6D Object Pose Tracking for Orthopedic Surgical Training using Visual-Inertial Sensor Fusion

Maarten Hogenkamp✉, Tobias Stauffer, Quentin Lohmeyer, and Mirko Meboldt

ETH Zurich, Zurich, Switzerland
{hmaarten, tobiasta, qlohmeyer, meboldtm}@ethz.ch

Abstract. Digital training simulators play a growing role in orthopedic surgery, offering realistic, standardized, and risk-free learning environments without the need for constant expert supervision. To enable simulators with realistic tactile feedback and haptic sensations, accurate tracking of surgical tools and anatomical structures in real-time is required. However, existing object tracking solutions are often expensive, difficult to integrate into training workflows, or lack robustness. To address these limitations, we propose a novel visual-inertial 6D object pose tracking system for orthopedic surgical training. Our approach features a custom fiducial object that combines multiple ArUco markers with an Inertial Measurement Unit, a dual-camera setup to improve occlusion robustness, and a sensor fusion algorithm that integrates high-frequency IMU data with vision-based tracking while ensuring precise coordinate and time synchronization. In our evaluation, we achieve a fiducial object pose accuracy of 0.9 mm/0.5° and extract drill hole metrics in a mock surgical procedure with average position, angle, and length errors of 1.7 mm, 2.0°, and 1.0 mm, respectively, while demonstrating low occlusion rates. Our cost-effective and easily integrated solution meets clinical training requirements and marks a step towards scalable and widely accessible digital orthopedic simulators. The tracking code is available at `https://github.com/MountainCoot/fusionpose`.

Keywords: Digital Surgical Training · Orthopedic Surgery · 6D Pose Estimation · Sensor Fusion · Inertial Measurement Unit

1 Introduction

Surgical training increasingly relies on digital simulation methods, which provide realistic, standardized, and risk-free training environments with balanced case mix and reduced need for expert supervision [17, 24]. In orthopedic surgery, digital simulators with tactile feedback and haptic sensations have received significant attention [5, 31, 21, 20], as they support the development of psychomotor skills essential in tasks such as bone drilling, screw insertion, and fracture fixation [24, 23, 10]. A fundamental cornerstone of these simulators is the tracking of surgical tools and anatomical structures, enabling realistic rendering with real-time guidance and overlays using Augmented Reality (AR) [31], skill assessment based on tool movements [14, 21], and even complete digitization of surgical

procedures via digital twins [26, 13]. Currently, orthopedic simulators primarily utilize commercially available Optical Tracking Systems (OTSs) and electromagnetic tracking systems [31, 21, 15]. They allow for straightforward tracking of position and orientation, also referred to as the 6D pose, with sub-millimeter accuracy at high frame rates using optical fiducial markers and electromagnetic sensors, respectively [27]. However, the integration of commercial tracking systems into digital simulators is held back by several factors: In addition to their high cost, OTSs require an unobstructed view of the markers, making them susceptible to signal losses, while electromagnetic tracking systems are prone to interference from metallic objects [27, 17]. Therefore, to facilitate scale-up of digital orthopedic training, there is a need for cost-effective tracking solutions that offer high robustness while maintaining ease of integration, accuracy, and acquisition rates comparable to commercial systems [24].

Recently, computer-vision-based methods using RGB cameras have gained popularity for affordable object tracking. Several authors propose markerless pose estimation in surgical contexts with convolutional neural networks [4, 12]. However, limited accuracy, high computational cost, and poor generalization hinder its widespread adoption. In contrast, marker-based motion tracking with RGB cameras largely circumvents these issues by leveraging fiducial markers with known size, geometry, and appearance. A notable example intended for drawing applications in AR is the DodecaPen stylus [32], which achieves accurate 6D pose estimation using ArUco markers. The same principle has been extended to the surgical domain [28], with some works incorporating multi-camera setups to reduce occlusions [29, 30, 6]. Nevertheless, beyond these proof-of-concept studies for general surgery applications, no setup for orthopedic training that enables straightforward multi-object tracking has been proposed. Moreover, acquisition rates of camera-based methods are significantly below commercial tracking systems and might be insufficient to capture fast movements.

A cost-effective option for motion tracking at high rates are Inertial Measurement Units (IMUs), sensors that measure acceleration and angular velocity without suffering from any form of interference. While they cannot be directly used for pose estimation due to signal drift, they enable the extraction of other motion metrics such as acceleration and jerk for skill evaluation [21]. In addition, IMUs are valuable when fused with additional sensors, with a notable example by Enayati et al. [8], who use an IMU in combination with an OTS to upsample 6D pose estimates and bridge short line-of-sight interruptions. However, their work does not address the issue of coordinate calibration and time synchronization between the IMU and the OTS, which is crucial for accurate pose estimation.

In summary, neither existing commercial tracking systems nor camera or IMU-based alternatives fully address the requirements of scalability, accuracy, and robustness necessary for practical use in digital orthopedic training. To fill this gap, we propose a novel visual-inertial tracking system tailored for orthopedic surgical training. It provides 6D object poses, video streams, and IMU data in real-time, facilitating applications such as AR guidance, skill assessment, and digital twin generation. Our key contributions are:

– A visual-inertial tracking method that integrates a custom fiducial object combining multiple ArUco markers with an IMU sensor, enabling tracking of interacting tools and anatomical structures in a surgical setup.
– A two-stage pose estimation pipeline, combining dual-camera-based vision pose estimation for occlusion robustness with a sensor fusion step that augments pose estimates with IMU data and effectively addresses sensor drift.
– An integrated spatial-temporal calibration framework embedded within the tracking pipeline, simultaneously handling coordinate alignment and IMU-camera time synchronization, crucial for precise sensor fusion.

Our approach uses off-the-shelf hardware, ensuring easy integration into surgical training setups. In our evaluation, we assess the system's accuracy and occlusion rates by quantifying the 6D pose accuracy of the fiducial object, highlighting the benefits of a multi-camera setup, and illustrating the system's ability to extract meaningful motion metrics in a mock orthopedic surgery scenario.

2 Method

Figure 1 provides an overview of the proposed tracking system, illustrating the individual components and the data flow through the processing pipeline.

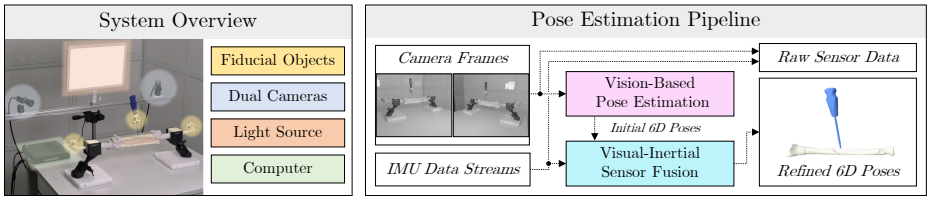


Fig. 1: Overview of the proposed visual-inertial tracking system. The 6D poses of the fiducial objects are estimated using a two-step processing pipeline. The output consists of real-time 6D pose estimates, video streams, and IMU data.

2.1 Custom Fiducial Object

Figure 2 shows the custom fiducial object. We designed it in a dodecahedron shape with ArUco markers glued onto its faces, similar to the DodecaPen [32]. This configuration ensures that the fiducial object is detectable from all viewing angles and enhances pose estimation accuracy by providing multiple ArUco markers on different planes simultaneously. Our 3D-printed design is lightweight, easily attachable to objects using screws, and features a removable lid secured by a snap-fit mechanism. Inside of the fiducial object, we integrate a battery and an IMU that captures acceleration and angular velocity data. We use Bluetooth Low Energy to transmit the IMU data and the battery state of charge to a computer.

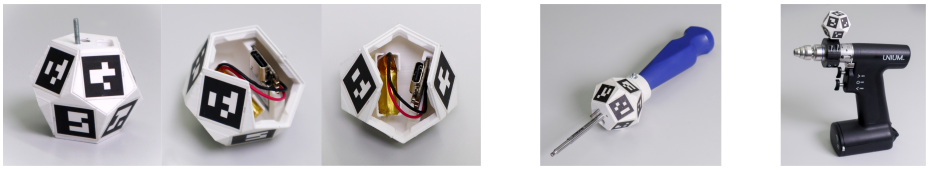


Fig. 2: Overview of the custom-designed fiducial object with and without lid, and mounted on a surgical drill and a screwdriver.

Since the ArUco markers are manually attached, we perform fiducial object calibration to obtain an accurate 3D representation of the marker corners by using a similar procedure as described in [32]. Specifically, we capture the fiducial object from multiple viewpoints and perform bundle adjustment to jointly refine both camera poses and marker corner positions to minimize reprojection error of the ArUco marker corners. Unlike [32], we enforce unit-length edges and right angles in the marker representations, as the primary source of calibration error lies in marker placement rather than marker printing. This additional constraint simplifies the calibration process while ensuring correct marker geometry.

2.2 Vision-Based Pose Estimation

Figure 3 illustrates the steps of the vision-based pose estimation pipeline. We calculate the poses for each camera independently to maintain tracking during occlusions of one camera. First, we crop the camera frames based on the last known positions of the fiducial objects to reduce computational cost. We then process each crop to detect all visible ArUco markers, which gives us the unique marker IDs along with their 2D corner pixel coordinates. Using a standard Perspective-n-Point (PnP) algorithm, we estimate the initial 6D poses of the fiducial objects by matching the 2D corners with the known 3D marker configuration obtained during fiducial calibration. To avoid the known issue of pose ambiguities [19], at least two detected markers are required per object. We filter out any markers that have a high reprojection error, usually caused by inaccurate corner localization due to motion blur or partial occlusions. To detect fiducial objects that newly enter the camera frame, we perform full-frame analysis at predefined intervals.

2.3 Visual-Inertial Sensor Fusion

After obtaining the 6D pose estimates of the fiducial object from the vision-based pose estimation step, we fuse them with the IMU data. The corresponding pipeline is shown on the right side of Figure 3. We formulate the fusion task as a state estimation problem, where the state vector $\mathbf{x}$ of the fiducial object is estimated using the sensor measurements $\mathbf{z}_{\text{inertial}}$ and $\mathbf{z}_{\text{vision}}$ from the IMU and cameras, respectively.

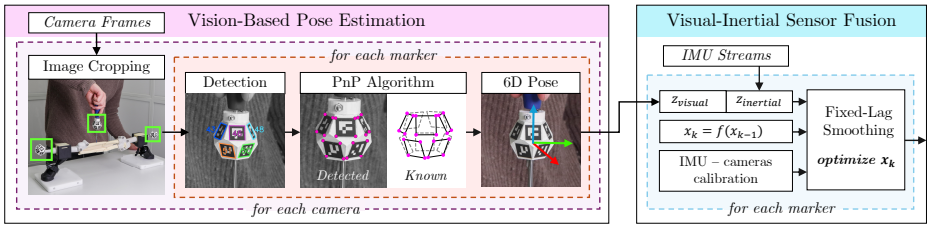


Fig. 3: Overview of the two separate pose estimation components. In the vision-based step, the camera frames are processed to obtain initial 6D poses. They are then integrated with IMU data for pose refinement in the sensor fusion step.

State Description We model the state $\mathbf{x}$ of the fiducial object as [8]:

$$\mathbf{x} = \begin{bmatrix} \mathbf{p}^C & \mathbf{q}^C & \mathbf{v}^C & \mathbf{b}_a^I & \mathbf{b}_\omega^I \end{bmatrix} \tag{1}$$

where $\mathbf{p}^C$ and $\mathbf{v}^C$ are the position and the velocity of the fiducial object, respectively, and $\mathbf{q}^C$ is its orientation expressed using quaternion notation, all given in the main camera frame. $\mathbf{b}_a^I$ and $\mathbf{b}_\omega^I$ are the accelerometer and gyroscope biases in the IMU frame, respectively, and account for measurement drift of the IMU. The state is propagated in time using the function f that follows a first-order continuous-time dynamics model, equivalent to the approach in [8].

Sensor Models The measurements of the IMU are described as:

$$\mathbf{z}_{\text{inertial}} = \begin{bmatrix} \tilde{\mathbf{a}}^I \\ \tilde{\mathbf{w}}^I \end{bmatrix} = \begin{bmatrix} \mathbf{a}^I + \mathbf{b}_a^I + \mathbf{n}_a^I - R_W^I \mathbf{g}^W \\ \mathbf{w}^I + \mathbf{b}_\omega^I + \mathbf{n}_\omega^I \end{bmatrix} \tag{2}$$

where $\tilde{\mathbf{a}}^I$ and $\tilde{\mathbf{w}}^I$ are the measured linear acceleration and angular velocity of the fiducial object, respectively. $\mathbf{a}^I$ and $\mathbf{w}^I$ are the true linear acceleration and angular velocity, and $\mathbf{n}_a^I$ and $\mathbf{n}_\omega^I$ are Gaussian noise terms. The accelerometer measures gravity, which is transformed from the gravity-aligned world frame to the IMU frame using the rotation matrix R_W^I. The measurements of the vision-based pose estimation are only corrupted by Gaussian noise such that

$$\mathbf{z}_{\text{vision}} = \begin{bmatrix} \tilde{\mathbf{p}}^C \\ \tilde{\mathbf{q}}^C \end{bmatrix} = \begin{bmatrix} \mathbf{p}^C + \mathbf{n}_p^C \\ \mathbf{q}^C \otimes \delta\mathbf{q}^C \end{bmatrix} \tag{3}$$

where $\tilde{\mathbf{p}}^C$ and $\tilde{\mathbf{q}}^C$ are the measured position and orientation of the fiducial object, respectively, and $\mathbf{n}_p^C$ and $\delta\mathbf{q}^C$ are Gaussian noise terms, with $\otimes$ denoting quaternion multiplication. The main camera frame is arbitrarily selected, and the measurements from the other camera are transformed using the extrinsics.

Sensor Calibration To fuse measurements from different sensors and coordinate systems, we perform sensor calibration following the framework of Geneva and Huang [11]. It estimates the IMU to fiducial object frame and main camera

to world frame transformations, as well as the time offset between the IMU and the cameras. We integrate the calibration directly into our pipeline and perform it by collecting camera and IMU data of the fiducial object for a short period while exciting all degrees of freedom. The calibration is performed once before tracking starts and is stored for future use.

State Estimation We estimate the state vector $\mathbf{x}$ using a fixed-lag smoother, which refines older states using future measurements. In comparison to a pure filter-based approach, such as a Kalman filter, a smoother improves accuracy and rejects outliers by introducing a slight delay [9]. We implement the smoother by leveraging factor graph optimization [16], which efficiently estimates the state vector using preintegration of the IMU measurements [9].

3 Evaluation

3.1 Implementation Details

For our setup, we use two monochrome cameras (Baumer VCXU.2-57C) with 6mm f/4 lenses (Edmund Optics) and a light source. The cameras are rigidly mounted at a 60° angle and calibrated intrinsically and extrinsically using a ChArUco board. The resulting trackable volume fits a general phantom setup as shown in Figure 1. Each camera records at 20 Hz with a 5 MP resolution and an exposure time of 10 ms to minimize motion blur. The frame rate is selected such that real-time inference is possible. The IMUs are part of microcontrollers (Seeed Studio XIAO) and sample at 200 Hz. A computer running Windows 11 with an Intel Core i7-14700T CPU and 16 GB of RAM processes all data in real-time. The software pipeline is implemented using the Robot Operating System (ROS) in Python and C++ with OpenCV for image processing [3] and GTSAM for sensor fusion [7]. The lag of the smoother is user-defined and set to 100 ms.

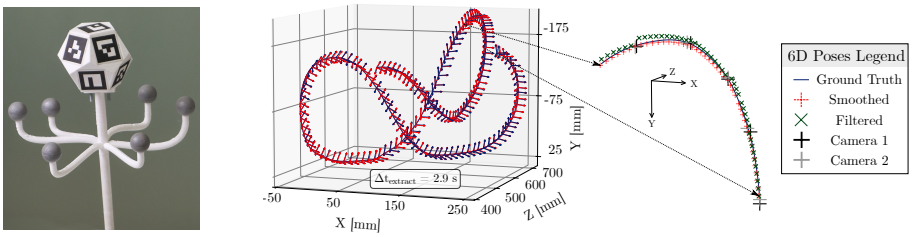
3.2 Experiment Design

The primary goal of the evaluation is to quantify the accuracy and occlusion rates of the proposed tracking system. For this purpose, we conduct two experiments. First, we evaluate the fiducial object by tracking it simultaneously using the proposed visual-inertial tracking system and an OTS (Atracsys fusionTrack 500), serving as ground truth. The OTS achieves a Root Mean Square Error (RMSE) of 80 µm up to 2 m at 335 Hz [1]. To this end, we mount an additional OTS marker on the fiducial object and extrinsically calibrate the OTS and the main camera, as well as the two markers [25]. The setup is illustrated in Figure 4a. In the second experiment, we simulate a mock surgical procedure by drilling five holes into a biomechanical foam testing block (Synbone, 30 PCF) using a surgical drill (Unium, 2 mm diameter) and tracking both the drill and the foam block with the proposed system. Subsequently, we mount OTS markers on the setup and trace the holes using the drill. Figure 5a depicts the two configurations. Finally, we

generate a digital twin from both recordings in a collider-based Unity simulation environment and evaluate drill hole metrics. For this purpose, we determine the drill tip position through pivot calibration [33] and sample points on the foam block surface with a pointer marker to align its mesh with the marker frame [2].

3.3 Results and Discussion

Fiducial Object Tracking We recorded multiple datasets, lasting a total of 144 s, that include linear and rotational movements with varying velocities of the fiducial object. Figure 4b shows an extract from one dataset with the fiducial object trajectory compared to the OTS ground truth for different configurations. Table 1a quantifies the accuracy and the camera occlusion rates, i.e., the ratio between the summed occlusion durations and total tracking time. The best results are achieved when using both cameras and the IMU, with a RMSE of 0.9 mm for position and 0.5 degrees for orientation and no occlusions. The occlusion rate of the ground truth was quite high at 7%, which might result from the fact that at least five out of six spheres of the OTS marker needed to be visible for successful tracking. The proposed system outperforms a single-camera setup in terms of accuracy and occlusion rates, demonstrating the benefits of the multi-camera setup and the sensor fusion algorithm.

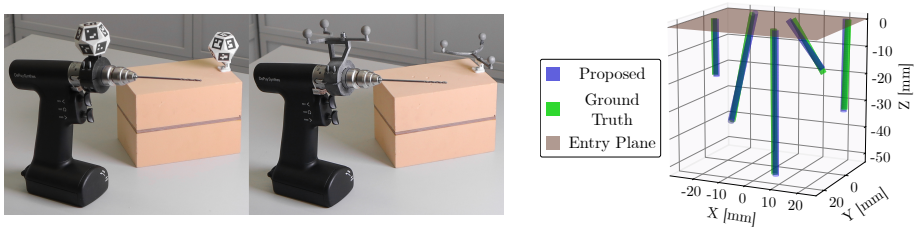


(a) (b)

Fig. 4: (a) Fiducial object with ground truth markers and (b) a dataset extract of the smoothed fiducial object trajectory compared to the ground truth, lasting 2.9 s. In the close-up on the right, the value of smoothing is visible: during the rapid movement, the purely filter-based result drifts away from the ground truth between camera updates, while the fixed-lag smoother constrains the pose estimates with future measurements.

Digital Twin Figure 5b depicts the drill holes obtained from the proposed system compared to the ground truth in a digital twin representation. The holes are accurately generated as demonstrated quantitatively in Table 1b. Overall, the accuracy values are inferior to the previous experiment and have a higher variance. This is likely due to calibration errors of the drill tip and the foam block,

and uncertainties such as the flexion of the drill shaft, which are not directly attributable to the proposed system. We observed low camera occlusion rates during the procedure, indicating the system's robustness. However, as equally low occlusion rates are observed for the ground truth, we cannot conclude that the proposed system is superior in this regard and the small amount of data in a controlled environment might be insufficient to demonstrate a difference. Notably, the IMU does not suffer from any occlusions and measurements are continuously provided at 200 Hz.



(a) (b)

Fig. 5: (a) Mock surgical procedure with a drill and foam block and (b) digital twin of the drill holes from the proposed system compared to the ground truth.

Table 1: Evaluation results of proposed tracking system. (a) Position and orientation RMSE and camera occlusion rates of fiducial object pose for different system configurations. (b) Drill hole entry position, angle, and length offsets in the digital twin, and camera occlusion rates during the procedure.

(a)

Configuration #cams	IMU	p_{rmse} [mm]	θ_{rmse} [◦]	occl. [%]
1	×	1.4	0.7	2
1	✓	1.0	0.5	2
2	✓	0.9	0.5	0
Ground truth		-	-	7

(b)

Metric	Min	Max	Avg
Δp [mm]	0.9	2.3	1.7
$\Delta\theta$ [◦]	0.4	5.4	2.0
Δl [mm]	0.3	2.0	1.0
Occl. (proposed) [%]			0
Occl. (ground truth) [%]			3

Overall, our system enables extraction of meaningful motion metrics and accurate digital twin generation, with a RMSE of approximately 1-2 mm and 2° in the mock surgical scenario. When comparing its performance to estimated accuracy requirements reported in the literature for orthopedic procedures, such as angular offsets up to 5° for hip implant placement, 3° for knee arthroplasty, and combined limits of around 1 mm and 5° for pedicle screw insertion [18, 22],

our system lies within or below these thresholds. These results demonstrate the system's suitability for training across a wide range of orthopedic interventions.

4 Conclusion

This paper presents a novel visual-inertial 6D object tracking system for orthopedic surgical training. By fusing data from ArUco markers and an IMU with dual-camera image streams, our system delivers real-time 6D pose estimates of surgical tools and anatomical structures. Cost-effective components and straightforward integration into existing training workflows make the system a valuable tool for scalable orthopedic surgical training applications. Our evaluation demonstrated the system's high accuracy and robustness in a simple surgical scenario. Future work will focus on analyzing the system's performance and usability in a more sophisticated training setting with medical students.

Acknowledgments. This project was funded by the Swiss Innovation Promotion Agency (Innosuisse) (Grant No. 102.079 IP-ICT).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Atracsys LLC: Fusiontrack 500 - high-performance optical tracking system. `https://atracsys.com/fusiontrack-500/` (2025), accessed: 2025-06-25
2. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures. vol. 1611, pp. 586–606. Spie (1992)
3. Bradski, G.: The opencv library. Dr. Dobb's Journal of Software Tools (2000)
4. Burton, W., Myers, C., Rutherford, M., Rullkoetter, P.: Evaluation of single-stage vision models for pose estimation of surgical instruments. International Journal of Computer Assisted Radiology and Surgery **18**(12), 2125–2142 (2023)
5. Cecil, J., Ramanathan, P., Rahneshin, V., Prakash, A., Pirela-Cruz, M.: Collaborative virtual environments for orthopedic surgery. In: 2013 IEEE international conference on automation science and engineering (CASE). pp. 133–137. IEEE (2013)
6. Chen, L., Ma, L., Zhang, F., Yang, X., Sun, L.: An intelligent tracking system for surgical instruments in complex surgical environment. Expert Systems with Applications **230**, 120743 (2023)
7. Dellaert, F., Contributors, G.: borglab/gtsam (May 2022). https://doi.org/10.5281/zenodo.5794541, `https://github.com/borglab/gtsam`
8. Enayati, N., De Momi, E., Ferrigno, G.: A quaternion-based unscented kalman filter for robust optical/inertial motion tracking in computer-assisted surgery. IEEE Transactions on Instrumentation and Measurement **64**(8), 2291–2301 (2015)
9. Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D.: On-manifold preintegration for real-time visual–inertial odometry. IEEE Transactions on Robotics **33**(1), 1–21 (2016)

10. Gani, A., Pickering, O., Ellis, C., Sabri, O., Pucher, P.: Impact of haptic feedback on surgical training outcomes: a randomised controlled trial of haptic versus non-haptic immersive virtual reality training. Annals of Medicine and Surgery **83**, 104734 (2022)
11. Geneva, P., Huang, G.: vicon2gt: Derivations and analysis. Technical Report (2020), `https://udel.edu/~ghuang/papers/tr_vicon2gt.pdf`
12. Hein, J., Cavalcanti, N., Suter, D., Zingg, L., Carrillo, F., Calvet, L., Farshad, M., Pollefeys, M., Navab, N., Fürnstahl, P.: Next-generation surgical navigation: Marker-less multi-view 6dof pose estimation of surgical instruments. arXiv preprint arXiv:2305.03535 (2023)
13. Hein, J., Giraud, F., Calvet, L., Schwarz, A., Cavalcanti, N.A., Prokudin, S., Farshad, M., Tang, S., Pollefeys, M., Carrillo, F., et al.: Creating a digital twin of spinal surgery: A proof of concept. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2355–2364 (2024)
14. Howells, N.R., Brinsden, M.D., Gill, R.S., Carr, A.J., Rees, J.L.: Motion analysis: a validated method for showing skill levels in arthroscopy. Arthroscopy: The Journal of Arthroscopic & Related Surgery **24**(3), 335–342 (2008)
15. Johns, B.D.: The creation and validation of an augmented reality orthopaedic drilling simulator for surgical training. Ph.D. thesis, Citeseer (2014)
16. Kschischang, F.R., Frey, B.J., Loeliger, H.A.: Factor graphs and the sum-product algorithm. IEEE Transactions on information theory **47**(2), 498–519 (2001)
17. McKnight, R.R., Pean, C.A., Buck, J.S., Hwang, J.S., Hsu, J.R., Pierrie, S.N.: Virtual reality and augmented reality—translating surgical training into surgical technique. Current reviews in musculoskeletal medicine **13**, 663–674 (2020)
18. Mor, A., Jaramaz, B., DiGioia, A.: Accuracy and validation. Computer and Robotic Assisted Hip and Knee Surgery, OUP (2004)
19. Oberkampf, D., DeMenthon, D.F., Davis, L.S.: Iterative pose estimation using coplanar feature points. Computer Vision and Image Understanding **63**(3), 495–511 (1996)
20. Pastor, T., Cattaneo, E., Pastor, T., Gueorguiev, B., Beeres, F.J., Link, B.C., Windolf, M., Buschbaum, J.: Digitally enhanced hands-on surgical training (dehst) enhances the performance during freehand nail distal interlocking. Archives of orthopaedic and trauma surgery **144**(4), 1611–1619 (2024)
21. Pourkand, A., Salas, C., Regalado, J., Bhakta, K., Tufaro, R., Mercer, D., Grow, D.: Objective evaluation of motor skills for orthopedic residents using a motion tracking drill system: outcomes of an abos approved surgical skills training program. The Iowa orthopaedic journal **36**, 13 (2016)
22. Rampersaud, Y.R., Simon, D.A., Foley, K.T.: Accuracy requirements for image-guided spinal pedicle screw placement. Spine **26**(4), 352–359 (2001)
23. Riehl, J., Widmaier, J.: A simulator model for sacroiliac screw placement. Journal of surgical education **69**(3), 282–285 (2012)
24. Ruikar, D.D., Hegadi, R.S., Santosh, K.: A systematic review on orthopedic simulators for psycho-motor skill and surgical procedure training. Journal of medical systems **42**, 1–21 (2018)
25. Shah, M.: Solving the robot-world/hand-eye calibration problem using the kronecker product. Journal of Mechanisms and Robotics **5**(3), 031007 (2013)
26. Shu, H., Liang, R., Li, Z., Goodridge, A., Zhang, X., Ding, H., Nagururu, N., Sahu, M., Creighton, F.X., Taylor, R.H., et al.: Twin-s: a digital twin for skull base surgery. International journal of computer assisted radiology and surgery **18**(6), 1077–1084 (2023)

27. Sorriento, A., Porfido, M.B., Mazzoleni, S., Calvosa, G., Tenucci, M., Ciuti, G., Dario, P.: Optical and electromagnetic tracking systems for biomedical applications: A critical review on potentialities and limitations. IEEE reviews in biomedical engineering **13**, 212–232 (2019)
28. Stenmark, M., Omerbašić, E., Magnusson, M., Andersson, V., Abrahamsson, M., Tran, P.K.: Vision-based tracking of surgical motion during live open-heart surgery. Journal of Surgical Research **271**, 106–116 (2022)
29. Wang, J., Meng, M.Q.H., Ren, H.: Towards occlusion-free surgical instrument tracking: A modular monocular approach and an agile calibration method. IEEE Transactions on Automation Science and Engineering **12**(2), 588–595 (2015)
30. Wang, J., Song, S., Ren, H., Lim, C.M., Meng, M.Q.H.: Surgical instrument tracking by multiple monocular modules and a sensor fusion approach. IEEE Transactions on Automation Science and Engineering **16**(2), 629–639 (2018)
31. Wu, L., Seibold, M., Cavalcanti, N.A., Hein, J., Gerth, T., Lekar, R., Hoch, A., Vlachopoulos, L., Grabner, H., Zingg, P., et al.: A novel augmented reality-based simulator for enhancing orthopedic surgical training. Computers in Biology and Medicine **185**, 109536 (2025)
32. Wu, P.C., Wang, R., Kin, K., Twigg, C., Han, S., Yang, M.H., Chien, S.Y.: Dodecapen: Accurate 6dof tracking of a passive stylus. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. pp. 365–374 (2017)
33. Yaniv, Z.: Which pivot calibration? In: Medical imaging 2015: Image-guided procedures, robotic interventions, and modeling. vol. 9415, pp. 542–550. SPIE (2015)