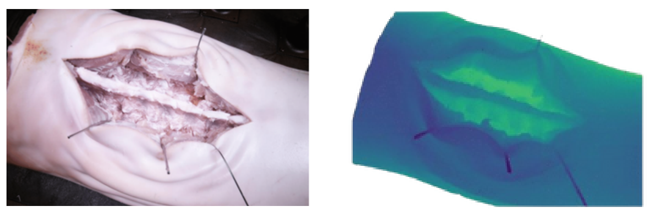
ACQUIRING SUBMILLIMETER-ACCURATE MULTI-TASK VISION DATASETS FOR COMPUTER-ASSISTED ORTHOPEDIC SURGERY

Emma Most, Jonas Hein, Frédéric Giraud, Nicola A. Cavalcanti, Lukas Zingg, Baptiste Brument, Nino Louman, Fabio Carillo, Philipp Fürnstahl, Lilian Calvet

IPCAI 2025

ABSTRACT

This work proposes a methodology to acquire **realistic** and *ex vivo* datasets with **accurate 3D ground truth** tailored 3D reconstruction and feature matching in open orthopedic surgery. We also provide a pilot dataset on an *ex vivo* pig spinal surgery.

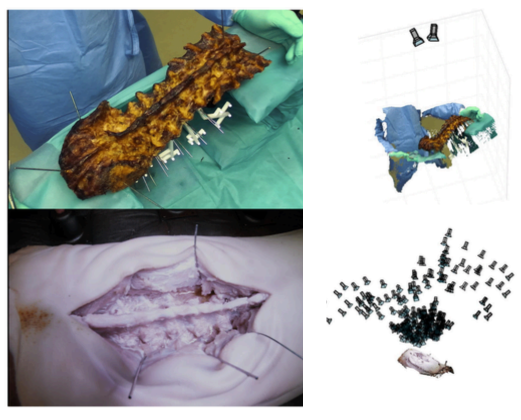


Left: A sample image in our pilot dataset. Right: The associated depth map derived from our 3D ground truth

MOTIVATION

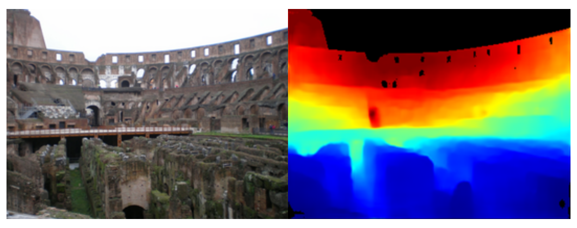
Advances in computer vision, especially in 3D reconstruction and feature matching are opening the door to various applications in computer assisted orthopedic surgery like marker-less surgical navigation, surgical digital twins and robotic assisted surgery. Evaluating and training these methods motivates the need to acquire large, accurate, and real datasets with 3D ground truth. While suitable datasets exist for indoor and outdoor scenes, there is only one dataset for open orthopedic surgery with 3D ground truth and camera poses and it presents certain limitations:

- Overly exposed and therefore unrealistic
- Limited to two viewpoints
- Reported accuracy is 1.5 mm



Top: Current state of the art (SpineDepth [1]) featuring images of dissected lumbar spines. Bottom: Ours, featuring an ex vivo pig spine surgery

Dataset sample and depth map from Omnidata (Eftekhar et al. 2021)



Dataset sample and depth map from Megadepth (Li et al. 2018)

METHOD

Our methodology aims to acquire a set of posed images and an accurately registered GT surface mesh of the anatomy. We propose a data acquisition framework consisting of three main steps:

- Anatomy surface reconstruction
- Posed image capture
- Registration.

We compare different methods for each step and acquire our pilot dataset using the optimal pipeline.

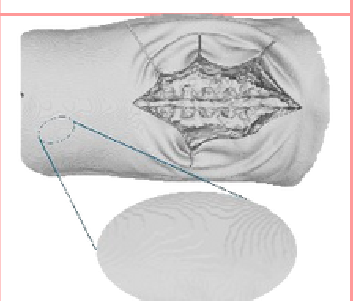
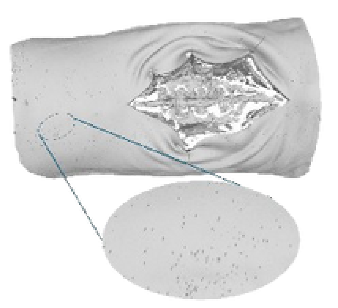
Setup

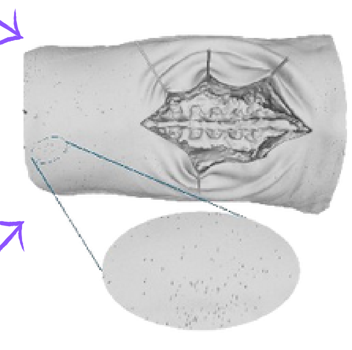
Our setup consists of a specimen fixated onto a wooden board, placed on an operating table, along with a set of 10 spherical *registration markers*. A clinician makes an incision that is kept open with k-wires. Experiments were made in a fully equipped operating room replica (https://or-x.ch/) to allow for realistic illumination conditions.

Images of the setup in the operating room, during data capture of our pilot dataset

Scene surface reconstruction

The first step consists in reconstructing the external surface of the anatomy.

Technique	Pros and Cons	Qualitative Results
CT scanner 	+ Complete reconstruction, including concave regions − Transportation of the anatomy from OR to CT can lead to noisy data	
Handheld optical scanner 	+ Excellent reconstruction of high frequency details (up to 0.05 mm) − Fails to fully reconstruct deep concave regions	



Fusion of the CT scan and optical scan. We manually segment the concave regions in the CT scan and the rest in the optical scan and combine the two co-registered models.

Posed image capture

The second step consists in capturing posed images.

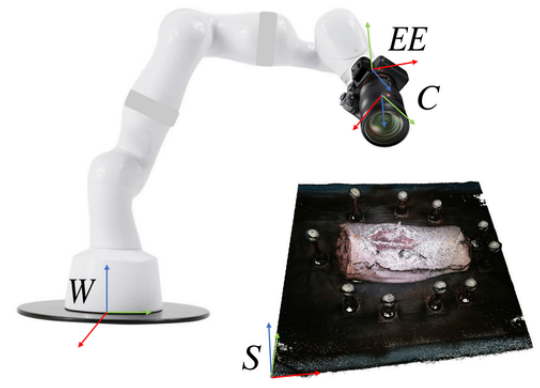
Question 1: How should the images be captured?

Manual Image Capture	• Requires minimal hardware • Versatile
Capture with Robotic Arm	• Automated • Delivers poses

→ Chosen for scalability

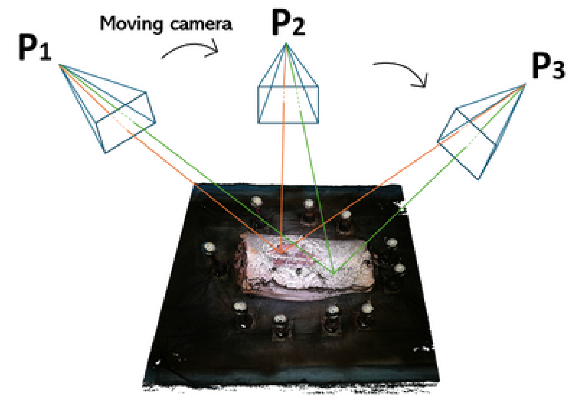
Question 2: How should the poses be estimated?

Option 1:
Using robot forward kinematics



- Hand-eye calibration
- Use robot forward kinematics to obtain camera poses expressed in the robot base frame

Option 2:
Using image-based Structure-from-Motion (SfM)

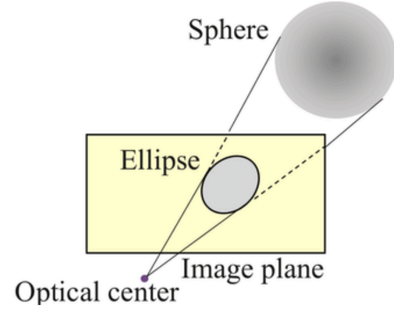


- Input all 216 captured images into SfM method (e.g. COLMAP)
- Use output poses as GT poses for our dataset
- High-resolution images lead to the highest pose accuracy out of both methods

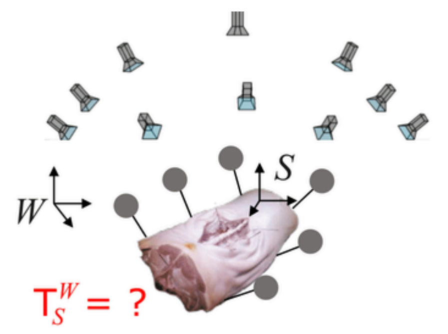
→ Chosen method

Scene registration

The third step of our method consists in expressing the image poses and the 3D mesh in the same reference frame.



- Spheres project into the image plane as ellipses
- Ellipses are segmented in the images
- Spheres are segmented in the scanned 3D mesh
- Least squares optimization minimizing the distance between the extracted ellipses and the imaged spheres



Let P be the camera projection matrix. A sphere can be represented as a quadric Q and projects into the image as an ellipse E of equation:
$E^{-1} = PQ^{-1}P^T$

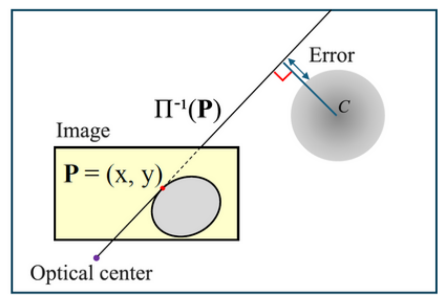
EVALUATION

The methodology was evaluated on a simulated *ex vivo* scoliosis spine surgery using a pig spine.

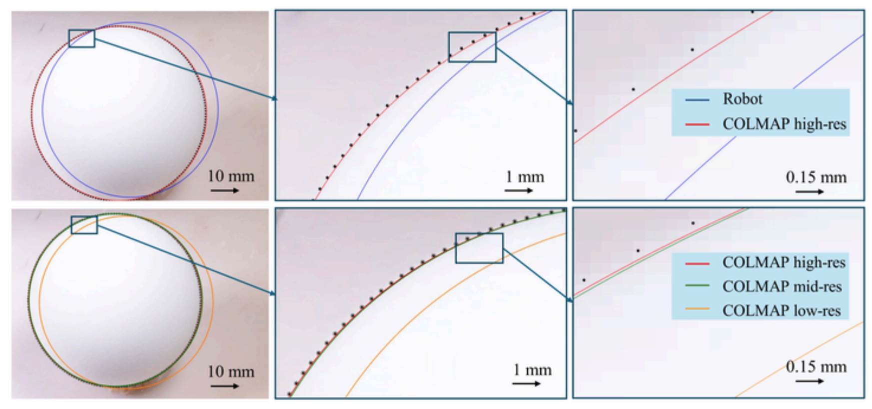
To evaluate the accuracy of the method:

- Spherical *control markers* were placed on top of the specimen and inside the incision, distinct from the 10 registration markers around the scene
- N=16 images of the evaluation setup were captured from different viewpoints
- The evaluation scene was scanned with the optical scanner and 3D marker locations were extracted
- In each captured image and for each control marker, points were sampled on the corresponding ellipse outline
- We defined the *radial error* as the distance between a back-projected ray from a point on the ellipse and the corresponding 3D sphere.
- The *reprojection error* in pixels was calculated between the ellipse obtained by projecting the sphere into the image with the estimated scene registration solution and the detected ellipse (point-to-ellipse distance from Rosin, 1998).

Evaluation setup



Radial error



Top: Comparison of reprojection errors obtained when using robot poses vs COLMAP. Bottom: Comparison of reprojection errors for different input image resolutions to COLMAP

	Robot	COLMAP	GLOMAP
Mean radial error (mm)	0.89	**0.35**	0.44
Mean reprojection error (px)	8.99	**3.71**	4.67

Quantitative radial and reprojection errors for robot poses vs COLMAP and GLOMAP, two SfM methods

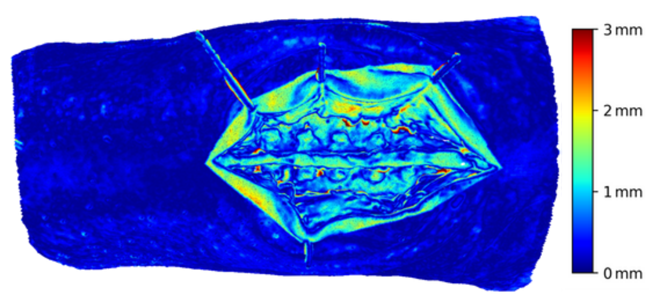
Both radial and reprojection errors errors are lower when using SfM poses rather than robot poses. This can be explained by the very high image resolution used in the SfM (9504 x 6336 px).

Our dataset as benchmark for 3D reconstruction

We demonstrate the use of our pilot dataset as a benchmark to compare different 3D reconstruction methods from posed images. We evaluated four methods that reconstruct the surface of a scene from posed RGB data in both sparse (N=8) and dense (N=216) settings. Four methods were evaluated against our GT mesh:

- COLMAP (traditional multi-view stereo)
- Neus2 (NeRF-based)
- Instant-NGP (NeRF-based)
- SuGaR (Gaussian splatting-based)

Surface reconstruction error in chamfer distance (mm)			
COLMAP	Instant-NGP	NeuS2	SuGaR
Dense: 0.68 Sparse: -	Dense: 2.31 **Sparse: 3.89**	Dense: 1.23 Sparse: 4.27	Dense: 2.99 Sparse: 5.35



Heatmap showing the Chamfer distance between the model reconstructed with COLMAP and our 3D ground truth

PERSPECTIVES

Our method to acquire highly accurate realistic surgical datasets will be used to acquire large scale datasets to train and evaluate computer vision tasks in the surgical field. The proposed methodology will be extended to RGB-D.

References:

- Liebmann, F. et al. SpineDepth: Multi-modal data collection for automatic labelling and intraoperative spinal shape reconstruction. J. Imaging, 7, 164 (2021). doi:10.3390/jimaging7090164
- Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.-M. COLMAP: Pixelwise view selection for unstructured multi-view stereo. ECCV (2016).
- Wang, Y. et al. NeuS2: Fast learning of neural implicit surfaces for multi-view reconstruction. ICCV (2023).
- Müller, T., Evans, A., Schied, C., Keller, A. Instant-NGP: Neural graphics primitives with multiresolution hash encoding. ACM Trans. Graph., 41(4), 102 (2022). doi:10.1145/3528223.3530127
- Guédon, A., Lepetit, V. SuGaR: Surface-aligned Gaussian splatting for efficient 3D mesh reconstruction. CVPR (2024).

ETH zürich Balgrist Universitätsklinik OR-X IRIT Institut de Recherche en Informatique de Toulouse