

Multi-Camera AR Guidance System for Surgical Instrument Handling and Assembly: Investigating Workload and Efficiency

Shiyu Li^{*1-3}, Julian Kreimeier^{*1-4}, and Hannah Schieber¹⁻⁴, Dirk Müller², Bernhard Kainz^{4,5}, Rüdiger von Eisenhart-Rothe², and Daniel Roth¹⁻³

¹ Technical University of Munich,

Human-Centered-Computing and Extended Reality Lab

² TUM University Hospital, Orthopedics and Sports Orthopedics

³ Munich Institute of Robotics and Machine Intelligence (MIRMI)

⁴ FAU Erlangen-Nürnberg, Erlangen, Germany

⁵ Department of Computing, Imperial College London, London, UK
{shiyu.li,julian.kreimeier,daniel.roth}@tum.de

Abstract. The handling and assembly of instruments during surgery impose high cognitive demands on scrub nurses, particularly when instruments are unfamiliar. We present a supporting guidance system for surgical instrumentation combining multi-camera marker-free 6D pose estimation with augmented reality (AR) in-situ visualization on a head-mounted display. Pose estimation and consecutive camera calibration are achieved through known objects. Our 6D pose estimation is purely trained on synthetic data, aiming for better generalizability and real-world applicability. The AR guidance displays tooltip localization cues and step-wise assembly animations. Via gaze-based selection and a foot pedal, users can switch assembly steps in intraoperative use.

In a technical evaluation, we outperform state-of-the-art 6D pose estimation. A user study with 29 scrub nurses conducted in a surgical simulation of knee arthroplasty compares our system against a paper manual. Our AR system significantly reduced the perceived workload. Objectively, AR guidance reduced task completion time by 21.3% (4.76 minutes). Specifically, less-experienced scrub nurses benefited from our approach. Error frequencies were comparable between conditions. Qualitative feedback highlighted improved process clarity, reduced information overload, and perceived independence. To summarize, our marker-free multi-camera AR guidance approach for surgical instruments can, subjectively and objectively, improve intraoperative instrumentation performance, particularly for untrained scrub nurses.

Keywords: Augmented Reality · Pose Estimation · Guidance.

* equal contribution

1 Introduction and Related Work

In the Operating Room (OR), scrub nurses sterilely support surgeons in a sensitive environment under time pressure. Endoprosthetic instruments, as a typical example, which are used for replacing damaged or diseased joints with artificial implants, have become very complex and manufacturer-dependent. Due to this and a general workforce shortage, inexperienced staff may need to handle unfamiliar instruments [5]. Yet flawless handling remains essential for efficiency and patient safety, as pauses during surgery pose unnecessary risks.

The WHO's Patient Health and Safety plan [27] requires operating manuals and safety instructions to be available at the point of use. Currently, when problems occur, PDF or paper manuals are requested, often from outside of the OR, and reviewed, commonly by additional personnel. This results in longer interruptions [10]. Linear digital manuals or tools for assembly guidance provided by different manufacturers are not broadly embedded, nor are they context dependent. Augmented Reality (AR) guidance with in-situ instrument detection could serve as a solution to fulfill the WHO criteria.

Augmented Reality Guidance recently focuses on surgeons' real-time navigation and visualization [1], instrumentation [23], and registration with deformable tissue [29]. Also, marker-free situs registration is being extensively researched [20]. However, prior to the surgeons' activities, handling and assembly of the instruments by the scrub nurse has not been extensively investigated, though this process is often a pain point for inexperienced staff and promising to support by AR guidance [16]. Using relevant objects' pose information, an AR system can show in-situ tooltips ("pick those") and assembly animations ("assemble them like this"), providing step-by-step guidance [15], reducing cognitive load, errors, and task completion time [3].

Surgical Instrument Detection is predominantly studied in endoscopic videos [6]. Extending detection to 6D pose estimation often relies on fiducial markers [12,22] or is limited to two surgical tools [11,2]. Given this context, a recent approach introduced nine surgical tools being used to perform static and dynamic camera calibration on-the-fly without a calibration board [19]. Also, the potential of synthetically-trained deep learning-based approaches is not fully leveraged [25,17]. It remains unclear how multi-camera, entirely marker-free pose estimation performs in a full-fidelity surgical simulation to inform an AR guidance system for inexperienced scrub nurses. To address these gaps, we contribute:

- An AR guidance system for surgical instrument handling assembly connected to deep learning-based 6D pose estimation.
- A multi-camera pose estimation method trained on synthetic data, overcoming challenges in data annotation, allowing for generalization to manifold instrument sets.
- A technical evaluation of the system's performance.
- An expert user study with scrub nurses comparing paper-based manuals and our AR guidance system in a simulated OR scenario.



Fig. 1: Informed by marker-free multi-camera pose estimation, our AR guidance system visualizes tooltips and an assembly animation for knee arthroplasty.

Our work is the first to benchmark multi-camera AR assembly guidance in an OR scenario. We focus on adapting and extending pose estimation methods for AR-guided surgical workflow support. Our guided AR system improved overall user experience and reduced task completion time by 21.3% or 4.76 minutes, respectively, in a sample of scrub nurses with different levels of experience.

2 Method

We propose an end-to-end approach for multi-object, multi-camera 6D pose estimation connected to an in-situ AR application building upon user requirements [5], allowing users to assemble challenging medical instruments. Our guidance system, purely trained on synthetic data, estimates poses of a) surgical trays and instruments on the assembly table, and b) provides visual guidance by means of AR in-situ tool tips and assembly visualizations. The backend extends [19] to a hierarchical marker-free multi-camera pose estimation to comply with sterility and space constraints in the OR. We leverage object pose estimation to calibrate the head-mounted display (HMD) with non-overlapping static cameras for maximum coverage. The HMD runs a Unity application (MRTK2) highlighting the object locations detected by our pose estimator.

2.1 Synthetic Data Generation

We scan each component⁶ with a Shining EinScann-SP 3D Scanner with up to 0.05 mm accuracy in a scanning volume of $1200 \times 1200 \times 1200 \text{ mm}^3$. We mitigate texture artifacts via MeshLab raster layers. The post-processed scans are used

⁶ DePuy Synthes TFN-ADVANCED™ Proximal Femoral Nailing System and Implant-cast GenuX MK implant

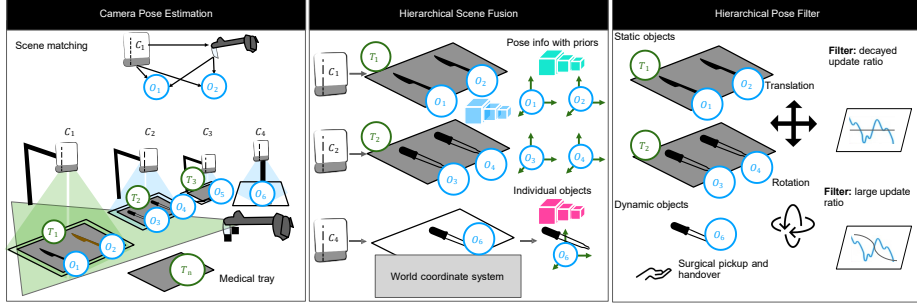


Fig. 2: System overview: Four static top-down cameras observe the surgical trays and deposit table (left), with RGB data fused and processed on a server for marker-free pose estimation and camera calibration (center, right).

in BlenderProc [7], adding domain randomization (varying backgrounds, lighting, camera poses, distractor objects [24]) to generate randomized and realistic training data, which helps to overcome the sim-to-real gap.

2.2 Hierarchical Multi-Camera Pose Estimation

Model Configuration Our pose detection model integrates a real-time capable object detector and pose estimator to balance speed and accuracy. We use object and keypoint detection modules within the YOLOX architecture [9]. We integrate DCC from RTMO [21] to increase keypoint accuracy from regressed bounding boxes (bbox). Object poses are recovered via RANSAC-based Perspective-n-Point (PnP) from predicted bbox and keypoints [25,18].

Training We exclusively use synthetic images (Sec. 2.1). To optimize the convolutional neural network (CNN), we use the 2D MLE, 3D keypoint [13] and proxy loss following MultiCam [19]. Hyperparameter selection follows RTMO [21].

Inference We use the same configuration as during training, but extend the model with filtering, optimizing robustness for real-world OR conditions, and reducing pose misalignment [14]. We combine individual instrument pose estimation and surgical instrument trays via the HMD camera and Kinect cameras [19]. Instrument poses are first estimated on the instrument table; if unsuccessful, the instrument is assumed to remain in its tray. Since relative poses within trays are predefined, the world pose of an instrument can be approximated.

Hierarchical Scene Fusion We calculate each camera pose using spatiotemporal overlapping object candidates following [17,19]. Then, we define our hierarchical scene structure with camera and object nodes. Given each camera pose, we fuse the object poses from trays and individual objects in each camera view.

Hierarchical Pose Filter All pose estimates are processed by a two-stage filter: First, a plausibility check rejects physically implausible poses (e.g., excessive distance, camera proximity, or inverted trays). Second, valid poses are smoothed using an exponential moving average, with linear filtering for translations and spherical linear interpolation for rotations as follows:

$$\mathbf{t}_{\text{smooth}} = \alpha \mathbf{t}_{\text{curr}} + (1 - \alpha) \mathbf{t}_{\text{prev}}, \quad \mathbf{q}_{\text{smooth}} = \text{slerp}(\mathbf{q}_{\text{prev}}, \mathbf{q}_{\text{curr}}, \alpha) \quad (1)$$

where $\mathbf{t}_{\text{curr}} \in \mathbf{R}^3$ and $\mathbf{q}_{\text{curr}}$ are the current translation and rotation, $\mathbf{t}_{\text{prev}}$ and $\mathbf{q}_{\text{prev}}$ are the previous ones, and $\alpha \in [0, 1]$ is the filter coefficient.

Considering static and dynamic objects in the scene, we apply a hierarchical pose filter using different filter coefficients, i.e., a fixed coefficient $\alpha = 0.2$ for individual dynamic instruments and an exponentially decayed coefficient $\alpha = e^{-\frac{i}{300}}$ for static trays to converge to a stable pose, where i is the number of frames. We assume static cameras and trays for the technical evaluation and the user study. To enable the use of moving trays and cameras, static constraints can be disabled. Given top-down cameras, these filters also make pose data more robust in terms of occlusions by hands picking objects from the trays.

2.3 System Architecture

We use RGB feeds from four Azure Kinect cameras and a Microsoft HoloLens 2 HMD (Fig. 2). We use the HoloLens because it is an optical-see-through HMD, which is good for the medical environment, as video-see-through can fail. Our C++ server builds upon OpenMMLab [21,4], achieving an average runtime around 200 ms per frame on a workstation with an RTX 4080 GPU (16 GB VRAM) and an AMD Ryzen 9 9900X CPU with 64 GB RAM. The HMD sends video streams to the workstation, and after the multi-camera, multi-object 6D pose estimation, the workstation sends pose information back using WiFi via TCP [8]. The measured end-to-end latency for data streaming from the workstation to the HMD is approx. 60 ms. The server analyzes the video streams from four static and one dynamic camera, each with 15 frames per second (tradeoff between computation speed and precision). Our HMD visualization pose refreshment rate is 3 seconds to account for battery life and system heating.

3 Technical Evaluation

To technically evaluate our system, we calculate runtime and pose estimation accuracy in ADD(-S) using 3785 frames of images in 720P resolution from one HMD (HoloLens 2) and four static external RGB cameras (Azure Kinect). Our ground truth data is annotated by the Optitrack motion capture system with Motive 3.1.4 and six infrared Primex 13 cameras. Furthermore, to assess the quality of the multi-view multi-object evaluation, we compare our method to CosyPose [17], as shown in Table 1. Our method shows overall better accuracy

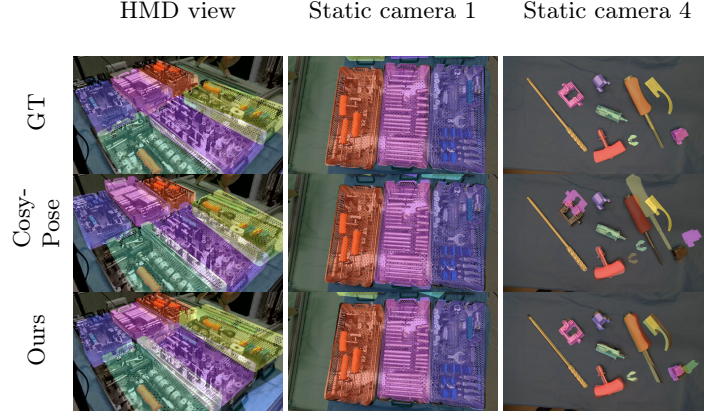


Fig. 3: Visualization of pose estimation for surgical trays and instruments (real data). Our approach shows less visualization error compared to CosyPose [17].

Table 1: Camera and object pose evaluation on our real test data.

Assets	Methods	Condition pose filter	Object pose $ADD(-S)AUC \uparrow$	Camera pose $e_{trans} \downarrow e_{rot} \downarrow$		Overall runtime ms
Entire trays	CosyPose [17]	X	38.26	84.46	9.79	183.47
	Ours	X	21.80	73.81	7.12	186.92
		✓	44.34	66.95	5.20	184.95
Instruments in trays	CosyPose [17]	X	16.94	94.14	8.37	167.37
	Ours	X	24.24	72.56	6.13	111.59
		✓	31.73	69.66	5.79	118.08

$ADD(-S)AUC$ = average distance metric under the accuracy-threshold curve [28],
 e_{trans} = translation error in mm, e_{rot} = rotation error in degree

with $ADD(-S)AUC$ 44.34 for trays and 31.73 for instruments due to improved scene matching and pose filter. The average camera pose error is less than 70 mm and rotation error less than 6 degrees, see Figure 3 for qualitative results.

4 User Study

Our system was compared to a paper manual (current standard) in a within-subjects study using a simulated surgical scenario, see Fig. 1. We measured assembly time, error rate, and workload (raw NASA TLX) [26]. Ethical approval was obtained from the ethics board of the Technical University of Munich (approval 2025-106-S-NP), following the declaration from Helsinki.

Sample, Procedure, and Task We recruited 29 scrub nurses (25 female, 4 male; 30.9 ± 10.9 years; OR experience from 0.5 to 32 years). 22 used the tested Implantcast GenuX MK implant ≤ 3 times (Fig. 4 mid shows the distribution

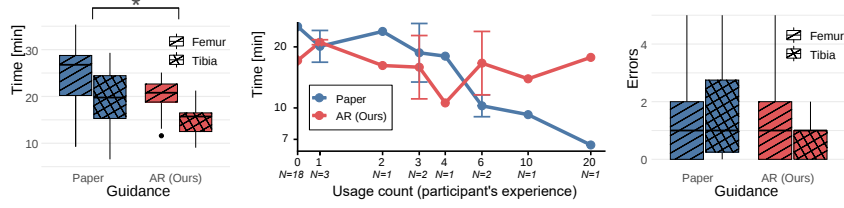


Fig. 4: Simulated surgery times for each guidance and surgery type (left), participants’ experience (middle), and errors (right). The results show that our system benefits practitioners with up to four prior OR experiences with the sets.

Table 2: Subjective results for paper and AR condition.

Raw NASA-TLX	P_M	P_{SD}	AR_M	AR_{SD}	$p \leq$	Effect
Mental	75.86	17.88	43.28	26.33	.001	$r = 0.84$
Physical	35.35	27.61	27.24	23.78	.266	$r = 0.21$
Temporal	73.28	20.63	33.45	24.28	.001	$r = 0.96$
Performance	53.45	24.39	30.17	18.54	.001	$d = 1.07$
Effort	63.45	22.96	32.07	15.15	.001	$d = 1.59$
Frustration	68.28	20.37	33.97	23.16	.001	$d = 1.57$

P = paper manual, AR = AR guidance, M = mean, SD = standard deviation

of experience). Having obtained informed consent, participants first completed an AR tutorial using a subset of the TFNA instruments for familiarization. Second, they assisted two parts of a simulated knee arthroplasty surgery (tibia and femur preparation, each 25–30 parts and 15–40 min), comparing paper manuals and our AR system. Counterbalancing was complete for 28 participants. An additional participant introduces only a negligible imbalance without biasing condition comparisons.

Quantitative Results Our system significantly outperformed the paper manual in time to completion ($F(1) = 12.036, p = .001, \eta_p^2 = .108$), see Figure 4. We found main effects for the type of bone ($F(1) = 16.989, p < .001, \eta_p^2 = .167$) and experience with the instruments ($F(3) = 5.455, p = .003, \eta_p^2 = .160$). This underlines that specifically less experienced staff benefit from the AR guidance, whereas a turning point can be identified between 4 and 6 previous uses. With AR, the timing differences between the femur and tibia in terms of complexity become clearer than with paper. No significant differences were observed in the error rates. Our system further resulted in a significantly lower perceived workload (mental, temporal, performance, effort, and frustration; see Table 2).

Qualitative Results The participants expressed an overall positive attitude toward our approach due to the step-wise instructions, clear menu structure, and the combination of assembly animations with tooltips on parts to pick.

In contrast to the paper manual, selective AR information presentation was associated with effective visualization and reduced information overload. The system was described as particularly suitable for training and supporting users with limited prior experience. Both gaze and hand-based interaction (moveable user interface) were described as supportive of system use.

5 Discussion

We introduce an AR guidance system driven by real-time 6D pose estimation trained on synthetic data. We outperform CosyPose[17] in evaluation metrics. Given manufacturers' CAD models, our approach could be scaled to manifold instrument sets. We support a flexible, multi-component instrument setup without the need of camera calibration. Our non-overlapping multi-camera setup is used for field of view (FoV) extension, rather than a multi-view pose accuracy improvement [17,11,2]. Our approach is leveraging known object knowledge [19] and combines hierarchical fusion and dynamic filtering. To our knowledge, this is the first AR system for pick, assemble, and handover tasks for scrub nurses, rather than surgeons' tasks [29,1].

The present system was validated by scrub nurses in a simulated surgery, demonstrating a clear advantage over traditional paper-based instructions for instrumentation. As intended, our system is especially beneficial for novices and inexperienced scrub nurses with a turnover point between 4 and 6 previous uses (Fig. 4, middle). Error frequencies remained similar. Time differences for femur and tibia preparation become more apparent as the guided handling of more complex parts scales. Overall, handling our AR system was perceived as positive throughout.

Limitations noted by participants included ergonomic constraints, limited field of view, sterility concerns, and acclimatization requirements (learning curve). Technical challenges included rare gaze calibration problems, delayed feedback, and tooltip inaccuracies. The fact that guidance text instructions have to be implemented manually represents a remaining challenge for the future.

Future Work will focus on improving pose estimation accuracy, generalization, incorporating new instruments, adaptive and context-aware visualization, and validation under real OR conditions across additional instrument sets. Overall, we established a baseline for integrating AR guidance into routine surgical practice, potentially expanding to monitoring systems for experienced users.

6 Conclusion

We present a multi-view pose-estimation-driven AR guidance system for surgical instrumentation. Our approach demonstrates measurable improvements in task performance and usability, particularly for inexperienced scrub nurses. Our combination of marker-free pose estimation and spatially registered visualization

provides context-aware, situated guidance that supports efficiency, reduces uncertainty, and enhances procedural confidence in surgical workflows. Our findings indicate that the system itself is technically robust and clinically promising.

Acknowledgments. This work was funded by the German Federal Ministry of Education and Research (BMBF) in the grant program "AI-based assistance systems for in-process health applications", grant number 16SV8973. Shiyu Li acknowledges the financial support from the China Scholarship Council. This work was partially supported by the TU Munich, through its MIRMI Seed Fund program, and by the Robotics Institute Germany (RIG) under grant 16ME0997K. The authors thank implantcast GmbH for lending the surgical equipment. We acknowledge HPC resources (NHR@FAU b143dc, b180dc), funded by federal and Bavarian state authorities, and Gerhard Wellein's HPC approach. NHR@FAU hardware is partially funded by DFG 440719683. Additional support by ERC projects MIA-NORMAL 101083647, DFG 513220538 and 512819079, and the state of Bavaria (HTA and the Bavarian Foundation Model Initiative).

Disclosure of Interests. The authors have no relevant competing interests.

References

1. Acar, A., Atoum, J., Connor, P.S., Pierre, C., Lynch, C.N., Kavoussi, N.L., Wu, J.Y.: Navius: navigated augmented reality visualization for ureteroscopic surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 433–443. Springer (2025)
2. Agethen, N., Roskamp, J., Koller, T.L., Klein, J., Zachmann, G.: Recurrent multi-view 6dof pose estimation for marker-less surgical tool tracking. *International Journal of Computer Assisted Radiology and Surgery* pp. 1–11 (2025)
3. Blattgerste, J., Renner, P., Streng, B., Pfeiffer, T.: In-Situ Instructions Exceed Side-by-Side Instructions in Augmented Reality Assisted Assembly. In: Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference. pp. 133–140. ACM, Corfu Greece (Jun 2018)
4. Contributors, M.: Openmmlab pose estimation toolbox and benchmark. <https://github.com/open-mmlab/mmpose> (2020)
5. Cramer, E., Kucharski, A., Kreimeier, J., Andreß, S., Li, S., Walk, C., Merkl, F., Högl, J., Wucherer, P., Stefan, P., others: Requirement analysis for an AI-based AR assistance system for surgical tools in the operating room: stakeholder requirements and technical perspectives. *International Journal of Computer Assisted Radiology and Surgery* pp. 1–10 (2024), publisher: Springer
6. Demir, K.C., Schieber, H., Weise, T., Roth, D., May, M., Maier, A., Yang, S.H.: Deep learning in surgical workflow analysis: a review of phase and step recognition. *IEEE Journal of Biomedical and Health Informatics* **27**(11), 5405–5417 (2023)
7. Denninger, M., Sundermeyer, M., Winkelbauer, D., Olefir, D., Hodan, T., Zidan, Y., Elbadrawy, M., Knauer, M., Katam, H., Lodhi, A.: BlenderProc: Reducing the Reality Gap with Photorealistic Rendering (2020)
8. Dibene, J.C., Dunn, E.: Hololens 2 sensor streaming. *arXiv preprint arXiv:2211.02648* (2022)
9. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430* (2021)

10. Göras, C., Olin, K., Unbeck, M., Pukk-Härenstam, K., Ehrenberg, A., Tessma, M.K., Nilsson, U., Ekstedt, M.: Tasks, multitasking and interruptions among the surgical team in an operating room: a prospective observational study. *BMJ open* **9**(5), e026410 (May 2019)
11. Hein, J., Cavalcanti, N., Suter, D., Zingg, L., Carrillo, F., Calvet, L., Farshad, M., Navab, N., Pollefeys, M., Fürnstahl, P.: Next-generation surgical navigation: Marker-less multi-view 6dof pose estimation of surgical instruments. *Medical Image Analysis* **103**, 103613 (2025)
12. Hogenkamp, M., Stauffer, T., Lohmeyer, Q., Meboldt, M.: 6d object pose tracking for orthopedic surgical training using visual-inertial sensor fusion. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 13–23. Springer (2025)
13. Hu, Y., Speierer, S., Jakob, W., Fua, P., Salzmann, M.: Wide-depth-range 6d object pose estimation in space. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 15870–15879 (2021)
14. Khuong, B.M., Kiyokawa, K., Miller, A., La Viola, J.J., Mashita, T., Takemura, H.: The effectiveness of an AR-based context-aware assembly support system in object assembly: 21st IEEE Virtual Reality Conference, VR 2014. 2014 IEEE Virtual Reality, VR 2014 - Proceedings pp. 57–62 (2014), publisher: IEEE Computer Society
15. Kleinbeck, C., Schieber, H., Andress, S., Krautz, C., Roth, D.: ARTFM: Augmented Reality Visualization of Tool Functionality Manuals in Operating Rooms. In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. pp. 736–737 (Mar 2022)
16. Kreimeier, J., Schieber, H., Li, S., Martin-Gomez, A., Roth, D.: Visual Guidance for Assembly Processes. In: *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. pp. 652–653 (Oct 2024), iSSN: 2771-1110
17. Labbé, Y., Carpentier, J., Aubry, M., Sivic, J.: Cosypose: Consistent multi-view multi-object 6d pose estimation. In: *European Conference on Computer Vision*. pp. 574–591. Springer (2020)
18. Li, S., Schieber, H., Corell, N., Egger, B., Kreimeier, J., Roth, D.: GBOT: Graph-Based 3D Object Tracking for Augmented Reality-Assisted Assembly Guidance. In: *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. pp. 513–523 (Mar 2024), iSSN: 2642-5254
19. Li, S., Schieber, H., Waldow, K., Busam, B., Kreimeier, J., Roth, D.: Multicam: On-the-fly multi-camera pose estimation using spatiotemporal overlaps of known objects. *IEEE Transactions on Visualization and Computer Graphics* (2026)
20. 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)
21. Lu, P., Jiang, T., Li, Y., Li, X., Chen, K., Yang, W.: RTMO: Towards high-performance one-stage real-time multi-person pose estimation (2023)
22. Martin-Gomez, A., Li, H., Song, T., Yang, S., Wang, G., Ding, H., Navab, N., Zhao, Z., Armand, M.: Sttar: surgical tool tracking using off-the-shelf augmented reality head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics* **30**(7), 3578–3593 (2023)
23. Müller, F., Roner, S., Liebmann, F., Spirig, J.M., Fürnstahl, P., Farshad, M.: Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3d imaging. *The Spine Journal* **20**(4), 621–628 (2020)

24. Schieber, H., Demir, K.C., Kleinbeck, C., Yang, S.H., Roth, D.: Indoor synthetic data generation: A systematic review. *Computer Vision and Image Understanding* p. 103907 (2024)
25. Schieber, H., Li, S., Corell, N., Beckerle, P., Kreimeier, J., Roth, D.: ASDF: Assembly State Detection Utilizing Late Fusion by Integrating 6D Pose Estimation. In: 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 190–199 (Oct 2024), iSSN: 2473-0726
26. Tubbs-Cooley, H.L., Mara, C.A., Carle, A.C., Gurses, A.P.: The NASA Task Load Index as a measure of overall workload among neonatal, paediatric and adult intensive care nurses. *Intensive and Critical Care Nursing* **46**, 64–69 (Jun 2018)
27. (WHO), W.H.O.: Global Patient Safety Action Plan 2021-2030, <https://www.who.int/teams/integrated-health-services/patient-safety/policy/global-patient-safety-action-plan>
28. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. *arXiv preprint arXiv:1711.00199* (2017)
29. Yang, Q., Li, F., Xu, J., Liu, Z., Sridhar, S., Jin, W., Du, J., Heiselman, J., Miga, M., Topf, M., et al.: Augmented reality-based guidance with deformable registration in head and neck tumor resection. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 45–54. Springer (2025)