

Comparative Validation of Single-Shot Optical Techniques for Laparoscopic 3-D Surface Reconstruction

L. Maier-Hein*, A. Groch, A. Bartoli, S. Bodenstedt, G. Boissonnat, P.-L. Chang, N. T. Clancy, D. S. Elson, S. Haase, E. Heim, J. Hornegger, P. Jannin, H. Kennigott, T. Kilgus, B. Müller-Stich, D. Oladokun, S. Röhl, T. R. dos Santos, H.-P. Schlemmer, A. Seitel, S. Speidel, M. Wagner, and D. Stoyanov

Abstract—Intra-operative imaging techniques for obtaining the shape and morphology of soft-tissue surfaces *in vivo* are a key enabling technology for advanced surgical systems. Different optical techniques for 3-D surface reconstruction in laparoscopy have been proposed, however, so far no quantitative and comparative validation has been performed. Furthermore, robustness of the methods to clinically important factors like smoke or bleeding has not yet been assessed. To address these issues, we have formed a

joint international initiative with the aim of validating different state-of-the-art passive and active reconstruction methods in a comparative manner. In this comprehensive *in vitro* study, we investigated reconstruction accuracy using different organs with various shape and texture and also tested reconstruction robustness with respect to a number of factors like the pose of the endoscope as well as the amount of blood or smoke present in the scene. The study suggests complementary advantages of the different techniques with respect to accuracy, robustness, point density, hardware complexity and computation time. While reconstruction accuracy under ideal conditions was generally high, robustness is a remaining issue to be addressed. Future work should include sensor fusion and *in vivo* validation studies in a specific clinical context. To trigger further research in surface reconstruction, stereoscopic data of the study will be made publicly available at www.open-CAS.com upon publication of the paper.

Index Terms—Biomedical imaging, endoscopes, optical imaging, stereo image processing, stereo vision, surgery.

Manuscript received February 28, 2014; revised May 07, 2014; accepted May 08, 2014. Date of publication May 23, 2014; date of current version September 29, 2014. This work was supported by the European Research Council (ERC) under Grant 242991, and NIHR i4i Grant I13A110910038, as well as the Royal Academy of Engineering/EPSRC Research Fellowship, Imperial College Junior Research Fellowship, the European Social Fund of the State Baden-Wuerttemberg, and the German Research Foundation (DFG; PD 15577, Research Training Group 1126 and Transregional Collaborative Research Centre 125). L. Maier-Hein and A. Groch contributed equally to this work. *Asterisk indicates corresponding author.*

*L. Maier-Hein is with the Junior Group Computer-Assisted Interventions, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany (e-mail: l.maier-hein@dkfz.de).

A. Groch, E. Heim, T. Kilgus, and A. Seitel are with the Junior Group Computer-Assisted Interventions, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

A. Bartoli is with the Advanced Laparoscopy and Computer Vision Group, Université d'Auvergne and CNRS, 63000 Clermont-Ferrand, France.

S. Bodenstedt, S. Röhl, and S. Speidel are with the Institute for Anthropomatics, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany.

G. Boissonnat is with the Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, Imperial College London, SW7 2AZ London, U.K., and also with the Department of Bioengineering, Imperial College London, SW7 2AZ London, U.K.

P.-L. Chang and D. Stoyanov are with the Centre for Medical Image Computing (CMIC) and Department of Computer Science, University College London, WC1E 6BT London, U.K.

N. T. Clancy and D. S. Elson are with the Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, Imperial College London, SW7 2AZ London, U.K., and also with the Department of Surgery and Cancer, Imperial College London, SW7 2AZ London, U.K.

S. Haase and J. Hornegger are with the Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg, 91058 Erlangen, Germany.

P. Jannin is with INSERM, LTSI U1099, Faculté de Médecine and University of Rennes 1, 35043 Rennes, France.

H. Kennigott, B. Müller-Stich, and M. Wagner are with the Department of General, Visceral and Transplant Surgery, University of Heidelberg, 69120 Heidelberg, Germany.

D. Oladokun is with the Department of General Surgery, Kingston Hospital, KT2 7QB London, U.K.

T. R. dos Santos is with the SENAI Institute for Innovation in Embedded Systems, Florianópolis, SC, 88034-001, Brazil.

H.-P. Schlemmer is with the Department of Radiology, DKFZ, 69120 Heidelberg, Germany.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2014.2325607

I. INTRODUCTION

LAPAROSCOPIC surgery provides numerous advantages compared to open surgery, including reduced surgical trauma, earlier convalescence, better cosmetic results, and shorter hospitalization. However, due to the limited field of view (FoV), the difficult hand-eye coordination as well as the loss of depth perception and tactile feedback, laparoscopic interventions generally require a lot of skill and experience to be performed successfully. In particular, mental fusion of the partially visible anatomy with high-resolution pre-operative tomographic images and/or surgical planning data in the presence of organ motion is extremely challenging. Recent developments in medical imaging, medical image computing and robotics, have opened the way for computer-assisted interventions (CAI) in which computer systems provide highly precise localization information about the patient's anatomy during the procedure [1]. One of the main difficulties in soft-tissue CAI is the accurate, fast and robust acquisition of the 3-D structure of the patient's anatomy during surgery. To allow for augmented reality (AR) visualization of subsurface anatomical detail, the acquired 3-D data is typically registered to 3-D pre-procedural planning images and models [2]–[4]. The modalities used for intra-operative patient localization include ultrasound (US), intra-operative computed tomography (CT), and interventional magnetic resonance imaging (iMRI). However, these are often

not well-suited to real-time image acquisition in a surgical theater with traditional instrumentation, provide poor image quality or are associated with radiation exposure and high costs.

In minimally-invasive surgery, an attractive approach involves 3-D reconstruction of soft-tissue surfaces using the endoscope itself [5]. As the latter serves as a natural interface between the surgeon and the patient, there is no need for additional displays to visualize the computed anatomical information. Optical techniques for 3-D surface reconstruction can be divided into two categories [5]: *passive* methods that only require images, and *active* methods that require controlled light to be projected into the environment. Passive methods include stereoscopy, monocular shape-from-X (SfX), and simultaneous localization and mapping (SLAM) while the most well-known active methods are based on structured light (SL) and time-of-flight (ToF). Both active and passive technologies have found successful applications in a wide spectrum of fields including domestic and industrial robotics, as well as the film and games industries, but have only recently been applied in minimally-invasive surgery. Reconstruction of the patient anatomy, however, poses several specific challenges. First, reconstruction algorithms face a *dynamic* and *deformable* environment. Second, tissue may have *homogeneous texture* making automatic salient feature detection and matching, which is required by most passive techniques, difficult. Finally, reconstruction methods must prove to be robust despite complex illumination featuring *specular highlights*, *smoke*, and *blood* as well as surgical instruments that occlude and manipulate the patient anatomy.

So far, quantitative validation of the different reconstruction methods has typically been performed under (close to) ideal conditions using simulated data [6]–[8] and phantom models with known ground truth geometry and motion characteristics [6], [9]–[11] or with ground truth data obtained by scanning techniques [6], [8]–[13]. Although crucial for clinical application, robustness to factors like smoke or bleeding, has not yet been assessed. Furthermore, the different techniques and methods have generally been validated on different data, making comparison of reconstruction results impossible. In response to this, we have formed a joint international initiative with leading institutions in the field to 1) assess the accuracy and robustness of state-of-the-art surface reconstruction methods in a comparative manner and 2) to provide a publicly available data set for assessing current and future reconstruction algorithms according to a common protocol. Our focus is on *single-shot techniques*, i.e., methods that do not require movement of the endoscope, namely stereoscopy, SL and ToF. These techniques were tested on the same objects under identical conditions with CT data serving as reference. To characterize the strengths and weaknesses of each technique, we investigated reconstruction accuracy in a comprehensive *in vitro* study, involving different organs with various shape and texture and also tested reconstruction robustness with respect to the pose of the endoscope and the amount of blood or smoke present in the scene. To ensure reproducibility of our results, we applied a previously published model for defining and reporting reference-based validation protocols in medical image computing [14].

The remaining part of this paper is structured as follows: Section II reviews all the surface reconstruction methods used in this study. Section III describes the validation methodology in detail. Next, the study results are presented in Section IV, followed by a discussion of our findings and perspectives on future work in Section V. Finally, Section VI provides a conclusion.

II. 3-D RECONSTRUCTION METHODS

This section briefly reviews the basic working principle of the different 3-D reconstruction methods investigated in this study.

A. Stereoscopy

Stereo reconstruction is one of the most established optical techniques for recovering the shape of an object. It can be broken down into the following steps: 1) camera calibration from images, 2) acquisition of two images of the scene, 3) establishing stereo correspondences of points in the images, and 4) structure triangulation using the known geometric properties of the cameras.

Based on a PubMed search and a recent review article on optical surface reconstruction in surgery [5] we decided to include stereo algorithms designed for laparoscopy and published within the last four years.

Stereo-KIT: The first stereo reconstruction technique, developed at Karlsruhe Institute of Technology (KIT), is based on a modified version [8] of the Hybrid Recursive Matching (HRM) algorithm [15]. The method uses spatial and temporal information to recursively generate a dense disparity map with subpixel precision in real-time. As input, the rectified and undistorted images from the calibrated stereo endoscope, the disparity image from the previous time step and the already calculated disparities from the current time step are required. In a two-stage process, the block recursion and the pixel recursion, a new disparity value is calculated for the current pixel by evaluating potential correspondence candidates. While the block recursion ensures a smooth disparity distribution especially in textureless regions, the pixel recursion introduces new values in regions of discontinuity. In a postprocessing step, disparity correction including specular highlights and bilateral smoothing is performed. The disparity image is then used to calculate the 3-D coordinates of the image points with a standard triangulation method. To achieve real-time, the components of the workflow have been ported to the graphics processing unit (GPU). A detailed description of the algorithm can be found in [8].

Stereo-UCL: The second stereo reconstruction technique, developed at University College London (UCL), is based on seed propagation and was presented recently [13], [16]. The method initially recovers a sparse set of matches (*seed points*) across the stereoscopic images using a feature-based technique described in [17]. It applies commonly used image features [18] implemented in the OpenCV library¹ though any detection and matching approach (cf. e.g., [19]) can be used just as effectively. The seed points are then processed to generate a semi-dense disparity map by exploiting the fact that points in the neighborhood of an already assigned pixel should correspond to a pixel close to that point's corresponding pixel. The

¹<http://opencv.org/>.

propagation algorithm works using a priority queue structure to populate the correspondence list on a best-first basis ranked by the zero mean normalized cross correlation (ZNCC) of a match. Uniqueness is enforced by keeping the first valid match to each pixel and not permitting new matches to that pixel in later iterations. Disparity continuity is maintained locally by limiting the propagation window around each match. The details of the algorithm are presented in [13], [16], and the source code is available online.²

Stereo-IC: The third stereo reconstruction technique, developed at Imperial College London (IC), uses a 3-D cost-volume to combine local matching with global smoothness optimization [20]. The 3-D volume consists of n slices that have the same dimensions as the stereoscopic images, where n is the range of searching disparity. A global optimization follows the cost-volume construction for achieving the smoothness assumption of the reconstructed disparity and maintaining the discontinuity at the same time. Since the global optimization happens in continuous space, the reconstructed model can reach sub-pixel accuracy without any postprocessing. Details of the technique can be found in [20].

B. Structured Light

SL techniques aim to recover the 3-D surface information of an object in a similar way to stereoscopy but using an artificial pattern of light. The principle is based on parallax and the use of the geometry of triangles and triangulation. In the case of monocular camera detection, a simple trigonometric relationship can be established between the projection system and a single camera. In this case the line-of-sight rays projected from the camera and the SL source intersect at the tissue surface. Similarly to stereoscopy, SL requires knowledge of the camera projection properties. A detailed review of SL in the context of laparoscopic surgery can be found in [5].

The SL technique implemented in this paper is based on the concept of *active stereo* [9], [21]. The SL pattern consists of a circular randomized multispectral pattern of spots [cf. Fig. 6(a)] generated by a dispersed supercontinuum laser that is coupled into a 1.9-mm-diameter optical fiber bundle with a GRIN lens attachment at the tip. The hardware used to generate this pattern, reported in [22], ensures that each spot has a unique wavelength. The narrow spectral profile of the spots means that their color is independent of the optical properties of the surface they are being projected onto, increasing the robustness of the system. Calibration of the device is based on the pinhole camera model [23] and the use of a planar calibration object printed with a grid pattern of known size.

C. Time-of-Flight

The ToF technique is an active reconstruction method based on measuring the run-time of light emitted from the camera and reflected by the object under observation [24], [25]. This enables the acquisition of corresponding distance (range) and intensity images of the observed scene with video frame rates. Typically, the scene is illuminated with intensity-modulated near infrared light (NIR) emitted from one or more illumination units, and

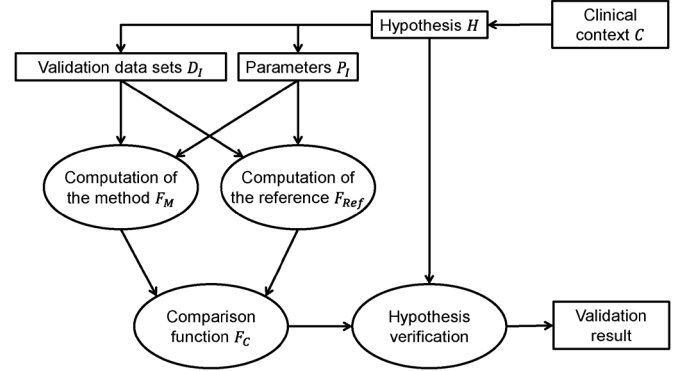


Fig. 1. Main steps of reference-based validation procedures for medical image processing according to Jannin *et al.* [14].

the phase difference ϕ between emitted and reflected light is determined by an on-chip correlation of the reflected signal with a reference signal. Based on the measured phase difference ϕ , the distance in a pixel is then obtained by $d = (c/4\pi f_m)\phi$, where c is the speed of light and f_m is the modulation frequency of the emitted light.

In this study, we apply the first industrial prototype endoscope developed by the company Richard Wolf GmbH (Knittlingen, Germany). It features both a white light source as well as a ToF illumination unit and simultaneously generates range images, corresponding gray-scale amplitude images and standard definition RGB images at a framerate of about 30 frames/s. It operates at a modulation frequency of up to 25 MHz with a lateral resolution of 50×64 .

III. COMPARATIVE VALIDATION STUDY

Although the importance of validation in the context of computer-assisted diagnosis and therapy is well-established [26], standard terminology and methodology that allow for reproducibility and comparability of results have long been lacking. To address this issue, Jannin *et al.* [14] proposed a framework for reporting reference-based validation studies. The idea of reference-based validation is to compare the result of a method with a reference (e.g., *gold standard*) that is assumed to be very close or equal to the ideal expected solution (*ground truth*). The principle of the validation process is illustrated in Fig. 1. It is initiated with the specification of the validation objective, which includes the clinical context (C) in which the method is aimed to be applied, and the specification of a hypothesis (H), relying on expected values required within the considered clinical context. The validation process then proposes an experiment to test the hypothesis. This requires definition of the validation data sets D_I and a set of input parameters P_I , which includes parameters related to the validation data sets and to the method F_M to be validated. The result of F_M on D_I is then compared to the result of a reference method F_{Ref} via a comparison function F_C . Finally, the comparison results are tested against the validation hypothesis in order to provide the validation result.

The aim of our study was a comparative validation of the accuracy and robustness of state-of-the-art surface reconstruction methods in the context of laparoscopic surgery. The difference to a conventional validation study was that we had *mul-*

²<http://www0.cs.ucl.ac.uk/staff/Dan.Stoyanov/software.html>.

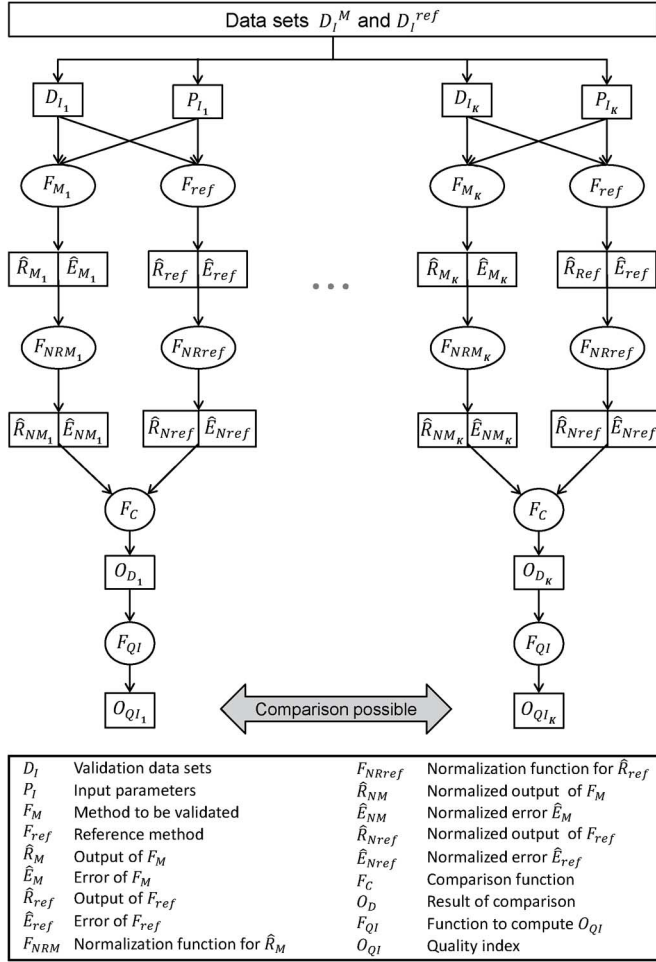


Fig. 2. Model of reference-based validation applied in this work and adapted from Jannin *et al.* [14]. Different state-of-the-art laparoscopic surface reconstruction methods F_{M_i} are applied to a common set of validation objects, from which endoscopic video images D_I^M and CT reference data D_I^{ref} is extracted. Both, the methods to be validated as well as the reference method F_{ref} are applied with a fixed parameter set P_{I_i} and produce an output $\hat{R}_{M_i} / \hat{R}_{ref}$ associated with some error $\hat{E}_{M_i} / \hat{E}_{ref}$. Output is further processed via the normalization functions F_{NRM_i} / F_{NRMref} (e.g., transformed to a common coordinate system), and a comparison function F_C provides a quality measure O_{D_i} for each method F_{M_i} . The latter can be further processed via function F_{QI} to provide a final quality index O_{QI} to be compared across techniques.

multiple methods to be validated. Our focus was the comparability of results of different methods rather than the performance of a specific method in a specific clinical procedure. Because the clinical context is broad, we decided against the definition of a hypothesis and used the protocol of Jannin *et al.* [14] mainly for a comprehensive description of our study that would ensure reproducibility of our work. The detailed validation process is provided in Fig. 2. The following sections provide an instantiation of the various components.

A. Validation Data Sets D_I^M and D_I^{ref}

In a clinical setting, high reconstruction accuracy must be achieved for very different shapes and textures. Furthermore, the methods applied should be robust with respect to parameters, such as the pose of the endoscope as well as the amount of smoke and the amount of blood in the FoV. To take these different aspects into account, a set of various validation objects

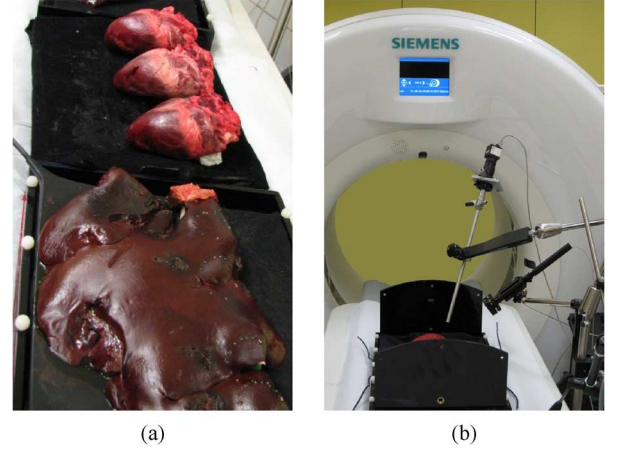


Fig. 3. Experimental setup: Porcine organs in laparoscopic trainer (a) and holding device (b) to ensure approximately identical poses of the different endoscopes relative to the validation objects. (Figure best viewed in color.)

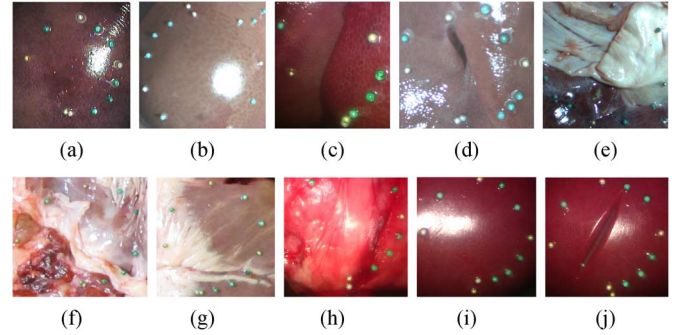


Fig. 4. (a)–(e) Different shapes of the liver: flat (a), convex (b), with discontinuity (c), with hole (d), with fat (e). (f), (j) Different shapes of the heart (f/g), fatty tissue (h), and a kidney without (i) and with (j) cut. Circles represent the region of interest, and the markers were used for registration with CT data. (Figure best viewed in color.)

was prepared, comprising several porcine livers, kidneys and hearts as well as fatty tissue (cf. Fig. 3(a) and Fig. 4) in a commercially available laparoscopic trainer (Pelvi, Karl Storz Endoscopy America, Culver City, CA, USA). The organs were purchased at a butcher's (*Fleischversorgungszentrum Mannheim*, Mannheim, Germany). To generate a variety of different shapes, the tissue was deformed and manipulated with surgical instruments. Different parts of the organs with varying texture were imaged to vary the tissue appearance. Fig. 4 shows the entire set of tissue parts. The validation data sets D_I^M comprised endoscopic images D_I^M and corresponding CT reference data D_I^{ref} of the organ.

Prior to image acquisition a set of Teflon balls (Engineering Laboratories, Inc., Oakland, NJ, USA) with a diameter of 1.6 mm were attached to each organ piece such that they enclosed a circular region on the tissue surface (cf. Fig. 4). Depending on the distance of the object to the camera (4, 5.5, 7 cm) the diameters of this circle were 3, 4.5, and 6 cm to ensure that the spheres were located approximately at the boundaries of the endoscopic images. To allow for an unambiguous registration with CT reference data, the markers were arranged in a unique pattern as shown in Fig. 4. The balls were painted with different colors



Fig. 5. Creation of smoke with a surgical coagulator. (Figure best viewed in color.)

(green and gold) to enable registration even if only a subset of markers could be detected in the video images.

All acquisitions were performed with the theater lights turned off in an interventional radiology suite on the stretcher of a CT scanner (Somatom Definition Flash, Siemens, Erlangen, Germany) as shown in Fig. 3. This minimized the risk of causing changes in the shape of the validation objects during our experiments. To ensure approximately identical poses of all endoscopes relative to the objects under observation, a holding device was affixed to the stretcher [cf. Fig. 3(b)]. To prevent the organs from drying out, they were kept in a box with water and ice as long as possible, and the experiments were performed in an air-conditioned room to keep the temperature below 20 °C. Multiple CT scans of the same organ were acquired to keep the time between endoscopic image acquisition and reference CT acquisition under 30 min.

Using a high definition (HD) stereo endoscope, a standard definition (SD) stereo endoscope, the SL endoscope presented in Section II-B and the ToF endoscope introduced in Section II-C, D_I^M was acquired by engineers accustomed to operation of the respective hardware. The poses of the endoscopes relative to the validation objects were varied to cover a range of clinically realistic settings. For validation of the reconstruction methods, the following data sets were used.

D-Distance: The objects used to investigate reconstruction accuracy as a function of distance are shown in Fig. 4 and comprise five parts of a liver (flat, convex, with discontinuity, with a hole, with fatty tissue), two parts of a kidney (homogeneous, with cut), two parts of a porcine heart (little texture, high texture), and fatty tissue, each acquired with a direct view of the endoscopes on the objects and reconstructed for two different distances: 4 and 7 cm. For each reconstruction method, this yielded a total of $n = 20$ surfaces (10 for each distance).

D-Angle: The validation objects used to investigate reconstruction accuracy as a function of the orientation of the endoscope were the five different liver shapes as described in *D-Distance*, each reconstructed for two different distances (5.5 and 7 cm) and two orientations of the endoscopes: Angles of 0° and 30° measured between the view direction of the endoscope and the approximated mean surface normal. For each reconstruction method, this yielded a total of $n = 20$ surfaces (10 for each angle).

D-Smoke: The validation objects used for investigating the robustness to smoke were again the five shapes of the liver described in *D-Distance* (camera distance: 5.5 cm; angle: 30°). A surgical coagulator (Autocon II 400 electrosurgical unit, Karl Storz GmbH & Co. KG, Tuttlingen, Germany) was used to create smoke (at 40 W) using a part of the liver that was located underneath the camera, as shown in Fig. 5. Note that this procedure only allows an angled pose of the endoscope. For each reconstruction method, this yielded a total of $n = 10$ surfaces (five with and five without smoke).

D-Blood: The validation objects used for investigating the robustness to blood were a liver and a heart, each imaged without and with artificial blood (*Kunstblut*, Metamorph GmbH, Berlin, Germany), as shown by means of example in Fig. 15(a).

To speed up the acquisition process, we always prepared several organs at a time [cf. Fig. 3(a)] and moved the CT stretcher to image the organs sequentially while the hardware remained at a fixed pose. To ensure comparability of results, we recorded the different relative positions of the CT stretcher for subsequent acquisitions with different devices. To ensure that the markers cannot serve as features in stereoscopic reconstruction, we manually detected them in all stereoscopic images. We then generated binary masks representing pixels not to be considered during correspondence search.

B. 3-D Reconstruction Methods F_{M_i}

The specific hardware and settings P_{I_i} used for the different 3-D reconstruction methods F_{M_i} reviewed in Section II are described in the following paragraphs:

Stereoscopy: For stereo image acquisition, a commercial SD stereo endoscope (resolution 720 × 576; baseline: ≈4 mm; optics: 30°) and a prototype HD stereo endoscope (resolution: 1920 × 1080; baseline: ≈3.5 mm; optics: 0°), both from the company Richard Wolf GmbH (Knittlingen, Germany), were used. Images were captured with 25 frames/s utilizing two Blackmagic Intensity Pro cards. Synchronization of the stereo channels was performed via timestamps. Prior to the experiments, the stereo endoscopes were calibrated with a checkerboard-based method [27], as detailed in Röhl *et al.* [8]. Randomized subsets of all recorded checkerboard patterns were generated and a calibration evaluation based on 3-D reconstruction error and rectification error was used to select the calibration with the highest accuracy. Calibrating the HD stereo endoscope yielded a large rectification error, which may partly be due to chromatic aberration, for example. Downsampling the images to 960 × 540 reduced this error significantly. It was therefore decided to only use the downsampled HD images for validation, since the *Stereo-KIT* and the *Stereo-IC* method requires rectified images for correspondence analysis. During validation it was observed that the rectification error, or *y-offset*, varied with time in the case of the HD images, potentially due to *jitter* of the prototype stereoscope. To correct for this variation, an automatic rectification correction using Scale-invariant feature transform (SIFT) features [28] was implemented. SIFT features were detected and matched in each stereo image pair, and the average difference in y-values of the matched points was then used to correct the y-offset of the rectification. The

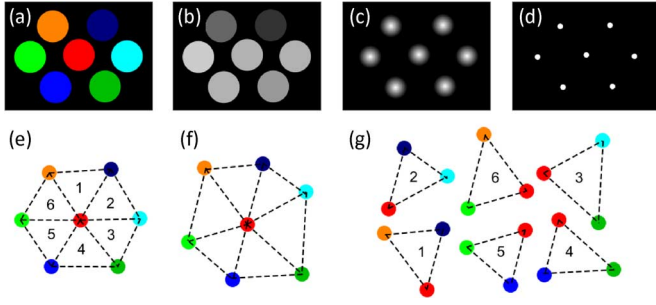


Fig. 6. Spot detection and tracking algorithm for the SL system. (a) RGB image of spot pattern. (b) Total intensity image. (c) Total intensity convolved with 2-D Gaussian kernel. (d) Centroids calculated from local maxima. (e) Reference image showing spot labelled by its RGB values and six surrounding triangles. (f) Possible match for reference spot in subsequent image. (g) Labelling of triangles identified in (f) that match the reference (e). (Figure best viewed in color.)

stereo algorithms were then applied to the masked original and rectified endoscopic images (cf. Section III-A).

The *Stereo-KIT* method was used as described in a previous publication [8], but using only one image frame and without postprocessing of the reconstructed surface. For the *Stereo-UCL* method, the parameters for feature detection in the OpenCV `cvGoodFeaturesToTrack` function with `qualityLevel = 1e - 10` and `minDistance = 30` pixels and all other parameters left to their default values. Initial matching was performed with Lucas–Kanade (LK) [29] iterations in a pyramidal implementation using the OpenCV `calcOpticalFlowPyrLK` function with parameters `winSize = Size(11,11)` and `maxLevel = 5` and all other parameters left to their default values. Parameters for the propagation used a growing neighborhood of 1 pixel around seed matches, correlation minima of 0.5 for seed matches and a windows size of 11. For the *Stereo-IC* method, the ZNCC patch was set to 11×11 , and the global optimization was run for 1000 iterations. The setting for the parameters of the primal-dual algorithm was the same as in Chang *et al.* [20]. The disparity range was from 10 to 120.

SL: For this study, laser spot detection was performed as follows (cf. Fig. 6): An integrated intensity image (b) was formed from the red, green and blue color planes of the RGB image (a), and background noise was eliminated using a low threshold mask before convolving with a 2-D Gaussian kernel of 20 pixel diameter (c), which returns high values close to each spot's brightest part and allows the centroid of each spot to be found (d). The normalized RGB values for each spot in a 7 pixel radius region-of-interest around the centroids were calculated. To define a unique label for each spot, a Delaunay triangulation was performed and each spot was identified by its surrounding triangles and the colors of their vertices (e). Triangles in the test image were matched to corresponding locations in a reference image based on a comparison of their internal angles and the RGB values of the nodes. The closest matching reference triangles to those surrounding a particular test spot were then examined (f). The reference spot that appeared (or was “counted”) most often in the matched triangles was selected as having the highest probability of being the correct one, with a count value

of 4 defined as being a reliable match (g). However, some spots in the matched reference triangles could receive an equally high count value or a value less than 4. To decide on such ambiguous spots, a labelling enhancement step was preformed. For a particular matched reference triangle, following the steps above, the sum of the count values on the three nodes (“group count”) was calculated. If this exceeded a threshold value ($\text{group count} = 9$) then the value of each node was increased by one. This procedure was then repeated for neighboring triangles. In this way, areas of the projected pattern containing reliably matched spots could be used to improve matching of spots with lower count values and eliminate weak correspondences.

In our experiments, the prototype SL projector probe was secured and mounted in a metal tube so that it could be positioned separately to the laparoscope using a clamp. For each set of tissue samples, the laparoscope was positioned first according to the angles and working distances specified in Section III-A. The projector was then clamped beside the laparoscope at an angle of approximately 30° and its working distance adjusted so that the spot pattern occupied an area that would cover as many of the CT markers as possible. This position was then locked and the system was calibrated as described in Section II-B using 12 different poses. Both white light and SL images were collected at each pose to allow calibration of the camera intrinsic parameters and ray propagation directions.

ToF: As mentioned in Section II, the first prototype endoscope developed by the company Richard Wolf GmbH (Knittlingen, Germany) was used in this study. The endoscope was initially calibrated with a checkerboard-based method similar to that proposed by Zhang [27]. As the ToF amplitude images have a very low resolution and limited quality, checkerboard corners can be challenging to detect, especially when the whole checkerboard needs to be in the FoV of the endoscope. In an adapted approach proposed previously [30], checkerboard fields are filled with unique patterns, such that an arbitrary number of fields may be imaged by the endoscope. The modulation frequency and the integration time were set to maximum values of 25 MHz and 945 ms to obtain the best possible SNR. Note that this was sensible because the scene was static and no over-exposed pixels occurred with these settings. As the hardware was still premature and the images feature a low SNR, temporal averaging of the pixel values was performed on a sequence of $n = 1000$ consecutive distance images. Furthermore, a bilateral filter [31] was applied as edge-preserving smoothing filter ($\sigma_{\text{domain}} = 2$, $\sigma_{\text{range}} = 20$). In the current prototype, the distance errors related to various factors such as temperature, intensity, integration time are still severe. As they even change over time (e.g., due to the temperature-related error) a distance calibration prior to every single image acquisition would be required, which, however, was impossible to conduct due to the time restrictions for this study. To obtain meaningful results despite these issues, an individual offset value, to be subtracted from all pixels in the range image, was determined for each acquired image. For this purpose, the markers were detected in the amplitude image and their 3-D position was calculated with the corresponding distance values based on the camera intrinsics. We then computed the offset that—when added to all distance

values—minimized the 3-D fiducial registration error (FRE) between reconstructed fiducials and the fiducials extracted from corresponding CT images.

C. Error \hat{E}_{M_i} of 3-D Reconstruction Methods

As shown in Fig. 2, the output of the reconstruction method F_{M_i} , when applied to D_I with parameters P_{I_i} is referred to as \hat{R}_{M_i} . The errors E_{M_i} related to the different reconstruction methods F_{M_i} are complex, and a thorough description is beyond the scope of this paper. However, the most important sources of error are provided in the following.

Stereoscopy: Errors in reconstruction from stereo images result mainly from mismatches in the established correspondences and inaccurate calibration. Correspondence search in endoscopic images is extremely challenging due to a number of different reasons, as discussed at length in a previous paper [5]. Among others, homogeneous or periodic textures, poor illumination and varying light conditions make salient feature detection and matching difficult. Specular reflections are view dependent and hence are not reliable image details for matching. Occlusions due to the surgical instruments, as well as smoke and areas covered with blood impeded reconstruction performance.

Even if corresponding pixels are found correctly, 3-D reconstruction may be prone to error. As accuracy decreases with a decreasing baseline length, it has been established that subpixel accuracy in correspondence establishment is required for acceptable results [32]. Without subpixel accuracy the so-called staircase effect occurs, meaning that the reconstructed 3-D points are located in discrete layers.

Another source of error is the calibration of the intrinsic and extrinsic camera parameters. Wide angle lenses of stereo endoscopes mean that barrel lens distortion is often present, especially at the image edges. Jitter in the coupling between the scope and the camera heads may occur, rendering a previously performed calibration less accurate. Jitter is especially crucial if a stereo algorithm requires rectified images (here: *Stereo-KIT* and *Stereo-IC*).

A direct estimation \hat{E}_{M_i} of the 3-D reconstruction error was not available, but the calibration error could be approximated by determining the distance between the estimated checkerboard corner positions, obtained by reprojection, with the computed feature positions.

SL: The main sources of error associated with the active stereo implementation of the SL technique can be broadly divided into those related to detection of the spot pattern and those related to the pinhole camera model calibration. For spots incident on a surface at a high angle, their shape may be distorted and stretched beyond this average value resulting in errors in centroid position. Similarly, errors in centroid calculation may also arise when the reflectance properties of the tissue within the spot are heterogeneous or when the spot is partially occluded. High absorption of blue and green light by hemoglobin decreases the visibility of these colors in blood-rich tissue. When there is sufficient blood in the FoV some of these spots will no longer be detectable. As well as being regularly reflected from the tissue surface, light will also penetrate and be scattered multiple times due to the turbid nature of biological

tissue. This has the effect of reducing the contrast between the projected spots and the background tissue surface between them. If the amount of diffuse light between two adjacent spots is sufficient it can result in the detection of a false centroid at the midpoint between the real ones. Further sources of error are found in the estimation of the camera's intrinsic parameters arising from detection of the corner features in the calibration object. This also leads to uncertainties in the points which are used to calculate the best-fit rays. Following the calibration routine there is a final error associated with reprojection of the detected spot onto the unknown tissue surface. Due to low density of the reconstruction possible with the current system further errors arise from the interpolation of the tissue surface between projected spots.

ToF: ToF cameras in general, and the prototype ToF endoscope in particular, are subject to numerous systematic errors and noise. The systematic error in a ToF pixel comprises a pixel-specific offset as well as a scene-dependent error that can be modelled as a function of the intensity measured, the integration time used, the temperature and run-time of the camera, and the measured distance itself. Such systematic errors can theoretically be compensated for by a depth calibration [33], [34], but a practical approach for endoscopic data has not yet been proposed. Due to the challenge in transmitting enough light through the endoscope channel, noise is another crucial issue in the context of ToF endoscopy [5]. In fact, the inhomogeneous illumination with a radial attenuation of the light leads to a particularly bad SNR at the boundaries of the images. Further important errors in the context of laparoscopic surgery arise from depth discontinuities, scattered light, multi-path reflections and tissue penetration. Depth discontinuities result in so-called *flying pixels*, which can be regarded as outliers because they typically incorporate range data from more than one object. The light scattering effect results from multiple light reflections between the camera lens and its sensor and causes a depth underestimation over the affected pixels [33]. The amount of interference increases with a decreasing distance to the objects under observation, which makes this error extremely relevant in the context of laparoscopic surgery. Multi-path errors are caused by interference of multiple light reflections captured at each sensor's pixel [33]. As these errors occur mainly with concave objects, such as surfaces from abdominal organs, they are also highly relevant in the context of endoscopic applications. Finally, NIR light may be absorbed by or penetrate into the tissue, which may lead to an overestimation of depth.

D. Reference Method F_{Ref} and Associated Error \hat{E}_{Ref}

The true organ surface R is approximated by a reference method F_{Ref} that can be expected to be more accurate than F_M and yields an estimate \hat{R}_{Ref} of the ideal surface. In this study, surfaces extracted from high-resolution CT data (0.6 mm slices, 0.2 mm overlap; Somatom Definition Flash) served as reference. The reference data set corresponding to a particular endoscopic data set was defined to be the one with the smallest difference in time between CT acquisition and endoscopic data acquisition. As the organs used were surrounded by air, a threshold filter (threshold = 800 HU) was sufficient for segmenting the organs. The corresponding surface was obtained by applying the

marching cubes algorithm to the segmented shapes and decimating the resulting surface to obtain a number of vertices manageable by our software.

Although the reference method has been chosen to be as accurate as possible it does yield errors, represented by the symbol E_{Ref} . In this particular case, these errors included errors related to segmentation (partial volume effects, slice thickness, user interaction), surface generation and organ deformation, resulting from the fact that the endoscopic data and the CT data were not acquired simultaneously. To obtain a rough estimate of the error \hat{E}_{Ref} of the reference for a particular tissue part, we computed a point-based registration of the first and last CT that was acquired from the corresponding organ based on the markers enclosing the region-of-interest (cf. Section III-A), and determined the corresponding FRE.

E. Normalization Function for the Reference F_{NRef} and Normalized Reference \hat{R}_{NRef}

To allow for a meaningful comparison between the reference surfaces and the reconstructed surfaces, the former ones were transferred into the camera coordinate systems. To compute the normalized reference \hat{R}_{NRef} for a given endoscopic data set, the following information I was extracted.

- Reference markers: The markers were manually located in the CT image by setting a point onto that part of the sphere that could be expected to correspond to the center of the marker in the endoscopic video data.
- Reconstructed markers: Initially, the markers were detected in the 2-D endoscopic video images using either an automatic blob/circle detector, (OpenCV `MinEnclosingCircle`) (when image quality allowed) or a manual procedure. In the case of the stereo algorithms, these manual reference correspondences were refined by two methods: An LK search with radius 11 (SD) and 31 (HD) pixels and a search along the epipolar line using the *Stereo-KIT* method. Next, the reconstruction algorithm was applied to compute the 3-D marker positions relative to the camera coordinate system for all methods. In the case of the stereo algorithms, this was done for all variants (purely manual, manual refined by LK search and manual refined by search along epipolar line), and the method that provided the best FRE was chosen to compute the reference registration (separately for each image). Outliers resulting from a suboptimal reconstruction were then manually removed from the set, and corresponding markers were deleted from the reference marker set. In this step, the SL method required special treatment. Since the density of the spot pattern was not high enough to ensure that each CT marker's 3-D position could be calculated directly an alternative approach had to be employed. Firstly, the centroids of each marker were found using white light images of the same FoV. These coordinates were converted to metric space from pixel space using an empirical relation derived from the white light view and the x-y plane of the reconstructed point cloud (since the camera axis was aligned with the z-axis). The z-coordinate of the markers was then calculated as the point of intersection between the interpolated surface and a line

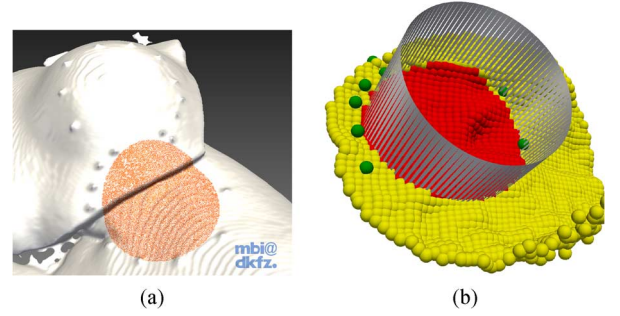


Fig. 7. Extraction of a sphere-shaped region (red) from the reference computed CT surface (a) and the reconstructed point cloud (yellow) (b) using a cylinder as bounding object. Reconstructed markers are shown in green. (Figure best viewed in color.)

parallel to the z-axis that passes through the marker's 2-D centroid. Furthermore, for some of the tissue images the spot tracking algorithm failed due to occlusion of several spots and had to be partially manually supervised. This was particularly evident in the images where blood and smoke occupied the FoV.

- Reference surface data: The reference method was applied to extract the reference surface that the reconstructed surface was to be validated against. Next, the method by Horn [35] was applied to the reference markers and the reconstructed markers to determine the rigid transformation that minimized the FRE between corresponding fiducials. The transformation was then used to register the reference surface to the camera coordinate system. To be able to validate the surface quality in the interior of the sphere defined by the markers, a cylinder-shaped bounding object was defined by manually fitting a circle to the set of reference markers extracted from the CT images. The intersection of the cylinder with the reconstructed surface was then manually processed in order to keep only the outer parts of the organ that were facing the endoscopic camera (cf. Fig. 7).

Additional errors related to the reference that were not included in \hat{E}_{Ref} are the marker localization errors. As the camera pose relative to the CT image is only known approximately, the point on the sphere surface that corresponds to the center of the marker visible in the endoscopic image is only known approximately. Furthermore, errors in the 2-D marker localization as well as errors related to the reconstruction method lead to erroneous marker positions in the camera coordinate system. As a consequence, the registration from the CT coordinate system to the camera coordinate system(s) is erroneous.

F. Normalization Function for the Reconstruction Methods F_{NRM_i} and Normalized Output \hat{R}_{NRM_i}

The point clouds resulting from intersection of the cylinder-shaped bounding objects described in Section III-E with the reconstructed point clouds served as reconstructions to be validated (cf. Fig. 7).

While this approach cannot introduce additional errors to the reconstructed surface, it may—in theory—lead to cutting out a wrong piece of the surface (i.e., not the interior of the marker

circle). Due to visual inspection of *all* reconstructed surfaces, this issue is not of practical interest.

In the case of the stereo algorithms, wrong correspondences may lead to severe outliers, as shown by means of example in Fig. 14. If not located in the interior of the cylinder, such points would not be considered in our validation. To address this issue, we applied an alternative approach to determine those reconstructed points corresponding to the reference surface: We projected the intersection of the reference surface with the cylinder to the endoscopic images to generate masks that determined which pixels to use for validation. In some cases, manual correction was necessary to ensure that the masks were fully contained in the sphere defined by the fiducials.

G. Validation Criterion V_C and Comparison Function F_C

The information I , extracted from \hat{R}_{NM_i} and \hat{R}_{NRef} was processed to determine reconstruction accuracy, point density, surface coverage and robustness as follows:

Accuracy: For each reconstructed surface point on all reconstructed (transformed) surfaces, the closest point (*corresponding point*) on the (transformed) reference surface was determined. Reconstruction accuracy, which is the result of the comparison function $O_D = F_C(\hat{R}_{NM_i}, \hat{R}_{NRef})$, was defined as the root mean square (rms) distance between corresponding points.

Point Density: The total number of reconstructed points within the circle of interest was determined. This is an important criterion besides reconstruction accuracy because it determines the resolution of the 3-D surface reconstructed.

Surface Coverage: Even when the number of reconstructed points is high, only parts of the object under observation may be covered, as shown by means of example in Fig. 12 (liver). To address this, we generated a subset $V_{CT(\delta)}$ of the set of all CT vertices for each reference surface, such that the inter-vertex distance was above a certain threshold $\delta = 3$ mm. CT vertices located closer than 5 mm to the cylinder were also removed to compensate for the fact that the markers were masked and hence, a certain region around each marker could not be reconstructed by the stereo methods. For each of the reconstructed points $v \in V_E$ for a given organ part and reconstruction method, we determined the closest point in $V_{CT(\delta)}$

$$\hat{v} = C(v, V_{CT(\delta)}) \quad (1)$$

$$= \operatorname{argmin}_{\tilde{v} \in V_{CT(\delta)}} \|v - \tilde{v}\|. \quad (2)$$

The percentage of CT vertices, which served as a nearest neighbor for any of the reconstructed points provides a measure for the surface coverage

$$Cvg = \frac{|\hat{V}_{CT(\delta)}|}{|V_{CT}|} \quad (3)$$

$$\hat{V}_{CT(\delta)} = \{\hat{v} \in V_{CT(\delta)} | f(\hat{v}) = true\} \quad (4)$$

$$f(\hat{v}) = \exists v \in V_E : \quad (5)$$

$$(\hat{v} = C(v, V_{CT(\delta)}) \wedge \|\hat{v} - v\| < 1 \text{ cm}). \quad (6)$$

Robustness: To investigate the robustness of the surface reconstruction methods to clinically relevant changes in the setting, we determined accuracy and point density as a function of

the distance to the object of interest, as well as the angle between the view direction of the camera and the surface normal of the object of interest. We further compared reconstruction accuracy for images with and without smoke or blood in the scene.

H. Quality Index O_{QI}

The validation data sets described in Section III-A were evaluated separately for the different reconstruction techniques as follows: Descriptive statistics regarding the rms surface reconstruction error (as defined via the comparison function V_C) as well as the point density were determined for all reconstructions. . .

D-Distance: . . . corresponding to the same distance

D-Angle: . . . corresponding to the same angle of the endoscope relative to the object surface

D-Smoke: . . . with and without smoke

As we did not have a sufficient amount of reconstructions with blood to allow for a meaningful quantitative comparison, this data set was intended to be inspected visually.

In addition, the overall performance of each technique was determined by computing descriptive statistics for the rms using all reconstructions without blood and smoke ($n = 35$). The same data sets were used to check for statistically significant performance differences of the different techniques.

Finally, as we did not have a sufficient amount of reconstructions for each shape/texture type to allow for a meaningful quantitative comparison, we performed a visual inspection of the reconstructed surfaces, color-coded with the distance to the nearest point on the reference surface.

IV. RESULTS

The mean FRE for the point-based registration of the first and last CT that was acquired for a certain tissue part based on the markers enclosing the region-of-interest (cf. Section III-A) was 0.6 ± 0.4 mm (max: 1.2 mm) ($n = 25$).

The mean reprojection error was 0.3 ± 1.4 (max: 45.3) pixels for the stereo SD images and 1.4 ± 0.4 (max: 7.2) pixels for the (half-) HD images. For the SL system, the mean rms ray fit error was 0.4 ± 0.1 (max: 0.6) mm. For the ToF endoscope, the reprojection error was 0.3 ± 0.2 (max: 1.1) pixels.

The median reconstruction error, measured as the rms distance to CT reference data (cf. Section III-H) ranged from 1.2 mm (*Stereo-KIT*) to 4.0 mm (ToF) when considering all shapes and textures without smoke and blood ($n = 35$). A box plots for the different methods is shown in Fig. 9. When neglecting reconstruction outliers by considering the median distance to the reference surface (as opposed to the rms), accuracy dropped below 1 mm for all stereo methods and to 1.2 mm (SL) and 2.8 mm (ToF) for the active methods. The decrease in error was particularly high for the passive methods (Stereo SD: 45%; Stereo HD: 51%, SL: 31%; ToF: 30%). The corresponding FREs for the marker-based registration of the reconstructed surface with the reference surface were 1.0 ± 0.5 mm (Stereo SD), 1.5 ± 0.9 mm (Stereo HD), 2.0 ± 1.3 mm (SL), and 2.9 ± 1.9 mm (ToF). Fig. 12 and Fig. 13 visualize the reconstruction error on organ parts with different shape and texture. The number of data sets excluded from evaluation due to poor marker detection was 0 (Stereo SD), 1 (Stereo HD), 5 (SL), and 2 (ToF).

TABLE I

NUMBER OF RECONSTRUCTED POINTS DETERMINED FOR THE ENTIRE SET OF EXPERIMENTS USED IN THIS STUDY. THE 3-D RECONSTRUCTION METHODS ARE DESCRIBED IN DETAIL IN SECTION II (SD: STANDARD DEFINITION; HD: HIGH DEFINITION)

Method	Mean \pm SD
Stereo-KIT(SD)	$(3.1 \pm 1.0) \cdot 10^4$
Stereo-KIT(HD)	$(1.2 \pm 0.4) \cdot 10^5$
Stereo-UCL(SD)	$(3.0 \pm 1.1) \cdot 10^4$
Stereo-UCL(HD)	$(9.4 \pm 5.3) \cdot 10^4$
Stereo-IC(SD)	$(3.2 \pm 1.0) \cdot 10^4$
Stereo-IC(HD)	$(1.3 \pm 0.4) \cdot 10^5$
SL	$(4.5 \pm 1.2) \cdot 10^1$
ToF	$(4.9 \pm 2.3) \cdot 10^2$

For all methods, reconstruction accuracy decreased with an increasing distance of the camera with increases in median rms error up to 179% [Stereo-UCL(HD)] [(Stereo(SD): $64 \pm 47\%$; Stereo(HD): $89 \pm 64\%$; SL: 37% ; ToF: 11%]. Most methods performed better when the endoscopic camera directly faced the tissue with increases in error of up to 84% (*Stereo-UCL(HD)*) (Stereo(SD): $22 \pm 31\%$; Stereo(HD): $23 \pm 44\%$; SL: 23% ; ToF: -27%). While the stereo approaches performed poorly in the presence of smoke, accuracy was not affected in the cases of the active methods (cf. Fig. 10). Figs. 14 and 15 illustrate the decrease in performance when the tissue was covered by smoke or blood.

Point density ranged from 45 ± 12 (SL) to $1.3 \cdot 10^5 \pm 0.4 \cdot 10^5$ [Stereo-IC(HD)], as shown in Table I. In general, the number of points decreased in the presence of smoke or blood, as shown in example in Figs. 14 and 15. Fig. 11 shows the surface coverage, as defined in Section III-G, for a radius of 3 mm.

In addition to these general observations, the following method-specific conclusions can be drawn from our study.

Stereoscopy-based methods performed extremely well when recovering both global and local shape, even when reconstructing homogeneous tissue. For all stereo methods, point density was extremely high compared to the active methods (cf. Table I), although in some cases, only very few points were reconstructed in the case of the semi-dense method *Stereo-UCL*. Moreover, the distribution of points was not always optimal, meaning that parts of the tissue were not reconstructed at all. This holds again true especially for *Stereo-UCL*, as shown in Fig. 12. While point density was generally higher for the HD images, surface coverage was better for the SD images. With respect to this measure, Stereo-KIT performed best with a median coverage of 100% (SD/HD). Although the median reconstruction error of all methods was lower on the SD images, the difference did not reach statistical significance according to a Wilcoxon rank sum test. The p-values for *Stereo-KIT* [Median: 1.2 mm (SD), 1.5 mm (HD)], *Stereo-UCL* [Median: 1.7 mm (SD), 2.1 mm (HD)], and *Stereo-IC* [Median: 1.3 mm (SD/HD)] were 0.07, 0.07, and 1.0, respectively. The *Stereo-IC* method yielded significantly better accuracy (p-value < 0.05) than both other stereo methods on the HD images and also better accuracy than *Stereo-UCL* on the SD images while the difference in error was not statistically different when comparing *Stereo-UCL* and *Stereo-KIT*.

The SL technique yielded very good results in terms of both, accuracy under ideal conditions and robustness. The coverage

of the surface was generally good because of the regular distribution of projected points (cf. Fig. 13). However, the current sparsity of points resulted in a rough reconstruction of the surface, and fine structures could not be recovered. This was apparent in the calculation of CT marker coordinates, which had to be done by interpolation, as explained above.

The variance in reconstruction accuracy was extremely high for the ToF endoscope. Fig. 16 shows two qualitatively very different reconstruction results for similar shapes. As an intrinsic characteristic of the technique, point density and framerate are invariant to the scene.

V. DISCUSSION

Optical imaging methods are an appealing modality for recovering 3-D geometry from the surgical site without exposure to harmful radiation. To date, however, neither passive nor active reconstruction methods have found widespread use in clinical practice. Based on a preliminary feasibility study [36], this work presented a comprehensive evaluation of the major state-of-the-art surface reconstruction techniques *in vitro* with a focus on 1) comparability of techniques and 2) robustness with respect to a variety of clinically relevant parameters, such as organ shape and texture, endoscope pose, as well as smoke and blood in the FoV of the endoscope. Previous works primarily validated individual techniques in close to ideal settings, and the few comparative validation studies were published in nonmedical contexts [37] or in the context of open surgery [37]–[39]. The following sections discuss the performance of the individual reconstruction methods (Section V-A) as well as the validation method applied (Section V-B).

A. Reconstruction Methods

With the exception of ToF, all reconstruction methods investigated yielded a median reconstruction error of 1–2 mm in the absence of blood or smoke. Considering that the accuracy of the reference (FRE typically 1–2 mm) cannot be expected to be in the submillimeter range, these are very promising results. The stereo methods outperformed the remaining techniques with respect to point density and reconstruction of fine structures, while SL yielded the best performance in the presence of smoke. The quality of the reconstructed surfaces generally decreased with an increasing distance of the endoscope to the object under observation and with an increasing angle of the endoscope's viewing direction to the surface normal.

In the following paragraphs, the performance of the individual techniques is discussed in detail.

1) *Stereoscopy*: Although point density was higher for the HD images, reconstruction accuracy was better on the SD images for all stereo methods. The authors attribute this to the calibration of the prototype HD stereo endoscope: Artefacts due to chromatic aberration, for example, impeded the procedure. This was especially problematic for finding the image rectification parameters. As the correspondence analysis is sensitive to large rectification errors and the rectification error is smaller on lower image resolution in general, lower resolutions yield more robust results. Furthermore, we observed a change in the rectification parameters over time, which can be attributed to jitter (cf. Section III-C) of the cameras. Although we were able

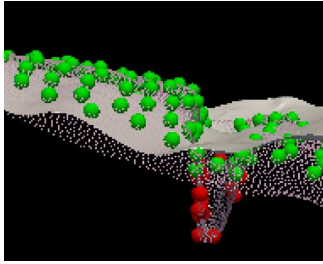


Fig. 8. Illustration of the validation criterion *surface coverage* as defined in Section III-G. Reference mesh (small points) is divided into regions represented by the green and red large points. Red points correspond to regions that have not been reconstructed by the method. In the provided example, the *Stereo-IC* method failed to fully reconstruct the cut in a kidney (gray surface) despite a large number of points. (Figure best viewed in color.)

to automatically compensate for this effect to some extent, we think that it could be a relevant issue in practice. We also investigated manual tuning of the parameters for individual data sets and were able to obtain better reconstruction results.

The better performance of the stereo algorithms for smaller distances and angles can be explained by negative correlation of disparity and depth as well as the reduced saliency of features with increasing angle and distance. It was not surprising that performance generally decreased in the presence of smoke (\rightarrow noise) and blood (\rightarrow homogeneous texture, specular highlights) because of the challenging correspondence analysis in this case. It is worth mentioning in this context that several outliers of the *Stereo-KIT* method occurred in the images without smoke within the *D-Smoke* data set, which explains the decreasing error in the presence of smoke in this case (cf. Fig. 10).

Considering point density, surface coverage and accuracy, the dense reconstruction algorithms (*Stereo-KIT* and *Stereo-IC*) yielded better performance in our study compared to the semi-dense approach (*Stereo-UCL*). However, the median surface coverage was surprisingly good for this method. One major advantage of the sparse method is that it does not require rectification of the stereo images. We attribute the comparably low accuracy of the *Stereo-KIT* method on some of the high resolution images to this effect. Manual observation of the results showed that high errors occurred primarily at image boundaries.

When comparing the two dense stereo methods, we observed that the *Stereo-IC* algorithm showed higher reconstruction accuracy and robustness while the surface coverage was higher for *Stereo-KIT*. In this context, it should be kept in mind that we ran all stereo algorithms on single image pairs with a fixed parameter set without postprocessing the resulting meshes. As we determined accuracy as the rms surface reconstruction error, we included outliers that are usually removed *after triangulation* in the case of *Stereo-KIT*. The fact that the coverage of *Stereo-IC* was lower can partially be attributed to the smoothness constraint: In the case of cuts, for example, the regions of the deeper parts within the cuts had no corresponding point for this method, as shown in Fig. 8.

The run-time of the stereo algorithms has been assessed in separate studies. For the *Stereo-KIT* approach, the computed framerates were 14 fps and 10 fps for standard resolution (640

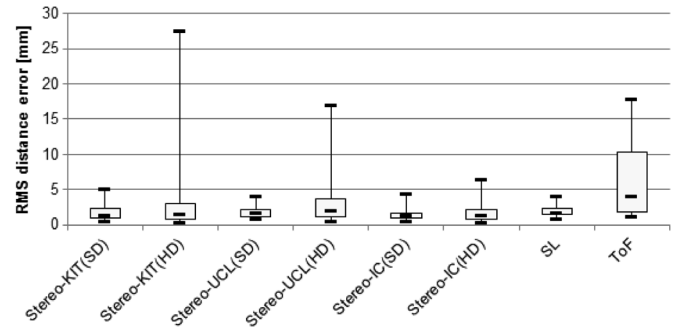


Fig. 9. Box plot (median, first and third quartiles, minimum and maximum) of the rms surface reconstruction error according to Section III-H, determined for all shapes without blood and smoke ($n = 35$) for all reconstruction techniques investigated in this study (cf. Section II).

$\times 480$) and half-HD resolution (960×540), respectively (Machine: i7 930, 2.8 GHz, 24 GB RAM Tesla C2050 graphics card) [8]. The run-times for the *Stereo-UCL* algorithm were 1 fps (SD; 720×576) and 0.1 fps (HD; 1920×1080) (Machine: i7-620 M processor with 2 cores at 2.66 GHz, 4 GB RAM) [13], [16]. According to a recent study [20], the run-time for the *Stereo-IC* algorithm was 20 fps with a cost-volume size $360 \times 288 \times 32$ (Machine: i7-3930 K, 3.2 GHz, 16 GB RAM, and GeForce GTX 670 2 GB graphics card). It also shows linear run-times with respect to the image resolution, e.g., 5 fps for 720×576 and 4 fps for 960×540 with the depth range 32.

Although all stereo reconstruction methods proposed can be implemented in real-time on the GPU, performance remains an important issue in practice. This is particularly essential when considering a clinical application that requires 3-D information in order to guide the surgeon. Video augmentation should be sufficiently fast to maintain video frame rates without lag or impedance to the normal visualization of the surgical site. Also, computational tasks such as registration or biomechanical modelling potentially need further computational resources. The hardware loop to perform overlay of such information is practically feasible but requires customized equipment. Video hardware created for broadcasting has been adapted for intra-operative use, but despite acceptable performance, lag remains an issue. To date, processing HD images in real-time is particularly difficult though with graphics processing units this problem can be overcome with appropriate parallelization.

In general, potential point density of the stereo methods is directly linked to the image resolution of the cameras. Therefore, with the trend in the microelectronics industry of further improving the resolution of imaging chips, point density is likely to increase. Defining the required distribution of points will then be important for optimizing computational performance versus data density.

Despite the issues related to run-time and robustness, stereoscopy is currently the most well-placed technique for translation into clinical practice because stereoscopic hardware for both imaging and display is already used in the operating room, for example with the da Vinci surgical system (Intuitive Surgical, Inc., Sunnyvale, CA, USA). Other stereo systems are either already available or in development, e.g., by Karl Storz GmbH (Tuttlingen, Germany), Richard Wolf GmbH

(Knittlingen, Germany), ConMed Corporation (Utica, NY, USA) and Olympus (Shinjuku, Tokyo, Japan), with more manufacturers looking at entering the market in the near future. It is worth noting that stereo endoscope systems developed in the early days of minimally-invasive surgery were not adopted into widespread clinical use, most likely due to the poor ergonomics of head-mounted display systems. However, with recent developments in display technologies and the popularity of 3-D video in the entertainment industry, it is likely that the re-emergence of stereo endoscope systems will have a wider impact in surgical practice.

2) *SL*: *SL* performed comparably to stereo, however, the FRE, which can be regarded as an indicator for the quality of the reference, was higher on average. This can be attributed to the fact that the marker positions needed to be interpolated due to the sparsity of the projected pattern. Hence, we believe that the determined error overestimates the true error more than in the case of stereo.

One advantage of the *SL* system compared to passive reconstruction techniques is that it does not rely on the automatic detection of intrinsic features. In theory, this makes the surface coverage approximately constant (for a certain projector/camera distance), but we observed a relatively high variance for reasons described below, as shown in Fig. 11. The prototype *SL* system evaluated in this paper uses a multispectral pattern to solve the active stereo correspondence problem. In comparison to other *SL* approaches such as phase shifting this has the advantage of acquiring 3-D data on a scene in a single snapshot. Also, the spectral sharpness of the color features increases robustness, making them immune to modulation by the background tissue. The probe is compatible with current endoscopic technology, either through scope instrument channel or trocar insertion.

A disadvantage of the prototype system investigated here is the sparsity of projected points which led to a low point density and the need to interpolate most parts of the surface. This can potentially be improved in future device prototypes by increasing the number of fibers or investigating different projection designs. There are also limitations when imaging conditions are less than ideal, for instance blob distortion resulting from an angular pose or a large projector to tissue distance leads to lower reconstruction accuracy. In regions where a significant amount of blood is in the field of view, strong absorption of blue light leads to poor detection of spots in this wavelength range. The colors of the projected spots were also not truly randomized, such that a large area was typically covered by blue spots which were not detected, giving a high variance.

In the presence of smoke, scattering of the projected pattern and a decrease in the number of detected and accurately identified spots potentially result in significant degradation of the reconstructed surface. However, the *SL* reconstruction results show robust operation in terms of average error when compared to stereo. Due to the heterogeneous nature of the smoke generated, some sections of the tissue were more obscured than others. Thick smoke caused severe distortion of projected spots, higher segmentation errors and detection of numerous false centroids, resulting in a decreased overall detection rate. However, if a spot could be detected, the only observed effect was a slight shift in its actual position, similar to the variability under ideal

conditions. In contrast under white light illumination the smoke becomes the dominant salient foreground feature and, as such, is detected by stereo techniques as the surface when the background is difficult to distinguish. Stereo also triangulates a large number of 3-D points in a dense reconstruction while the *SL* system has a much smaller number of more easily recognized large blobs to identify.

Another issue of the active stereo technique as implemented here is the requirement for triangulation, which means that the projection axis must be offset from the imaging axis, or a stereo camera system must be used. A related limitation is the need for a separate calibration for each arrangement of the laparoscope and projector since, unlike stereo laparoscopes, their relative positions are not permanently fixed and may change significantly during experiments. For robust and reproducible results future designs will incorporate the illumination scheme into the laparoscope itself either as a separate fiber-delivered system or by launching the pattern into one channel in a stereo endoscope in a manner similar to that proposed by Chan *et al.* [21].

For the prototype system described here the computation time was approximately 2–3 s in total, almost entirely taken up by the spot detection and labelling algorithm. However, this code has not yet been optimized for speed and could, in principle, operate at real-time rates as demonstrated by current nonsurgical commercial systems.

Future work related to the *SL* prototype system investigated here will focus on three main areas: improving spot detection accuracy, optimizing the projected features themselves and making the *SL* pattern imperceptible to the user.

3) *ToF*: The *ToF* endoscope yielded the worst reconstruction results out of all techniques investigated in this study, which can be attributed to a number of different reasons. Most importantly, we used only a preliminary prototype. Many of the problems that we have encountered have already been resolved or are currently being addressed. For example, the SNR and thus the measurement precision in camera direction [25] were very low due to the challenge in transmitting enough light to the tissue. In fact, the inhomogeneous illumination with a radial attenuation of the light led to a particularly bad SNR at the boundaries of the images, which resulted in a rather poor reconstruction accuracy for the markers and thus a relatively high mean FRE of almost 3 mm for the registration with the reference. We partially addressed this issue by temporal averaging over successive frames, but the standard deviation in each pixel was still much higher than that obtained with conventional (nonendoscopic) *ToF* cameras. In the meantime, major improvements have been undertaken. In particular, the current prototype provides a higher number of lasers and operates at a modulation frequency of 60 MHz, thus yielding a much better SNR. Note that this is possible because in contrast to standard *ToF* cameras, endoscopic cameras do not need to ensure unambiguous depth ranges of several meters. Another endoscope-related issue we encountered during the experiments was the fact that rotation of the light cable led to an inhomogeneous and changing illumination of the scene because the lasers were continuously assigned to different fibers. In the meantime, a fluid-filled cable is applied, which ensures a much more homogeneous illumination even when the cable is rotated.

Our measurements were further subject to severe systematic errors (cf. Section III-B). Pixel offsets, resulting from fixed pattern noise and fixed pattern phase noise, for example, are clearly visible in the range images. A comprehensive calibration would involve acquisition of a large amount of reference data in a high-dimensional space incorporating pixel ID, distance, amplitude, integration time and temperature prior to operation. While this was practically not feasible for the present study, current work with the prototype focuses on practical depth calibration based on state-of-the-art ToF calibration methods [34].

Further issues that potentially contributed to the large error include scattered light, penetration of IR light into the tissue, and multi-path reflections. Although initial works have been proposed to compensate some of these errors (cf. Foix *et al.* [33]), much more research needs to be directed towards these issues.

On the positive side, ToF is the only technique investigated here that does not require a scanning component or baseline and provides a very high frame rate without relying on sophisticated image processing software. Furthermore, the effect of smoke on ToF reconstruction performance in our study was negligible. We attribute this to the usage of NIR light because smoke light absorption is lower for longer wavelengths [40] and organic smoke particles typically exhibit a lower optical scattering than for visible light [41]. It should further be pointed out that with the release of the Kinect 2 for the Microsoft Xbox 360, the first low-cost ToF camera is now available on the market. This development has led to further improvements of the underlying chip technique, and even more progress can be expected in the future.

In summary, due to the prototype stage of the ToF endoscope applied, no broad conclusions should be drawn from this initial study with respect to the accuracy achievable with ToF endoscopy in general. As a young technique that provides a compact design it has a lot of potential for various applications relying on miniaturization, such as endoscopy.

B. Study Design

As a unique characteristic of this work, a previously proposed validation protocol was applied to enable a comprehensive description including a detailed error analysis and thus ensure reproducibility of our work. The following paragraphs discuss our study design in detail.

1) *Study Conditions*: To our knowledge the present study is the first to assess not only accuracy of a given reconstruction technique but also robustness with respect to factors like smoke and blood. While *in vivo* data could provide clinically realistic endoscopic video sequences, generation of adequate reference data is a major challenge, especially due to tissue dynamics. Furthermore, such a setting would not allow for imaging identical objects with different techniques. For these reasons, we chose an *in vitro* setting for this validation with high-resolution CT images serving as reference. This gave us more controllability at the cost of realism.

To minimize the risk of causing changes in shape due to transport of the validation objects our experiments were conducted on the CT stretcher in an interventional radiology suite rather than in an operating room (OR). While this enabled us to provide higher-quality reference images than we could have obtained

with an intra-operative CT, for example, this also implied that we did not have the specific conditions (e.g., temperature) an OR would provide. Furthermore, our systems were operated by engineers as opposed to medical staff, and the endoscopes were not hand-held, as they typically are during surgery, but affixed to a mounting device. The reason for this was three-fold. Firstly, this allowed us to acquire comparable images with the different hardware applied. Secondly, the prototype SL system required a calibration of the projector and the camera prior to image acquisition and thus a stable pose of the projector relative to the endoscope at this stage. Finally, temporal averaging for the ToF endoscope data was necessary to acquire meaningful results at this stage.

Another practical issue to be discussed in this context is the calibration of the endoscopes. We observed a small increase in FRE over time, which indicates that the calibration parameters changed over time. However, a full calibration of the endoscopes prior to each image acquisition would not be feasible. Furthermore, while change of focus occurs quite regularly in clinical laparoscopy, all calibration techniques applied in this study assume a constant focus during operation. Although some initial research has been conducted with respect to *online* calibration of optical devices without calibration objects, the issue can be regarded as an unsolved technical challenge. Hence, evaluation studies at a higher level should investigate whether the benefits of 3-D reconstruction in a particular clinical context can balance the potential need for leaving the focus unchanged. However, emerging scopes with chip-on-tip designs can solve this issue by having infinite focus.

In this study, smoke was generated with a surgical coagulator. While this yielded realistic images, it should be pointed out that the amount of smoke was not easily controllable. In fact, we initially aimed to generate different amounts of smoke to assess reconstruction accuracy and point density as a function of the amount of smoke present in the scene but then decided against this due to poor reproducibility. Hence, reconstruction results obtained from different endoscopic images in the presence of smoke are not directly comparable. In the case of the stereo methods, for example, we believe that the higher increase in error for the SD images compared to the HD images (cf. Fig. 10) can be attributed to a larger amount of smoke in these images.

To further obtain clinically realistic and challenging endoscopic images, we covered our validation objects with artificial blood. In contrast to smoke, the artificial blood could be distributed in a manner that allowed for imaging of almost identical scenarios with the different techniques. Real blood would have yielded even more realistic absorption/reflection properties, due to, for example, the presence of hemoglobin which is particularly relevant in the context of SL reconstruction. Porcine blood with anticoagulant or out-of-date human blood bags could be a solution to this problem although other forms of haemoglobin such as methaemoglobin may form which also have different absorption spectra. Overall, more realistic studies with larger data sets are required for a more comprehensive evaluation of reconstruction robustness in the presence of blood.

We focused on reconstructing organs of different shape and texture. Future studies could incorporate additional clinically

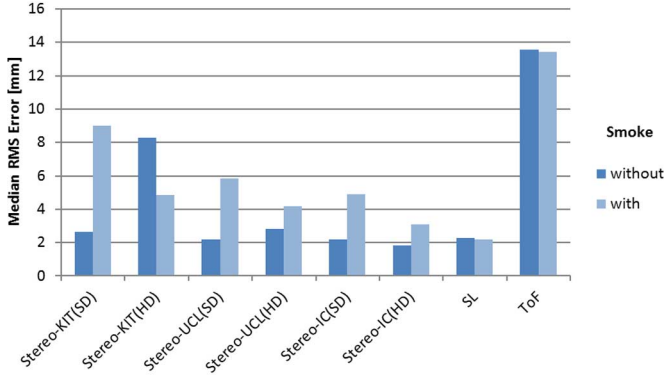


Fig. 10. Results for *D-Smoke* (cf. Section III-A): Median rms surface reconstruction error for different liver shapes with ($n = 5$) and without ($n = 5$) smoke.

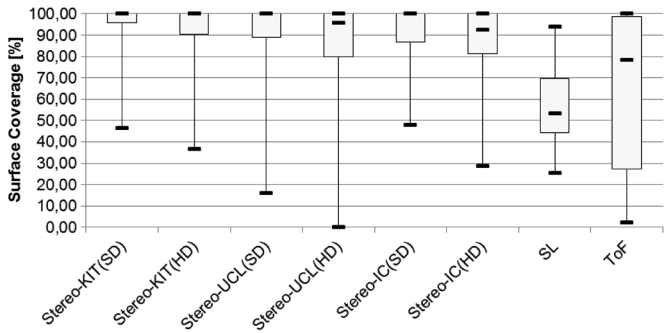


Fig. 11. Box plot (median, first and third quartiles, minimum and maximum) of the surface coverage as defined in Section III-G for an inter-vertex distance of $\delta = 3$ mm and all data sets used in this study, including those with blood and smoke.

relevant objects such as medical instruments. In particular, reconstruction accuracy in the presence of fine structures such as needles or threads could be interesting and could allow us to understand performance in regions affected by shadows.

We had originally planned acquisition of much more data sets within the two days scheduled for the experiments. However, despite several test runs we underestimated acquisition time. Based on our experience and considering the immense resources required for such a comparative validation study (in our case more than ten researchers in addition to medical staff for two days for the raw data acquisition in addition to extensive image postprocessing times), we do not consider an even larger *in vivo/in vitro* validation study of this kind feasible. Standardized phantoms (cf. e.g., <http://www.open-CAS.com>) could be a solution to the problem because they would allow for acquisitions to be performed at different locations and times. This would also overcome the issue of advanced hardware; a new device would not require repetition of the entire experiment. While phantoms are not as realistic as real organs this approach would at least allow for comparability at acceptable costs.

Future validation studies with advanced devices that are ready for clinical application should focus on clinically realistic conditions, with medical staff operating the equipment in the operation room, even if this comes at the cost of loss of control. In this context, practical aspects such as sterilization, workflow issues related to calibration of devices and human factor issues

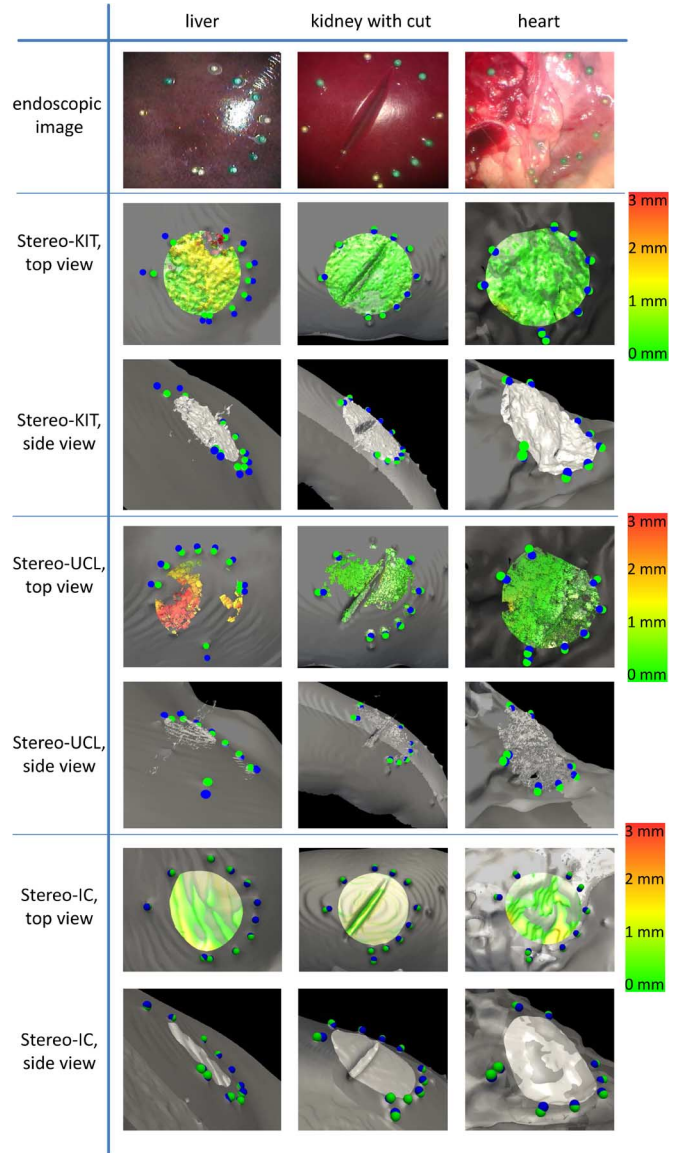


Fig. 12. Reconstruction results obtained with the stereo methods (cf. Section II-A) for different shapes: a porcine liver with homogeneous shape and texture, a porcine kidney with a cut and a porcine heart with complex shape and texture. Colors on the reconstructed surface represent the distance to the closest point on the (registered) reference surface. Large green/blue spheres represent the reference markers and reconstructed markers respectively. (Figure best viewed in color.)

should also be investigated. Furthermore, a specific hypothesis in a specific clinical context (cf. [14]) should be tested in this case.

2) *3-D Reconstruction Methods*: This was the first study to validate the main optical laparoscopic surface reconstruction techniques that work on single images in a comparative manner. We consciously excluded methods that require movement of the scope, such as structure-from-motion or SLAM for two reasons: to focus on those techniques that are suitable for nonrigid environments at present and to ensure comparable acquisition conditions across techniques.

A disadvantage of our study could be seen in the fact that we investigated only one specific SL endoscope and one specific ToF endoscope and that most of our hardware was still in

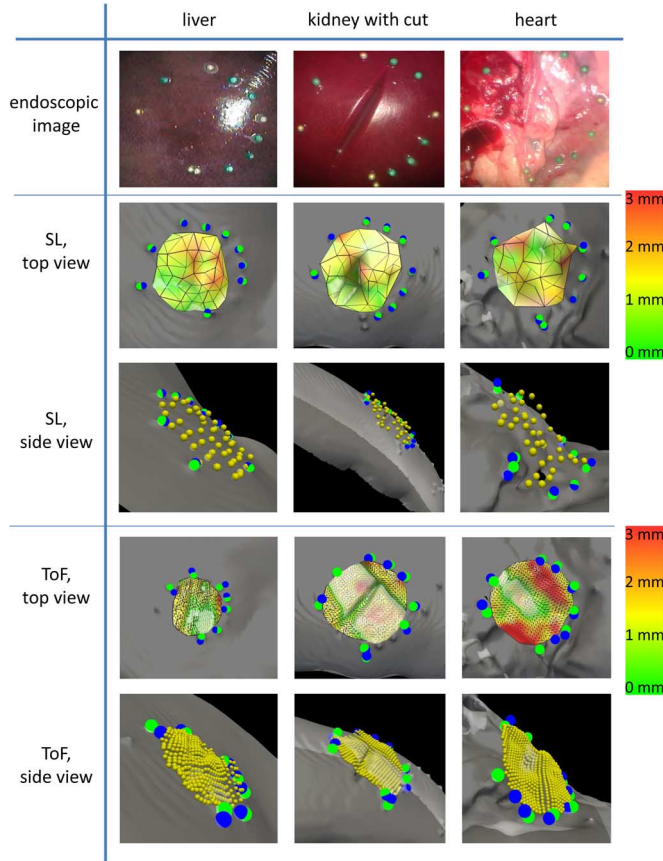


Fig. 13. Reconstruction results obtained with the active reconstruction methods SL (cf. Section II-B) and ToF (cf. Section II-C) for different shapes: a porcine liver with homogeneous shape and texture, a porcine kidney with a cut and a porcine heart with complex shape and texture. Colors on the reconstructed surface represent the distance to the closest point on the (registered) reference surface. Large green/blue spheres represent the reference markers and reconstructed markers, respectively. (Figure best viewed in color.)

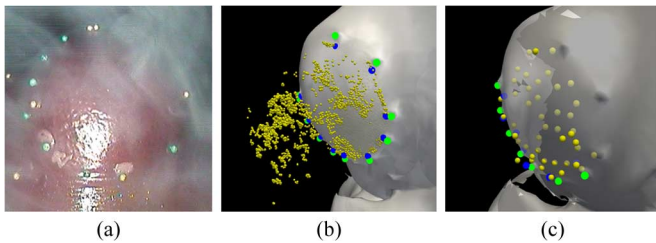


Fig. 14. Reconstruction performance in the presence of smoke. Endoscopic image of a liver (a) and surface reconstruction obtained with the Stereo-KIT algorithm (b) (cf. Section II-A) and with the prototype SL endoscope (c) (cf. Section II-B). (Figure best viewed in color.)

a prototype stage. Hardware modifications would require us to repeat all the image acquisitions. Still, with the only ToF endoscope prototype that we are aware of, one of the very few SL endoscopic systems, and one of the first HD stereo endoscope systems available, we did evaluate state-of-the-art techniques in optical surface reconstruction. A general solution to this issue would be to design standardized phantoms and acquisition protocols as discussed above. In this case, however, the measurement would not be performed on real tissue.

Comparative assessment of passive methods only would be much more straightforward due to the reduced hardware

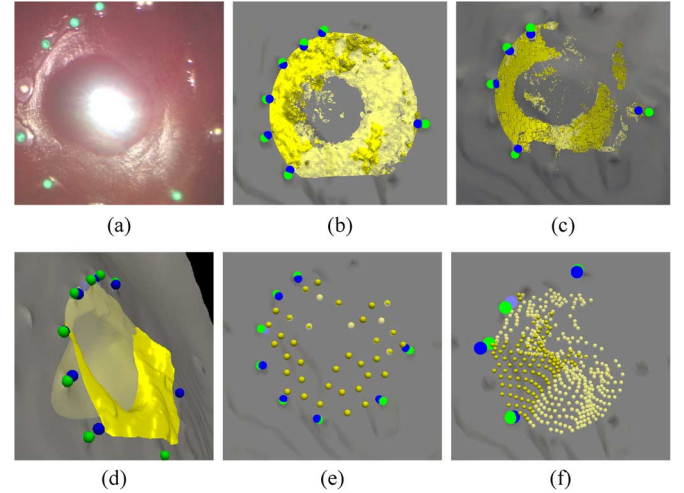


Fig. 15. Reconstruction performance in the presence of blood. Endoscopic image of a liver covered with blood (a) and surface reconstruction obtained with the Stereo-KIT algorithm (b) (cf. Section II-A), the Stereo-UCL algorithm (c) (cf. Section II-A), the Stereo-IC algorithm (d) (cf. Section II-A), the prototype SL endoscope (e) (cf. Section II-B) and the prototype ToF endoscope (f) (cf. Section II-C). (Figure best viewed in color.)

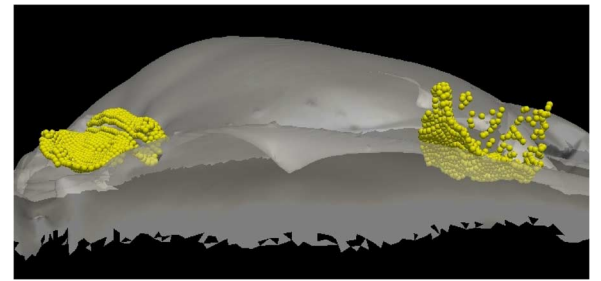


Fig. 16. Two qualitatively very different reconstruction results (yellow point sets) obtained with the ToF endoscope. While the left point cloud fits very well to the reference, the right one is erroneous. In the latter case, the standard deviation obtained from temporal averaging over time, was extremely high, indicating a poor SNR. (Figure best viewed in color.)

complexity, as demonstrated by Seitz *et al.* [37]. In this case, however, care should be taken when designing the data acquisition protocol for calibration. Algorithms based on shape-from-shading (SfS) [42], shape-from-focus [43], and photometric stereo [44], for example, require more data than that used for stereo calibration. In fact, we considered including SfS algorithms after conducting the experiments for this study but failed because of the missing calibration of the light source.

In this work, reprojection/ray fitting errors were used as an indicator for the quality of camera calibration, but direct estimates \hat{E}_{M_i} of the error of the different reconstruction techniques were not available. For the stereo based approaches, such errors could in theory be estimated by monitoring the value of the similarity metric used for correspondence. For example high correlation values could be associated with reliable matches. However, this alone is insufficient when dealing with homogeneous surfaces and potentially other image characteristic such as local texture measures would need to be incorporated. Obtaining a meaningful quantitative prediction of the 3-D reconstruction error remains subject of further research. For the SL system investigated here, the size and shape (e.g., circle versus ellipsoid) of

the blobs could be used as indicators for reconstruction accuracy, but again, inferring a quantitative estimate of error requires further research. In the case of ToF, initial work on estimating the pixel quality has been conducted. Reynolds *et al.* [45], for example, applied a random forest based approach to assign a confidence value between 0 and 1 to each ToF pixel taking into account a variety of different parameters such as the measured amplitude and the object's shape. Future work should be directed to explicitly estimating the error (e.g., in mm) for each surface point.

3) *Reference*: One issue related to reference generation is that *in silico* validation is not possible due to the *comparative* validation strategy. Instead, real objects with known shape and a known pose relative to the camera coordinate systems are required. While generating the reference shape using CT data or scanning devices is feasible, computing the pose remains a great challenge. One possible approach involves calibration of the reference image modality (e.g., a CT scanner or a laser range scanner) with the endoscopes. This could be done by optically tracked markers attached to the different devices, for example, but would add even more complexity to the hardware setup and the acquisition protocol. Maybe most importantly, we did not manage to generate a reliable hand-eye calibration with the ToF endoscope due to the temperature dependent systematic error.

We used CT data as reference and applied a point-based registration of the camera coordinate system with the CT coordinate systems. One drawback of this method is that the quality of the reference depends on the quality of the reconstruction technique, because it relies on the 3-D reconstruction of points detected in camera images using the techniques evaluated here. This was particularly crucial in the case of the active reconstruction techniques. Due to the radial attenuation of light in the ToF images, the SNR at the boundaries and thus the reconstruction accuracy of the markers was rather poor. In fact, we obtained a mean FRE of almost 3 mm. In the case of SL, one issue was the sparsity of reconstructed points. The 3-D marker positions needed to approximate (cf. Section III-E), which comes at the cost of reduced accuracy of the reference. While the stereo methods generally yielded a good FRE of the order of magnitude of 1 mm, one issue is that marker reconstruction is potentially prone to error due to specular reflections on the marker surfaces. Hence, we recommend markers with duller surfaces for obtaining an even better reference.

A further issue is that the reference CT and the endoscopic images could not be acquired simultaneously. Ensuring that there is no deformation between reference data acquisition and observing the tissue with a laparoscope turned out to be a practical challenge. However, the relatively low FRE of 0.6 mm for registering different CTs of the same organ suggests that only minor changes in shape occurred.

4) *Validation Criterion, Comparison Function and Quality Index*: We measured the rms surface distance and the number of points to quantify reconstruction performance. For this purpose, the nearest neighbor function based on the Euclidean distance was applied to establish correspondences between a reconstructed surface and the reference surface. We have also considered accounting for anisotropic reconstruction errors by using a weighted distance, as proposed in [46], [47], but this

would have implied estimating a covariance matrix of localization error for each reconstructed point and thus led to a much higher complexity.

Reconstruction robustness to a number of factors, such as the endoscope pose and the presence of smoke in the scene, was assessed for the first time. Future studies with more mature hardware (in the case of SL and ToF) should further investigate robustness to tissue and endoscope motion as well as to varying lighting conditions.

Run-time, although an important factor for intra-operative applications, was not separately assessed in this study because the different techniques were all implemented in different programming languages and run on different machines. Still, previous publications allowed for a comparison of the different techniques with respect to this parameter.

We also considered comparing surface descriptors, such as vertex normals, or curvature-based measures like mean and Gaussian curvature, curvedness, or shape index. Typically, these measures are computed directly on the mesh data based on vertex connectivity, which, however, leads to an extremely high sensitivity to noise, as discussed in [48]. An alternative approach to computing surface descriptors involves the interpolation of a simple shape, such as a quadric, in the local neighborhood of each vertex, and to use the curvature properties of this shape [49]. This approach is much more robust to noise [48] but raises the issue of choosing an appropriate size for the neighborhood, which—on the one hand—should contain a sufficient number of points to do the interpolation and—on the other hand—should not be too complex due to the required approximation by a simple shape. As we were not able to generate coherent descriptor values for varying numbers of points and noise levels even for identical underlying meshes, we decided against surface descriptors as validation criterion.

Currently, novel technical solutions often lack widespread acceptance among physicians who tend to be reluctant to change their established techniques. The lack of acceptance for clinical use is partly due to a suboptimal integration of the proposed systems into the clinical workflow. Future studies should thus focus on validation criteria related to workflow integration and human factor issues. Among others, time for training and setup as well as usability of user interfaces and displays should be assessed. A comprehensive evaluation would further require a detailed cost analysis.

VI. CONCLUSION

To our knowledge, this is the first study that provides an in-depth insight into the potential and limitations of state-of-the-art laparoscopic surface reconstruction techniques. Based on our results as well as the recently published review on the different techniques [5], the strengths and weaknesses of the individual techniques investigated here can be summarized as follows.

The main advantages of stereoscopy is that it can be used with standard laparoscopic equipment and can therefore already be applied in the clinical setting. Both accuracy and point density are high, and even fine structures can be reconstructed. A current drawback is that stereo-based approaches are computationally

demanding, and performance depends highly on the scene and the paucity of textural information.

In contrast, active techniques require additional light to be introduced at the surgical site and as a result can reliably deliver dense depth maps at high update rates because they do not rely on natural features. A common limitation is the required hardware equipment adaptation, which becomes important in the context of workflow integration and costs. The prototype SL device showed that reconstruction accuracy close to that of stereo could be achieved, although with much lower density. Further iterations of the device and the shape of the projected features will allow for further increases in accuracy and reconstruction density. ToF, as the youngest technique, is still subject to severe systematic errors and noise. Although it is not yet ready for clinical application, it has potential as the only method that can generate dense depth maps in real-time without requiring a baseline.

Common to all techniques is that robustness (with respect to smoke and bleeding, for example) must be improved. Also, online calibration to cope with changes in focus, for example, remains a technical challenge to be addressed though newly emerging scopes with chip-on-tip designs can solve this by having infinite focus.

In conclusion, the study suggests that robustness to the widely varying clinical challenges is practically not likely in the short term with any of the methods reviewed here. Despite the potential of the relatively new laparoscopic active surface reconstruction methods based on SL and ToF, stereo is currently the only practical and feasible approach to obtaining 3-D information during surgery because stereo endoscopes are already used in clinical practice.

ACKNOWLEDGMENT

The authors would like to thank R. Wolf GmbH for providing the ToF endoscope and the HD stereo endoscope for this study. The authors would also like to thank M. Jochim (Department of Radiology, DKFZ) and A. Franz (computer-assisted Interventions, DKFZ) for helping with the data acquisition and processing.

REFERENCES

- [1] K. Cleary and T. M. Peters, "Image-guided interventions: Technology review and clinical applications," *Annu. Rev. Biomed. Eng.*, vol. 12, pp. 119–142, Aug. 2010.
- [2] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, "Deformable medical image registration: Setting the state of the art with discrete methods," *Annu. Rev. Biomed. Eng.*, vol. 13, no. 1, pp. 219–244, 2011.
- [3] P. Markelj, D. Tomanzevic, B. Likar, and F. Pernus, "A review of 3-D/2-D registration methods for image-guided interventions," *Med. Image Anal.*, vol. 16, no. 3, pp. 642–661, 2010.
- [4] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [5] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov, "Optical techniques for 3-D surface reconstruction in computer-assisted laparoscopic surgery," *Med. Image Anal.*, vol. 17, pp. 974–996, 2013.
- [6] M. Hu, G. Penney, P. Edwards, M. Figl, and D. Hawkes, "3-D reconstruction of internal organ surfaces for minimal invasive surgery," in *Proc. MICCAI*, 2007, vol. 10, no. 1, pp. 68–77.
- [7] P. Mountney and G.-Z. Yang, "Motion compensated SLAM for image guided surgery," in *Proc. MICCAI*, 2010, pp. 496–504.
- [8] S. Röhl, S. Bodenstedt, S. Suwelack, H. Kennigott, B. P. Müller-Stich, R. Dillmann, and S. Speidel, "Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration," *Med. Phys.*, vol. 39, no. 3, pp. 1632–1645, 2012.
- [9] T. T. Wu and J. Y. Qu, "Optical imaging for medical diagnosis based on active stereo vision and motion tracking," *Opt. Exp.*, vol. 15, pp. 10421–10426, Aug. 2007.
- [10] R. Richa, P. Poignet, and C. Liu, "Efficient 3-D tracking for motion compensation in beating heart surgery," in *Proc. MICCAI*, 2008, pp. 684–691.
- [11] D. P. Noonan, P. Mountney, D. S. Elson, A. Darzi, and G.-Z. Yang, "A stereoscopic fibroscope for camera motion and 3-D depth recovery during minimally invasive surgery," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3274–3279.
- [12] D. Burschka, M. Li, M. Ishii, R. H. Taylor, and G. D. Hager, "Scale-invariant registration of monocular endoscopic images to CT-Scans for sinus surgery," *Med. Image Anal.*, vol. 9, no. 5, pp. 413–426, 2005.
- [13] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G.-Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI*. New York: Springer, 2010, Lecture Notes Comput. Sci., pp. 275–282.
- [14] P. Jannin, C. Grova, and C. Maurer, "Model for defining and reporting reference-based validation protocols in medical image processing," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 1, pp. 63–73, 2006.
- [15] N. Atzpadin, P. Kauff, and O. Schreier, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 321–334, Mar. 2004.
- [16] D. Stoyanov, "Stereoscopic scene flow for robotic assisted minimally invasive surgery," in *Proc. MICCAI*, 2012, vol. 15, no. 1, pp. 479–486.
- [17] D. Stoyanov, G. Mylonas, F. Deligianni, A. Darzi, and G. Yang, "Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures," in *Proc. MICCAI*, 2005, vol. 3750, pp. 139–146.
- [18] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 593–600.
- [19] G. Puerto and G.-L. Mariottini, "A comparative study of correspondence-search algorithms in MIS images," in *Proc. MICCAI*, 2012, pp. 625–633.
- [20] P.-L. Chang, D. Stoyanov, A. Davison, and P. E. Edwards, "Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery," in *Proc. MICCAI*, 2013, pp. 42–49.
- [21] M. Chan, W. Lin, C. Zhou, and J. Y. Qu, "Miniaturized three-dimensional endoscopic imaging system based on active stereovision," *Appl. Opt.*, vol. 42, pp. 1888–1898, 2003.
- [22] N. T. Clancy, D. Stoyanov, A. Groch, L. Maier-Hein, G. Z. Yang, and D. S. Elson, "Spectrally-encoded fibre-based structured lighting probe for intraoperative 3-D imaging," *Biomed. Opt. Exp.*, vol. 2, no. 11, pp. 3119–3128, 2011.
- [23] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [24] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight cameras in computer graphics," *Comput. Graph. Forum*, vol. 29, no. 1, pp. 141–159, 2010.
- [25] R. Lange, "3-D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology," Ph.D. dissertation, Univ. Siegen, Siegen, Germany, Sep. 2000.
- [26] P. Jannin, J. M. Fitzpatrick, D. J. Hawkes, X. Pennec, R. Shahidi, and M. W. Vannier, "Validation of medical image processing in image-guided therapy," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1445–1449, Dec. 2002.
- [27] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1330–1334, Nov. 2000.
- [28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, Nov. 2004.
- [29] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.
- [30] S. Haase, C. Forman, T. Kilgus, R. Bammer, L. Maier-Hein, and J. Hornegger, "ToF/RGB sensor fusion for augmented 3-D endoscopy using a fully automatic calibration scheme," in *Bildverarbeitung für die Medizin (BVM)*, 2012.
- [31] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.*, Washington, DC, 1998, p. 839.

- [32] M. Shimizu and M. Okutomi, "Significance and attributes of subpixel estimation on area-based matching," *Syst. Comput. Jpn.*, vol. 34, no. 12, pp. 1–10, 2003.
- [33] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sensors*, vol. 11, no. 9, pp. 1917–1926, Sep. 2011.
- [34] S. Mersmann, A. Seitel, M. Erz, B. Jähne, F. Nickel, M. Mieth, A. Mehrabi, and L. Maier-Hein, "Calibration of Time-of-Fight cameras for accurate intraoperative surface reconstruction," *Med. Phys.*, vol. 40, no. 8, p. 082701, 2013.
- [35] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Am.*, vol. 4, p. 629, 1987.
- [36] A. Groch, A. Seitel, S. Hempel, S. Speidel, R. Engelbrecht, J. Penne, K. Höller, S. Röhl, K. Yung, S. Bodenstedt, F. Pflaum, T. dos Santos, S. Mersmann, H.-P. Meinzer, J. Hornegger, and L. Maier-Hein, "3-D surface reconstruction for laparoscopic computer-assisted interventions: Comparison of state-of-the-art methods," in *Proc. SPIE Med. Imag., Visualizat., Image-Guided Procedures, Model.*, 2011, p. 796415.
- [37] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 519–528.
- [38] A. N. Kumar, T. S. Pfeiffer, A. L. Simpson, R. C. Thompson, M. I. Miga, and B. M. Dawant, "Phantom-based comparison of the accuracy of point clouds extracted from stereo cameras and laser range scanner," in *Proc. SPIE Med. Imag., Image-Guided Procedures, Robotic Intervent., Model.*, 2013, p. 867125.
- [39] A. L. Simpson, J. Burgner, C. L. Glisson, S. D. Herrell, B. Ma, T. S. Pfeiffer, R. J. W. , III, and M. I. Miga, "Comparison study of intraoperative surface acquisition methods for surgical navigation," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1090–1099, Apr. 2013.
- [40] K. Lewis, W. P. Arnott, H. Moosmüller, and C. E. Wold, "Strong spectral variation of biomass smoke light absorption and single scattering albedo observed with a novel dual-wavelength photoacoustic instrument," *J. Geophys. Res-Atmos.*, vol. 113, no. D16, 2008.
- [41] R. K. Chakrabarty, H. Moosmüller, W. P. Arnott, M. A. Garro, J. G. Slowik, E. S. Cross, J.-H. Han, P. Davidovits, T. B. Onasch, and D. R. Worsnop, "Light scattering and absorption by fractal-like carbonaceous chain aggregates: Comparison of theories and experiment," *Appl. Opt.*, vol. 46, no. 28, pp. 6990–7006, 2007.
- [42] T. Collins and A. Bartoli, "Live monocular 3-D laparoscopy using shading and specular information," in *Proc. Int. Conf. Inf. Process. Comput.-Assist. Intervent.*, 2012, vol. 7330, pp. 11–21.
- [43] T. Takeshita, M. Kim, and Y. Nakajima, "3-d shape measurement endoscope using a single-lens system," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 8, no. 3, pp. 451–459, 2013.
- [44] T. Collins and A. Bartoli, "3-D reconstruction in laparoscopy with close-range photometric stereo," in *Proc. MICCAI*, 2012, vol. 15, no. 2, pp. 634–642.
- [45] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, and G. Brostow, "Capturing Time-Of-Flight data with confidence," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 945–952.
- [46] L. Maier-Hein, M. Schmidt, A. Franz, T. dos Santos, A. Seitel, B. Jähne, J. Fitzpatrick, and H. Meinzer, "Accounting for anisotropic noise in fine registration of time-of-flight range data with high-resolution surface data," in *Proc. MICCAI*, 2010, pp. 251–258.
- [47] L. Maier-Hein, A. Franz, T. dos Santos, M. Schmidt, M. Fangerau, H.-P. Meinzer, and J. M. Fitzpatrick, "Convergent iterative closest-point algorithm to accommodate anisotropic and inhomogeneous localization error," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1520–1532, Aug. 2012.
- [48] T. R. dos Santos, "Multi-modal partial surface matching for intraoperative registration," Ph.D. dissertation, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany, 2012.
- [49] F. Cazals and M. Pouget, "Estimating differential quantities using polynomial fitting of osculating jets," in *Proc. 2003 Eurograph./ACM SIG-GRAPH Symp. Geometry Process.*, 2003, pp. 177–187.