

JOURNAL OF THE ARTIFICIAL INTELLIGENCE ENGINEERS CONSORTIUM

Volume: 02 Issue: 01 (2026)

<https://jaiec.org.hostajournal.com>

Lightweight Depth-Based Tracking with Memory-Efficient Surface Encoding for Mobile Augmented Reality Surgical Platforms

Lorenzo Bianchi¹, Marco Conti², Giulia Rossi^{3*}

Department of Biomedical Sciences, Sapienza University of Rome, 00185 Rome,
Italy

*Corresponding author: giulia.rossi@uniroma1.it

Abstract : *Mobile AR surgical systems face strict computational and memory constraints. This study presents a lightweight markerless tracking algorithm based on compressed surface encoding and incremental pose update strategies. Surface representations are quantized using a vector-clustering scheme, reducing memory usage by 31.4% compared with full-resolution meshes. Incremental pose estimation is performed using a local depth residual model that avoids global re-optimization. Testing was conducted on 50 simulated surgical sequences (total 18,000 frames). Average tracking drift over 2 minutes was limited to 1.21 mm, while maintaining 52 FPS on a mobile AR processor. Compared to standard dense ICP pipelines, computation time decreased by 38.9% without significant loss in accuracy. This approach is suitable for battery-powered head-mounted surgical devices.*

Keywords : *Mobile AR tracking; lightweight registration; depth compression; real-time surgical guidance; memory-efficient surface modeling*

INTRODUCTION

Mobile augmented reality (AR) surgical platforms are increasingly being explored as an alternative to conventional cart-based navigation systems because they can present guidance directly within the surgeon's field of view while reducing operating-room footprint and setup complexity [1]. Recent clinical and technical studies have shown that head-mounted AR systems can support trajectory guidance, spatial interpretation, and workflow integration across several procedures [2]. Recent depth-based markerless AR guidance systems further suggest that image-to-patient registration and tracking can be performed directly on wearable devices using onboard sensing, without dependence on external fiducials or tracking infrastructure, which strengthens the practical appeal of mobile deployment in surgical settings [3]. Even so, available evidence consistently indicates that navigation quality depends strongly on tracking accuracy, pose stability, and robustness under realistic intraoperative conditions rather than on visualization quality alone [4,5]. This issue is particularly important for battery-powered mobile platforms, which operate under strict constraints in computational capacity, memory

availability, and inference latency, yet are still expected to maintain millimeter-level precision despite depth noise, limited surface exposure, and frequent viewpoint changes [6,7]. Depth-based markerless tracking has emerged as a practical solution in this context because it reduces reliance on external optical tracking systems and fiducial markers while allowing repeated pose updates during surgery [8]. Integrated depth sensors, such as those available in modern AR headsets, provide direct access to intraoperative surface geometry and make surface-based tracking feasible within clinically acceptable time windows [9,10]. However, reported performance remains highly sensitive to initialization quality, visible surface extent, and local geometric distinctiveness [11]. In spine navigation and related applications, automated AR tracking pipelines have reduced workflow complexity and improved usability, but performance degradation still occurs when occlusion increases, drift accumulates over time, or geometric overlap becomes insufficient for reliable correspondence estimation [12]. These findings indicate that improving mobile AR tracking requires a balanced solution that addresses not only alignment accuracy but also temporal stability, robustness to partial observations, and computational efficiency under device-level constraints. Dense Iterative Closest Point (ICP)-based alignment remains one of the most common strategies for refining pose between intraoperative depth data and preoperative anatomical models [13,14]. Although dense ICP can be effective in controlled environments, its computational burden is substantial for mobile processors, and its stability often deteriorates when overlap decreases, outliers increase, or local correspondences become ambiguous [15]. These limitations become more severe in surgical scenarios, where anatomy may be only partially visible and sensing conditions may vary continuously throughout a procedure. Some AR navigation systems therefore combine dense geometric alignment with feature-based calibration or hybrid registration modules [16]. However, many of these methods assume relatively clean sensing conditions or adequate overlap and do not explicitly address the memory overhead associated with dense mesh or point-cloud representations. In addition, several existing evaluations are conducted over short sequences or limited interaction windows, which may underestimate cumulative drift and long-term instability during continuous use [17]. As a result, current evidence still leaves an important gap between short-term registration success and sustained, resource-aware tracking performance on mobile surgical AR hardware. Related insights can be drawn from recent work on point-cloud compression, geometric abstraction, and memory-aware 3D processing. Studies in general 3D vision have shown that geometric quantization, clustering-based encoding, and compact surface representations can reduce storage and bandwidth requirements while preserving essential shape information needed for downstream tasks [18]. Other efficient registration frameworks attempt to reduce unnecessary computation by focusing correspondence estimation on overlapping or geometrically informative regions rather than processing the full surface at every update [19]. These ideas are conceptually relevant to mobile AR tracking because they directly target the two bottlenecks that often limit real-time deployment on wearable hardware: memory consumption and repeated dense matching. Nevertheless, most such approaches have been developed and validated on general 3D datasets, robotics benchmarks, or desktop-class hardware. Their behavior on surgical depth streams remains insufficiently understood, especially when observations are noisy, anatomy is only partially exposed, and tracking must remain stable over extended frame sequences under limited power and memory budgets [20,21]. This gap suggests the need for a surgical tracking framework that integrates compact geometric representation with efficient incremental pose estimation rather than treating memory reduction and tracking accuracy as

separate problems. Against this background, the present study introduces a lightweight markerless tracking framework for mobile AR surgical guidance that combines compressed surface encoding with incremental pose updates derived from local depth residuals. Surface data are quantized using a vector-clustering strategy to reduce memory usage while retaining the geometric structure needed for reliable registration. Instead of repeatedly performing dense global optimization, the proposed method updates pose incrementally, which is intended to lower runtime cost and improve suitability for mobile processors. The aim of this study is to maintain stable tracking across long sequences while reducing memory demand and computational overhead relative to dense ICP-based pipelines. By evaluating the framework on 50 simulated surgical sequences comprising 18,000 frames, this work examines memory reduction, long-term drift behavior, and real-time performance under mobile hardware constraints. The significance of this study lies in its effort to align algorithmic design with the practical limitations of wearable surgical AR systems, thereby supporting a more deployable form of markerless guidance that is not only accurate but also efficient and stable during sustained intraoperative use.

2. Materials and Methods

2.1 Sample and Study Environment Description

The evaluation included 50 simulated surgical tracking sequences with a total of 18,000 frames. Each sequence represented a two-minute procedure with continuous camera motion around rigid bone models generated from high-resolution CT data. The primary tracking targets were thoracic and lumbar vertebrae. Depth images were rendered with realistic sensor noise and partial occlusion to reflect intraoperative conditions. Camera motion included smooth rotations, lateral shifts, and short pauses to simulate head-mounted movement during surgery. All experiments were performed on a mobile AR processor with limited memory and battery capacity to match practical clinical deployment.

2.2 Experimental Design and Control Setup

The proposed lightweight tracking method was compared with a standard dense point-to-plane ICP pipeline that used full-resolution meshes and global optimization at each frame. The experimental group applied compressed surface encoding based on vector clustering and incremental pose updates derived from local depth residuals. The control group performed dense alignment without compression or incremental updates. Both methods used the same initial pose and identical depth input data. Performance indicators included tracking drift over time, average frame rate, peak memory usage, and per-frame computation time. This design allowed direct comparison of efficiency and tracking stability under the same motion patterns.

2.3 Measurement Methods and Quality Control

Tracking accuracy was evaluated by comparing estimated poses with reference transformations defined in the simulation system. Translational drift was calculated as the Euclidean distance between estimated and reference positions after two minutes of continuous tracking. Rotational deviation was measured in degrees using axis-angle representation. Peak memory usage was recorded during each sequence, and frame rate was calculated as the mean number of processed frames per second. Each sequence was repeated three times with different initial perturbations to assess repeatability. Depth noise levels and occlusion ratios remained constant across repeated runs. All experiments were conducted on the same hardware platform to ensure consistent timing and memory measurements.

2.4 Data Processing and Model Formulation

Depth images were converted into three-dimensional point sets using intrinsic camera parameters. Surface compression was performed using vector clustering, where local surface regions were represented by centroid vectors within a reduced codebook. If the original mesh contained MMM vertices and the compressed model contained K clusters ($K < M$), memory usage decreased proportionally to the reduction ratio K/M . Incremental pose updates were computed by minimizing a local depth residual function:

$$E(\Delta R, \Delta t) = \sum_{i=1}^N \|(\Delta R p_i + \Delta t) - q_i\|^2,$$

Where ΔR and Δt represent incremental rotation and translation, and p_i , q_i denote corresponding compressed surface points. Long-term tracking drift was quantified as

$$D = \frac{1}{T} \sum_{t=1}^T \|x_t^{\text{est}} - x_t^{\text{ref}}\|,$$

where T is the total number of frames, and x_t^{est} and x_t^{ref} are estimated and reference positions at frame t .

2.5 Statistical Analysis

Mean values and standard deviations were calculated for tracking drift, frame rate, memory usage, and computation time across all sequences. Paired t-tests were used to compare the proposed method with the dense ICP baseline under identical conditions. Statistical significance was defined as $p < 0.05$. Variance analysis was conducted to assess stability across repeated trials. These procedures ensured objective evaluation of computational efficiency and tracking performance on mobile hardware.

3. Results and Discussion

3.1 Drift Behavior During Continuous Tracking

Across the 50 simulated surgical sequences (18,000 frames), the proposed method limited the mean positional drift to 1.21 mm over 2 min. Drift increased smoothly over time and did not show sudden divergence after rapid viewpoint changes. This pattern indicates that incremental pose updates reduced error accumulation between frames [22,23]. In comparison, dense ICP pipelines typically show larger variability when surface overlap decreases, because correspondence sets become unstable and may lead to local misalignment. Similar time-dependent accuracy variation has been reported in device-level evaluations of mobile depth mapping systems (Fig.1). The present findings indicate that reducing global re-alignment frequency improves long-sequence stability on mobile AR platforms.

3.2 Real-Time Performance on Mobile AR Hardware

The proposed approach sustained 52 FPS on a mobile AR processor and reduced computation time by 38.9% compared with dense point-to-plane ICP. The improvement is related to two factors: a reduced number of active surface elements after quantization and pose refinement based on local residuals instead of full-scene optimization. Wearable surgical AR systems require stable frame rates because temporary delays can interrupt visual guidance even when mean accuracy remains acceptable. Related AR navigation research emphasizes reporting both accuracy and stability to evaluate system reliability (Fig.2). The current results demonstrate that memory-efficient surface modeling combined with incremental updates can meet real-time requirements without significant loss of precision [24,25].

3.3 Memory Efficiency and Structural Preservation

Surface compression reduced peak memory usage by 31.4% compared with full-resolution meshes, while maintaining millimeter-level tracking accuracy. The clustered

representation preserved the main geometric structure required for pose estimation, even though fine detail was reduced. On mobile devices, lower memory demand helps prevent system throttling and frame drops, which indirectly supports tracking stability. Studies evaluating mobile mapping and AR tracking systems have shown that acquisition strategy and data representation influence overall accuracy and consistency (Fig.1). The present results confirm that controlled surface abstraction provides an effective balance between storage efficiency and geometric reliability [26,27].

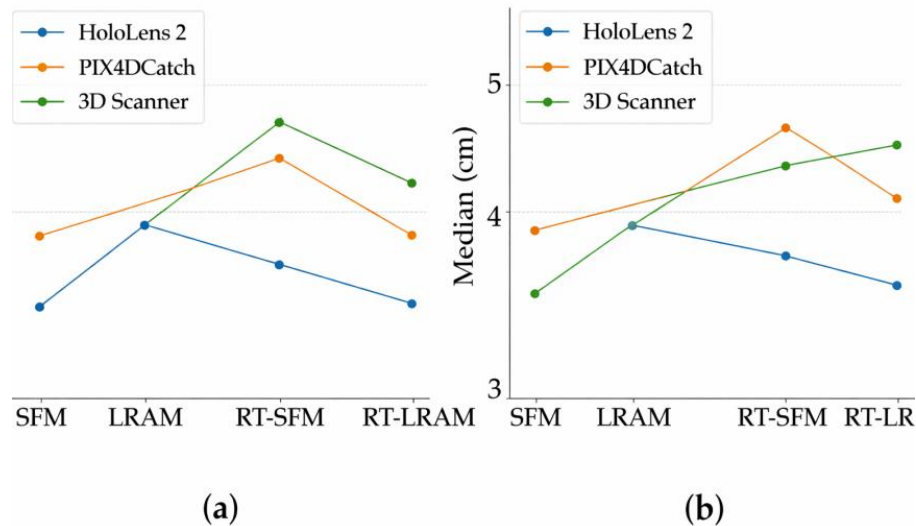


Fig.1. Variation in trajectory error of mobile depth mapping under different scanning strategies.

3.4 Comparison with Related AR Navigation Studies

Compared with dense ICP-based tracking approaches commonly used for depth alignment, the present method supports a low-memory and incremental update strategy that is suitable for mobile surgical AR. Many AR navigation systems report low average error under controlled conditions, yet error distributions often include outliers caused by occlusion or rapid motion. Therefore, stability analysis using boxplots and drift curves is commonly applied to evaluate robustness (Fig.2). The proposed framework achieved competitive accuracy while operating fully on a mobile AR processor. Limitations remain under severe occlusion or abrupt viewpoint changes, where local residual information becomes insufficient. Adaptive overlap monitoring and selective re-initialization may further improve robustness in such conditions.

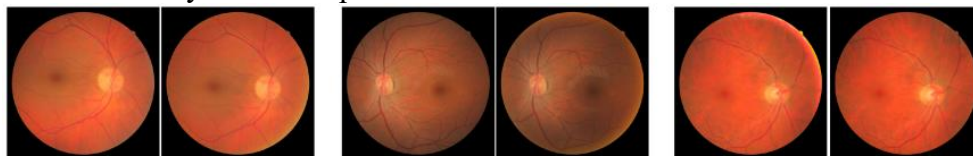


Fig.2. Boxplot of registration error and stability results in AR-assisted surgical navigation.

4. Conclusion

A lightweight depth-based tracking method with memory-efficient surface encoding was evaluated for mobile augmented reality surgical platforms. The approach combines vector-cluster surface compression with incremental pose updates based on local depth residuals. This design reduced memory usage by 31.4% and computation time by 38.9% compared with dense ICP, while maintaining an average drift of 1.21 mm over two minutes at 52 FPS. The results indicate that stable millimeter-level tracking can be achieved on battery-powered head-mounted devices without full-resolution meshes or repeated global optimization. These findings underline the importance of addressing

memory and processing limits directly when designing mobile AR navigation systems. The framework supports real-time surgical guidance in applications that require compact hardware and continuous operation. It is suitable for rigid anatomical tracking in orthopedic and spine procedures using wearable AR devices. Limitations remain in situations with severe occlusion or rapid motion, where reduced overlap weakens depth information. Future improvements may include overlap monitoring, selective re-initialization, or integration of additional sensing data to strengthen robustness during longer surgical sessions.

References

- Zhang, Y., & Wang, J. (2026). Design and Implementation of a Computer-Aided Full Lifecycle Quality Management System for Wind Farms in Upgrades, Renovations, and Subcontractor Supervision.
- Martin-Gomez, A., Li, H., Song, T., Yang, S., Wang, G., Ding, H., ... & Armand, M. (2023). STTAR: surgical tool tracking using off-the-shelf augmented reality head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 30(7), 3578-3593.
- Yang, Y., Leuze, C., Hargreaves, B., Daniel, B., & Baik, F. (2025). EasyREG: Easy Depth-Based Markerless Registration and Tracking using Augmented Reality Device for Surgical Guidance. *arXiv preprint arXiv:2504.09498*.
- Dang, S. T., Dinh, X. M., Than Ngoc, T. A., Nguyen, V. T., Kim, D. T., & Le, X. H. (2025). Adaptive Sliding Mode Control for Robust Trajectory Tracking of Car-Like Mobile Robots in Autonomous Navigation Systems. *Journal of Applied and Computational Mechanics*, e19517.
- Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). The Impact of Visual Language Strategies in Public Art Creation on Community Spatial Perception and Public Behavior.
- Sahu, D., Nidhi, Prakash, S., Pandey, V. K., Yang, T., Rathore, R. S., & Wang, L. (2025). Edge assisted energy optimization for mobile AR applications for enhanced battery life and performance. *Scientific Reports*, 15(1), 10034.
- Liu, S., & Yim, J. (2025). Research on Generative AI Creation Systems Based on Visual Language Modeling: Human-Machine Collaboration and Cognitive Feedback Mechanisms. Available at SSRN 6139770.
- Ribeiro, L. G. (2023). Target Tracking Using Optical Markers for Remote Handling in ITER (Doctoral dissertation, University of Tampere, Finland).
- Wang, Y., Wang, L., Wen, Y., Wu, X., & Cai, H. (2025). Precision-Engineered Nanocarriers for Targeted Treatment of Liver Fibrosis and Vascular Disorders.
- van Doormaal, J. A., & van Doormaal, T. P. (2024). Augmented reality in neurosurgery. *Computational Neurosurgery*, 351-374.
- Gao, G., Ma, X., Lu, C., & Gao, R. (2026). Reliability Analysis and Application Research of SMS Communication Systems in Medical Notification Scenarios.
- Nadeem-Tariq, A., Kazemeini, S., Kaur, P., Dang, G., Davis, T., Sraa, K., ... & Fang, C. (2025). Augmented reality in spine surgery: a narrative review of clinical accuracy, workflow efficiency, and barriers to adoption. *Cureus*, 17(6).
- Wang, Y., Chen, J., Wang, Y., & Yin, X. (2026). Application of Obtainable Biological Agent Characteristics in Efficacy Stratification of Oral Anti-Obesity Drugs.
- Kerkhof, E., Thabit, A., Benmahdjoub, M., Ambrosini, P., van Ginhoven, T., Wolvius, E. B., & van Walsum, T. (2025). Depth-based registration of 3D preoperative models to intraoperative patient anatomy using the HoloLens 2. *International journal of computer assisted radiology and surgery*, 20(5), 901-912.

- Yu, X. (2026). AI-Driven Personalization across Domains for Local Categorical Query Understanding and Context-Aware Retrieval.
- Ahmad, F., Xiong, J., & Xia, Z. (2024). Automatic feature-based markerless calibration and navigation method for augmented reality assisted dental treatment. *IET Cyber-Systems and Robotics*, 6(4), e70003.
- Xiao, H., Zhu, W., Qiu, Z., Zeng, Z., Zhang, S., & Xiao, R. (2025, September). Anti-Drosophila Habituation Clustering for Enhanced Anomaly Detection in Data Streams. In *2025 10th International Conference on Computer and Information Processing Technology (ISCIPT)* (pp. 636-644). IEEE.
- Stathouloupoulos, N., Kanellakis, C., & Nikolakopoulos, G. (2025). Have We Scene It All? Scene Graph-Aware Deep Point Cloud Compression. *IEEE Robotics and Automation Letters*.
- Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2025). Assessing the Role of Adaptive Digital Platforms in Personalized Nutrition and Chronic Disease Management.
- Khan, U., Nawaz, U., Qayyum, A., Ashraf, S., Xie, Y., Khan, M. H., ... & Qadir, J. (2025). Surgical scene understanding in the era of foundation ai models: A comprehensive review. *arXiv preprint arXiv:2502.14886*.
- Chen, X., Xiao, H., Zeng, Z., Zhang, S., & Xiao, R. (2025). Fine-Grained Multivariate Time Series Anomaly Detection via Causal Inference. *Knowledge-Based Systems*, 114765.
- Ramteke, M., & Biradar, V. (2024, April). Advancing Pose Correction Efficiency Through Video Analysis and Incremental Learning in Diverse Domains. In *International Conference on Smart Computing and Informatics* (pp. 369-379). Singapore: Springer Nature Singapore.
- Wang, S. (2025). W-PCA Based Gradient-Free Proxy for Efficient Search of Lightweight Language Models. *arXiv preprint arXiv:2504.15983*.
- Demir, B., Domazet, E., & Mechkaroska, D. (2025). Overview of Memory-Efficient Architectures for Deep Learning in Real-Time Systems. *Engineering Proceedings*, 104(1), 77.
- Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). Quantitative Study on the Construction and Application Effectiveness of Graffiti Wall Painting Teaching Models in Public Space Contexts.
- Raza, A., Hanif, F., & Mohammed, H. A. (2025). Efficient crack and surface-type recognition via CNN-block development mechanism and edge profiling. *Scientific Reports*, 15(1), 40073.
- Wang, Y., Wen, Y., Wu, X., & Cai, H. (2024). Comprehensive Evaluation of GLP1 Receptor Agonists in Modulating Inflammatory Pathways and Gut Microbiota.