EasyChair Preprint

№ 8113

Dense 3D Reconstruction Through Lidar: a New Perspective on Computer-Integrated Surgery

Guido Caccianiga and Katherine J. Kuchenbecker

May 29, 2022

Dense 3D Reconstruction Through Lidar: A New Perspective on Computer-Integrated Surgery

Guido Caccianiga and Katherine J. Kuchenbecker

Haptic Intelligence Department, Max Planck Institute for Intelligent Systems
{caccianiga, kjk}@is.mpg.de

INTRODUCTION

Technical innovations in sensing and computation are quickly advancing the field of computer-integrated surgery. From one side, spectral imaging and biophotonics are strongly contributing to intraoperative diagnostics and decision making. Simultaneously, learning-based algorithms are reshaping the concept of assistance and prediction in surgery. In this fast-evolving panorama, we strongly believe there is still a need for *robust geometric reconstruction of the surgical field* whether the goal is traditional surgical assistance or partial or full autonomy.

3D reconstruction in surgery has been investigated almost only in the space of mono and stereoscopic visual imaging because surgeons always view the procedure through a clinical endoscope. Compared to using traditional computer vision, deep learning has made significant progress in creating high-quality 3D reconstructions and dense maps from such data streams, especially for monocular simultaneous localization and mapping (SLAM) [1]. The main limitations are linked to reliability, generalization, and computational cost.

Meanwhile, *lidar* (light detection and ranging) has greatly expanded in use, especially in SLAM for robotics, terrestrial vehicles, and drones. Lidar sensors explicitly measure the depth field rather than inferring it from camera images. The technology is evolving quickly thanks to the upsurge of mixed and augmented reality in consumer mobile devices [2]: high-resolution, short-range, miniaturized lidar sensors are expected soon.

In parallel to these developments, the concept of multiple-viewpoint surgical imaging was proposed in the early 2010's in the context of magnetic actuation and micro-invasive surgery [3]. In routine clinical practice, however, the use of multiple trans-abdominal cannulae still limits the kinematics of the camera and instruments to have a fixed pivot point at the body wall. For this reason, here we propose an approach in which *each surgical cannula can potentially hold a miniature lidar*. We envision that exploring this powerful sensing technology and enabling multi-viewpoint imaging without disrupting the current surgical workflow will yield far more accurate and complete 3D reconstructions of the surgical field, which opens new opportunities for the future of computer-integrated surgery.

MATERIALS AND METHODS

We chose a simulated surgical scene composed of custom silicone organs arranged to reproduce a human abdomen [4] (Fig. 1). A ground-truth (GT) measurement of this scene was meticulously acquired using an Artec Eva HD (0.1 mm accuracy) 3D scanner (Fig. 3d). An Intel L515 high-resolution (1024×768, 30 fps) lidar camera was then attached to each of the four cannulae of an Intuitive Surgical da Vinci Si (Fig. 1). We programmed the L515s' lasers and infrared receivers to produce point clouds at close range (≥5 cm); they interface with our Ubuntu 20.04 workstation through the Realsense SDK and ROS wrappers. A Raspberry Pi 4B externally triggers the four lidars to sequentially capture their point clouds. We record the stereo images of the da Vinci endoscope (1920×1080, 30 fps) using a DeckLink Duo 2 with custom ROS drivers [5]. We performed 3D reconstruction from the endoscope images using the pre-trained HSM network because it performed best on the SERV-CT surgical dataset [6].

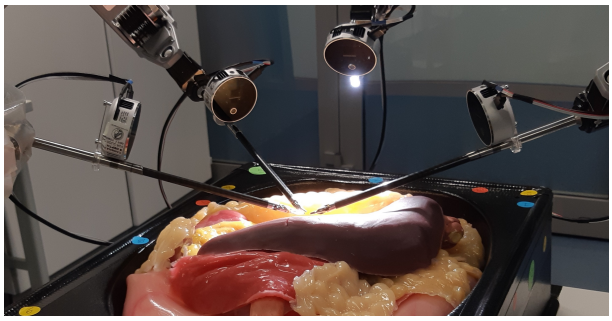

Fig. 1 Our experimental setup featuring a da Vinci Si, four puck-shaped Intel L515 high-resolution lidar cameras attached to the cannulae, and custom organs [4] located within an EASIE-R whole-abdomen surgical simulator.

All five resulting point clouds (endoscope and four lidars) were imported into Artec Studio 16 and independently registered to the GT surface via a robust proprietary alignment procedure based on geometry features. Then the registered point clouds were exported to CloudCompare for computational evaluation. For each cloud, we first calculated a scalar field (*point-to-GT*) representing the signed distance of each recorded point to the GT surface (Fig. 2a-f); non-zero values show sensing errors. We then evaluated how well each sensor covers the surgical scene by sampling the entire GT surface (constant density of one point per mm^2) and calculating a second set of scalar fields (*GT-to-point*) for the signed distance from each GT vertex to the closest point in a surface created from the selected point cloud (Fig. 2g-l); unseen GT points have high distance.

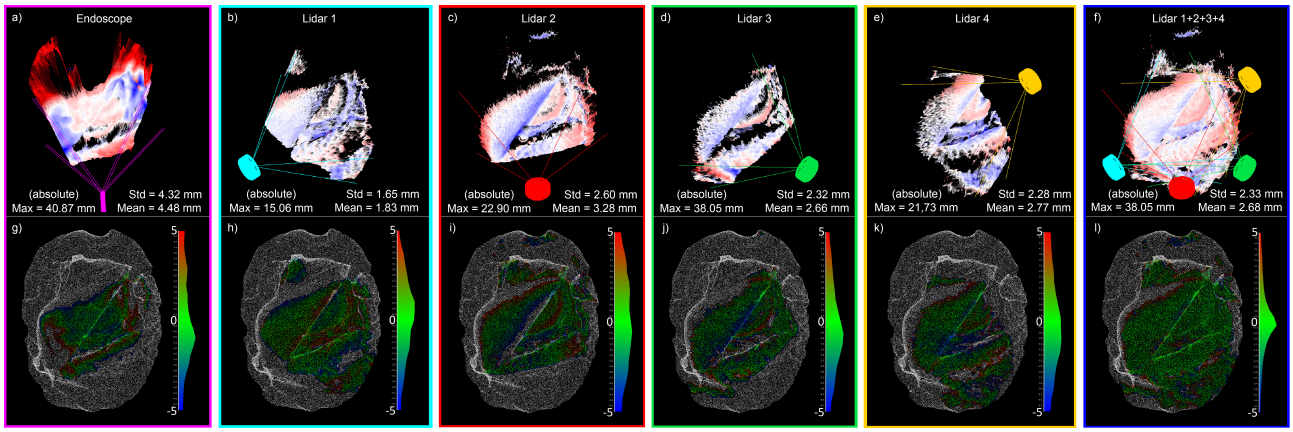

Fig. 2 The point clouds (a–f) from each sensor color-mapped with the *point-to-GT* scalar fields (red: positive error, blue: negative). Statistics are calculated from the absolute scalar values. For each point cloud (a–f), the respective *GT-to-point* scalar field distribution is shown (g–l) in the ±5 mm range. Green areas highlight the portion of the ground truth surface that is reconstructed accurately.

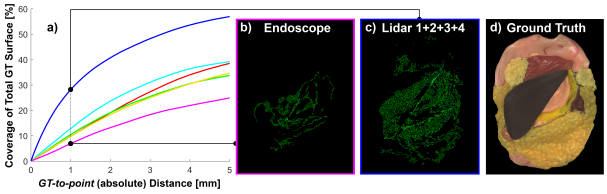

Fig. 3 a) Percentage of total ground truth surface covered by each sensor for an increasing *GT-to-point* error range. b, c) areas covered by the endoscope (magenta) and the four lidars combined (blue) for the ±1 mm error range. d) Ground truth.

RESULTS

The *point-to-GT* scalar fields depicted in Fig. 2(a-e) show the overall accuracy of each camera with respect to the GT. Maximum, mean, and standard deviation of the distances for each point cloud are reported. Figure 2(f) shows the result of overlapping the four lidar point clouds to characterize the total potential coverage of a multi-viewpoint approach. To better address scene coverage while also considering point-placement accuracy, the *GT-to-point* scalar fields are plotted in the ±5 mm range and shown in Fig. 2(g-l) with their error distributions.

Figure 3 shows the percentage of the total surface of the GT seen by each sensor for an allowed error range between 0 mm and ±5 mm. Each lidar accurately sees about 50% more of the GT surface than the endoscope. The inset images show, in comparison, the coverage of the endoscope and the four combined lidars at a conservative error tolerance of ±1 mm. For each GT vertex covered by multiple lidars, we chose the point with minimum error to provide an upper bound on performance that could be expected from this method.

DISCUSSION

We proposed a novel approach to dense 3D reconstruction using lidars attached to the surgical cannulae. Direct comparison between lidar from different viewpoints and a state-of-the-art 3D reconstruction method on stereoendoscope images showed that lidar-generated point clouds achieve better accuracy and scene coverage. Furthermore, lidars have the advantage of directly producing measured point clouds, removing the computational effort and delay of estimating depth from the pixel disparity. This experiment especially hints at the potential of lidar imaging when deployed in a multiple-viewpoint approach.

Aligning point clouds to the GT a posteriori allowed quantitative analysis, but continuous cloud-to-cloud alignment should be implemented for real-world usage. Another limitation for clinical testing and translation of the proposed setup is the physical size of the lidars we used (36 mm × 30 mm minimum chip size). We envision new surgical procedures where multiple miniaturized sensors could be directly embedded in cannulae, as [7] did with cameras, or, if untethered, grasped and assembled on instruments or magnetically controlled from the outside. Such technologies will enable novel computer-integrated systems that could help surgeons during the critical steps of complex procedures by providing omnidirectional visualization, AR preoperative overlays, safety enhancement, and real-time pseudo-haptic cues.

REFERENCES

[1] D. Recasens, J. Lamarca, J. M. Fácil, J. M. M. Montiel, and J. Civera, "Endo-depth-and-motion: Reconstruction and tracking in endoscopic videos using depth networks and photometric constraints," *IEEE RA-L*, vol. 6, no. 4, pp. 7225–7232, 2021.

[2] M. Vogt, A. Rips, and C. Emmelmann, "Comparison of iPad Pro®'s LiDAR and TrueDepth capabilities with an industrial 3D scanning solution," *Technologies*, vol. 9, 2021.

[3] Y.-H. Su, K. Huang, and B. Hannaford, "Multicamera 3D viewpoint adjustment for robotic surgery via deep reinforcement learning," *JMRR*, vol. 06, p. 2140003, 3 2021.

[4] E. D. Gomez, "The role of haptic tool vibrations in skill acquisition and assessment in minimally invasive robotic surgery," Master's thesis, University of Pennsylvania, 2013.

[5] M.-P. Forte, R. Gourishetti, B. Javot, T. Engler, E. T. Gomez, and K. J. Kuchenbecker, "Design of interactive AR functions for robotic surgery and evaluation in dry-lab lymphadenectomy," *IJMRCAS*, pp. 1–21, Nov. 2021.

[6] 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.

[7] A. Afifi, C. Takada, Y. Yoshimura, and T. Nakaguchi, "Real-time expanded field-of-view for minimally invasive surgery using multi-camera visual simultaneous localization and mapping," *Sensors*, vol. 21, pp. 1–20, 3 2021.