3D Reconstruction for Minimally Invasive Surgery: Lidar versus Learning-based Stereo Matching

Author(s):
Caccianiga, Guido; Nubert, Julian; Hutter, Marco; Kuchenbecker, Katherine J.

Publication date:
2023-05-29

Permanent link:
https://doi.org/10.3929/ethz-b-000659440

3D Reconstruction for Minimally Invasive Surgery: Lidar versus Learning-based Stereo Matching

Guido Caccianiga[1,2], *Student Member, IEEE*, Julian Nubert[2,1], *Student Member, IEEE*, Marco Hutter[2], *Member, IEEE*, Katherine J. Kuchenbecker[1], *Fellow, IEEE*

***Abstract*— This work investigates real-time 3D surface reconstruction for minimally invasive surgery. Specifically, we analyze depth sensing through laser-based time-of-flight sensing (lidar) and stereo endoscopy on ex-vivo porcine tissue samples. When compared to modern learning-based stereo matching from endoscopic images, lidar achieves higher precision, lower processing delay, higher frame rate, and superior robustness against sensor distance and poor illumination. Furthermore, we report on the negative effect of near-infrared light penetration on the accuracy of time-of-flight measurements across different tissue types.**

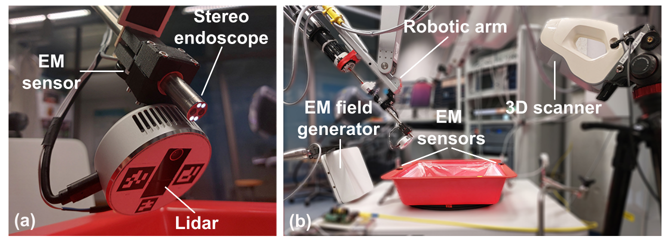

Fig. 1. (a) Lidar and electromagnetic (EM) motion sensors rigidly attached to the da Vinci stereo endoscope; (b) overall hardware setup.

I. Introduction and Related Work

Accurate 3D imaging of the patient's anatomy and its efficient intraoperative deployment are known to be required for the integration of stereotaxis in fields like orthopedics and neurosurgery [1]. While robotics and image guidance are becoming the backbone of abdominal minimally invasive surgery (MIS), obtaining reliable and complete intraoperative 3D reconstruction of the soft organs from endoscopic images remains a challenge. Once obtained, an accurate real-time map of the abdominal cavity could improve the surgeon's spatial awareness as well as open the way for partially or fully autonomous robotic procedures.

Learning-based monocular [2] and stereo [3] depth estimation in the visible light spectrum represent the state of the art for shape sensing in computer vision. While they achieve satisfactory performance in a variety of research applications, the output of each model highly depends on the image content and the availability of training data, limiting their deployment in scope and generalization. Furthermore, RGB imaging in the abdominal cavity frequently suffers from the presence of blood, smoke, and reflective surfaces, as well as a lack of visual features [4]. Finally, processing time remains a bottleneck for large state-of-the-art neural network (NN) models, rendering high-resolution depth estimation difficult for real-time applications.

In the last decades, laser-based time-of-flight sensing, also known as light detection and ranging (lidar), has emerged as a leading technology for indoor and outdoor robotic perception. The use of lidar in surgery has been mostly limited by hardware size and heat production [5], as well as its limited field of view and distortions introduced by the fiber optic coupling [6]. Previous work by Caccianiga and Kuchenbecker [7] proposed the use of a small commercial lidar sensor as an additional viewpoint attached to a robotic MIS cannula, achieving promising results in a dry-lab experiment. In this study, we report on the use of time-of-flight sensing for imaging two different ex-vivo porcine tissues at surgically relevant working distances (8–16 cm). The 3D output of the lidar sensor is carefully compared to that of a state-of-the-art learning-based stereo matching algorithm, with special attention to real-time performance and the accuracy of the reconstructed surface. We refer to [8] for a more detailed (statistical) analysis report.

[1] Haptic Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, Germany.
[2] Robotic Systems Lab, ETH Zurich, Switzerland.
Corresponding author: Guido Caccianiga, caccianiga@is.mpg.de

II. Setup and Experiments

For the experiments, we used two fresh samples of ex-vivo porcine tissue: *i)* a square cut of the abdominal wall and *ii)* a whole liver with gallbladder and bile ducts. Liquid porcine blood was poured on the tissue samples to simulate intra-operative bleeding. Our imaging devices were an Intuitive Surgical da Vinci Si Surgical System equipped with a 0° HD stereo endoscope and an Intel RealSense L515 lidar camera rigidly attached parallel to the end of this endoscope. An Artec Eva 3D scanner was used to obtain an accurate ground-truth reconstruction of the ex-vivo samples. Furthermore, a Northern Digital Aurora electromagnetic tracking system was used to kinematically link the robotic arm holding the cameras to the imaged scene. The setup is shown in Fig. 1. An Ubuntu 20.04 workstation with an RTX 3080 GPU was used for hardware control, signal processing, and data recording. Stereo matching was implemented with a custom ROS wrapper around RAFT-Stereo [9]. The lidar 3D output was acquired through the RealSense SDK (v2.50) and the RealSense ROS wrappers (v2.3.2).

First, we compared the temporal performance of the two video pipelines in terms of frame rate and processing delay. Then, we investigated the effect of four surgery-relevant experimental conditions (imaging distance, type of tissue,

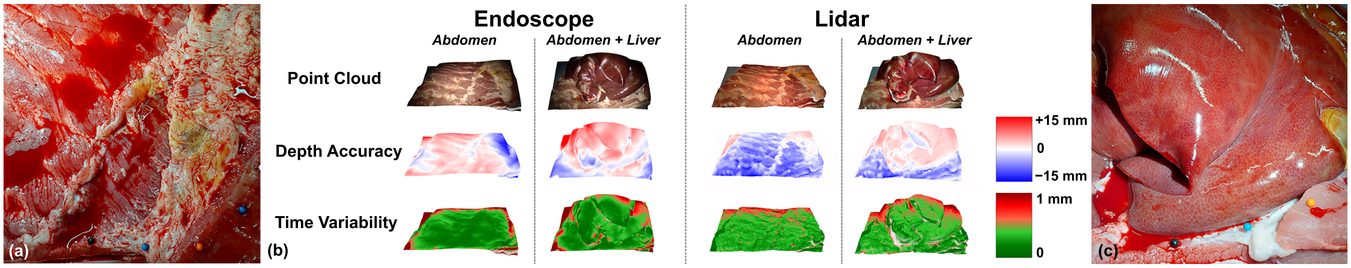

Fig. 2. (a) A view of the ex-vivo porcine abdomen. (b) Examples of point clouds are shown with the original RGB texture and color-mapped with mean (*Depth Accuracy*) and standard deviation (*Time Variability*) of the signed distance from ground truth. Error metrics are computed at each point from a batch of 125 static video frames. A clear *Depth Accuracy* offset is seen for lidar between abdominal muscle fibers (blue), fat (white), and liver tissue (light red). *Time Variability* highlights different patterns of depth measurement fluctuation across the imaging modalities. (c) A view of the ex-vivo porcine liver.

illumination, and presence of blood) on the reconstruction error by comparing each setting to the ground-truth scans.

III. Results and Discussion

When processing the da Vinci stereo images (downsampled to 360i), we measured a total delay from capture to 3D visualization (RViz) of 185 ms (39 ms for disparity inference) which is 65% larger than for the lidar (112 ms) at equal resolution. Furthermore, stereo-matching outputs 3D point clouds at 15 Hz; around 2× slower. Stereo disparity inference time increases to 89 ms for full HD frames, limiting the output to 5 Hz while increasing the delay (235 ms).

Both 3D reconstruction techniques showed an average absolute error in the 1–5 mm range, with a grand mean of 2.8 mm for the stereo reconstruction and 2.0 mm for the lidar. Performances degrade greatly to 4.9 mm for the endoscope and more moderately to 2.9 mm for the lidar in the presence of blood and low illumination. Imaging distance affects the 3D stereo output, with better depth and shape accuracy at close range. Differently, the lidar exhibited little or no change in accuracy due to target distance. Overall, the two cameras had comparable time variability, with an average standard deviation of the signed error below 1 mm. For the stereo reconstruction, high standard deviation is seen in areas with low illumination (e.g., at the edge of the frame), while the lidar's time variability is highest at the most distant points.

We found an offset of about 5 mm (object estimated as more distant) in the point clouds generated by the lidar imaging the abdominal wall sample. When looking at the respective signed distance fields (Fig. 2), we see that portions of the tissue with high fat concentration are estimated with close to zero error (white areas), while exposed muscle fibers show a negative error (blue areas). The observed depth shift between muscle, fat, and liver tissue is likely caused by the interaction between the laser light (860 nm wavelength) and the specific tissue density and biological composition. Additional investigation might include different tissue thicknesses, multiple laser wavelengths, and pathological tissue samples. Furthermore, areas of accumulated blood led to a positive shape error in lidar point clouds (surface estimated as closer). The opposite effect was noticed in the 3D volumes generated through stereo matching, where blood pools led to negative shape error. This last aberration was particularly strong in stereo imaging when blood was considered together with local light-reflection effects. The nature of such artifacts needs to be investigated further, especially in conjunction with more realistic lighting and hemorrhagic conditions.

IV. Conclusions and Future Work

In addition to its better precision in 3D shape reconstruction, lidar shown clear advantages over image-based stereo matching such as higher frequency, lower latency, and improved robustness concerning target distance and illumination. Simultaneously, we found that lidar accuracy is affected by biological tissue type. Proper engineering and fine-tuning of a miniaturized time-of-flight sensor for endoscopy could be a promising pathway for future high-fidelity real-time 3D reconstruction in surgery.

Our next steps will concentrate on the development of robust real-time registration and fusion of multiple 3D RGB-D streams captured from different perspectives. We believe that computer-integrated surgery will benefit from the abundance of intraoperative 3D imaging sources and the efficient processing and visualization of their outputs.

References

[1] J. Martínez and F. Cardinale, *Robotics in Neurosurgery: Principles and Practice*. Springer International Publishing, 2022.

[2] Y. Ming, X. Meng, C. Fan, and H. Yu, "Deep learning for monocular depth estimation: A review," *Neurocomputing*, vol. 438, pp. 14–33, 2021.

[3] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, "A survey on deep learning techniques for stereo-based depth estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 1738–1764, 2022.

[4] P. J. E. Edwards, D. Psychogyios, S. Speidel, L. Maier-Hein, and D. Stoyanov, "SERV-CT: A disparity dataset from cone-beam cone-beam CT for validation of endoscopic 3D reconstruction," *Medical Image Analysis*, p. 102302, 2021.

[5] A. Roberti, N. Piccinelli, F. Falezza, G. De Rossi, S. Bonora, F. Setti, P. Fiorini, and R. Muradore, "A time-of-flight stereoscopic endoscope for anatomical 3D reconstruction," in *Proc. IEEE International Symposium on Medical Robotics (ISMR)*, 2021.

[6] R. Stolyarov, V. Buharin, M. Val, C. Beurskens, M. DeMaio, S. Srinivasan, T. Calef, and P. Kim, "Sub-millimeter precision 3D measurement through a standard endoscope with time of flight," in *Proc. SPIE*, vol. 11949, 3 2022, p. 119490E.

[7] G. Caccianiga and K. J. Kuchenbecker, "Dense 3D reconstruction through lidar: A new perspective on computer-integrated surgery," in *Proc. Hamlyn Symposium on Medical Robotics*, 1 2022, pp. 63–64.

[8] G. Caccianiga, J. Nubert, M. Hutter, and K. J. Kuchenbecker, "Dense 3d reconstruction through lidar: A comparative study on ex-vivo porcine tissue," *arXiv preprint arXiv:2401.10709*, 2024.

[9] L. Lipson, Z. Teed, and J. Deng, "RAFT-Stereo: Multilevel recurrent field transforms for stereo matching," in *Proc. International Conference on 3D Vision (3DV)*, 2021.