{i,j} = \frac{\mathbf{h}_{i,j} \times \mathbf{v}_{i,j}}{\|\mathbf{h}_{i,j} \times \mathbf{v}_{i,j}\|} \right\}. \tag{3}$$

Due to the normalization each normal vector has only two degrees of freedom. The 3D surface coordinates are given in a left handed Euclidean coordinate system, whose origin coincides with the optical center of the ToF camera. Therefore, the two degrees of freedom are the angle α_x between the x coordinate and the z axis and the angle α_y between the y coordinate and the z axis. As we are representing the normals

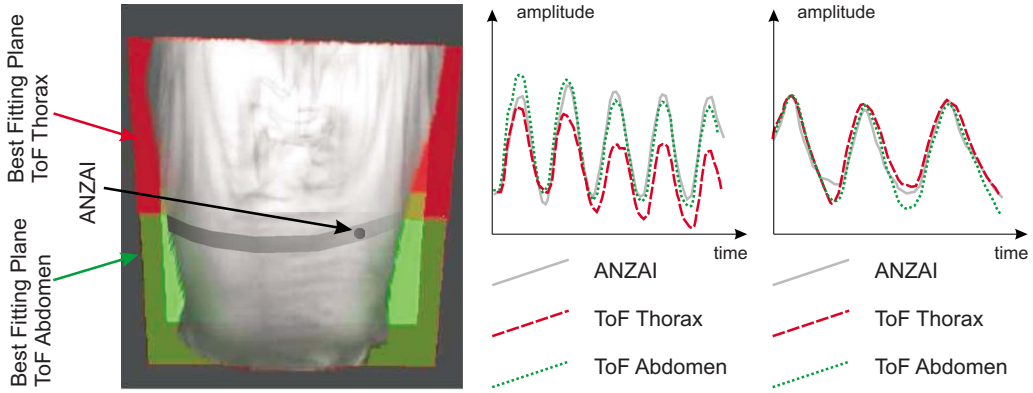

FIG. 2. On the left: 3D model of the automatically segmented upper part of the body of a patient. The colored planes are clipping planes used to determine two independent respiratory signals (chest; abdomen). On the right: Derived 1D signals. Vertical axis: Time in seconds. Horizontal axis: Motion amplitude of the planes representing the respiratory motion during breathing of chest and abdomen and the ANZAI belt (gray).

in a different parameter space, this can be considered as a Hough Transformation like approach.

Using these two angles we compute a 2D $K_h \times L_h$ histogram $\mathbf{H}$ with a bin size of δ_B degrees, where $h \in \{0,1,\ldots,\lceil 2\pi/\delta_B \rceil\}$ and δ_B is the range size of similar normal vectors we want to group. The histogram is initialized with zeros. If the angles α_x and α_y associated with the unit normal vector are in the range of the bin, each bin $h_{i,j} \in \mathbf{H}$ is incremented by the associated intensity value $a_{i,j}$. Thus, 3D coordinates of points with similar normals are associated with each bin. Using a threshold δ_H we can select corresponding cluster peaks out of the 2D histogram. After this step, k planes $\mathbf{E_i}$ with $i \in \{0,1,\ldots,k-1\}$ are identified and a certain number n_i of 3D points have been associated with each plane $\mathbf{E_i}$,

$$\mathbf{E_i} = \{\mathbf{e}_{i,j} | \mathbf{e}_{i,j} \in \mathbf{R}^3, j \in \{0,1,\ldots,n_i - 1\}\}. \quad (4)$$

In order to connect these points, 3D point coordinates of $\mathbf{E_i}$ are associated with a histogram bin $h_{k,m}$, where $k,m \in \{0,1,\ldots,\lceil 2\pi/\delta_B \rceil\}$ whose value is greater than δ_H are averaged,

$$\bar{\mathbf{e}}_i = \frac{1}{n_i} \sum_{j=0}^{n_i-1} \mathbf{e}_{i,j}. \quad (5)$$

The resulting 3D point $\bar{\mathbf{e}}_i$ is used as a seed point for an optimized flood-fill algorithm in order to identify connected points belonging to a certain plane. The obtained 3D point set of connected 3D points with similar normal vectors is denoted $\mathbf{E}'_i$. The normal vector $\mathbf{n}_i$ determined by the corresponding histogram bin $h_{k,m}$ is used as a reference normal vector.

The 3D points of $\mathbf{E}'_i$ which are gathered to a plane are used to compute a best fitting plane. Each 3D point $\mathbf{e}'_{i,j} \in \mathbf{E}'_i$ is then projected on $\hat{\mathbf{V}}_i$ along the normal vector $\mathbf{n}'_i$ of the computed plane. The resulting points are denoted $\mathbf{e}^p_{i,j} \in \mathbf{R}^3$ and the set of these points is denoted $\mathbf{E}^p$.

As the plane $\hat{\mathbf{V}}_i$ is an infinite plane by definition, we compute a bounding box to find the real physical boundaries for this plane.

Using the introduced algorithm it is possible to identify the patient table with the absence of a patient. Once the best fitting plane corresponding to the patient table is known, a patient lying on the table can be segmented. Furthermore, best fitting planes for the thorax and abdomen area can be computed automatically. The determination of the thorax and abdomen region is based on the fact, that the orientation of the patient is known and only the upper body of the patient is visible to the camera. As an initial segmentation the point cloud of the upper body is divided into two equal parts. Experiments showed that this initial guess is valid for most datasets. These planes have an additional constraint as they have to be parallel to the patient table. We can now compute the distance d of each plane to the optical center of the ToF sensor by using the Hessian normal[8] form, where $\mathbf{n}'_i$ is the unit normal vector of the plane and $\mathbf{e}^p$ is an arbitrary point on the plane.

$$d = \mathbf{n}'_i \cdot \mathbf{e}^p. \quad (6)$$

By observing d over time for each plane we can derive multiple respiratory signals (see Fig. 2). Figure 3 shows the principle of the plane computation, where the respiratory signal is derived from.

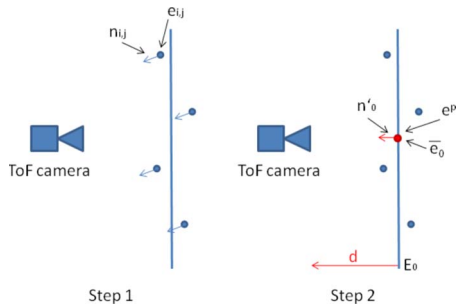

FIG. 3. Pictorial representation of the plane computation: Please note, that for clarity this figure depicts only a 2D slice of the 3D process. Step 1 shows a selection of points with similar normals [Eqs. (3) and (4)]. Within the second step the average of all points $\bar{\mathbf{e}}_0$ is computed [Eq. (5)] and the distance d is computed using the plane normal $\mathbf{n}'_0$ [Eq. (6)].

TABLE I. Experimental results showing average, median, variance of the correlation ANZAI to ToF signal of 13 test persons.

			average	median	variance	ratio[a]
	Breathing: ANZAI: ToF:	Abdomen Abdomen Abdomen	0.91	0.94	0.01	79%
	Breathing: ANZAI: ToF:	Thorax Thorax Thorax	0.85	0.86	0.01	53%
	Breathing: ANZAI: ToF:	Thorax Abdomen Thorax	0.62	0.76	0.17	54%
	Breathing: ANZAI: ToF:	Abdomen Thorax Abdomen	0.67	0.75	0.06	79%

[a]Average abdominal breathing percentage.

To evaluate our method we measured the systematic error of the system by computing the root mean square (RMS) error of the signal while looking at a rigid phantom. We observe a stable RMS of about 0.1 mm. We also compared the ToF based signal (both thorax and abdomen) with a respiratory signal acquired by an external gating system (AZ-733V, ANZAI Medical Co.). Therefore, we acquired both signals synchronized and computed the correlation coefficient. Based on 13 different persons, we computed mean, median, and variance of the computed correlation coefficients for two different setups (see Table I). We placed the ANZAI belt at the abdomen of the test person for the first test scenario. The test person was required to do abdominal breathing first and breast breathing afterwards. In the second setup, we modified the position of the ANZAI belt and put it near the breast of the test person, which was also required to perform both breathing types successively. As we acquire two signals, it is also possible to measure the ratio between abdominal breathing and breast breathing by comparing the amplitude of both signals for the last 20 seconds (see Table I).

We could show that the deployment of a ToF camera system within respiratory tracking applications is comparable to existing systems. Furthermore, using a ToF sensor it is possible to acquire multiple respiratory signals for different anatomical regions in real time. As an example we showed the simultaneous acquisition of two respiratory signals, one for abdominal breathing and one for breast breathing. In general, the generic method is also feasible to partition the body into more regions of interest, if necessary. The computation time on a standard single core computer (Pentium M, 2 GHz) is below 50 ms per frame, therefore we are able to sample respiratory signals with a frequency of about 20 Hz. A demonstration video of the prototype is deposited at EPAPS.[13]

ACKNOWLEDGMENTS

This work was supported by the International Max Planck Research School for Optics and Imaging (IMPRS-OI), Erlangen, Germany, the Erlangen Graduate School in Advanced Optical Technologies (SAOT), and the Sonderforschungsbereich 603 (SFB603).

[a)]Author to whom correspondence should be addressed. Electronic mail: christian.schaller@informatik.uni-erlangen.de

[1]Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, and T. Ringbeck, "Smart pixel—photometric mixer device (PMD)/new system concept of a 3D-imaging-on-a-chip," in *5th International Conference on Mechatronics and Machine Vision in Practice*, pp. 259–264, Nanjing, China, 1988.

[2]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*, pp. 86–93, Heidelberg, Germany, 2007.

[3]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," Radiother. Oncol. **81**(1), S209 (2006).

[4]P. R. Devarakota, M. Castillo-Franco, R. Ginhoux, B. Mirbach, and B. Ottersten, "Occupant classification using range images," IEEE Trans. Veh. Technol. **56**(4), 1983–1993 (2007).

[5]J. Penne, K. Höller, S. Krüger, and H. Feussner, "NOTES 3D: Endoscopes learn to see 3-D; basic algorithms for a novel endoscope," in *Second International Conference on Computer Vision Theory and Applications*, Barcelona (Insticc Press, 2007), pp. 134–139.

[6]J. Biederer, J. Dinkel, H. Bolte, T. Welzel, B. Hoffmann, C. Thierfelder, U. Mende, J. Debus, M. Heller, and H. U. Kauczor, "Respiratory-gated helical computed tomography of lung: Reproducibility of small volumes in an ex vivo model," Int. J. Radiat. Oncol., Biol., Phys. **69**(5), 1642–1649 (2007).

[7]M. von Siebenthal, G. Szekely, U. Gamper, P. Boesiger, A. Lomax, and P. Cattin, "4D MR imaging of respiratory organ motion and its variability," Phys. Med. Biol. **52**(6), 1547–1564 (2007).

[8]I. N. Bronshtein, K. a. Semendyayev, G. Musiol, and H. Muhlig, *Handbook of Mathematics*, 4th ed. (Springer, New York, 2004).

[9]R. A. Cook, G. Carnes, T. Lee, and R. G. Wells, "Respiration-averaged CT for attenuation correction in canine cardiac PET/CT," J. Nucl. Med. **48**(5), 811–818 (2007).

[10]H. Ue, H. Haneishi, H. Iwanaga, and K. Suga, "Nonlinear motion correction of respiratory-gated lung SPECT images," IEEE Trans. Med. Imaging **25**(4), 486–495 (2006).

[11]Y. Seppenwoolde, R. I. Berbeco, S. Nishioka, and B. Heijmen, "Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: A simulation study," Med. Phys. **34**(7), 2774–2784 (2007).

[12]D. Ionascu, S. B. Jiang, S. Nishioka, H. Shirato, and R. I. Berbeco, "Internal-external correlation investigations of respiratory induced motion of lung tumors," Med. Phys. **34**(10), 3892–3903 (2007).

[13]See EPAPS Document No. E-MPHYA6-35-026807 for a demonstration video of the prototype. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.