

Bridging the Markerless Tracking Validation Gap: Intraoperative X-ray Workflow Integration

Connor Daly^{1(✉)}, Salvatore Russo², Jinendra Ekanayake³, Daniel S. Elson¹,
and Ferdinando Rodriguez y Baena¹

¹ The Hamlyn Centre for Robotic Surgery, Imperial College London, London, UK,
cd1723@ic.ac.uk

² Department of Neurosurgery, Imperial College Healthcare NHS Trust, London, UK

³ Stanford University School of Medicine, Stanford University, California, USA

Abstract. Markerless intraoperative tracking of spinal anatomy via RGB-D sensors is a promising alternative to invasive bone-anchored fiducials, yet no existing clinical dataset provides independent subsurface ground truth for validating these methods. We address this gap with a clinically viable pipeline that derives per-vertebra 6-DoF pose from routine intraoperative C-arm fluoroscopy, co-registered to the RGB-D camera frame through a multimodal fiducial array detectable in both imaging modalities. Critically, the pipeline requires no invasive instrumentation contacting the patient, no additional radiation, or deviation from standard surgical workflows, relying solely on the fluoroscopic images already acquired during routine level checks. A calibrated transform chain links the vertebral pose recovered via multi-view 2D–3D CT-to-fluoroscopy registration to the RGB-D coordinate frame, enabling direct comparison of any markerless registration estimate against a subsurface reference. Ex vivo validation against optically tracked ovine vertebrae confirms end-to-end accuracy of 2.06 ± 0.29 mm and $3.35 \pm 0.93^\circ$, sufficient to serve as independent ground truth for methods with substantially larger errors. We demonstrate the pipeline’s utility by evaluating two markerless registration strategies on clinical data spanning 11 vertebrae across 1,500 frames from three surgical sequences: the first such evaluation against an independent subsurface reference in a clinical setting. The proposed framework provides reusable infrastructure for benchmarking emerging markerless tracking methods on real surgical data.

Keywords: RGB-D Tissue Tracking · Orthopedic Surgery · 2D–3D Registration

1 Introduction and Motivation

Markerless intraoperative tracking of orthopedic anatomy via RGB-D sensors is a promising emerging direction for surgical guidance [10,14]. Current surgical navigation systems depend on invasive bone-anchored fiducials to register (and subsequently track) preoperative imaging with the surgical scene [25,26], introducing added operative time and complication risk [7]. Markerless approaches

instead extract anatomically meaningful correspondences directly from intraoperative point clouds via learned segmentation, enabling registration and tracking without physical instrumentation [10,14]. Daly *et al.* [4] advanced this direction with the first clinical markerless tracking dataset for spine surgery. However, their ground truth is inherently approximate: alignment relies on surgeon-indicated surface landmarks refined against intraoperative topology, neither of which references the pose of the subsurface anatomy. For clinical adoption, markerless systems must be validated against a gold-standard measurement offering quantifiable translational and rotational error metrics [21,15], a requirement no existing clinical dataset satisfies. Cadaveric studies have obtained accurate vertebral pose through rigidly fixed optical tracking arrays [16,15]. Liebmman *et al.* [15], for example, fixed 3D-printed constructs to vertebral bodies via K-wires, achieving sub-2 mm accuracy but requiring anterior exposure incompatible with standard posterior approaches, additional tracking hardware within the operative field, and an intraoperative CT. These requirements are not permissible within a clinical observational study, leaving a methodological gap: there is currently no clinically viable means of obtaining precise vertebral pose ground truth for evaluating markerless registration in real surgical settings. To address this, we present a pipeline that leverages routine intraoperative C-arm imaging to derive ground-truth vertebral pose within the RGB-D camera frame. Our contributions include:

1. A multimodal data capture and labelling pipeline that co-registers intraoperative RGB-D recordings with ground-truth vertebral poses derived from routine intraoperative fluoroscopy, requiring no additional radiation, no invasive instrumentation contacting the patient, or deviation from standard surgical workflows.
2. The first quantitative evaluation of markerless spine registration methods on clinical data against an independent subsurface reference, reporting per-vertebra translational and rotational error across three surgical sequences.
3. A clinically grounded evaluation framework, validated *ex vivo* to 2.06 ± 0.29 mm and $3.35 \pm 0.93^\circ$, which can serve as reusable infrastructure for benchmarking markerless registration methods against an independent subsurface reference. CAD files for the intraoperative marker array and calibration phantom, along with the evaluation code and annotated data can be found in the accompanying repository BMTVG.

This work reports initial results from a large-scale ongoing clinical data collection study; the pipeline and evaluation framework are designed to scale as further cases are captured, enabling progressively larger-scale benchmarking of markerless methods.

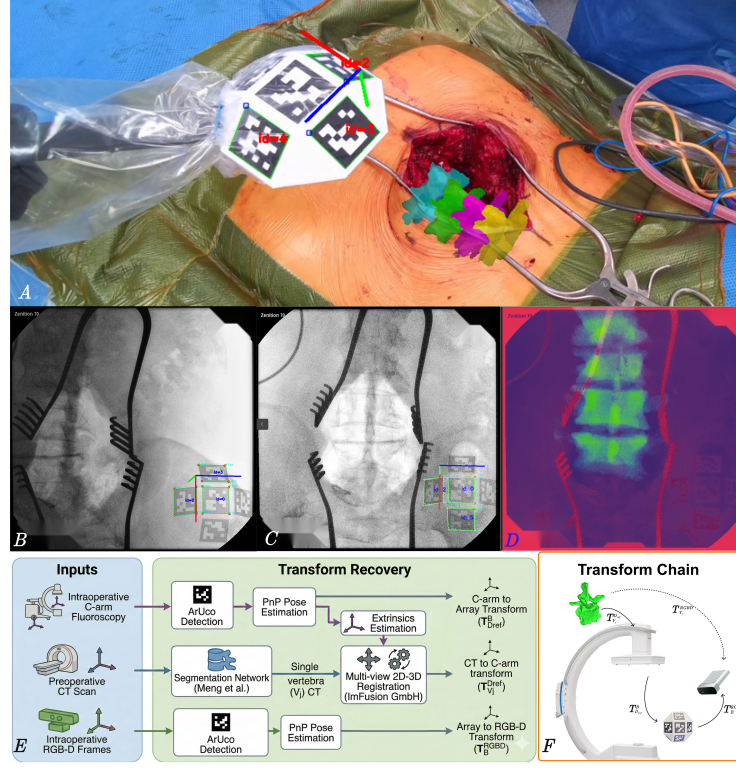


Fig. 1: (A) Intraoperative setup with the marker in a sterile cover, secured taut with a rubber band by the scrub nurse to maintain detectability; overlay shows automatic pose detection and projection of the preoperative CT into the scene via the transform chain in (F). (B, C) Automatic marker pose detection in C-arm images from a fixation case, with projection errors (pixels) overlaid. (D) AP DRR: CT registered to C-arm images in ImFusion. (E) Data-processing overview: acquisition of the transforms that project preoperative CT into the RGB-D stream. (F) Solid lines: our full transform chain projecting vertebra j 's CT into the Azure Kinect RGB-D frame.

2 Methodology

2.1 Data Capture

In contrast to the use of coarse surface features to approximate the underlying pose of the spine by Daly *et al.* [4], we integrated routine intraoperative C-arm fluoroscopy, taken alongside the RGB-D recording, to locate the patient's preoperative anatomy within the intraoperative scene. C-arm fluoroscopy is broadly accessible across clinical settings [2,18], ensuring our pipeline generalizes across institutions without requiring intraoperative CT [13]. We used established multi-view 2D–3D registration algorithms, as implemented in the ImFusion software

suite (ImFusion GmbH, Germany) [11], requiring at least two fluoroscopic images from separate poses with known projection geometry. We modeled each C-arm source as a pinhole camera [22,9,17] and calibrated intrinsics using a custom multi-plane phantom [3,5] with nine radio-opaque ArUco markers [8] at known depths, estimating parameters via bundle adjustment [1,22]. Calibration images were acquired prior to surgery from AP (Anterior Posterior) and oblique angles within $\pm 30^\circ$, reflecting the range of intraoperative C-arm positioning.

2.2 Intraoperative Capture

Ethics approval was given by the Research Ethics Committee of the Solihull Health Research Authority, Solihull, UK (IRAS 341461). Informed consent was obtained from all participants as per the study protocol. The cases we captured within our surgical study were taken from the clinical caseload of a tertiary neurosurgery department. The captured procedures are all lumbar-thoracic fixation cases. For recording RGB-D sequences, we used an Azure Kinect camera (Microsoft Corporation, USA), positioned directly over the incision using a boom arm fixed to a tripod. To co-register fluoroscopic and RGB-D coordinate frames without external tracking hardware, we designed a multimodal fiducial array, automatically detectable in both imaging modalities. Five radiopaque ArUco markers, each 30×30 mm, are inset on a radio-translucent PETG body. The insets were 3D-printed using PLA XRS filament (Nanovia Tech, France). The non-coplanar array layout ensured no marker overlap occurs at either AP or oblique viewing angles. The array was held on a boom arm adjacent to the exposure, beneath the C-arm detector and within the RGB-D camera field of view, and covered with an ultrasound probe cover to maintain the sterile field. An intraoperative C-arm is routinely used to provide confirmation of the vertebra level being operated on, typically from an AP view with adjacent oblique views. As the marker array was adjacent to the exposure, it is visible in the X-ray field of view during these level check shots, allowing both the determination of relative C-arm pose and the co-registration of X-ray and RGB-D coordinate frames to be performed with the same array. For each clinical case, we also collected the accompanying preoperative CT scan. We segmented the relevant bone tissue from CT using the network developed by Meng *et al.* [20].

2.3 CT Imaging to RGB-D Registration

For each C-arm shot, using the acquired intrinsic parameters from our multi-view calibration, the radiopaque ArUco markers are automatically detected within the image, and we then recovered the board-to-detector transform via PnP pose estimation [6,12]. Designating the first acquired image as the reference frame, we computed relative transforms between shots through the common marker array coordinate system, and combined these with the previously calibrated intrinsic parameters to yield projection matrices suitable for multi-view 2D–3D registration. Multi-view 2D–3D X-ray to CT registration is a well-established technique capable of achieving sub-millimeter accuracy [19]. We then performed

Table 1: Marker-array tracking accuracy per modality, isolated from the 2D–3D registration stage. The array was moved between four fixed positions; per transition, each modality’s 6-DoF pose change is differenced against an NDI Polaris optical tracker (the reference), giving a residual against the tracker, not a modality-to-modality comparison. RGB-D uses the sterile probe cover. Transitions 1→2, 1→3 are in-plane; 1→4 is along the optical axis.

Modality	1 → 2		1 → 3		1 → 4	
	Trans. (mm)	Rot. (°)	Trans. (mm)	Rot. (°)	Trans. (mm)	Rot. (°)
RGB-D tracking	0.40	0.31	0.70	0.61	0.10	1.85
X-ray tracking (AP)	0.58	0.03	1.19	0.07	0.02	0.35
X-ray tracking (oblique)	0.46	0.18	0.96	0.50	0.66	0.38

intensity-based 2D–3D registration, optimizing over six rigid-body parameters by maximizing mutual information between digitally reconstructed radiographs (DRRs) and the acquired fluoroscopic images, using the ImFusion software suite. This process was repeated for each segmented vertebra $\mathbf{V}_j$, yielding the transform from $\mathbf{V}_j$ to the detector coordinate frame, $\mathbf{D}_{ref}$, at the reference position. The transforms $\mathbf{T}_B^{\mathbf{D}_{ref}}$ and $\mathbf{T}_B^{\mathbf{RGBD}}$, locating the marker array in the reference detector frame and the RGB-D camera frame respectively, were both recovered via ArUco detection and PnP pose estimation.

Thus, we obtained the complete transform chain from the coordinate system of the vertebra $\mathbf{V}_j$ to the camera $\mathbf{C}$:

$$\mathbf{T}_{\mathbf{V}_j}^{\mathbf{RGBD}} = \mathbf{T}_B^{\mathbf{RGBD}} \cdot (\mathbf{T}_B^{\mathbf{D}_{ref}})^{-1} \cdot \mathbf{T}_{\mathbf{V}_j}^{\mathbf{D}_{ref}}$$

3 Evaluation and Results

3.1 *Ex Vivo* Validation

To demonstrate the accuracy and repeatability of our data collection and labeling pipeline, we conducted several validation experiments. We first established the level of error introduced by each individual component of the transform chain. To do this, we first rigidly fixed a passive infrared tracking array to the multi-modal marker array. We then directly compared how well the marker array itself could be tracked in both the C-arm imaging, as well as the Azure Kinect RGB frames, with respect to a highly accurate optical tracker (Polaris Vicra, Northern Digital Inc, Canada; 0.25 mm RMS volumetric accuracy, 0.50 mm 95% CI) [23]. For each imaging system, the marker array was moved between 4 distinct points in front of the sensor. We present the results in Tab. 1. Subsequently, to validate the end-to-end accuracy of the full transform chain, we compared the difference in pose of four excised ovine vertebrae determined by our pipeline, with that found by directly tracking ArUco markers, rigidly fixed to the vertebrae. We mounted two plastic ArUco markers to each of the vertebrae with

Table 2: End-to-end accuracy of the full transform chain vs. the fiducial-tracked pose of each excised ovine vertebra (the reference). Unlike Tab. 1 (marker tracking only), these include the 2D–3D fluoroscopy-to-CT registration stage; the increase over Tab. 1 (≈ 0.9 mm, $\approx 1.5^\circ$) characterizes that stage’s contribution. Combined TRE is the mean over both landmarks.

Vertebra	Combined TRE (mm)	Rotational Error ($^\circ$)
T1	1.96	4.06
T2	2.32	3.40
T3	1.62	3.94
T4	2.32	1.98
Agg.	2.06 ± 0.29	3.35 ± 0.93

plastic screws, minimizing the markers’ visibility within the C-arm imaging. We then performed CT scans of each vertebra. Within the coordinate space of the generated scans, we then localized the ArUco positions, meaning then that each vertebra could be tracked directly via the RGB-D camera, by-passing the transform chain shown in Fig. 1 (F) entirely. To compute registration differences, we selected two points (the tip of the spinous process, and a point on the superior laminar margin) for establishing a Total Registration Error (TRE), with the optically tracked pose representing the target. The vertebrae and radio-opaque marker array were positioned between the source and detector of a TAU 2020 C-arm (Ziehm-Orthoscan Inc, USA), as described in section 2.2. With the radio-opaque marker array visible from the RGB-D camera, we captured 2 C-arm shots with an angular disparity of 30° . This is comparable to the mean angular disparity of 25.6° seen in clinical cases we captured. Following the process described in 2.3, we projected the vertebrae from the CT coordinate space to the RGB-D camera coordinate space. We present the TRE for the point landmarks together with the rotational error in Tab. 2. Rotational error is the geodesic angle on $SO(3)$, $\theta = \arccos(\frac{\text{tr}(R^T \hat{R}) - 1}{2})$, between the reference rotation R and the estimate $\hat{R}$, reported in degrees.

The independent validation of the RGB-D and X-ray pose tracking against an optical tracker reference confirms sub-millimeter translational agreement and sub-degree rotational agreement. The end-to-end pipeline, validated *ex vivo* against fiducial-tracked vertebral pose, yields a combined TRE of 2.06 ± 0.29 mm and rotational error of $3.35 \pm 0.93^\circ$, approaching the ≤ 2 mm and $\leq 5^\circ$ thresholds commonly cited for spinal navigation systems [21]. Critically, this level of precision is sufficient to serve as independent ground truth for evaluating markerless tracking methods, whose registration errors are substantially larger, particularly in rotation.

3.2 Clinical Data Registration

With the pipeline validated in the lab, we demonstrated its utility by evaluating two markerless registration strategies on clinical data from three surgical se-

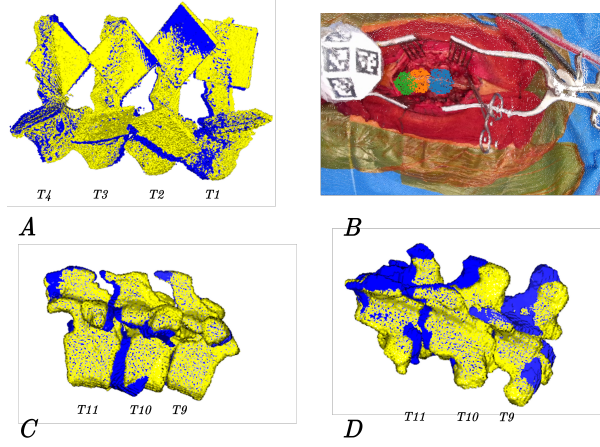


Fig. 2: **(A)** The pose of ovine CT meshes directly tracked via the Azure Kinect in blue, compared to those tracked via fluoroscopy and projected into the Kinect camera space in yellow; **(B)** Per vertebra patches predicted by the network in Sec. 3.2; **(C, D)** Intraoperative pose estimates via X-ray in yellow, through per vertebra ICP in blue.

quences, encompassing 11 vertebrae across 1,500 frames. To our knowledge, this represents the first evaluation of markerless spine registration against an independent subsurface gold standard reference in a clinical setting. The first strategy follows [4], deforming the preoperative spine mesh via a kinematic chain model before rigidly registering it to segmented intraoperative points using ICP. This approach relies on semi-automated surface-based optimization of the kinematic chain parameters, with manual refinement, to align the deformed mesh with the intraoperative point cloud: a workflow difficult to scale to real-time intraoperative use. We therefore additionally trialed a fully automated approach inspired by [14], in which CT meshes of individual vertebrae are independently registered to per-vertebra segmentation patches. These patches were obtained via a two-stage pipeline: a PointNet++ [24] heatmap regression network, trained on manually annotated spinous process tip positions, first localizes tips within the point cloud (recall 0.95, F1 0.89, localization error 3.98 ± 1.99 mm), after which a PointNet segmentation network isolates points within 40 mm of each detection (IoU 0.71, F1 0.83). Per-vertebra segmentation labels were generated by computing mesh-to-point-cloud proximity for individual vertebral CT meshes from [4], following the same approach used therein for global spine masks. The results of both registration approaches, evaluated against our fluoroscopic pose estimates with a correspondence distance of 3 mm for both methods, are shown in Tab. 3, with per-vertebra and per-axis breakdowns in Tab. 4.

The rigid method of [4] outperforms per-vertebra ICP on rotation ($4.52^\circ \pm 1.00^\circ$ vs. $8.36^\circ \pm 3.69^\circ$) and exhibits substantially lower variance across all metrics, while mean translation errors are comparable with overlapping distribu-

Table 3: Markerless registration accuracy on clinical data: per-vertebra subsurface pose error vs. the fluoroscopically-derived vertebral pose (the reference). Two strategies are compared over three 500-frame sequences: the rigid full-spine method of [4] and per-vertebra ICP. ‘Exposure’ is the exposed-vertebra count.

Seq.	Exposure	Subsurface Pose Error			
		Per Vertebra ICP		Rigid ICP	
		Trans. (mm)	Rot. ($^{\circ}$)	Trans. (mm)	Rot. ($^{\circ}$)
1	5	5.15 ± 3.65	8.21 ± 3.74	5.70 ± 1.59	4.61 ± 1.55
2	3	3.98 ± 1.81	9.10 ± 3.68	4.66 ± 0.82	3.53 ± 0.78
3	3	4.02 ± 3.10	7.86 ± 3.63	3.38 ± 0.58	5.36 ± 0.30
Overall		4.52 ± 3.00	8.36 ± 3.69	4.78 ± 1.10	4.52 ± 1.00

tions. The per-vertebra approach, though fully automated, lacks the geometric constraints afforded by registering the full spine as a single structure, limiting its ability to resolve orientation. While both methods achieve mean translational errors below 5 mm, neither yet meets the ≤ 2 mm and $\leq 5^{\circ}$ thresholds cited for spinal navigation [21], underscoring both the value of independent pose evaluation and the headroom remaining for emerging methods.

4 Discussion

The clinical registration comparison in Tab. 3 demonstrates the core contribution of this work: for the first time, competing markerless registration strategies can be evaluated on real surgical data against an independent subsurface reference. Prior clinical evaluation was limited to surrogate surface metrics [4]; our pipeline provides per-vertebra 6-DoF pose estimates from routine intraoperative fluoroscopy without additional hardware or workflow disruption. Registration errors in Tab. 3 are measured against our fluoroscopic estimates, which carry an intrinsic uncertainty of 2.06 ± 0.29 mm and $3.35 \pm 0.93^{\circ}$ (Tab. 2); these values thus represent upper bounds on true registration error. Alas, the rotational difference between the two methods (3.84°) exceeds our reference uncertainty, confirming that the rigid method of [4] is meaningfully more accurate in rotation. Translational errors are not statistically different. Overall, the rigid approach is clearly superior, but its reliance on manual deformation tuning remains a barrier to clinical deployment, and thus an avenue for future work. Key limitations remain. The study reports three clinical sequences and ovine *ex vivo* validation; further labeled cases, drawn from a larger ongoing clinical study, will broaden the evaluation. Accuracy is also constrained by the 2D–3D registration stage, which depends on image quality, angular separation, and the number of fluoroscopic views. Clinically, classical surface registration is intrinsically limited: the posterior vertebral surface exposed during surgery is geometrically ambiguous, leaving ICP unable to robustly resolve pose, particularly rotation. This is unlikely to improve with initialization alone. With sufficient fluoroscopic ground truth, a learned mapping from intraoperative point clouds to vertebral pose

could instead leverage appearance, anatomical context, and priors beyond ICP’s scope.

Table 4: **Appendix.** Additional breakdown of the registration errors in Table 3.

(a) Per vertebra error.

Seq.	Vert.	Per Vertebra ICP		Rigid ICP [4]	
		Trans. (mm)	Rot. ($^{\circ}$)	Trans. (mm)	Rot. ($^{\circ}$)
1	T7	3.97 ± 3.88	8.69 ± 3.36	4.80 ± 2.00	3.92 ± 1.56
	T8	3.33 ± 2.54	6.41 ± 3.94	4.76 ± 1.54	2.82 ± 1.27
	T9	8.89 ± 3.74	5.69 ± 3.24	6.24 ± 1.44	3.68 ± 1.44
	T10	5.42 ± 4.73	9.69 ± 4.95	6.78 ± 1.40	5.22 ± 2.05
	T11	4.15 ± 3.36	10.59 ± 3.19	5.91 ± 1.56	7.39 ± 1.41
2	T9	3.22 ± 1.47	6.62 ± 3.01	3.10 ± 0.85	4.34 ± 0.66
	T10	3.83 ± 1.84	6.73 ± 3.20	4.44 ± 0.79	3.24 ± 0.99
	T11	4.88 ± 2.11	13.95 ± 4.83	6.43 ± 0.81	3.02 ± 0.69
3	L3	2.59 ± 2.18	3.19 ± 1.79	3.07 ± 0.56	4.86 ± 0.35
	L4	4.90 ± 2.14	11.68 ± 4.85	1.87 ± 0.48	4.14 ± 0.33
	L5	4.58 ± 4.97	8.71 ± 4.26	5.21 ± 0.71	7.08 ± 0.21

(b) Rotation per axis (mean-absolute camera-frame Euler-XYZ, deg).

Seq.	Per Vertebra ICP			Rigid ICP		
	R_x	R_y	R_z	R_x	R_y	R_z
1	4.98 ± 4.40	3.95 ± 3.77	2.94 ± 3.25	1.85 ± 2.43	3.04 ± 1.40	2.04 ± 2.53
2	4.45 ± 2.54	6.55 ± 2.22	3.09 ± 2.84	0.80 ± 0.69	3.02 ± 0.71	1.16 ± 0.80
3	4.37 ± 3.39	4.73 ± 2.92	2.77 ± 2.98	2.78 ± 0.37	3.41 ± 0.27	1.40 ± 0.40

Acknowledgments. This work was supported by the UKRI CDT in AI for Healthcare <http://ai4health.io> (Grant No. EP/S023283/1) and the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Bradski, G., Kaehler, A.: Opencv: Open source computer vision library. <https://opencv.org/> (2000), accessed: 2026-02-08
2. Chan, Y., Banza, L., Martin Jr, C., Harrison, W.J.: Essential fracture and orthopaedic equipment lists in low resource settings: consensus derived by survey of experts in africa. *BMJ open* **8**(9), e023473 (2018)
3. Cho, Y., Moseley, D.J., Siewerdsen, J.H., Jaffray, D.A.: Accurate technique for complete geometric calibration of cone-beam computed tomography systems. *Medical physics* **32**(4), 968–983 (2005)
4. Daly, C., Marconi, E., Riva, M., Ekanayake, J., Elson, D.S., Rodriguez y Baena, F.: Towards markerless intraoperative tracking of deformable spine tissue. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 627–637. Springer (2025)
5. Daly, M.J., Siewerdsen, J.H., Cho, Y.B., Jaffray, D.A., Irish, J.C.: Geometric calibration of a mobile c-arm for intraoperative cone-beam ct. *Medical physics* **35**(5), 2124–2136 (2008)

6. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**(6), 381–395 (1981)
7. Gao, Z., Zhang, X., Xu, Z., Jiang, C., Hu, W., Zhang, H., Hao, D.: Mazor x robot-assisted upper and lower cervical pedicle screw fixation: a case report and literature review. *BMC Geriatrics* **24**(1), 916 (2024)
8. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* **47**(6), 2280–2292 (2014)
9. Gorges, S., Kerrien, E., Berger, M.O., Troussset, Y., Pescatore, J., Anxionnat, R., Picard, L.: Model of a vascular c-arm for 3d augmented fluoroscopy in interventional radiology. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 214–222. Springer (2005)
10. Hu, X., Nguyen, A., Baena, F.R.y.: Occlusion-robust visual markerless bone tracking for computer-assisted orthopedic surgery. *Transactions on Instrumentation and Measurement* **71**, 1–11 (2022). <https://doi.org/10.1109/TIM.2021.3134764>
11. ImFusion: ImFusion Suite: Visual computing for medical imaging. <https://imfusion.com/products-overview/imfusion-suite/> (2025), accessed: 2025-11-14
12. Jain, A.K., Mustafa, T., Zhou, Y., Burdette, C., Chirikjian, G.S., Fichtinger, G.: Ftrac—a robust fluoroscope tracking fiducial. *Medical physics* **32**(10), 3185–3198 (2005)
13. Ji, Z.W., Shen, M.J., Sun, J.J., Wang, J.L., Deng, Y.K., Zhang, Y., Wu, X.X., Chen, K.W., Mao, H.Q.: Enhanced safety and efficacy of o-arm navigation over c-arm guidance in percutaneous kyphoplasty for patients with osteoporotic vertebral compression fractures and spinal deformity: A comparative study. *Orthopaedic Surgery* **17**(1), 125–132 (2025)
14. Liebmann, F., von Atzigen, M., Stütz, D., Wolf, J., Zingg, L., Suter, D., Cavalcanti, N.A., Leoty, L., Esfandiari, H., Snedeker, J.G., et al.: Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery. *Medical Image Analysis* **91**, 103027 (2024)
15. Liebmann, F., Stütz, D., Suter, D., Jecklin, S., Snedeker, J.G., Farshad, M., Fürnstahl, P., Esfandiari, H.: Spinedepth: a multi-modal data collection approach for automatic labelling and intraoperative spinal shape reconstruction based on rgb-d data. *Journal of Imaging* **7**(9), 164 (2021)
16. Liu, H., Baena, F.R.Y.: Automatic markerless registration and tracking of the bone for computer-assisted orthopaedic surgery. *IEEE Access* **8**, 42010–42020 (2020)
17. Liu, W.P., Otake, Y., Azizian, M., Wagner, O.J., Sorger, J.M., Armand, M., Taylor, R.H.: 2d–3d radiograph to cone-beam computed tomography (cbct) registration for c-arm image-guided robotic surgery. *International journal of computer assisted radiology and surgery* **10**(8), 1239–1252 (2015)
18. MacKechnie, M.C., Shearer, D.W., Verhofstad, M.H., Martin, C., Graham, S.M., Pesantez, R., Schuetz, M., Hüttel, T., Kojima, K., Bernstein, B.P., et al.: Establishing consensus on essential resources for musculoskeletal trauma care worldwide: a modified delphi study. *JBJS* **106**(1), 47–55 (2024)
19. Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3d/2d registration methods for image-guided interventions. *Medical image analysis* **16**(3), 642–661 (2012)
20. Meng, D., Boyer, E., Pujades, S.: Vertebrae localization, segmentation and identification using a graph optimization and an anatomic consistency cycle. *Computerized Medical Imaging and Graphics* **107**, 102235 (2023)

21. Miller, C.A., Ledonio, C.G., Hunt, M.A., Siddiq, F., Polly, D.W.: Reliability of the planned pedicle screw trajectory versus the actual pedicle screw trajectory using intra-operative 3d ct and image guidance. *International Journal of Spine Surgery* **10** (2016)
22. Navab, N., Heining, S.M., Traub, J.: Camera augmented mobile c-arm (camc): calibration, accuracy study, and clinical applications. *IEEE Transactions on Medical Imaging* **29**(7), 1412–1423 (2010)
23. Northern Digital Inc.: Polaris Vicra Technical Specifications (2018), <https://www.ndigital.com/products/polaris-vicra/>, p/N 10001248 (Rev 001)
24. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems* **30** (2017)
25. Stealthstation s8 surgical navigation system. <https://europe.medtronic.com/xd-en/healthcare-professionals/products/neurological/surgical-navigation-systems/stealthstation/stealthstation-s8.html>, accessed: 2026-02-26
26. Wilson Jr, J.P., Fontenot, L., Stewart, C., Kumbhare, D., Guthikonda, B., Hoang, S.: Image-guided navigation in spine surgery: from historical developments to future perspectives. *Journal of Clinical Medicine* **13**(7), 2036 (2024)