

3-D Operation Situs Reconstruction with Time-of-Flight Satellite Cameras Using Photogeometric Data Fusion

Sven Haase¹, Sebastian Bauer¹, Jakob Wasza¹,
Thomas Kilgus³, Lena Maier-Hein³, Armin Schneider⁴, Michael Kranzfelder⁴,
Hubertus Feußner⁴, Joachim Hornegger^{1,2}

¹ Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
sven.haase@fau.de

² Erlangen Graduate School in Advanced Optical Technologies (SAOT)

³ Div. Medical and Biological Informatics Junior Group: Computer-assisted
Interventions, German Cancer Research Center (DKFZ) Heidelberg

⁴ Minimally Invasive Therapy and Intervention, Technical University of Munich

Abstract. Minimally invasive procedures are of growing importance in modern surgery. Navigation and orientation are major issues during these interventions as conventional endoscopes only cover a limited field of view. We propose the application of a Time-of-Flight (ToF) satellite camera at the zenith of the pneumoperitoneum to survey the operation situs. Due to its limited field of view we propose a fusion of different 3-D views to reconstruct the situs using photometric and geometric information provided by the ToF sensor. We were able to reconstruct the entire abdomen with a mean absolute mesh-to-mesh error of less than 5 mm compared to CT ground truth data, at a frame rate of 3 Hz. The framework was evaluated on real data from a miniature ToF camera in an open surgery pig study and for quantitative evaluation with a realistic human phantom. With the proposed approach to operation situs reconstruction we improve the surgeons' orientation and navigation and therefore increase safety and speed up surgical interventions.

1 Introduction

Minimally invasive procedures gained a lot of attention, recently. In comparison to conventional surgery, endoscopic interventions aim at reducing pain, scars, recovery time and thereby hospital stay. Therefore, minimally invasive procedures hold benefits for both the patients and the hospital. Navigation and orientation are of particular relevance for the surgeon in minimally invasive surgery due to the limited field of view with conventional endoscopes. To improve both, different concepts to insert additional cameras have been proposed [1,2]. For instance, Cadeddu et al. describe a video camera that is positioned on the posterior abdominal wall and guided by an anterior magnetic device. Instead, we propose the concept of *3-D satellite cameras* as illustrated in Fig. 1(a). These cameras are inserted into the abdomen via a trocar and positioned at the top

of the pneumoperitoneum. Here, the imaging device can survey the operation field. Nevertheless, due to size limitations in endoscopic procedures, satellite cameras have shortcomings related to the hardware and optical systems. One of these is a narrow field of view. To expand the limited field of view the camera will reconstruct the entire situs initially by rotating and acquiring images from different areas for data fusion and then focus on the operation field. With no further repositioning of the patient the assumption of rigidity is acceptable for navigation assistance. Opposed to related work, our satellite camera delivers Time-of-Flight (ToF) 3-D surface and photometric information instead of pure 2-D video data. This enables a broad field of medical applications, e.g. collision detection, automatic navigation or registration with preoperative data.

Different approaches for data fusion with real-time capability have been proposed recently [3,4,5]. Warren et al. proposed a simultaneous localization and mapping based approach for natural orifice transluminal endoscopic surgery [3]. For stereo endoscopy, Röhl et al. presented a novel hybrid recursive matching algorithm that performs matching on the disparity map and the two input images [5]. Areas with little textural diversity are challenging scenarios for those color-based approaches regarding 3-D reconstruction. Instead of using conventional endoscopes we propose to navigate a 3-D satellite camera for reconstruction of the whole situs to enable a better orientation within the pneumoperitoneum. A ToF sensor acquires photogeometric data, i.e. both range data and intensity images encoding the amplitudes of the measured signal. By exploiting both complementary information we are able to reconstruct surfaces in areas with low textural diversity as well as areas with low topological diversity. In-vivo experiments on real data from a miniature ToF camera indicate the feasibility of using 3-D satellite cameras for situs reconstruction during minimally invasive surgery.

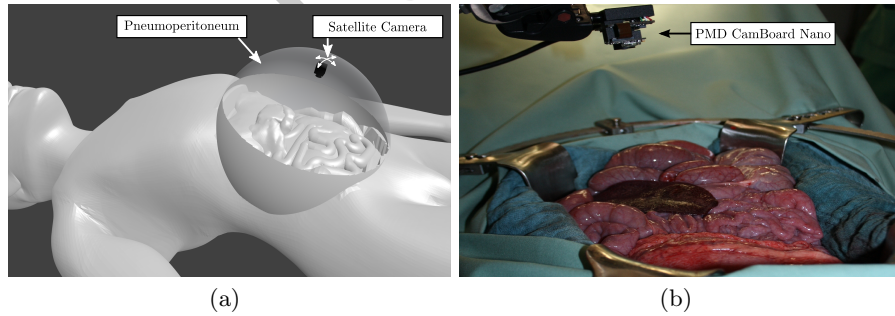


Fig. 1. (a) Illustration of the 3-D Time-of-Flight satellite camera hovering above the situs at the zenith of the pneumoperitoneum. (b) Experimental setup for acquiring in-vivo data in a pig study. Note the physical dimension of the miniature ToF camera.

2 Methods

We use a truncated signed distance function (TSDF) [6] to reconstruct the interior abdominal space. The advantage of this approach is threefold. First, by incorporating successive frames, details are refined. Second, the TSDF allows incorporating additional information for regions that were seen from different perspectives. This allows implicit denoising of data with lower quality. Third, the TSDF representation is computational efficient with both constant run time and memory. Inspired by the work of Whelan et al. [7], we enhanced the traditional TSDF from 3-D to 4-D to incorporate the amplitude domain. In this context, a major contribution is the incorporation of confidence weights derived from ToF characteristics into the TSDF reconstruction. To cope with real-time requirements in medical environments we apply a GPU-based photogeometric registration approach [8]. Below we detail the initial preprocessing for ToF data.

2.1 Time-of-Flight Data Processing

As ToF devices exhibit a low signal-to-noise ratio [9], preprocessing range data is an essential step. We apply a real-time capable framework that combines three processes. First, we interpolate invalid pixels based on a normalized convolution [10]. Second, we decrease temporal noise by averaging successive frames, allowed by the high acquisition speed of our sensor (see Sect. 3). Third, we perform bilateral filtering for edge-preserving denoising. The amplitude data depend not only on the material but also on the distance to the light source. Therefore, normalizing this data is necessary for incorporating the photometric domain into the registration process. We normalize amplitude data according to a simplified physical model $\tilde{a}(\mathbf{x}) = a(\mathbf{x})r^2(\mathbf{x})$ [11]. Here, a denotes the amplitude value at the pixel coordinate \mathbf{x} and r denotes the measured radial distance. Furthermore, we also apply edge-preserving denoising in the amplitude domain. Nevertheless, photometric registration would still be affected by glare lights. To cope with this, we detect glare lights by basic thresholding and label them as invalid pixels.

2.2 Photogeometric Data Fusion into a Volumetric TSDF Model

The preprocessed data deliver photogeometric information of the situs from different points of view. For estimating the rotation matrix \mathbf{R}_k and the translation vector \mathbf{t}_k between the camera coordinate system of frame k and the global world coordinate system we align two successive frames by applying an approximate iterative closest point (ICP) implementation [8]. The approach extends the traditional 3-D nearest neighbor search within ICP to higher dimensions, thus enabling the incorporation of additional complementary information, e.g. photometric data. It is based on the random ball cover acceleration structure for efficient nearest neighbor search on the GPU [12]. For 4-D data considered in this paper, the photogeometric distance metric d is defined as:

$$d(\mathbf{m}, \mathcal{F}) = \min_{\mathbf{f} \in \mathcal{F}} ((1 - \chi) \|\mathbf{f}_g - \mathbf{m}_g\|_2^2 + \chi \|f_p - m_p\|_2^2), \quad (1)$$

where $\chi \in [0, 1]$ is a non-negative constant weighting the influence of the photometric data. \mathbf{f}_g and \mathbf{m}_g denote the position of an individual 3-D point in the fixed point set \mathcal{F} and the moving point set \mathcal{M} , respectively. f_p and m_p denote the photometric scalar value given by the normalized amplitude data \tilde{a} .

Our reconstruction is based on a volumetric model defined by a TSDF along the lines of [6]. The TSDF is based on an implicit surface representation given by the zero level set of an approximated signed distance function of the acquired surface. For each position $\mathbf{p} \in \mathbb{R}^3$, the TSDF \mathcal{T}_S holds the distance to the closest point on the current range image surface w.r.t. the associated inherent projective camera geometry:

$$\mathcal{T}_S(\mathbf{p}) = \eta(\|\mathbf{S}(\mathbf{P}_s(\mathbf{p}_k))\|_2 - \|\mathbf{p}_k\|_2) C(\mathbf{P}_s(\mathbf{p}_k)), \quad (2)$$

where $\mathbf{p}_k = \mathbf{R}_k \mathbf{p} + \mathbf{t}_k$ denotes the transformation of \mathbf{p} from world space into the moving local camera space. $\mathbf{P}_s : \mathbb{R}^3 \mapsto \mathbb{R}^2$ performs the projection of each 3-D point \mathbf{p}_k into the image plane. \mathbf{S} reconstructs the 3-D surface point to a given range value in the sensor domain and η is a truncation operator that controls the support region, i.e. outside this region the distance function is cut off. C is a confidence weight that is introduced below.

For improved reconstruction, e.g. in terms of loop closures, we fuse our data in a frame-to-model manner [6], i.e. the current frame is not registered to the previous frame directly but to a raycasted image of the reconstructed model seen from the camera of the previous frame. Due to our high acquisition frame rate the rigid assumption for frame-to-model transformation estimation is tolerable.

To enable photogeometric reconstruction in a frame-to-model manner, our approach stores and fuses amplitude information. The amplitude value \mathcal{T}_A is described by:

$$\mathcal{T}_A(\mathbf{p}) = \tilde{a}(\mathbf{P}_s(\mathbf{p}_k)) C(\mathbf{P}_s(\mathbf{p}_k)). \quad (3)$$

For robust data fusion we assign a confidence weight to each TSDF value to describe the reliability of the new measurement. In particular, we introduce the confidence function C as:

$$C(\mathbf{x}) = e^{-\frac{\alpha}{\tilde{a}(\mathbf{x})}} e^{-\frac{\|\mathbf{x} - \mathbf{c}\|}{\beta}} v(\mathbf{x}), \quad (4)$$

with α and β controlling the influence of the first terms and \mathbf{c} denoting the pixel position of the center in the range image. Here, we exploit three characteristics of ToF cameras. With higher distances to the center or lower amplitude values the confidence decreases. The binary validity information $v(\mathbf{x})$ is provided by the ToF sensor and combined with the result of our glare light detection.

To provide temporal denoising we benefit from different frames that acquired the same spots by:

$$\tilde{\mathcal{T}}^t = \gamma \mathcal{T}^t + (1 - \gamma) \mathcal{T}^{t-1}, \quad (5)$$

where $\tilde{\mathcal{T}}^t$ denotes the temporal denoised result and \mathcal{T}^t denotes the current result of Eq. 2 and Eq. 3. The weight γ describes the influence of the previously reconstructed result \mathcal{T}^{t-1} .

3 Experiments

The experiments are split into two parts. For qualitative evaluation we acquired real in-vivo data in a pig study. For quantitative evaluation we acquired real data of a human abdomen phantom and compared it to CT ground truth data. In both experiments the satellite camera was moved across the situs at a typical measuring distance of 20 cm, while reconstructing the 3-D geometry of the operation field. For both experiments we applied the same preprocessing pipeline and used 25 frames for data fusion. In particular, we acquired a scene of 250 frames and fused every 10th frame to obtain a sufficient frame-to-frame movement. We averaged data over 3 successive frames to reduce temporal noise for the registration process. The parameters for the bilateral filter and the normalized convolution were set empirically. The temporal denoising parameter was set to $\gamma = 0.95$. The weightings of the confidence terms were set to $\alpha = 2000$ and $\beta = 100$. The photometric weighting was set to $\chi = 0.00025$. Regarding the scale of the parameter the maximum amplitude value of 40000 has to be taken into account. In the considered scenario, the texture is rather homogeneous. Hence, we set χ comparably low. Nonetheless, it guides the registration in flat regions. Our framework was implemented in CUDA and evaluated on an off-the-shelf laptop with an NVIDIA Quadro FX 1800M GPU and an i7-940XM CPU. For our experiments we used a CamBoard Nano miniature ToF camera from PMD Technologies GmbH, Siegen, Germany. It acquires ToF data at 60 Hz with a resolution of 160×120 px. The data is available online ¹.

Even though being a compact device the CamBoard Nano ($37 \times 30 \times 25 \text{ mm}^3$) exceeds the physical dimension needed for minimally invasive surgery. Hence, we performed our experiments in an open surgery scenario. For qualitative evaluation a pig was examined under artificial respiration, see Fig. 1(b). We compare the frame-to-frame data fusion to our frame-to-model approach and point out the benefits of our contributions - incorporating photometric data into the registration process and adding confidence weights to the TSDF. Furthermore, we reconstructed the whole operation situs from a ToF sequence of 25 frames.

Quantitative evaluation is performed by scanning an ELITE phantom [13] with CT and then acquiring data with the CamBoard Nano, while reconstructing the abdomen with our proposed framework. To compare the ToF reconstruction with the ground truth surface data, anatomical landmarks on both meshes were detected manually and registered. Then, we calculated the Hausdorff distance for a volume of interest to compare both surfaces in a mesh-to-mesh manner.

4 Results and Discussion

To investigate the performance on in-vivo data we reconstructed the abdomen of a pig, see Fig. 2. In addition, Fig. 3 illustrates the weakness of the frame-to-frame reconstruction compared to our frame-to-model result. Note that the blood vessel labeled in Fig. 3(c) is visible when reconstructing the scene using the proposed

¹ <http://www5.cs.fau.de/research/data/>

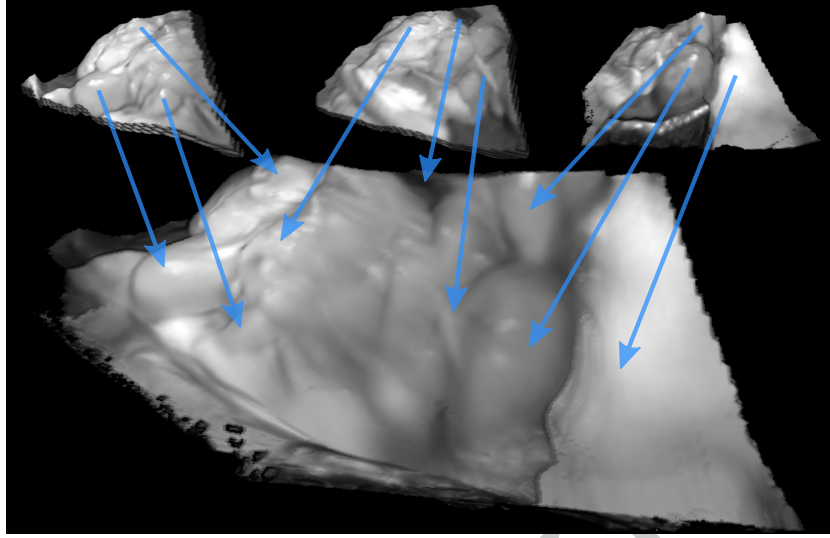


Fig. 2. The first row depicts 3 single frames of a 25-frames ToF sequence, acquired in an experimental in-vivo pig study. Below, the reconstructed operation situs is illustrated. Note that salient structures of individual frames are clearly visible in the reconstruction.

confidence weighting scheme and 4-D data, while it is blurred when considering the geometric data alone in the registration process, see Fig. 3(d). Compared to the frame-to-frame data fusion and the frame-to-model approach without confidence weights we achieve a smoother result while preserving the shape. In experiments with CT ground truth data we achieved a mean absolute mesh-to-mesh distance of 4.73 mm. The color-coded Hausdorff distance of the considered volume of interest is illustrated in Fig. 4(b). Note that substantial distance errors occur at the boundaries where only few data were available. The reconstruction accuracy in the central region was substantially better, see Fig. 4(c).

The experiments on real data illustrate that our framework is capable to provide reliable reconstructions even in the presence of severe noise ($\sigma_{ToF} \sim 5$ mm). Nevertheless, the distance map in Fig. 4(b) indicates that systematic errors in the ToF data and insufficient data at boundaries lead to locally imperfect reconstructions with higher mesh-to-mesh errors. The results on in-vivo data stress the benefits of our contributions. With our proposed 4-D frame-to-model data fusion we achieved more robust results compared to the 3-D approach. The introduced confidence weights for the TSDF produced a smoother reconstruction.

5 Conclusions

In this paper we proposed the use of a miniature ToF device as a 3-D satellite camera for minimally invasive surgery to reconstruct the operation situs. To extend the camera’s field of view, we introduced a fusion framework that allows to

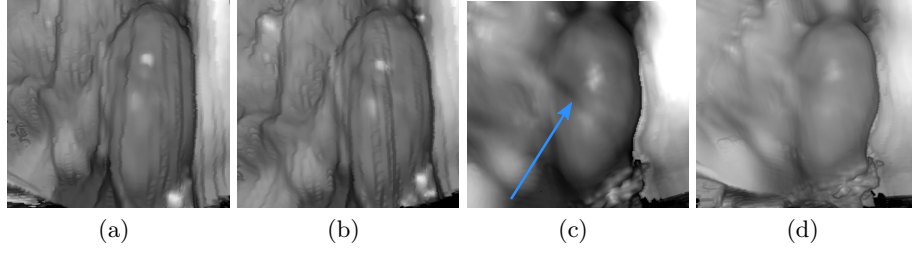


Fig. 3. Closeup of the reconstructed operation situs at the position of the upper right image in Fig. 2. Reconstruction results for (a) 4-D frame-to-frame data fusion, (b) 4-D frame-to-model fusion without confidence weights, (c) 4-D frame-to-model approach with confidence weights, (d) frame-to-model approach using geometric information only. Note that the blood vessel marked with the red arrow is best visible in (c).

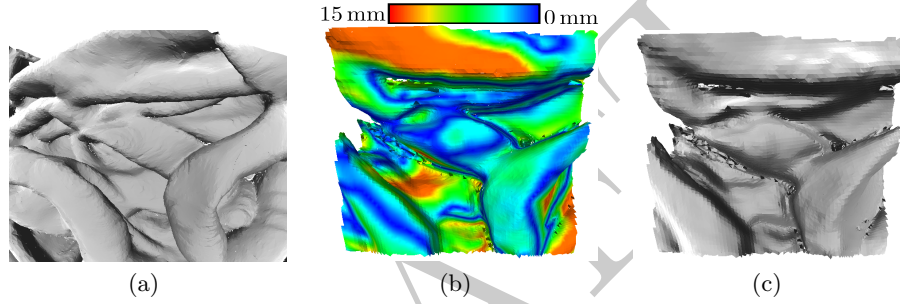


Fig. 4. (a) Ground truth data of the ELITE phantom. (b) Time-of-Flight surface with color coded Hausdorff distances. (c) Time-of-Flight reconstruction.

reconstruct the operation situs for better orientation and navigation using both geometric and photometric information. Our proof-of-concept GPU implementation runs at 3 Hz on an off-the-shelf laptop. Experiments on real data showed that we benefit from our proposed confidence weights and resulted in a mean absolute mesh-to-mesh distance of less than 5 mm compared to ground truth CT data. Future work will investigate the upcoming generation of miniaturized ToF cameras that are expected to feature a geometry that fits through a trocar.

Acknowledgments We gratefully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. HO 1791/7-1. This research was supported by the Graduate School of Information Science in Health (GSISH) and the TUM Graduate School. The authors acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the DFG in the framework of the German excellence initiative. We also thank the Dr. Pfleger Stiftung for their partial support. This work was supported in the context of the R&D program IuK Bayern under Grant No. IUK338/001.

References

1. Oleynikov, D., Rentschler, M., Hadzialic, A., Dumpert, J., Platt, S.R., Farritor, S.: Miniature robots can assist in laparoscopic cholecystectomy. *Surgical Endoscopy* **19**(4) (2005) 473–476
2. Cadeddu, J., Fernandez, R., Desai, M., Bergs, R., Tracy, C., Tang, S.J., Rao, P., Desai, M., Scott, D.: Novel magnetically guided intra-abdominal camera to facilitate laparoendoscopic single-site surgery: initial human experience. *Surgical Endoscopy* **23** (2009) 1894–1899
3. Warren, A., Mountney, P., Noonan, D., Yang, G.Z.: Horizon StabilizedDynamic View Expansion for Robotic Assisted Surgery (HS-DVE). *International Journal of Computer Assisted Radiology and Surgery* **7**(2) (2012) 281–288
4. Mountney, P., Yang, G.Z.: Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. In: *IEEE Engineering in Medicine and Biology Society*. Volume 1. (2009) 1184–7
5. Röhl, S., Bodenstedt, S., Suwelack, S., Kenngott, H., Mueller-Stich, B., Dillmann, R., Speidel, S.: Dense gpu-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. *Medical Physics* **39**(3) (2012) 1632–1645
6. Newcombe, R.A., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In: *10th IEEE International Symposium on Mixed and Augmented Reality*. (2011) 127–136
7. Whelan, T., Johannsson, H., Kaess, M., Leonard, J., McDonald, J.: Robust real-time visual odometry for dense RGB-D mapping. In: *IEEE Robotics and Automation*. (2013)
8. Bauer, S., Wasza, J., Lugauer, F., Neumann, D., Hornegger, J.: Real-Time RGB-D Mapping and 3-D Modeling on the GPU Using the Random Ball Cover. In Fossati, A., Gall, J., Grabner, H., Ren, X., Konolige, K., eds.: *Consumer Depth Cameras for Computer Vision - Research Topics and Applications*. (2013) 27–48
9. Kolb, A., Barth, E., Koch, R., Larsen, R.: Time-of-flight cameras in computer graphics. *Computer Graphics Forum* **29**(1) (2010) 141–159
10. Knutsson, H., Westin, C.F.: Normalized and Differential Convolution: Methods for Interpolation and Filtering of Incomplete and Uncertain Data. In: *IEEE Computer Vision and Pattern Recognition*. (1993) 515–523
11. Oprisescu, S., Falie, D., Ciuc, M., Buzuloiu, V.: Measurements with ToF Cameras and Their Necessary Corrections. In: *International Symposium on Signals, Circuits and Systems*. Volume 1. (2007) 1–4
12. Cayton, L.: Accelerating nearest neighbor search on manycore systems. *CoRR abs/1103.2635* (2011)
13. Gillen, S., Gröne, J., Kndgen, F., Wolf, P., Meyer, M., Friess, H., Buhr, H.J., Ritz, J.P., Feussner, H., Lehmann, K.: Educational and training aspects of new surgical techniques: experience with the endoscopiclaparoscopic interdisciplinary training entity (elite) model in training for a natural orifice transluminal endoscopic surgery (notes) approach to appendectomy. *Surgical Endoscopy* **26** (2012) 2376–2382