Journal of
Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Intraoperative on-the-fly organ-mosaicking for laparoscopic surgery

Daniel Reichard
Sebastian Bodenstedt
Stefan Suwelack
Benjamin Mayer
Anas Preukschas
Martin Wagner
Hannes Kenngott
Beat Müller-Stich
Rüdiger Dillmann
Stefanie Speidel

SPIE.

Intraoperative on-the-fly organ-mosaicking for laparoscopic surgery

Daniel Reichard,[a,*] Sebastian Bodenstedt,[a] Stefan Suwelack,[a] Benjamin Mayer,[b] Anas Preukschas,[b] Martin Wagner,[b] Hannes Kenngott,[b] Beat Müller-Stich,[b] Rüdiger Dillmann,[a] and Stefanie Speidel[a,*]
[a]Karlsruhe Institute of Technology, Institute for Anthropomatics and Robotics, Adenauerring 2, D-76131 Karlsruhe, Germany
[b]University of Heidelberg, Department of General, Abdominal and Transplantation Surgery, Im Neuenheimer Feld 110, D-69120 Heidelberg, Germany

Abstract. The goal of computer-assisted surgery is to provide the surgeon with guidance during an intervention, e.g., using augmented reality. To display preoperative data, soft tissue deformations that occur during surgery have to be taken into consideration. Laparoscopic sensors, such as stereo endoscopes, can be used to create a three-dimensional reconstruction of stereo frames for registration. Due to the small field of view and the homogeneous structure of tissue, reconstructing just one frame, in general, will not provide enough detail to register preoperative data, since every frame only contains a part of an organ surface. A correct assignment to the preoperative model is possible only if the patch geometry can be unambiguously matched to a part of the preoperative surface. We propose and evaluate a system that combines multiple smaller reconstructions from different viewpoints to segment and reconstruct a large model of an organ. Using graphics processing unit-based methods, we achieved four frames per second. We evaluated the system with *in silico*, phantom, *ex vivo*, and *in vivo* (porcine) data, using different methods for estimating the camera pose (optical tracking, iterative closest point, and a combination). The results indicate that the proposed method is promising for on-the-fly organ reconstruction and registration. [DOI: 10.1117/1.JMI.2.4.045001]

Keywords: endoscopic image processing; stitching; simultaneous localization and mapping; visualization; surgical vision; quantitative endoscopy.

Paper 15160PR received Aug. 3, 2015; accepted for publication Nov. 4, 2015; published online Dec. 10, 2015.

1 Introduction

The amount of minimally invasive surgeries performed yearly is increasing rapidly. This is largely due to the numerous benefits these types of intervention have on the patient side: shorter stay in hospital, less trauma, minimal scarring, and lower chance of postsurgical complications. There are several drawbacks for the surgeon, though: limited hand-eye coordination, no haptic feedback, no direct line of sight, and a limited field of view.

Computer-assisted surgery tries to alleviate some of these drawbacks by providing the surgeon with information relevant to the state of the intervention. Prior to the intervention, preoperative data are acquired for diagnosis and surgical planning. Elaborate equipment (e.g., CT or MRI) generates precise data and also allows imaging from the interior of the body. Three-dimensional (3-D) models created from this data can provide the surgeon with a virtual view inside the patient during surgery. To this end, the models have to be registered to the current surgical scene, i.e., the current location and orientation of the real structure have to match those of the virtual one. The available tools for intraoperative imaging (e.g., endoscope) are limited in image quality and field of view. But they can be used to create intraoperative surface models that enable the registration process with the preoperative data.

Many groups have explored ways to obtain intraoperative surface models. To sample an intraoperative surface, Herline et al.[1] used a probe in which the tip was moved over the visible parts of the liver. The probe was localized with an active position sensor. To avoid possible tissue damage, newer approaches commonly rely on ranged sensors. Laser range scanners used by Clements et al.[2] offer high reconstruction quality for conventional liver surgery. The downside is the need of additional hardware in the operating room. Dumpuri et al.[3] extended this approach to take intraoperative soft tissue deformation into account. After an initial rigid registration of the laser scan and CT surfaces, the residual closest point distances between the rigidly registered surfaces are minimized using a computational approach. The method was further refined by Rucker et al.[4] using a tissue mechanics model subjected to boundary conditions, which were adjusted for liver resection therapy.

For registering preoperative data in laparoscopic surgery, the organ surface can be observed with optical laparoscopic sensors that provide a 3-D-reconstruction of a single video frame. There are many methods for reconstructing 3-D surface structures.[5] The most commonly used methods rely on multiple view geometry. Through correspondence analysis between two or more images, a 3-D-reconstruction can be obtained via triangulation. Structure from motion (SfM) uses one camera with images from at least two different perspectives for triangulation. A similar approach is the stereo camera. It uses two image sensors, which can be calibrated to each other. The known transformation between the two stereo images allows a more precise reconstruction. Instead of using naturally given correspondences, shape from shading algorithms use structured light for active triangulation. The structured light has to be projected onto the scene, which is proving to be difficult in surgical practice. The methods mentioned previously only reconstruct a small

*Address all correspondence to: Daniel Reichard, E-mail: daniel.reichard@kit.edu; Stefanie Speidel, E-mail: stefanie.speidel@kit.edu

field of view, and due to the homogeneous structure of tissue, a single frame, in general, will not provide enough detail to rule out geometrical ambiguities (i.e., an intraoperative surface patch has multiple possible matches on the preoperative model surface) during registration.

To remedy this problem, Plantefève et al.[6] used anatomical landmarks to achieve a stable initial registration. The preoperative landmarks were labeled automatically while the intraoperative labeling required manual interaction. After the initial registration, a biomechanical model and the established correspondences between the landmarks were used to counteract intraoperative soft tissue deformation and movement.

To expand the reconstructed surface, methods to associate multiple frames are needed. One of these is the procedure of localizing the camera in the world while simultaneously mapping it, known as simultaneous localization and mapping (SLAM) in literature. SLAM is a well-known approach in robotic mapping and has also found its way into computer-assisted laparoscopic surgery. Mountney et al.[7] introduced an SLAM approach using a stereo endoscope to map the soft tissue of the liver. They worked with a sparse set of image texture features, which are tracked by an extended Kalman filter. In later work, the system was expanded to compensate breathing motions.[8]

To recover from occlusions or sudden camera movements, Puerto-Souza et al.[9,10] developed a robust feature matching, the hierarchical multiaffine (HMA) algorithm. In tests with real intervention data sets, the HMA algorithm exceeded the existing feature-matching methods in the number of image correspondences, speed, accuracy, and robustness.

SLAM can also be achieved through a single moving camera. With the previous mentioned SfM technique, reconstructing 3-D scene information is possible. In the work of Grasa et al.[11,12] this method is used to create a sparse reconstruction of a laparoscopic scene in real time. However, reconstructions from single camera solutions have the problem that they do not provide an absolute scale. To approach this problem, Scaramuzza et al.[13] used nonholonomic constraints. Recently, Newcombe et al.[14] introduced the KinectFusion method, which provides dense reconstructions of medium-sized (nonmedical) scenes in real time using a Microsoft Kinect for data acquisition. In the work of Haase et al.,[15] an extension of Newcombe et al.[14] is used to reconstruct the surgical situs with multiple views taken by a 160 × 120 pixels time-of-flight camera.

In this paper, we present a system that combines 3-D reconstructions generated online by a stereo endoscope from multiple viewpoints, while simultaneously segmenting structures on-the-fly. It is based on our previous work[16] and was extended by a detailed description of the method and an extensive evaluation on *in silico*, phantom, *ex vivo,* and *in vivo* data. In our system, the reconstructions and the segmentations are combined into one organ model. To compute a 3-D point cloud from a stereo image pair, the hybrid recursive matching (HRM) algorithm outlined by Röhl et al.[17] was used. It was compared with other 3-D surface reconstruction methods by Maier-Hein et al.[18] and achieved the best results. The segmentation of the organ of interest is done on the basis of color images. Using a random forest based classifier,[19] each pixel is labeled as part of an organ of interest or background. The resulting point clouds and their respective labels are then integrated into a voxel-volume using a KinectFusion based algorithm.[14] Given enough viewpoints, the voxel-volume will contain a combined model more suited for registration than the model generated from single shot.

The novelty of the approach presented in this work is the application of a stereo endoscope, a modality already available in the surgical workflow, to reconstruct an entire scene from multiple viewpoints online, while simultaneously segmenting one or more organs of interest. Our main contributions are as follows:

- Mosaicking of frame reconstruction parts using a frame-to-model registration with the possible use of a tracking device (e.g., NDI Polaris).
- Dense surface model that is generated online and is available after each image frame.
- Per-frame segmentation of organs is achieved through a fast graphics processing unit (GPU) random forest approach.
- Global segmentation allows accumulation of the single-frame segmentation probabilities for each global surface point. The combined segmentation results lead to a higher and more robust recognition rate.

In the following, we will present a more detailed description of our reconstruction workflow, followed by an evaluation using *in silico*, phantom, *ex vivo,* and *in vivo* data (porcine). Three methods for determining the camera pose are also evaluated: optical tracking, iterative closest point (ICP) tracking, and a combination of these two methods. The evaluation and workflow are described in the context of laparoscopic liver surgery.

2 Methods

Our system for reconstructing the scene consists of multiple steps (Fig. 1). First, we reconstruct a 3-D point cloud from stereo image frames. At the same time, the organs of interest are segmented in the video image. Afterward, the reconstruction is combined with the segmentation results and integrated into a truncated signed distance (TSD) volume. From this volume, a mosaicked model of the combined reconstructions can be retrieved. Using a TSD volume allows us to incorporate information from different viewpoints to create a larger model than from a single view, while simultaneously reducing noise in the model.

2.1 *Reconstruction and Segmentation*

The stereo endoscope provides left and right camera images, which are first preprocessed to remove distortion and to rectify the image pair. Using correspondence analysis,[17] we first calculate a disparity map between the two images and then triangulate those matches, resulting in a dense 3-D point cloud R_i in camera

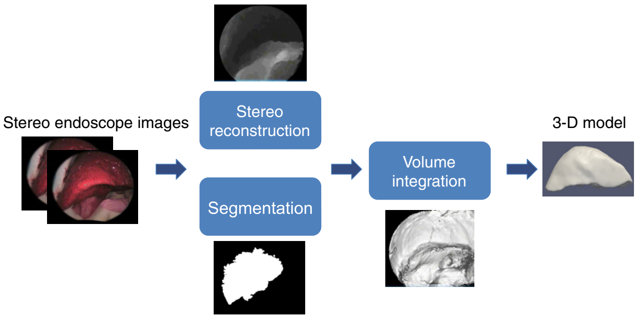

Fig. 1 System overview.

coordinates for each time step. The preprocessing and the correspondence analysis were both implemented on the GPU.

Every pixel in the scene is simultaneously classified using a random forest[19] into foreground, e.g., liver, and background. As features, the hue and saturation channels from the HSV color space and the color-opponent dimensions a and b from the LAB color space were used. The classifier thus provides a mapping $C_i(\mathbf{p}) \rightarrow \{1 \dots n\}$, $\mathbf{p} \in R_i$ from 3-D point to a class-label for each time step.

The random forest was trained on multiple previously labeled image. We trained a forest consisting of 50 trees with a maximum depth of 10. To allow real-time processing, the classification portion of the random forest was ported to the GPU.

2.2 *Integration into Truncated Signed Distance-Volume*

Assuming the pose P_i of the camera in each time step is known, the point clouds R_i can be transformed into the world coordinate system $R_i^W = P_i(R_i)$. At every time step, R_i^W is integrated into a TSD volume $S_i(\mathbf{p}) \rightarrow [F_i(\mathbf{p}), K_i(\mathbf{p}, j), W_i(\mathbf{p})]$, where $\mathbf{p}$ is a voxel in the volume. The TSD value $F_i(\mathbf{p})$ and the weight $W_i(\mathbf{p})$ are computed as suggested in Ref. 14.

$$F_i(\mathbf{p}) = \frac{W_{i-1}(\mathbf{p}) F_{i-1}(\mathbf{p}) + W_{R_i}(\mathbf{p}) F_{R_i}(\mathbf{p})}{W_{i-1}(\mathbf{p}) + W_{R_i}(\mathbf{p})}, \tag{1}$$

$$W_i(\mathbf{p}) = W_{i-1}(\mathbf{p}) + W_{R_i}(\mathbf{p}), \tag{2}$$

where $W_{R_i}(\mathbf{p})$ is the weight of voxel $\mathbf{p}$ in the current frame. It can be used to weight the TSD value computed for the current frame F_{R_i} correlated to the measurement uncertainty, or set uniformly to one. F_{R_i} can be computed as

$$F_{R_i}(\mathbf{p}) = \Psi[\lambda^{-1} \| t_i - \mathbf{p} \|_2 - R_i(x)], \tag{3}$$

$$\lambda = \| K^{-1} \dot{x} \|_2, \tag{4}$$

$$x = \lfloor K T_i^{-1} \mathbf{p} \rfloor, \tag{5}$$

$$\Psi(\eta) = \begin{cases} \min(1, \frac{\eta}{\mu}) \operatorname{sgn}(\eta), & \eta \geq -\mu \\ \text{undefined}, & \text{else} \end{cases}, \tag{6}$$

where K is the camera calibration matrix, $\dot{x}$ is the homogenized image coordinate x, $\lfloor . \rfloor$ is the nearest neighbor lookup, T_i is the camera transformation, and t_i is the translation part of T_i. λ^{-1} converts the ray distance $\| t_i - \mathbf{p} \|_2$ to a depth value in the camera coordinate system. The function $\Psi(\eta)$ specifies the area of influence of R_i over the voxels F_{R_i}. The parameter μ is responsible for the maximal distance before the influence of a point on a voxel is truncated.

We included $K_i(\mathbf{p}, j)$ in the volume to account for class membership of $\mathbf{p}$.

$$K_i(\mathbf{p}, j) = \frac{W_{i-1}(\mathbf{p}) K_{i-1}(\mathbf{p}, j) + W_{R_i^W}(\mathbf{p}) K_{R_i^W}(\mathbf{p}, j)}{W_{i-1} + W_{R_i^W}}, \tag{7}$$

$$K_{R_i^W}(\mathbf{p}, j) = \begin{cases} 1, & \text{if } C_i[R_i^W(\mathbf{p})] = j \\ 0, & \text{else} \end{cases}, \tag{8}$$

where $R_i^W(\mathbf{p})$ represents the point in R_i^W that lies in $\mathbf{p}$ and j stands for the classifier category (e.g., background and target structure).

The class membership $C_i(\mathbf{p})$ at the current time step can then be computed as

$$C_i(\mathbf{p}) = \underset{j \in \{1 \dots n\}}{\operatorname{argmax}} K_i(\mathbf{p}, j). \tag{9}$$

This way of smoothing class membership over time allows our system to cope with potential misclassifications.

2.3 *Camera Pose*

To integrate the point cloud R_i into the TSD volume, the pose P_i of the camera at time step i has to been known. In this paper, we consider three methods for estimating P_i.

1. ICP: We adopt the assumption of Newcombe et al.[14] that the pose of the camera changes only slightly between frames. By registering R_i with a ray cast of the TSD volume using the projective data association ICP algorithm,[20] we estimate P_i. With the small movement assumption and the special ICP variant, all pixels can be used in real time.
2. Polaris: We use the NDI Polaris optical tracking system to track both camera and the patient.
3. Mixed: We combine the two methods by using the tracking information as a seed for the ICP.

3 Results

We performed five experiments to evaluate our system using *in silico*, phantom, *ex vivo*, and *in vivo* livers. For each liver, a reference was computed by laser scan or CT. In each experiment, we moved the stereo endoscope over the liver and used the captured images to reconstruct and segment the liver simultaneously. For each experiment, three mosaicked models, each with a different method for tracking the camera pose, were constructed as described in Sec. 2.3. Afterward, we computed the average distance of each intraoperative reconstructed point to the reference for each model. To reduce the influence of tracking errors, the mosaicked porcine liver models were registered to the reference using ICP. For the purpose of comparison, we also computed the average distance of the unprocessed single frame point clouds R_i^W to the ground truth.

The camera pose used for transforming each point cloud into the world coordinate system was given by an NDI Polaris optical tracking system. For the two silicone and the first *ex vivo* experiment, a calibrated phase alternating line (PAL) stereo endoscope with a fixed camera unit and a PC workstation (Table 1, No. 1) were used. The second *ex vivo* and the *in vivo* experiment were conducted with a calibrated HD stereo endoscope with chip-on-the-tip technology (Table 1, No. 2).

Both configurations took, on average, ~0.25 s for one frame integration, implying a frame rate of ~4 fps. More run-time information is available in Table 1.

3.1 *In Silico*

In order to evaluate the mosaicking without the errors induced by the stereo matching (HRM), we used a simulation framework

Table 1 The PC and endoscopic hardware used for evaluation. Both stereo endoscopes were calibrated before the experiments and have no zoom and fixed focusing. The run-time analysis reveals that the higher computational cost caused through higher resolution can be compensated by faster hardware.

	PC configuration			
No.	CPU	Graphic	RAM	Endoscope
1	i7-2700k	GeForce GTX 650Ti	16 Gbyte	Richard Wolf – PAL (720 × 576 pixels)
2	i7-5820k	GeForce GTX 970	16 Gbyte	Storz three-dimensional TIPCAM®1 (1920 × 540 pixels)

	Average run-time				
No.	Hybrid recursive matching (HRM)	TSD volume integration	Segmentation	Overall	Frame rate
1	0.134	0.024	0.045	0.252	≈4
2	0.131	0.027	0.030	0.241	

to generate a circular image sequence of a textured CT-liver model (Fig. 2). For each of the 320 images, depth map and camera position were computed. With the simulated input data, an accurate mosaicked reconstruction of the model was achieved (Table 2).

The simulation was also used to create noisy depth data to evaluate the mosaicking behavior on imperfect data. The noise was generated through a Perlin noise model, as it is similar to the errors made by HRM. Three different noise levels, noise 1 (mean error 1.12 mm ± 0.86), noise 2 (2.30 mm ± 1.68), and noise 3 (3.36 mm ± 2.49), were used. The results are showing that the mosaicking is reducing the noise and is producing a more accurate model than the single-shot reconstructions (Fig. 3).

3.2 *Phantom Liver*

After verifying the method *in silico*, we performed five phantom experiments with three silicone livers (Fig. 4). The first two livers were recorded with both the Wolf and HD stereo endoscope, and the third only with the HD stereo endoscope (Fig. 5). The first (Wolf endoscope 1 and HD endoscope 1) and third (HD endoscope 3) liver were placed on a flat surface, whereas the second liver (Wolf endoscope 2 and HD endoscope 2) was placed inside a 3-D printed patient phantom (Fig. 4). As previously mentioned, an NDI Polaris optical tracking system was used for endoscope position tracking. To evaluate the ICP only approach, Polaris tracking data from the first image frame served as registration to the reference model.

The results show that the use of an HD stereo endoscope increases the quality and stability of the method (Table 3). In combination with the Wolf stereo endoscope, our method produces the best results with Polaris mode. With the HD stereo endoscope, the best results shift toward mixed mode. Figure 6 illustrates an example of a failed reconstruction using the ICP for frame-to-model registration. Multiple consecutive frames to model registrations with high errors in position or orientation usually lead to a fracture in the final reconstruction, i.e., the spatial relation of the reconstructed parts before and after the ICP failure(s) is erroneous.

To determine if the models created by our approach are suitable for registering a preoperative model in absence of soft tissue deformation, we transformed the model for silicone 1 using multiple random rigid transformations. Thereupon, we performed a rough registration of the model to the reference laser scan with fast point feature histograms[21] and fine-tuned it with the use of ICP. The average distance error for 600 random transformations was 13.19 mm ± 23.39, with 90% having an

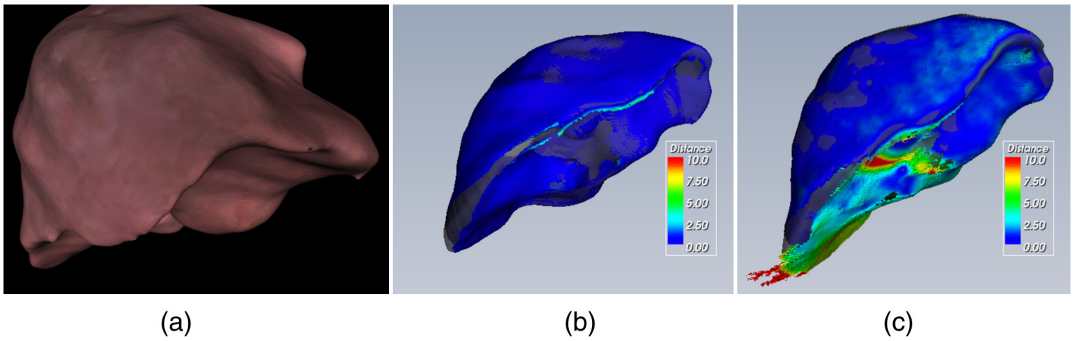

Fig. 2 (a) Textured simulation model, and (b) error distribution using reference depth data and (c) using hybrid recursive matching (HRM).

Table 2 The root mean square (RMS) error between the mosaicked models and the reference model in mm. The last column contains the RMS error, using all frames for evaluation separately. In the last row, hybrid recursive matching (HRM) is used for depth map creation instead of the reference depth data.

	Endoscope position			
Simulation	Reference	Iterative closest point (ICP)	Mixed	Single shot with reference position
Reference depth map	0.022 ± 0.009	0.016 ± 0.008	0.021 ± 0.009	—
With noise 1	0.021 ± 0.012	0.836 ± 0.354	0.187 ± 0.089	0.812 ± 0.363
With noise 2	0.100 ± 0.046	4.941 ± 2.149	1.137 ± 0.975	1.610 ± 0.731
With noise 3	1.156 ± 0.597	20.622 ± 9.780	3.862 ± 1.806	2.388 ± 1.383
HRM	0.468 ± 0.251	2.004 ± 0.920	1.820 ± 0.682	0.694 ± 0.307

Table 3 The RMS error between the mosaicked models and the reference models in mm. The use of the HD stereo endoscope caused an overall improvement of the results. The first and second recording (lines 1 and 2) illustrate the strong influence of the HRM quality in the final reconstruction. The error levels in single shot (single HRM reconstructions) are reflected in the results of the final reconstruction.

	Endoscope position			
Silicone	Polaris	ICP	Mixed	Single shot with Polaris position
Wolf endoscope 1	1.89 ± 1.97	5.04 ± 4.66	3.79 ± 3.37	2.88 ± 2.14
Wolf endoscope 2	3.64 ± 3.55	6.73 ± 5.97	6.14 ± 5.59	5.36 ± 5.33
HD endoscope 1	1.04 ± 0.98	1.64 ± 1.26	0.73 ± 0.61	2.12 ± 1.69
HD endoscope 2	2.13 ± 1.78	3.76 ± 3.42	1.36 ± 1.27	3.31 ± 2.58
HD endoscope 3	1.56 ± 1.29	4.32 ± 3.83	1.27 ± 0.94	2.48 ± 2.04

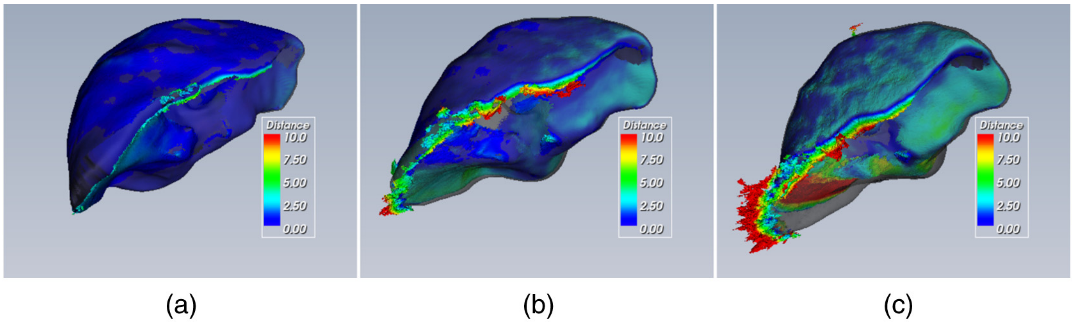

Fig. 3 Error distribution for (a) noise 1, (b) noise 2, and (c) noise 3.

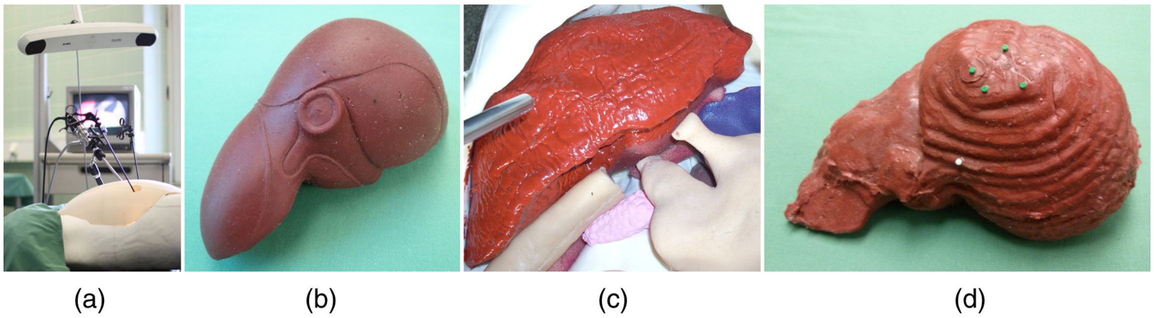

Fig. 4 (a) Experimental phantom setup: stereo endoscope, optical tracking system, and patient phantom. (b) Silicone liver 1, (c) silicone liver 2, and (d) silicone liver 3.

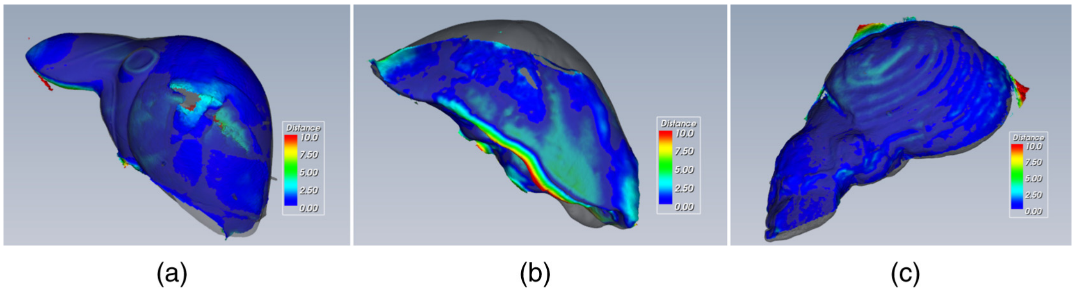

Fig. 5 Error distribution (HD stereo endoscope) in (a) liver 1, (b) liver 2, and (c) liver 3. The images for the Wolf stereo endoscope can be viewed in our previous work.[16]

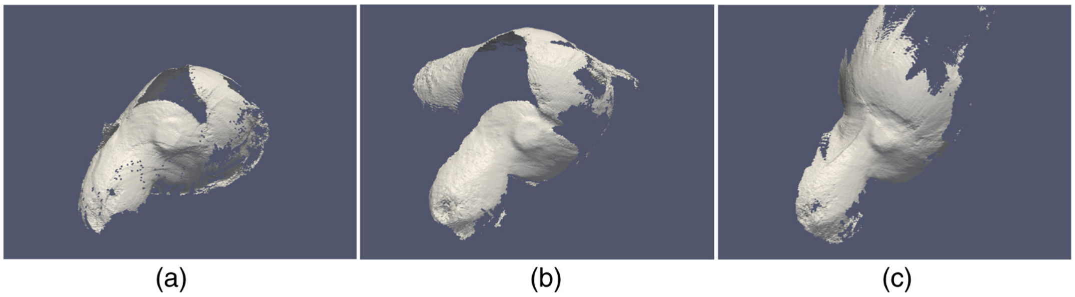

Fig. 6 Example of an iterative closest point (ICP) failure in silicone liver 1. A reconstruction using only Polaris is shown in (a). The hole in the model is due to missing viewpoints. A reconstruction using ICP is shown in (b), where the mosaicking failed, creating a larger hole. The mixed approach was used in (c), which closed the hole, connecting parts of the liver that do not belong together.

error <10 mm. In comparison, using only a single frame reconstruction had an error of 89.92 mm ± 20.48.

3.3 *Ex Vivo Porcine Liver*

As a first step into the real operating environment, two *ex vivo* porcine liver experiments were conducted. In the first experiment (liver 1), the Wolf stereo endoscope was used, and reference data were provided by a laser scan. For the second experiment (liver 2), we used the high-resolution HD Storz stereo endoscope and CT imaging as reference data (Fig. 7).

The results from liver 1 are comparable to the phantom data, both using the same hardware, showing that the HRM can cope with real liver texture. The second experiment using the HD Storz stereo endoscope reduced the root mean square error from 4.21 to 1.51 mm (Table 4). While slightly different experiment settings could cause small differences, the grave change is certainly due to the better image quality and resolution.

3.4 *In Vivo Porcine Liver*

To evaluate our system in an *in vivo* setting, we performed an animal experiment. At first, the pig was prepared for surgery and placed on the CT table (Fig. 8). After applying a pneumoperitoneum as well as placing ports for the endoscope and instruments, we recorded several image sequences featuring a sweep of the porcine liver. Shortly after each sequence, a CT scan was taken in order to evaluate the sequence, using the liver model acquired through the scan. To minimize breathing deformation between the two image modalities, respiration was paused between scan acquisitions.

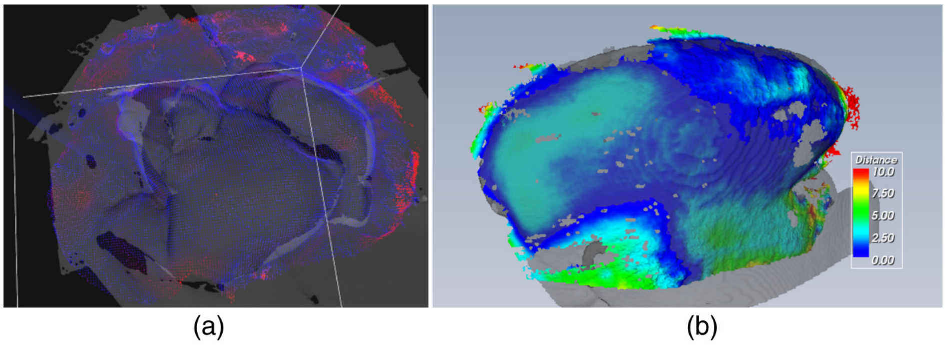

Fig. 7 Error distribution in *ex vivo* liver experiments: (a) *ex vivo* with Wolf stereo endoscope. Blue signals an error <2 mm and red an error >2 mm. (b) *Ex vivo* with HD stereo endoscope.

Table 4 The RMS error between the mosaicked models and the ground truth models in mm. The HD stereo endoscope outperforms the older Wolf stereo endoscope clearly. This demonstrates the importance of image quality and image resolution for the reconstruction result.

	Endoscope position			
Ex vivo	Polaris	ICP	Mixed	Single shot with Polaris position
Wolf endoscope	4.21 ± 3.78	10.37 ± 5.67	10.94 ± 5.48	6.88 ± 6.11
HD endoscope	1.57 ± 1.63	2.23 ± 1.97	1.51 ± 1.21	6.38 ± 4.79

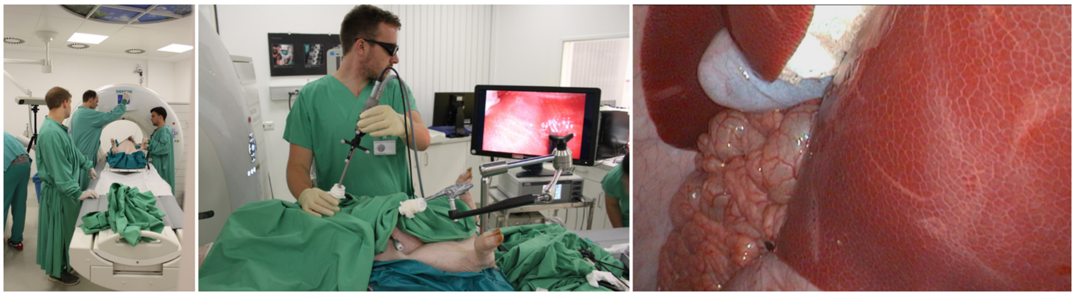

Fig. 8 Experimental setting for *in vivo* evaluation. The image sequences were taken directly on the CT table to minimize recording time between endoscopy and CT images. The last picture displays an exemplary *in vivo* endoscopic image.

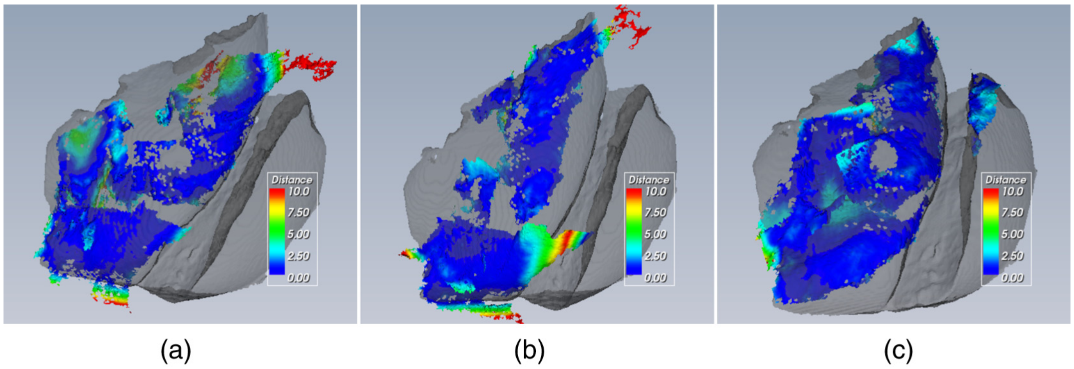

Fig. 9 Error distribution for *in vivo* liver experiment: (a) Polaris, (b) ICP, and (c) mixed.

Table 5 The RMS error between the mosaicked models and the ground truth models in mm. A reference CT scan was performed before and after each sequence. The sequences differ in endoscope movement and liver coverage. The different error values between sequences indicate that the endoscope handling plays a significant role for the reconstruction quality.

	Endoscope position			
In vivo	Polaris	ICP	Mixed	Single shot with Polaris position
Sequence 1	0.75 ± 0.65	1.97 ± 1.72	0.54 ± 0.36	5.11 ± 3.96
Sequence 2	1.87 ± 1.03	2.48 ± 1.33	0.97 ± 0.79	7.25 ± 5.53
Sequence 3	1.92 ± 1.54	2.21 ± 1.27	1.07 ± 1.02	6.03 ± 4.60

The previous results in the second *ex vivo* experiment agree with the *in vivo* results. Both were obtained using the HD Storz stereo endoscope (Fig. 9). As in the previous *ex vivo* experiment, the error in all three sequences was smallest in mixed mode (Table 5). The mean error of the three mixed mode results is 0.86 mm.

4 Discussion

In this paper, we presented an approach enabling the reconstruction and segmentation of organs from multiple viewpoints online during laparoscopic surgery. We have clearly demonstrated that mosaicking multiple reconstructions reduces the distance error when compared to single-shot reconstructions.

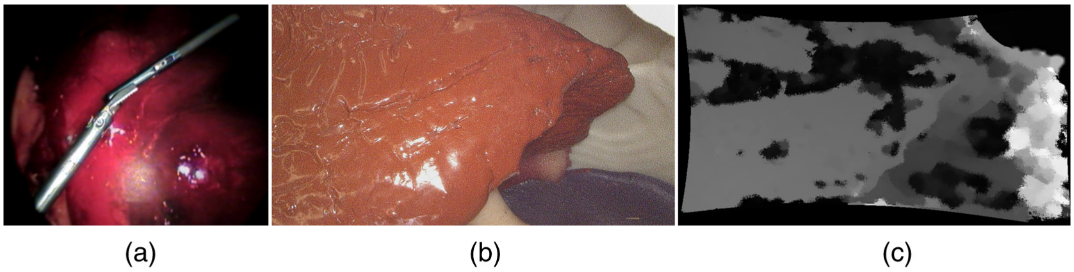

Fig. 10 Difficult situations: (a) our method would integrate the instruments into the volume destroying previously captured surface information near it. A poorly illuminated image is shown in (b) leading to (c) a poor HRM result.

Furthermore, we have shown that using a mosaicked model for rigid registration produces a significantly smaller error (dropping from 90 to 13 mm).

The comparison between results from the Wolf stereo endoscope and the HD stereo endoscope allows an insight into the correlation of image quality and final reconstruction result. The data suggest that image quality and image resolution are important for two steps. First, the HRM reconstruction needs a certain image quality to produce satisfactory results, e.g., good illumination, resolution, and little distortion. For the HRM, on the other hand, the increased sensor noise greatly reduces the reconstruction quality (as shown in Fig. 10). For the ICP-based methods (ICP only and mixed), the bad frame reconstruction not only affects the mosaicked model directly, but also the frame-to-model registration, as the ICP uses the frame 3-D reconstruction to register the frame to the model created so far. Without the use of Polaris tracking, multiple consecutive bad frame registrations usually lead to a complete fail of the mosaicking attempt. The Polaris localization method allows a higher HRM error tolerance since the patches are at least placed at the correct location.

In our experiments, Polaris tracking was necessary to achieve the best results. But advances in hardware, like HD stereo endoscopes, will make image-based tracking more robust. As shown in our *ex vivo* experiments, the ICP-only error dropped 78% due to the use of the better HD endoscope. Also, the mixed mode exceeds the pure Polaris method when used with the HD endoscope, meaning, the small localization errors were reduced by the ICP. This is a synergetic process as Polaris provides a good initial alignment needed for a stable ICP.

There are limitations of our work. Objects, like instruments, moving between camera and organ lead to reconstruction errors. Although the instruments are likely classified as background, they are still integrated into the voxel volume. This causes an erroneous morphing of the underlying previously captured organ surface. To fix this problem, the instruments have to be specifically classified in the image and the associated pixels then excluded from the integration process. We are currently working on a stable automatic classification of instruments to solve this problem. A general problem is the HRM reconstruction quality. Slight deviations from suitable illumination settings can lead to bad reconstruction results as shown in Fig. 10. Therefore, careful monitoring of the capture settings is needed. Since our method relies on surface sweeps, a sufficient space for endoscopic movement is required. Not enough surface area is captured for reconstruction otherwise. Finally, the frame-to-model ICP registration modes (mixed and only) are likely not suitable for organs with uniform appearances (e.g., prostate) or would at least create a higher error as with distinct shaped organs (e.g., liver or kidney).

Future research will focus on accounting for dynamic scenes, as currently only static scenes were considered, meaning that soft tissue deformation was not taken into account. Due to the shown limitations of the frame-to-model ICP registration, evaluating other methods for localization should be considered to lessen the dependency on optical tracking systems. Especially, feature-based approaches, taking advantage of the veined surface of organs and color information in general, are a promising addition to depth data only methods.

Acknowledgments

Special thanks to Rudolf Rempel and Patrick Mietkowski for the great clinical support. The present research is sponsored by the Klaus Tschira Foundation and the European Social Fund of the State Baden-Wuerttemberg. It was conducted within the setting of the research training group 1126 and the SFB/Transregio 125, Project A01 funded by the German Research Foundation. The study protocol was approved by the local ethics committee Heidelberg and by the regional committee Karlsruhe. Veterinary handling and care were provided by the staff of the Interfacultary Biomedical Research Facility at Heidelberg University. All animals were treated in compliance with the National Research Council's criteria for humane care as described in the Guide for the Care and Use of Laboratory Animals prepared by the National Institutes of Health.

References

1. A. J. Herline et al., "Surface registration for use in interactive, image-guided liver surgery," *Comput. Aided Surg.* **5**(1), 11–17 (2000).
2. L. W. Clements et al., "Robust surface registration using salient anatomical features for image-guided liver surgery: algorithm and validation," *Med. Phys.* **35**(6), 2528–2540 (2008).
3. P. Dumpuri et al., "Model-updated image-guided liver surgery: preliminary results using surface characterization," *Prog. Biophys. Mol. Biol.* **103**(2), 197–207 (2010).
4. D. C. Rucker et al., "A mechanics-based nonrigid registration method for liver surgery using sparse intraoperative data," *IEEE Trans. Med. Imaging* **33**(1), 147–158 (2014).
5. L. Maier-Hein et al., "Optical techniques for 3-D surface reconstruction in computer-assisted laparoscopic surgery," *Med. Image Anal.* **17**(8), 974–996 (2013).
6. R. Plantefève et al., "Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery," *Ann. Biomed. Eng.* 1–15 (2015).
7. P. Mountney et al., "Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 347–354, Springer (2006).

8. P. Mountney and G.-Z. Yang, "Motion compensated SLAM for image guided surgery," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 496–504, Springer (2010).
9. G. Puerto-Souza et al., "A fast and accurate feature-matching algorithm for minimally-invasive endoscopic images," *IEEE Trans. Med. Imaging* **32**(7), 1201–1214 (2013).
10. G. A. Puerto-Souza, A. Castaño-Bardawil, and G.-L. Mariottini, "Real-time feature matching for the accurate recovery of augmented-reality display in laparoscopic videos," in *Augmented Environments for Computer-Assisted Interventions*, C. A. Linte et al., Eds., pp. 153–166, Springer, Berlin, Heidelberg (2013).
11. O. G. Grasa, J. Civera, and J. Montiel, "EKF monocular SLAM with relocalization for laparoscopic sequences," in *2011 IEEE Int. Conf. on Robotics and Automation*, pp. 4816–4821, IEEE (2011).
12. O. G. Grasa et al., "Visual SLAM for handheld monocular endoscope," *IEEE Trans. Med. Imaging* **33**(1), 135–146 (2014).
13. D. Scaramuzza et al., "Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints," in *2009 IEEE 12th Int. Conf. on Computer Vision*, pp. 1413–1419, IEEE (2009).
14. R. A. Newcombe et al., "KinectFusion: real-time dense surface mapping and tracking," in *Proc. of the 2011 10th IEEE Int. Symp. on Mixed and Augmented Reality*, pp. 127–136, IEEE Computer Society, Washington, DC (2011).
15. S. Haase et al., "3-D operation situs reconstruction with time-of-flight satellite cameras using photogeometric data fusion," in *Medical Image Computing and Computer-Assisted Intervention*, K. Mori et al., Eds., pp. 356–363, Springer, Berlin Heidelberg (2013).
16. S. Bodenstedt et al., "Intraoperative on-the-fly organ-mosaicking for laparoscopic surgery," *Proc. SPIE* **9415**, 94151S (2015).
17. S. Röhl et al., "Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration," *Med. Phys.* **39**, 1632 (2012).
18. L. Maier-Hein et al., "Comparative validation of single-shot optical techniques for laparoscopic 3-D surface reconstruction," *IEEE Trans. Med. Imaging* **33**(10), 1913–1930 (2014).
19. F. Schroff, A. Criminisi, and A. Zisserman, "Object class segmentation using random forests," 2008, http://research.microsoft.com/pubs/72423/Criminisi_bmvc2008.pdf (13 November 2015).
20. G. Blais and M. D. Levine, "Registering multiview range data to create 3-D computer objects," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(8), 820–824 (1995).
21. R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3-D registration," in *IEEE Int. Conf. on Robotics and Automation*, pp. 3212–3217, IEEE (2009).

Daniel Reichard is pursuing a doctoral degree in computer science at the Institute for Anthropomatics and Robotics, KIT. He is performing research within the computer-assisted surgery junior research group. His research interests are in endoscopic three-dimensional (3-D) reconstruction and 3-D model registration for preoperative data.

Stefanie Speidel has a computer-assisted surgery junior research group at the Institute for Anthropomatics and Robotics, KIT. She received her PhD in 2009 in the context of the intelligent surgery research training group, a cooperation among KIT, the University of Heidelberg, and the German Cancer Research Center. Her research interests include endoscopic vision, intraoperative sensor analysis for context-aware assistance, and intraoperative registration with biomechanical models.

Rüdiger Dillmann received his PhD from University of Karlsruhe in 1980. Since 1987 he has been professor of the Department of Computer Science and is director of the Humanoids and Intelligence Systems Research Lab at the Karlsruhe Institute of Technology (KIT). 2002 he became director of an innovation lab at the Research Center for Information Science (FZI). Since 2009 he has been spokesman of the Institute of Anthropomatics and Robotics at KIT.

Biographies for the other authors are not available.