Robotic Arm Platform for Multi-View Image Acquisition and 3D Reconstruction in Minimally Invasive Surgery

Alexander Saikia[1,2], Chiara Di Vece[1], Sierra Bonilla[1], Chloe He[1], Morenike Magbagbeola[1], Laurent Mennillo[1], Tobias Czempiel[1,2], Sophia Bano[1] and Danail Stoyanov[1]

***Abstract*—Minimally invasive surgery (MIS) offers significant benefits, such as reduced recovery time and minimised patient trauma, but poses challenges in visibility and access, making accurate 3D reconstruction a significant tool in surgical planning and navigation. This work introduces a robotic arm platform for efficient multi-view image acquisition and precise 3D reconstruction in MIS settings. We adapted a laparoscope to a robotic arm and captured *ex-vivo* images of several ovine organs across varying lighting conditions (operating room and laparoscopic) and trajectories (spherical and laparoscopic). We employed recently released learning-based feature matchers combined with COLMAP to produce our reconstructions. The reconstructions were evaluated against high-precision laser scans for quantitative evaluation. Our results show that whilst reconstructions suffer most under realistic MIS lighting and trajectory, two matching methods achieve close to sub-millimetre accuracy with 0.80 and 0.76mm Chamfer distances and 1.06 and 0.98mm RMSEs for ALIKED and GIM respectively. Our best reconstruction results occur with operating room lighting and spherical trajectories. Our robotic platform provides a tool for controlled, repeatable multi-view data acquisition for 3D generation in MIS environments, which can lead to new datasets necessary for novel learning-based surgical models.**

I. INTRODUCTION

Surgery has experienced remarkable evolution, particularly with the rise of minimally invasive techniques like endoscopy and laparoscopy. These methods have transformed modern surgical practice by reducing patient recovery time and minimizing tissue damage. However, they also present challenges, such as limited visibility, reduced tactile feedback, and increased cognitive demands on surgeons. As surgical techniques continue to advance, there is an increasing need for innovative technologies to address these challenges, enhance precision, and improve patient outcomes [1].

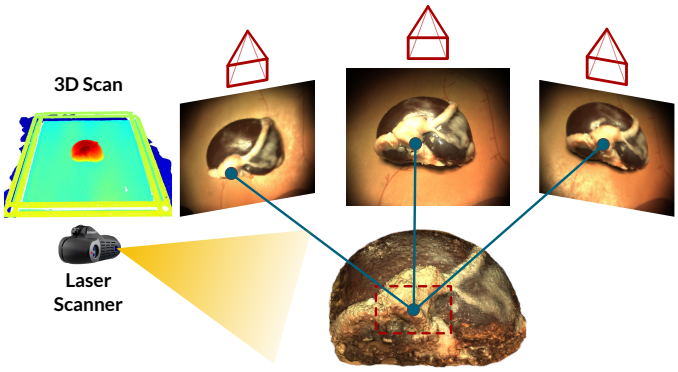

Fig. 1. This work creates 3D reconstructions from multi-view images collected by our robotic arm platform. We use a laser scanner to obtain GT data for comparison.

In minimally invasive surgery (MIS) and especially robot-assisted minimally invasive surgery (RAMIS), endoscopy is crucial for navigating the surgical field. Moreover, accurate and reliable 3-dimensional (3D) reconstruction of organs during surgery is vital for advancing downstream computer-assisted tasks, such as Augmented Reality (AR) and Virtual Reality (VR) [2], [3] and enhanced visualisation for surgical navigation [4]. Tools accomplishing these downstream tasks can augment the surgeon's understanding of the operating field and enable the identification of key structures such as blood vessels and tumours, improving both safety and precision in MIS and RAMIS [5]. However, the process of producing 3D reconstructions is challenging owing to the complexity of the surgical environment and the limited availability of image data with adequate 3D Ground Truth (GT).

This work presents a robotic arm-based platform that addresses these challenges by offering multi-view image acquisition and 3D reconstruction in MIS environments. By utilising a robotic arm we ensure more consistent, reliable scanning enabling a direct comparison of different acquisition settings as well as obtaining pose information. The system is designed to accommodate various imaging modalities and employs a laser scanner to capture highly accurate 3D GT, allowing a comprehensive evaluation of the reconstruction pipeline as seen in Figure 1. Our approach utilises state-of-

Manuscript received: September, 23rd, 2024; Revised December, 21st, 2024; Accepted February, 3rd, 2025.

This paper was recommended for publication by Editor Jessica Burgner-Kahrs upon evaluation of the Associate Editor and Reviewers' comments.

This work was supported in part by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145/Z/16/Z], the Department of Science, Innovation and Technology (DSIT) and the Royal Academy of Engineering under the Chair in Emerging Technologies programme; EPSRC Centre for Doctoral Training in Intelligent, Integrated Imaging In Healthcare (i4health) [EP/S021930/1]; EPSRC Optical and Acoustic imaging for Surgical and Interventional Sciences (OASIS) [UKRI145].

[1] UCL Hawkes Insitute, the Department of Medical Physics and Biomedical Engineering and the Department of Computer Science, University College London, London, United Kingdom

[2] EnAcuity Ltd., London, United Kingdom.

Corresponding author: alexander.saikia.21@ucl.ac.uk

Digital Object Identifier (DOI): see top of this page.

TABLE I
COMPARISON OF PUBLICLY AVAILABLE SURGICAL DATASETS WITH THEIR GT PROPERTIES AND SIZE. GT AVAILABILITY IS INDICATED WITH CHECK MARKS (✓). (MV) MULTI-VIEW, (CT) COMPUTED TOMOGRAPHY, (SL) STRUCTURED LIGHT,(MM) MULTI-MODALITY, (LS) LASER SCAN.

Dataset	Type	Scenes	Subjects	Frames	Pose	Depth	3D Model	GT Type	Platform
CV3D [6]	Mono MV	*In-vivo* Synthetic	22	30073	✓	✓	✓	Phantom/Synthetic	Synthetic
SCARED [7]	Stereo MV	*Ex-vivo* Porcine	9	17607	✓	✓		SL (Keyframes)	Da Vinci
SERV-CT [8]	Stereo MV	*Ex-vivo* Porcine	16	32		✓	✓	CT	Da Vinci
StereoMIS [9]	Stereo MV	Porcine/Human	6	14804	✓	✓		Stereo	Da Vinci
Ours	Mono MV	*Ex-vivo* Ovine	8	29806	✓		✓	LS	Robot Arm

the-art feature matching methods, such as ALIKED [10] and GIM [11], paired with LightGlue (LG) [12] for robust correspondence across frames, and COLMAP [13], [14] for dense 3D reconstruction. The platform is validated by collecting and processing multiple *ex-vivo* organ datasets and benchmarking its performance against GT laser scans to demonstrate the accuracy and robustness of the proposed system.

Our main contributions and 3D reconstruction pipeline for surgery are summarised as follows:

- We present a customisable robotic arm platform integrated with a laparoscopic system for precise multi-view data acquisition with 6-Degrees of Freedom (DoF) pose data;
- We demonstrate our acquisition protocol with different design choices, such as trajectories and lighting, to account for high realism and adaptability to the medical domain;
- We utilise a high-precision laser scanner for obtaining high fidelity GT suitable for comprehensive evaluation of 3D reconstruction pipelines;
- We demonstrate the use of multi-view images acquired by the system to generate 3D reconstructions, providing both qualitative and quantitative comparisons across established algorithmic choices using relevant performance metrics.

Our platform represents a step towards semi-automatic 3D reconstruction of organs, which has the potential of being translated to surgical vision enhancement such as registration of pre-operative data such as MRI or PET, surgical measurements and advanced visualisation and guidance to aid the surgeon whereby improving patient outcome.

II. RELATED WORK

A. Datasets

The growing field of RAMIS aims to improve the safety, ease, and effectiveness of procedures [15], [16], [17]. The development of robust computer-assisted surgical systems and 3D reconstruction pipelines rely on high-quality datasets. Notwithstanding, publicly available datasets with robot kinematics/poses and GT 3D information remain rare due to logistical and ethical constraints.

Current datasets in RAMIS (see Table I) often suffer from significant limitations in terms of size, realism and ground-truth accuracy, leading to concerns around overfitting and poor generalisability. In practice, the creation of datasets for RAMIS involves many trade-offs. Human data is difficult to obtain, so most works opt to use animal models such as pigs [7], [8], [9] or synthetic data [6]. While these substitutions are invaluable for prototyping, their generalisability to in-vivo human applications can be limited [6]. The methods employed to capture depth and 3D information also vary widely, each presenting unique benefits and drawbacks. For instance, while synthetic data generation provides reliable depth data at scale, it suffers from a realism gap [6]. Structured light [7], CT imaging [8] and stereo depth estimation [9], while computed on real data, face challenges arising from calibration and time synchronisation inaccuracies, tissue deformation, and visual artefacts arising from phenomena such as specularity.

In creating a new platform for dataset acquisition, we aim to address the significant complexity of creating a new surgical 3D dataset. Our platform allows for plug-and-play acquisition using preplanned trajectories and a laser scanner.

B. 3D Reconstruction for Surgery

After collecting multi-frame surgical data, 3D reconstruction is essential to enhance the surgeon's view and proceed with downstream tasks. There are numerous 3D reconstruction algorithms, including Structure from Motion (SfM), Simultaneous Localisation and Mapping (SLAM) and, if a stereo endoscope is used, stereo reconstruction [18], [19]. These algorithms rely on feature extraction and matching between 2-dimensional (2D) images for generating 3D point clouds, which are used to estimate depth and camera pose. Traditional feature extraction and matching methods like SIFT [20] have been widely used for this purpose. However, more sophisticated methods can generate more accurate matching in surgical environments, particularly when dealing with low-texture organs.

In [21], the authors demonstrated that combinations of ALIKED with LG and GIM with LG yielded the best results on out-of-domain data. Other learning-based detectorless techniques, such as Dust3r [22] and Mast3r [23], process the entire pipeline at once, making them highly computationally intensive and impractical for large datasets on standard machines.

III. ROBOTIC PLATFORM FOR 3D GENERATION

This section outlines our robotic arm platform for multi-view image acquisition and 3D reconstruction in MIS, as seen in Figure 2. We describe the design of the robotic arm, imaging and recording systems, introduce the various scanning trajectories implemented using the robot and discuss our hand-eye calibration method.

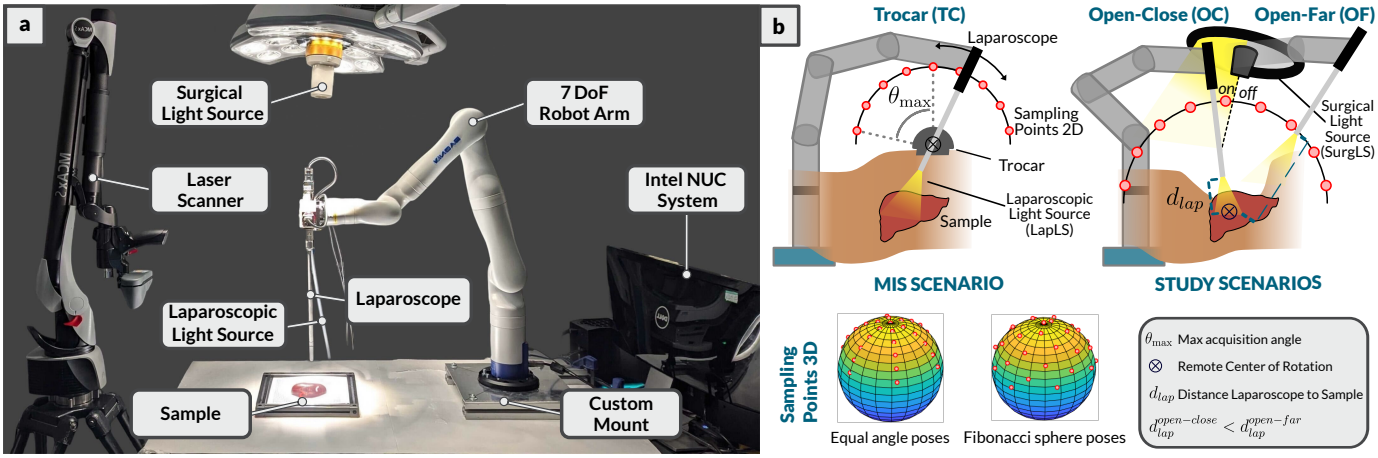

Fig. 2. **a)** Annotated figure showing the robotic platform. The robot is attached to the table using a custom mount. The laparoscope optical system is attached to the robot with 3D printed and laser-cut parts. An Intel NUC handles all trajectory control and data capture. **b)** Depiction of lighting conditions LapLS and SurgLS and the three trajectories: TC, OC and OF. Additionally, the sampling points in 3D are depicted for both equal angle poses, where poses are calculated by sequentially incrementing the azimuth and altitude angles by fixed amounts and Fibonacci sphere poses which are more even coverage over the imaging sphere.

A. Robotic Arm

The platform utilises a Kinova Gen3 7-DoF robotic arm[1] with a maximum reach of 902 mm and a maximum speed of 0.5 m/s. Unlike many other robotic arms, the actuators can rotate infinitely, allowing more of the workspace to be covered by the Kinova, making it ideal for our application of capturing the different organs from different angles.

We interface with the robot using a modified version of ros2-kortex package[2]. The platform uses the robotic manipulation platform MoveIt2[3] [24] for motion planning and kinematics control. A custom `moveit_config` was created to suit our robot-laparoscope configuration.

B. Imaging

Our imaging system uses a FLIR Blackfly S USB3 (BFS-U3-50S5C-C, Teledyne FLIR, Wilsonville, Oregon, USA) Red-Green-Blue (RGB) camera custom mounted to a (HOPKINS®Telescope 26003 AGA, Karl Storz SE & Co. KG, Tuttlingen, Germany). This 5-megapixel camera has a resolution of 2448x2048, giving us optimal data for further 3D reconstructions. As seen in Figure 2, the laparoscope is mounted directly to the Kinova robotic arm using 3D-printed and laser-cut brackets.

We consider two different light source settings for our study: the Storz laparoscopic light source (Storz D-LIGHT C 201336 20) attached directly to the laparoscope by fibre optic light cable (Storz 495 NCS), referred to as the LapLS and the overhead ceiling-mounted surgical lights (Maquet Volista Surgical Light, Maquet GmBH, Rastatt, Germany) in combination with the laparoscopic light source at low intensity, referred to as the SurgLS. The overhead surgical light source evenly illuminates the image and eliminates vignetting effects visible when using the laparoscopic light source only. Whilst the combination of light sources is not realistic for MIS, it gives an ideal set of lighting conditions for comparison.

C. Trajectories

The robot's laparoscopic multi-view acquisition aims to image the specimen organ from various angles, ensuring full surface coverage at least once. We represent this acquisition process as a sphere with the sample positioned at the centre. In Figure 2.b, we depict this using two spherical scanning trajectories: OC and OF. The key difference between them is the variation in the distance to the specimen, where the OC distance ($d_{lap}^{Open-Close}$) is shorter than the OF distance ($d_{lap}^{Open-Far}$). While this setup provides excellent coverage, it is only suitable for open surgery without constraints on the surgical field.

To achieve these spherical trajectories, we set a remote centre of motion (RCM) collinear but external to the laparoscope and place it at the centre of these virtual spheres. Thus, the tip of the laparoscope moves in a spherical trajectory around the RCM, and all poses can be defined as an orientation of the RCM.

In laparoscopic surgery, however, the endoscope's access is limited by the TC — a single entry point through the skin surface. Therefore, we define a third trajectory, TC, to replicate this scenario. This scenario is more clinically realistic but also introduces significant challenges for organ imaging due to reduced field of view (FoV).

Our platform relies on MoveIt2, which utilises the Open Motion Planning Library (OMPL) [25], [26] to plan these constrained trajectories, specifically using the RRTConnect planner, using the KDL solver. We change the kinematics solver seed state to the current joint states to counteract awkward solutions arising from non-unique solutions.

D. Pose Generation Using Fibonacci Sphere Sampling

Achieving an even distribution of points on a spherical surface is essential for optimal coverage and accurate 3D

[1] www.kinovarobotics.com
[2] https://github.com/Kinovarobotics/ros2_kortex
[3] https://moveit.ai

reconstruction. Traditional methods, like uniformly sampling azimuthal and altitude angles, often result in uneven distributions, causing point clustering near the poles leading to sampling bias. To address this, we employed the Fibonacci sphere sampling method to generate evenly distributed poses on a spherical surface [27], which gave us a uniform distribution of points across a sphere as seen in Figure 2.b.

To maintain focus on a region of interest, we limit the poses for all three trajectories to those within a specified angular range, defined by an angle limit θ_{max}. This ensures that only points with an elevation angle greater than $90^\circ - \theta_{max}$ are included, thereby restricting poses to the top of the sphere. For our data acquisition, θ_{max} was set to 40°.

E. Data Recording

We use our custom-developed Python app built on the FLIR Spinnaker Software Development Kit (SDK) to preview and capture data. Each video frame is saved as a numpy file, and frame information such as exposure time and current framerate is stored in a CSV frame log. The RCM and end-effector positions of the robot are also saved for analysis.

F. Hand-Eye Calibration

An additional dataset was captured on a ChAruco board using all three trajectories to ensure a good calibration. The intrinsic calibration of the camera was obtained using OpenCV, with a reprojection error of 0.88 pixels. The robot was then hand-eye calibrated using a dual quaternion method [28] to obtain the transform from the end-effector of the robotic arm to the principal point of the camera.

IV. Data Acquisition

For this work, we used a total of eight *ex-vivo* ovine organs: six kidneys and two livers. RGB frames from each organ set can be seen in Figure 3.a. Four of the kidneys were grouped together to give a total of six organ sets. Each organ was placed on a tray with a featured background to aid the SfM pipeline. Three different trajectories were used to image each organ set as visualised in Figure 2:

1) Trocar (TC): the RCM was placed $\sim$120 mm above the centre of the sample. This leaves the tip $\sim$80 mm above the sample, creating a realistic MIS scenario.
2) Open-Close (OC): the RCM is placed in the sample with d_{lap} = 80 mm.
3) Open-Far (OF): the RCM is placed in the sample with d_{lap} = 120 mm for a wider FoV.

The TC and OC trajectories took an average time of 126s and the OF trajectory took 140s due to its increased travel distance with the robot set to 10% of its maximum velocity for safety. The organs were imaged with the two different light source configurations: LapLS and SurgLS seen in Figure 2. For the kidney datasets, each of the trajectories was performed with both light source configurations, leading to six videos per organ set and twenty-four total videos. For the liver sample sets only the TC and OF trajectory were performed, because the livers were larger the OC trajectory was skipped to minimise the risk of collisions. This led to four videos per liver, eight in total and an overall dataset size of thirty-two videos. There were an average of 909 frames per video, each with pose information and RGB images. The average number of frames per trajectory was 886, 871 and 958 for the TC, OC and OF trajectories, respectively.

In tackling issues with GT as detailed in our related work, we implement a laser scanner to acquire highly accurate 3D meshes and point clouds from our samples. We use the Nikon H120 ModelMaker and MCAx S (Nikon Corporation, Tokyo, Japan) for this approach. The laser scanning system has an accuracy of down to 7μm [4] giving us the ability for to precisely evaluate our methods. This accuracy is greater than both structured light and the Computed Tomography (CT) GTs supplied by current datasets as structured light/stereo depth estimation is limited by the resolution of the cameras and feature matching algorithms and CT scanners have larger voxel sizes and requires a segmentation step for a 3D model for comparison that could introduce errors. A full 3D reconstruction GT allows evaluation of a whole pipeline in comparison to solely depth maps, as projection from depth maps can introduce errors. Depth maps can be extracted from a GT point cloud provided camera poses and intrinsics, which does, however, rely on good registration from laser scanner to point cloud.

V. 3D Reconstruction

This section details our pipeline for 3D reconstructions from our robotic arm pipeline as seen in Figure 3. To perform the 3D reconstruction, we utilised two image feature extraction and matching methods, GIM-LG and ALIKED-LG. GIM-LG uses a handcrafted feature extractor and a retrained matching component, LG, following the GIM training framework. ALIKED is a lightweight CNN-based keypoint detector and descriptor extractor that we paired with the LG matching component. These models were chosen due to their performance in [21], particularly in out-of-domain challenges. For comparison, we also paired LG with SIFT [20], a commonly used feature detector. All the data was processed on a workstation with an AMD Ryzen Threadripper PRO 5975WX CPU and two NVIDIA RTX A6000 GPUs.

1) Structure from Motion: The images are undistorted using calibration data, and each video is sub-sampled to 100 frames to reduce computational complexity due to high frame correlation. To generate a list of image pairs, we use the method from [21] by pairing sequential frames using a sliding window. We enhance the matching process by using a pre-trained DINOv2-SALAD [29] model to generate extra pairs. DINOv2 is employed for local feature extraction, while SALAD aggregates these local features into clusters and uses the optimal transport method to generate global features. Following this approach resulted in an average of 1169 pairs per dataset.

These pairs are processed by the external feature matchers (ALIKED-LG, GIM-LG and SIFT-LG), and the resulting

[4]https://industry.nikon.com/en-us/products/3d-laser-scanners/manual-3d-scanning/modelmaker-h120/

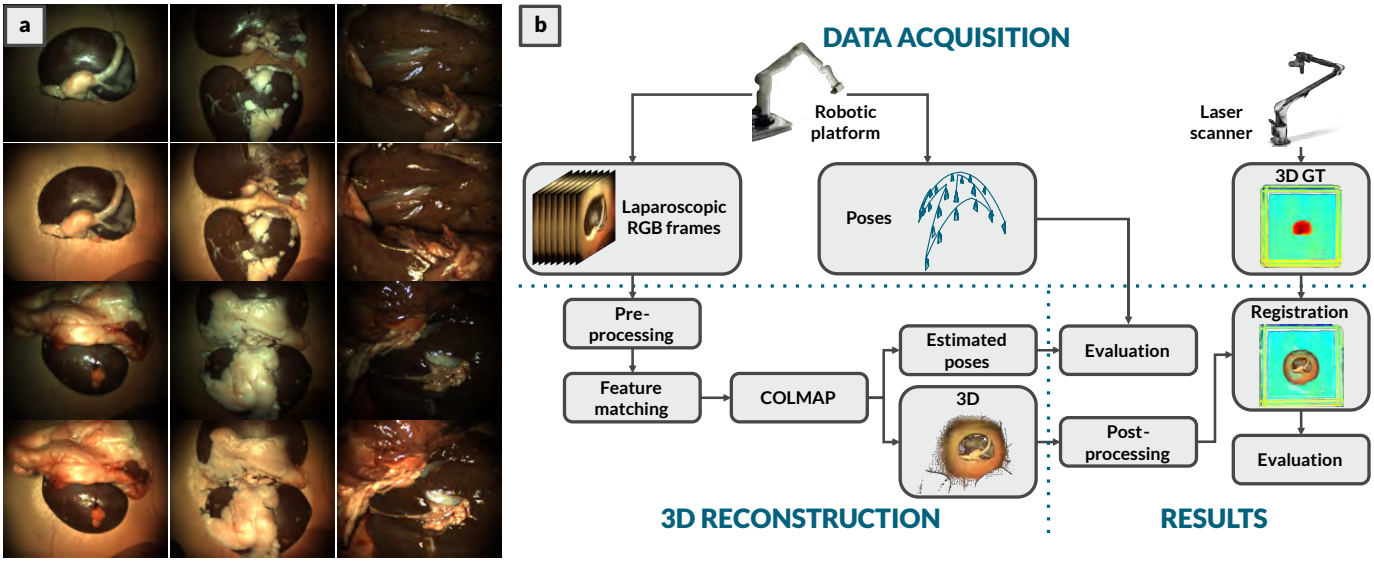

Fig. 3. **a)** RGB frames from each of the 6 organ sample sets under the different lighting scenarios. Columns 1&2 show different kidney samples. In column 3, we show a section of the two liver samples. Rows 1&3 show the organs under the LapLS and conversely, rows 2&4 display the SurgLS. **b)** Data acquisition: A summary of all the data collected by our platform. 3D Reconstruction: Our 3D reconstruction pipeline for processing our data. Results: Evaluation of the 3D reconstructions by comparing to acquired GT data.

matches are imported into the COLMAP database[5]. COLMAP is then used for geometric verification and triangulation via RANSAC to produce sparse point clouds and camera pose estimations. Dense 3D reconstruction is achieved in a subsequent step using COLMAP's Multi-View Stereo functionality. The hand-eye calibrated poses were not used in the pipeline as we use them to evaluate the predicted poses.

2) Post-processing: Prior to comparison, the dense point clouds undergo a post-processing step. Due to our pipeline returning relative point clouds, the point clouds are manually registered to the laser scan GT using Open3D [30] by manually selecting matching points in both point clouds before undergoing iterative closest point (ICP) registration with the laser scan to refine the registration.

Manually registered point clouds are downsampled to 0.5 mm and cleaned from statistical outliers using Open3D. The mean distance to its k-nearest neighbours is calculated using a KD-Tree for efficient neighbour queries for every point. Subsequently, the global mean and standard deviation are computed for all points. Points are considered outliers if their mean neighbour displacements exceed a threshold defined as the population mean plus a multiple of the standard deviation. This method removes isolated noise while preserving the point cloud structure, making it suitable for denoising in 3D reconstruction workflows. For this, we set `k = 20` and `std_ratio = 1.0` to maintain data quality and then points more than 60 mm from the centroid are removed. Finally, we use Open3D's Point-to-Plane ICP registration [31] to register our reconstruction with the GT with the default Tukey Loss with $k = 1$. This uses the normals of the target scan, obtained directly from the laser scanner, in its objective function to increase the fidelity of the ICP algorithm.

[5]https://github.com/surgical-vision/colmap-match-converter

VI. RESULTS AND DISCUSSION

A. Qualitative Results

In this section, we present our qualitative results and findings. In Figure 4, two representative cases are shown for a liver and a kidney. The first case involves a kidney acquired using the OF trajectory with SurgLS, processed using the ALIKED-LG matcher. The second case involves a liver acquired using the TC trajectory with LapLS, processed using the GIM-LG matcher. Both methods resulted in visually accurate 3D reconstructions despite the differences in acquisition trajectories and lighting conditions.

While the overall reconstructions appear good, the presence of minor gaps does not significantly affect how the 3D models can be used, but they may slightly reduce the visual quality. For improved visualisation, we used Poisson meshes, which helped fill in some of the gaps and gave us cleaner-looking models. By observing the location of these gaps, it appears that one of the main challenges in the 3D pipeline is dealing with the dark, smooth surfaces of the organ and specular highlights. These surfaces, especially under laparoscopic lighting, made it harder for the algorithms to detect enough matching features, particularly when we used the limited perspectives of the TC trajectory. With fewer camera angles, there were more instances of occlusions and missing data compared to the wider OC and OF trajectories, where the broader FoV provided a clearer picture. Despite these difficulties with unreconstructed surfaces, the registrations between our predicted models and the GT laser scans appear to be visually well aligned with only minor discrepancies as seen in the error maps in Figure 4.

B. 3D Reconstruction Results

1) Metrics: In evaluating the performance of point cloud registration to a GT laser scan, we employed three widely used distance metrics in millimetres: Chamfer distance, Hausdorff

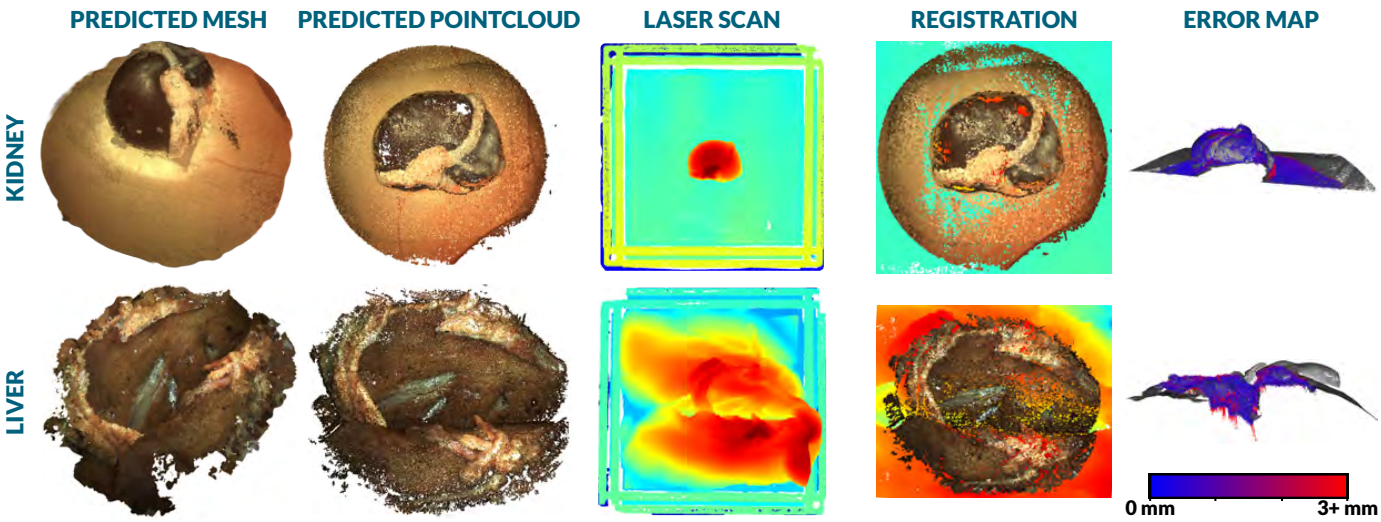

Fig. 4. 3D reconstructions of a kidney and liver obtained using different trajectories and lighting conditions. For the kidney (row 1), data was captured using the OF trajectory under LapLS and processed with the ALIKED-LG image matcher. The liver (row 2) was captured using the TC trajectory under sls and processed using the GIM-LG matcher. In both cases, the predicted 3D models (col 1&2) are compared to the GT laser scans (col 3) and the post-processed, aligned reconstructions (col 4). Error maps (col 5) display a cross-sectional view of the registration where the GT is coloured grey and each reconstruction point is coloured according to its individual error

distance, and Root Mean Squared Error (RMSE). Each metric provides a different aspect of the registration quality: Chamfer distance measures average minimum distances between points, Hausdorff distance captures maximum deviations, and RMSE reflects average point-wise errors. Together, they serve as an comprehensive view of how closely our reconstructions matched the GT. Due to incompleteness in some laser scans, we excluded the top 5% of distances from the metric calculations. This step prevents artificially inflated errors from points in the source cloud that are not covered by the target cloud.

Table II presents the quantitative evaluation of our point cloud registrations against the GT laser scans. It compares the performance of different methods across the two lighting conditions averaged across all trajectories for both kidney and liver datasets and conversely for the trajectories averaged across the light source conditions.

2) Anomalies: Of 96 total reconstructions (from 32 sequences and 3 feature matchers), 12 failed and were excluded from metrics. ALIKED-LG and GIM-LG each had one failure under MIS conditions with kidneys, while SIFT-LG accounted for the remaining 10 failures due to insufficient feature matching resulting in fewer than 50 points in the final pointclouds.SIFT-LG failures were distributed evenly across both lighting conditions and trajectories, indicating fundamental limitations in medical imaging rather than scenario-specific issues. Furthermore, in liver reconstructions, where SIFT-LG failed in 50% of attempts $(4/8)$, while ALIKED-LG and GIM-LG maintained perfect success rates, demonstrating the superior ability of modern feature matching methods in this domain.

3) Differences in Organ Reconstructions: Overall, the reconstruction results for the different organs and methods show consistent findings as shown in table II, with ALIKED-LG and GIM-LG consistently outperforming the SIFT-LG baseline method. Only for the Kidney dataset under the SurgLS the SIFT-LG shows better Chamfer results (-0.112) and RMSE (-0.166) compared to the ALIKED-LG method. The sharper transitions and irregularities at the kidney edges often make accurate feature matching more complex, and ALIKED-LG seems to struggle in these scenarios. The ALIKED-LG model outperforms GIM-LG and SIFT-LG by a large margin on the Liver dataset.

4) Lighting Conditions: Table II highlights the influence of lighting conditions on registration metrics.SurgLS, offering broader and more even illumination, generally produced better results, particularly with lower Chamfer, Hausdorff, and RMSE values. In contrast, LapLS, more focused and directional, led to slightly higher errors, especially in the Hausdorff distance, likely due to the introduction of shadows, uneven illumination, and pronounced vignetting. Despite these differences, both GIM-LG and ALIKED-LG exhibited strong performance across lighting conditions, with GIM-LG performing slightly better overall, particularly in the Chamfer (2.017 and 2.922) and RMSE (0.764 and 1.029) metrics on the kidney dataset. These results confirm that both methods are effective for 3D reconstruction in surgical environments, with GIM-LG showing especially consistent performance, without outliers, even in challenging lighting conditions.

5) Trajectories: The last three columns in Table II present the average metrics for each trajectory. Due to the larger size of the livers, no OC trajectory was recorded to avoid collisions with the organ. The metric differences between the OC and OF datasets highlight that the choice of d_{lap} significantly impacts reconstruction quality, emphasizing the need for careful parameter selection. Among the trajectories, the TC trajectory, the most realistic, performed the worst, likely due to the limited number of views.

6) Minimally Invasive Surgery Conditions: We compared optimal conditions (OC trajectory for kidneys and OF trajectory for livers, both with SurgLS) against clinically represen-

TABLE II
AVERAGE AND STANDARD DEVIATION METRICS IN MM OR SUCCESSES FOR LIGHTING AND TRAJECTORIES AGAINST EACH MATCHING METHOD.

Metric	Method	Light Source		Trajectory		
		SurgLS	LapLS	TC	OC	OF
Kidneys						
Chamfer	SIFT-LG	0.760±0.346	0.983±0.519	0.974±0.633	0.787±0.264	0.855±0.419
	GIM-LG	0.605±0.163	0.786±0.237	0.844±0.135	0.669±0.284	0.581±0.124
	ALIKED-LG	0.872±0.573	0.869±0.244	0.815±0.211	0.801±0.320	0.981±0.631
Hausdorff	SIFT-LG	2.626±1.119	3.783±1.895	3.363±2.097	3.158±1.547	3.093±1.455
	GIM-LG	2.017±0.482	2.922±1.184	2.814±0.604	2.651±1.453	1.930±0.357
	ALIKED-LG	3.090±2.200	3.261±0.982	2.972±0.893	3.051±1.283	3.452±2.440
RMSE	SIFT-LG	0.995±0.480	1.294±0.658	1.246±0.810	1.041±0.347	1.147±0.582
	GIM-LG	0.764±0.208	1.029±0.327	1.078±0.195	0.882±0.398	0.735±0.155
	ALIKED-LG	1.161±0.843	1.146±0.341	1.059±0.313	1.058±0.443	1.320±0.930
Success	SIFT-LG	9/12	9/12	6/8	6/8	6/8
	GIM-LG	12/12	11/12	7/8	8/8	8/8
	ALIKED-LG	12/12	11/12	7/8	8/8	8/8
Livers						
Chamfer	SIFT-LG	1.571±0.451	1.735	1.990	-	0.550±0.320
	GIM-LG	0.686±0.106	1.225±1.154	0.757±0.117	-	1.172±1.182
	ALIKED-LG	0.551±0.088	0.588±0.179	0.692±0.133	-	0.483±0.037
Hausdorff	SIFT-LG	5.430±2.467	6.072	6.902	-	2.112±1.578
	GIM-LG	2.499±0.451	4.347±3.747	2.721±0.499	-	4.181±3.840
	ALIKED-LG	2.030±0.297	2.244±0.526	2.391±0.505	-	1.973±0.303
RMSE	SIFT-LG	2.068±0.648	2.284	2.621	-	0.720±0.681
	GIM-LG	0.892±0.132	1.588±1.471	0.969±0.158	-	1.530±1.502
	ALIKED-LG	0.720±0.098	0.769±0.212	0.876±0.171	-	0.653±0.069
Success	SIFT-LG	3/4	1/4	1/4	-	3/4

TABLE III
COMPARISON OF RECONSTRUCTION METRICS IN MM UNDER OPTIMAL AND MIS CONDITIONS AND DEGRADATION IN ACCURACY (DIA)

Condition	Settings	Chamfer (↓)	Hausdorff (↓)	RMSE (↓)
Kidneys				
Optimal	OC+SurgLS	0.567	1.953	0.720
MIS	TC+LapLS	0.884	3.151	1.144
DiA (%)		55.9	61.3	58.9
Livers				
Optimal	OF+SurgLS	0.564	2.121	0.745
MIS	TC+LapLS	0.722	2.559	0.920
DiA (%)		28.0	20.7	23.5

tative MIS conditions using the TC trajectory with LapLS. For this analysis, we excluded SIFT-LG results to focus on recent methods, particularly given its consistent failures, especially under either the LapLS or the TC trajectory and its performance deficit to recent methods. These results can be seen in Table III. The only failed reconstructions for the ALIKED-LG and GIM-LG were under MIS conditions. The metrics indicate significant performance degradation when moving from optimal to MIS conditions. Qualitatively, these degradations arise from a more limited set of views and less uniform lighting conditions. These results demonstrate that even though current methods have performed well in this work, they still face substantial challenges under MIS conditions, suggesting the need for further research into domain-specific improvements.

C. Pose Evaluation

We employed the Umeyama algorithm [32] to align our predicted poses with the GT for comparison. Figure 5 illustrates a subsampled set of pose estimations from a kidney dataset using an OC trajectory, with the ALIKED-LG feature

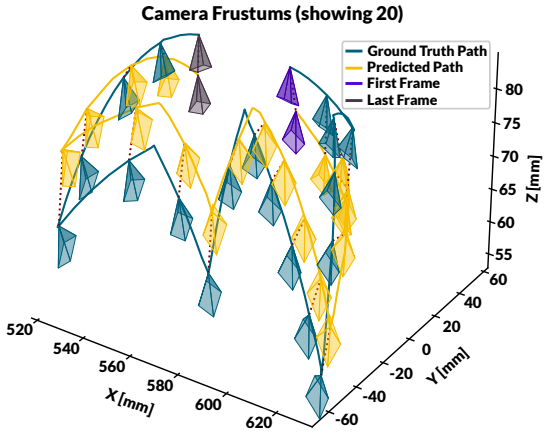

Fig. 5. A graph depicting a sub-sampled set of poses of the GT (blue) and predicted (yellow) poses for a kidney dataset with the SurgLS and the OF trajectory using the ALIKED-LG image matcher. The dotted red lines represent the error between the prediction and the GT pose.

matching version of the pipeline. Overall, the poses were accurate, with translation estimations appearing slightly better than rotations. The relative pose error (RPE) for rotation was 0.0290, 0.0032 and 0.0161 radians using SIFT-LG, GIM-LG and ALIKED-LG respectively. For translations, the RPEs were 0.285, 0.245 and 0.265 mm in the same order.

On average, 88.78% of frames were able to have predicted poses, with a standard deviation of 27.85%, which can be attributed to a few anomalous cases with fewer than 15 frames having predicted poses. Across all datasets, 81.25% successfully predicted poses for all frames, with GIM-LG slightly outperforming ALIKED-LG at 90.95% compared to 86.60%. This highlights the robustness of both methods, especially GIM-LG, in feature matching and pose prediction.

VII. CONCLUSIONS

In this work, we introduced a robotic arm platform that acquires multi-view image/video datasets and highly accurate GT laser scans for surgical 3D reconstruction research. Our results demonstrate that recent feature matching methods can achieve nearly sub-millimetre performance on *ex-vivo* organ data collected with our platform.

We show our platform is efficient stemming from our automated trajectory planning and execution which eliminates the need for manual camera positioning and ensures repeatable data collection across different trajectories and scenes. Through use of the laser scanner, the platform does not suffer from logistical issues with safety and setup that are present in other GT acquisition methods such as CT that can cause issues with efficiency both in effort and time spent.

We focused on a limited number of *ex-vivo* specimens under controlled conditions, these do not fully capture the complexity of *in-vivo* surgical environments with tissue deformation, smoke, blood, and tool occlusions. Therefore further focus on a wider variety and/or combination of organs, other connective tissue present in surgery and tools would be beneficial to address this limitation. The current pipeline also lacks real-time processing capabilities, which is crucial for clinical use but not a requirement for dataset creation. While our methods perform well on ideal data, they degrade under realistic MIS conditions, underscoring the need for more robust, surgery-specific algorithms.

Beyond direct 3D reconstruction, the collected data can serve as training material for feature detection, matching and 3D perception algorithms specifically designed for surgery. These improved methods can, in turn, be integrated into RAMIS, AR and assist surgeons with intraoperative decision-making and guiding towards more data-driven surgical interventions.

REFERENCES

[1] L. Maier-Hein, S. S. Vedula, S. Speidel, N. Navab, *et al.*, "Surgical data science for next-generation interventions," *Nature Biomedical Engineering*, vol. 1, pp. 691–696, 2017. [Online]. Available: https://doi.org/10.1038/s41551-017-0132-7

[2] S. Nicolau, L. Soler, D. Mutter, and J. Marescaux, "Augmented reality in laparoscopic surgical oncology," *Surgical Oncology*, vol. 20, no. 3, pp. 189–201, 2011, special Issue: Education for Cancer Surgeons. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960740411000521

[3] N. Chong, Y. Si, W. Zhao, Q. Zhang, *et al.*, "Virtual reality application for laparoscope in clinical surgery based on siamese network and census transformation." in *MICAD*, ser. Lecture Notes in Electrical Engineering, R. Su, Y.-D. Zhang, and H. Liu, Eds., vol. 784. Springer, 2021, pp. 59–70. [Online]. Available: http://dblp.uni-trier.de/db/conf/micad2/micad2021.html#ChongSZZYZ21

[4] S. Overley, S. Cho, A. Mehta, and P. Arnold, "Navigation and robotics in spinal surgery: Where are we now?" *Neurosurgery*, vol. 80, pp. S86–S99, 03 2017.

[5] L. Bianchi, U. Barbaresi, L. Cercenelli, B. Bortolani, *et al.*, "The impact of 3d digital reconstruction on the surgical planning of partial nephrectomy: A case-control study. still time for a novel surgical trend?" 2020.

[6] T. L. Bobrow, M. Golhar, R. Vijayan, V. S. Akshintala, *et al.*, "Colonoscopy 3d video dataset with paired depth from 2d-3d registration," *Medical Image Analysis*, p. 102956, 2023.

[7] M. Allan, J. Mcleod, C. Wang, J. C. Rosenthal, *et al.*, "Stereo correspondence and reconstruction of endoscopic data challenge," *arXiv preprint arXiv:2101.01133*, 2021.

[8] P. E. Edwards, D. Psychogyios, S. Speidel, L. Maier-Hein, *et al.*, "Serv-ct: A disparity dataset from cone-beam ct for validation of endoscopic 3d reconstruction," *Medical image analysis*, vol. 76, p. 102302, 2022.

[9] M. Hayoz, C. Hahne, M. Gallardo, D. Candinas, *et al.*, "Learning how to robustly estimate camera pose in endoscopic videos," *International journal of computer assisted radiology and surgery*, vol. 18, no. 7, pp. 1185–1192, 2023.

[10] X. Zhao, X. Wu, W. Chen, P. C. Y. Chen, *et al.*, "Aliked: A lighter keypoint and descriptor extraction network via deformable transformation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–16, 2023.

[11] X. Shen, Z. Cai, W. Yin, M. Müller, *et al.*, "Gim: Learning generalizable image matcher from internet videos," *arXiv preprint arXiv:2402.11095*, 2024.

[12] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "Lightglue: Local feature matching at light speed," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 17 581–17 592.

[13] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[14] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm, "Pixelwise view selection for unstructured multi-view stereo," in *European Conference on Computer Vision (ECCV)*, 2016.

[15] R. D. Howe and Y. Matsuoka, "Robotics for surgery," *Annual review of biomedical engineering*, vol. 1, no. 1, pp. 211–240, 1999.

[16] V. Vitiello, S.-L. Lee, T. P. Cundy, and G.-Z. Yang, "Emerging robotic platforms for minimally invasive surgery," *IEEE reviews in biomedical engineering*, vol. 6, pp. 111–126, 2012.

[17] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, *et al.*, "Autonomy in surgical robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 651–679, 2021.

[18] L. Maier-Hein, A. Groch, A. Bartoli, S. Bodenstedt, *et al.*, "Comparative validation of single-shot optical techniques for laparoscopic 3-d surface reconstruction," *IEEE transactions on medical imaging*, vol. 33, no. 10, pp. 1913–1930, 2014.

[19] M. Xu, Z. Guo, A. Wang, L. Bai, *et al.*, "A review of 3d reconstruction techniques for deformable tissues in robotic surgery," *arXiv preprint arXiv:2408.04426*, 2024.

[20] G. Lowe, "Sift-the scale invariant feature transform," *Int. J*, vol. 2, no. 91-110, p. 2, 2004.

[21] S. Bonilla, C. Di Vece, R. Daher, X. Ju, *et al.*, "Mismatched: Evaluating the limits of image matching approaches and benchmarks," *arXiv preprint arXiv:2408.16445*, 2024.

[22] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, *et al.*, "Dust3r: Geometric 3d vision made easy," in *CVPR*, 2024.

[23] V. Leroy, Y. Cabon, and J. Revaud, "Grounding image matching in 3d with mast3r," 2024.

[24] D. Coleman, I. Sucan, S. Chitta, and N. Correll, "Reducing the barrier to entry of complex robotic software: a moveit! case study," *arXiv preprint arXiv:1404.3785*, 2014.

[25] I. A. Sucan, M. Moll, and L. E. Kavraki, "The open motion planning library," *IEEE Robotics & Automation Magazine*, vol. 19, no. 4, pp. 72–82, 2012.

[26] Z. Kingston, M. Moll, and L. E. Kavraki, "Exploring implicit spaces for constrained sampling-based planning," *The International Journal of Robotics Research*, vol. 38, no. 10-11, pp. 1151–1178, 2019.

[27] B. Keinert, M. Innmann, M. Sänger, and M. Stamminger, "Spherical fibonacci mapping," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 1–7, 2015.

[28] K. Daniilidis, "Hand-eye calibration using dual quaternions," *The International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.

[29] S. Izquierdo and J. Civera, "Optimal transport aggregation for visual place recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 17 658–17 668.

[30] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.

[31] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992.

[32] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 13, no. 04, pp. 376–380, 1991.