Apple vs. Meta: A Comparative Study on Spatial Tracking in SOTA XR Headsets

Article · November 2024

DOI: 10.1145/3636534.3696215

4 authors, including:

Tianyi Hu
Duke University
13 PUBLICATIONS **35** CITATIONS
SEE PROFILE

Tim Scargill
Duke University
28 PUBLICATIONS **153** CITATIONS
SEE PROFILE

Maria Gorlatova
Duke University
105 PUBLICATIONS **1,702** CITATIONS
SEE PROFILE

Apple vs. Meta: A Comparative Study on Spatial Tracking in SOTA XR Headsets

Tianyi Hu Fan Yang Tim Scargill Maria Gorlatova
{tianyi.hu,fy62,ts352,maria.gorlatova}@duke.edu
Department of Electrical and Computer Engineering, Duke University
Durham, NC, USA

ABSTRACT

Inaccurate spatial tracking in extended reality (XR) headsets can cause virtual object jitter, misalignment, and user discomfort, limiting the headsets' potential for immersive content and natural interactions. We develop a modular testbed to evaluate the tracking performance of commercial XR headsets, incorporating system calibration, tracking data acquisition, and result analysis, and allowing the integration of external cameras and IMU sensors for comparison with open-source VI-SLAM algorithms. Using this testbed, we quantitatively assessed spatial tracking accuracy under various user movements and environmental conditions for the latest XR headsets, Apple Vision Pro and Meta Quest 3. The Apple Vision Pro outperformed the Meta Quest 3, reducing relative pose error (RPE) and absolute pose error (APE) by 33.9% and 14.6%, respectively. While both headsets achieved sub-centimeter APE in most cases, they exhibited APE exceeding 10 cm in challenging scenarios, highlighting the need for further improvements in reliability and accuracy.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**.

KEYWORDS

Extended reality, VI-SLAM, Performance characterization

ACM Reference Format:
Tianyi Hu Fan Yang Tim Scargill Maria Gorlatova. 2024. Apple vs. Meta: A Comparative Study on Spatial Tracking in SOTA XR Headsets. In *The 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '24), November 18–22, 2024, Washington D.C., DC, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3636534.3696215

1 INTRODUCTION

Recent advancements in XR headsets, including the Apple Vision Pro (AVP) and Meta Quest 3 (MQ3), have attracted significant interest from developers, consumers, and researchers. Most headsets rely on inside-out tracking, using onboard cameras and inertial measurement unit (IMU) sensors to estimate user movements. Specifically, the AVP uses visual-inertial odometry (VIO) [2, 15], while the MQ3 employs visual-inertial simultaneous localization and mapping (VI-SLAM) [7, 15]. Compared to outside-in tracking, which requires external beacons and strategic beacon placement in each new environment, inside-out tracking offers a simpler, more user-friendly experience [11–13]. However, certain user movements and environmental conditions are known to negatively affect the performance of VIO and VI-SLAM systems [10, 13]. For example, rapid rotations and low-light environments can significantly challenge accurate tracking, potentially resulting in hologram instability, which degrade the overall XR experience [21]. Despite considerable progress over the past decade [12, 19, 24], the latest XR headsets still exhibit tracking issues [6, 26], with users and developers reporting problems such as unexpected hologram drifting.

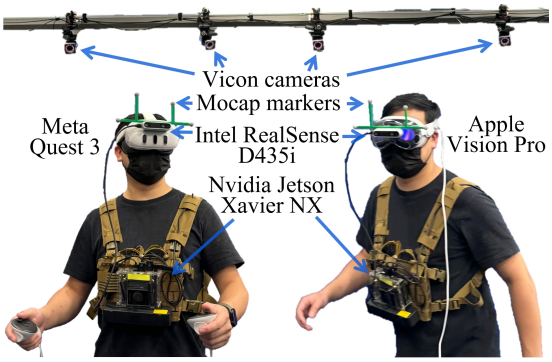


Figure 1: Our XR headset tracking performance testbed.

A comprehensive quantitative evaluation of tracking performance in XR headsets is essential for developers to optimize applications and for users to choose the best headsets for their needs. However, manufacturers like Apple [1] and Meta [16] do not disclose performance analysis results or provide APIs for accessing raw sensor data, nor do they make their spatial tracking algorithms public. This lack of transparency complicates the evaluation of tracking components, making it difficult for developers and consumers to make informed decisions. Previous studies [3, 10, 11, 13, 20] have validated the tracking performance of earlier headset models. Building on these works, we developed a modular testbed to evaluate the tracking accuracy of various XR headsets in comparison to open-source VI-SLAM baselines, under different environmental conditions and user movements.

To address these issues, we designed and implemented a testbed for evaluating XR headset tracking performance under various user movements and environmental conditions. The testbed includes a complete pipeline for system calibration, tracking data acquisition, and result analysis. It also allows for the mounting of an external camera on the headset, allowing comparisons with open-source VI-SLAM algorithms. We performed experiments across different environmental conditions and user motions on the latest XR headsets, the AVP and MQ3, and reported their tracking performance. The testbed calibration code, headset apps, and hardware designs are publicly available on GitHub[1].

Our main contributions are as follows:

- We design a modular testbed incorporating system calibration, tracking data acquisition, and result analysis for XR headset tracking performance evaluation with support of open-source VI-SLAM algorithms as baselines.
- We investigate factors that degrade tracking performance and design experiments under various user motions and environmental conditions.
- We evaluated two state-of-the-art XR headsets, the Apple Vision Pro and Meta Quest 3, across 108 trials on our testbed. Compared to the Meta Quest 3, the Apple Vision Pro demonstrated a 33.9% reduction in RPE and a 14.6% reduction in APE. To the best of our knowledge, we are the first to publicly report their tracking performance.

The rest of this paper is organized as follows: Section 2 reviews related work, while Section 3 presents our testbed system design. Section 4 details our experimental setups, followed by Section 5, which reports our experimental results. Finally, Section 6 provides our conclusions and future work.

2 RELATED WORK

Factors affecting headset tracking: Research on visual-inertial odometry (VIO) and visual-inertial simultaneous localization and mapping (VI-SLAM) has highlighted several factors that hinder tracking performance. Environmental conditions such as low-light scenarios, dynamic lighting changes, and textureless surfaces present significant challenges to these systems [4, 12]. Conversely, feature-rich environments enhance pose estimation robustness [8, 23]. Additionally, rapid camera movements leading to motion blur introduce further ambiguity in feature matching [14, 18]. Our experimental design incorporates these factors to evaluate headset performance under challenging conditions.

XR headset tracking evaluations: Due to the importance of accurate tracking, many studies have explored various methods for evaluating XR headset performance. A common approach is to mount headsets on robotic arms, allowing precise and reproducible tests with pre-programmed movements [10, 13, 20]. While this ensures controlled testing conditions, the constrained range of motion and inability to mimic the complexity of natural human movement limit its effectiveness.

Another approach focuses on hologram drifting, where studies [22, 25, 28] measure the deviation of a virtual object from its original position after user movements. Although drift is primarily caused by pose estimation errors and can affect user experience, this method is labor-intensive, restricts evaluation to virtual content, and only manifests pose errors in application-specific contexts, making it less suitable for comparing headset performance across different settings.

Finally, motion capture (mocap) systems, such as those used by [3, 11, 17], track infrared markers attached on headsets to provide ground truth data for tracking evaluation. While mocap is effective in capturing head movement, previous user studies have limitations. Holzwarth et al. [11] restrict movements to 2D by mounting the headsets on a trolley; Boulo et al. [3] only evaluated straight-line movements; and Monica et al. [17] tests under the same environmental conditions. These experimental results, while valuable, do not fully capture the complexity of both user movement and environmental factors that affecting tracking performance.

To the best of our knowledge, we are the first to report tracking evaluation results of the latest XR headsets, the Apple Vision Pro and Meta Quest 3. Our modular testbed can adapt to different headsets for tracking performance evaluations across various user motions and environmental conditions.

3 SYSTEM DESIGN

Our testbed system comprises three primary hardware components: a target XR headset for evaluation, a mocap station for ground truth trajectories and time synchronization, and a single-board computer running an open-source VI-SLAM algorithm as a baseline. Figure 4 illustrates our testbed pipeline, which can be divided into three main stages: system setup and initialization (§ 3.1), tracking data acquisition (§ 3.2), and tracking data analysis (§ 3.3).

3.1 System Setup and Initialization

We designed a 3D-printed frame to attach infrared markers and an external camera with an IMU sensor to the target headset for obtaining mocap ground truth trajectories and supporting the VI-SLAM baseline, as shown in Figure 2. The frame's rigid body is registered on a mocap tracker, with its center, denoted as $\mathbf{V}$, set on the external camera. The world coordinates of the mocap system are denoted by W, with $({}^{W}_{V}\mathbf{R}, {}^{W}_{V}\mathbf{P})$ representing the rotation and translation of center $\mathbf{V}$ relative to W. Before evaluating the tracking performance, two essential steps are required: XR headset extrinsic calibration and baseline system time synchronization.

[1]https://github.com/hu-tianyi/XRHeadsetTrackingTestbed

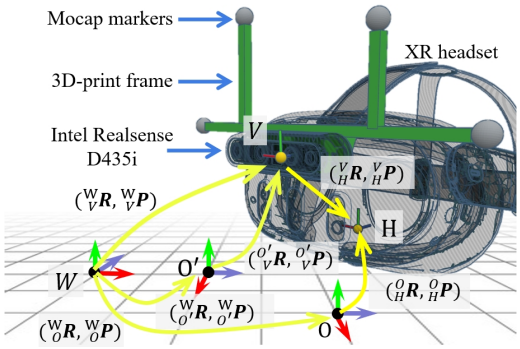


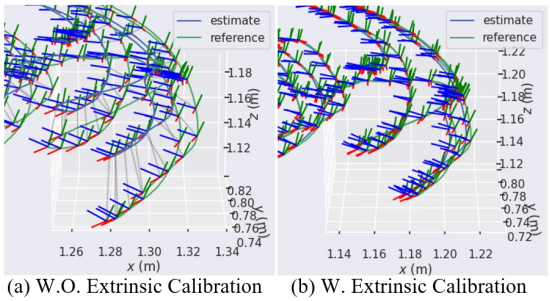
Figure 2: Reference coordinates used in our testbed, where W is the Vicon world coordinates, O is the headset's coordinate, and O' is the open-source SLAM's coordinates.

XR headset extrinsic calibration: As illustrated in Figure 2, the center of the headset, denoted as **H**, may differ from the center of the mocap rigid body, **V**. To derive the ground truth trajectory of **H** from that of **V**, a constant relative transform, represented by $({}^{V}_{H}\mathbf{R}, {}^{V}_{H}\mathbf{P})$, must be determined. However, manufacturers such as Apple and Meta do not disclose the exact physical locations of their headset centers **H**, complicating the determination of the relative transform. To compute this transform, we first record ground truth trajectories of **V**, denoted as $({}^{W}_{V}\mathbf{R}, {}^{W}_{V}\mathbf{P})$, and the headset's estimated trajectories of **H**, denoted as $({}^{O}_{H}\mathbf{R}, {}^{O}_{H}\mathbf{P})$ in a bright, feature-rich environment with slow motion to minimize drift. The calibration process is then formulated as an optimization problem, aiming to adjust the relative transform to minimize the discrepancy between the transformed ground truth trajectory and the estimated headset trajectory:

$$ {}^{V}_{H}\mathbf{P} = \underset{\mathbf{P}}{\operatorname{argmin}} \sum_{t} \left\| \left({}^{W}_{V}\mathbf{R}_{\mathbf{t}} \cdot {}^{V}_{H}\mathbf{R}_{\mathbf{t}} \cdot \mathbf{P} + {}^{W}_{V}\mathbf{P}_{\mathbf{t}} \right) - {}^{W}_{H}\mathbf{P}_{\mathbf{t}}{}' \right\|^2 \quad (1) $$

where t is timestamp, ${}^{V}_{H}\mathbf{R}_{\mathbf{t}} = ({}^{W}_{V}\mathbf{R}_{\mathbf{t}})^{-1} \cdot {}^{W}_{H}\mathbf{R}_{\mathbf{t}}{}'$ and ${}^{W}_{H}\mathbf{R}', {}^{W}_{H}\mathbf{P}'$ are the estimated headset rotation and position in world coordinates, obtained by aligning $({}^{O}_{H}\mathbf{R}, {}^{O}_{H}\mathbf{P})$ with the ground truth trajectory using Umeyama's alignment method [27]. The results of extrinsic calibration are illustrated in Figure 3, showing an average point distance of 0.3 cm. This marks a substantial improvement over the pre-calibration average distance of 11.0 cm, highlighting the critical role of the calibration process in ensuring reliable and accurate ground truth data.

Baseline system time synchronization: Time synchronization is crucial for accurately evaluating trajectories from different devices, as their local system clocks may differ. We synchronize the baseline device's local time with the mocap station using the Network Time Protocol (NTP) over a single hop in a local area network. The mocap station functions as the NTP server, and the baseline device operates as the client. After synchronization, the time discrepancy between the baseline device and the mocap station is reduced to less than 1 millisecond.



(a) W.O. Extrinsic Calibration (b) W. Extrinsic Calibration

Figure 3: Alignment between ground truth trajectory and headset estimated trajectory before and after extrinsic calibration.

3.2 Tracking Data Acquisition

During the tracking data acquisition stage, the testbed simultaneously collects three trajectories: the ground truth trajectory from the mocap station, the estimated trajectory from the target XR headset, and the estimated trajectory from the open-source VI-SLAM system used as a baseline.

Mocap system: The mocap system tracks the target headset using infrared markers attached to headset. These markers reflect infrared light, which is captured by Mocap cameras positioned around the tracking area. The mocap tracker triangulates the position of each marker in three-dimensional space, thereby precisely determining the orientation ${}^{W}_{V}\mathbf{R}$ and location ${}^{W}_{V}\mathbf{P}$ of the rigid body's center **V** in real time. In our setup, the center point **V** is positioned on the external camera, as shown in Figure 2. By applying the relative transform $({}^{V}_{H}\mathbf{R}, {}^{V}_{H}\mathbf{P})$—obtained during extrinsic calibration—we can derive the ground truth trajectory of the target headset's center **H**, denoted as ${}^{W}_{H}\mathbf{R}$ and ${}^{W}_{H}\mathbf{P}$.

XR headset: During runtime, the XR headset uses its tracking cameras and IMU sensors to estimate user movement. We develop customized apps that invoke headset APIs to obtain the estimated trajectory of the device center **H**, denoted as $({}^{O}_{H}\mathbf{R}, {}^{O}_{H}\mathbf{P})$. Depending on the operating system, different APIs are used: the ARKit API queryDeviceAnchor() for the AVP and the Unity Engine API ovrcamerarig.centerEyeAnchor for the MQ3. Since we do not have system privileges to synchronize the headset clock via NTP, we implement a synchronization method that allows the mocap station to transmit its latest timestamp to the headset over a single-hop wireless network. The headset apps we developed record the headset's estimated trajectory using the mocap timestamp instead of the headset's local timestamp. We measure the roundtrip time for sending the timestamp from the mocap station to the headset and back. The average roundtrip delay is 10.57 milliseconds for the AVP and 10.28 milliseconds for

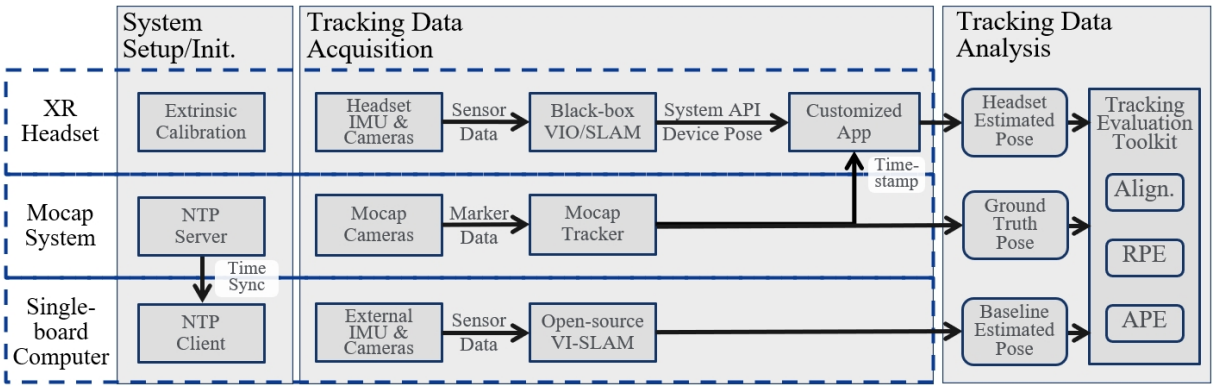


Figure 4: Testbed pipeline for XR headset tracking performance evaluation.

the MQ3, indicating a one-way latency of about 5 milliseconds, which is negligible in trajectory evaluation.

Baseline VI-SLAM system: During the tracking data acquisition stage, the user wears a tactical vest with a single-board computer running an open-source VI-SLAM system, serving as the baseline for tracking performance evaluation. As illustrated in Figure 1, the computer connects to an external camera and IMU mounted on the target headset. The baseline VI-SLAM system estimates its rotation and position, denoted as $({}^{O'}_{V}\mathbf{R}, {}^{O'}_{V}\mathbf{P})$ in its coordinate system $\mathbf{O}'$, at runtime. Since the mocap rigid body's center $\mathbf{V}$ is positioned on the external camera and time synchronization via NTP is performed during system setup, no further calibration or synchronization is required for the baseline system.

3.3 Tracking Data Analysis

We evaluate the estimated trajectories from both the baseline system and the XR headset against the original and extrinsic-calibrated ground truth trajectories. We use the commonly employed EVO toolkit [9] for trajectory alignment and performance assessment. We report two metrics: relative pose error (RPE) and absolute pose error (APE). RPE divides the estimated trajectory into fixed-length subtrajectories, aligning each starting point with the ground truth and measuring the pose error at the subtrajectory's endpoint [5, 29]. This metric captures local drift, preventing error accumulation by aligning each subtrajectory independently. APE, on the other hand, aligns the entire estimated trajectory with the ground truth and calculates the pose error at each timestamp [9, 29]. This metric shows the overall drift, as errors accumulate without periodic realignment.

4 EXPERIMENT SETUP

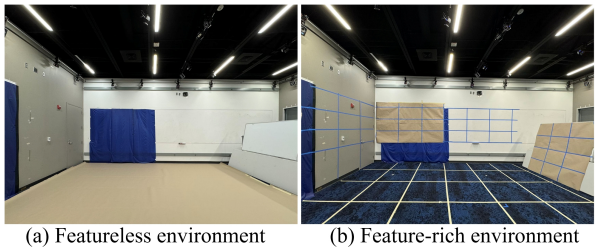
To assess XR headset tracking performance relative to an open-source VI-SLAM baseline, we design experiments involving various user motions and environmental conditions

As detailed in Table 1, our experiments are grouped into the four categories: Patrol, Inspection, Head Rotation, and Brightness—the first three focus on different movement patterns, while the last examines environmental brightness levels. Each movement pattern is evaluated with four trajectory settings, varying by speeds and environmental feature levels. The brightness experiments included six trajectory settings combining three brightness levels and two feature levels. Each setting was run in three trials to report the mean and standard deviation of tracking errors. We evaluated the AVP and MQ3 in these experiments (Figure 1), with a total of 108 trials. Our findings are discussed in Section 5.

Table 1: Experiment design that covers different user movement patterns, speeds, environment brightness levels, and feature levels.

Category	Traj. ID	Description	Note
Patrol Movement	*P1*	Low speed + Feature-rich	~0.75 m/s
	P2	Low speed + Featureless	
	P3	High speed + Feature-rich	~1.0 m/s
	P4	High speed + Featureless	
Inspection Movement	*I1*	Low speed + Feature-rich	~0.75 m/s
	I2	Low speed + Featureless	
	I3	High speed + Feature-rich	~1.0 m/s
	I4	High speed + Featureless	
Head Rotation Movement	*R1*	Low speed + Feature-rich	~0.08 m/s
	R2	Low speed + Featureless	
	R3	High speed + Feature-rich	~0.2 m/s
	R4	High speed + Featureless	
Brightness	*B1*	High light + Feature-rich	~358 lux
	B2	High light + Featureless	
	B3	Low light + Feature-rich	~133 lux
	B4	Low light + Less feature	
	B5	Dim light + Feature-rich	~24 lux
	B6	Dim light + Featureless	

All experiments were conducted in a 6 m × 6 m mocap room equipped with 24 Vicon Vero v2.2 cameras and 4 Vicon Vantage v5 cameras. The Vicon mocap system was calibrated to achieve an average error of less than 0.4mm, providing accurate ground truth for our measurements.



(a) Featureless environment (b) Feature-rich environment

Figure 5: Testing environment with different feature levels, where (a) the carpet is covered for a featureless environment (b) blocks applied to the carpet and walls for feature-rich environments.

4.1 XR Headset Specifications

Apple Vision Pro: Released in February 2024 [1], the AVP features two side cameras, four downward cameras, and four IMU sensors for spatial tracking. It is powered by an M2 processor and a dedicated R1 chip for processing camera and sensor data. We evaluated the AVP using Vision OS v1.2.

Meta Quest 3: Released in October 2023 [16], the MQ3 is equipped with four tracking cameras—two front-facing and two side-facing—and is powered by a Snapdragon XR2 Gen2 processor. We evaluated the MQ3 with Horizon OS v67.

Open-source VI-SLAM platform: We use a NVIDIA Jetson Xavier NX running an open-source VI-SLAM system, which employs an Intel RealSense D435i camera to capture forward-facing stereo frames and IMU data. ORB-SLAM3 [5] in stereo-inertial mode was used to estimate trajectories, serving as our baseline.

4.2 Environment Conditions

To evaluate the impact of environmental features, we created both featureless and feature-rich settings (Figure 5). The featureless environment included blank walls, whiteboards, curtains, and craft paper covering the carpet. In the feature-rich environment, we added distinct features using tape to create $1\text{m} \times 1\text{m}$ and $1\text{m} \times 0.5\text{m}$ blocks on the floor and walls. To assess the impact of brightness, we conducted low-speed patrol movements under three lighting conditions: high (358 lux), low (133 lux), and dim (24 lux).

4.3 User Movement

To evaluate the headsets, we employ user motion patterns established in previous works [12, 22], focusing on three common movements: patrol, inspection, and head rotation.

Patrol: This pattern reflects a user navigating through an environment while maintaining their gaze in the direction of movement. We employ a square trajectory with 3-meter sides in both clockwise and counterclockwise directions. This movement presents challenges due to frequent changes in visual content, especially during turns.

Inspection: This pattern replicates typical user behavior when examining a virtual object from various angles. To ensure reproducibility, we use a circular trajectory with a 3-meter diameter, moving in both clockwise and counterclockwise directions while maintaining gaze fixed at the circle's center.

Head rotation: In this pattern, a user explores virtual content while remaining stationary. The user stands still and rotates their head in three clockwise circles, followed by three counterclockwise circles.

To maintain consistent testing conditions across different headsets and trials, we employ a single experienced AR user to complete all 108 trials. We use small stickers on the floor and walls to guide the user's movement and gaze. For experiments involving varying movement speeds, such as slow walking at 0.75 m/s and fast walking at 1.0 m/s, we monitor the user's speed during each trial and provide real-time feedback to ensure the desired pace is maintained. This approach ensures consistent user movement speeds, as detailed in Table 2 in the Appendix.

5 EXPERIMENTAL RESULTS

We present our experimental results in bar charts in Figure 6 and provide the complete tabulated results in the appendix in Table 2.[2] Our results indicate that the AVP achieved an average RPE of 0.52 cm and APE of 6.98 cm, outperforming the MQ3, which recorded an RPE of 0.79 cm and APE of 7.99 cm. In experiment R1, involving slow head rotations in a bright, feature-rich environment, both headsets showed RPEs under 0.3 cm and APEs under 0.8 cm. However, under challenging conditions like experiment B6, conducted in a dim, featureless environment, both headsets exhibited APEs exceeding 10 cm, highlighting needs for improvement.

5.1 Device Specifications

Significant performance differences were observed among the XR headsets. The AVP consistently outperformed the MQ3, with a 33.9% lower RPE and 14.6% lower APE on average. The baseline VI-SLAM system exhibited the worst performance, often losing tracking in low-light and featureless environments.

The AVP's superior performance can be attributed to its six tracking cameras, compared to the MQ3's four and the baseline's two. This wider camera array provides enhanced environmental perception, improving robustness under challenging conditions. Additionally, the AVP's R1 chip enables faster sensor data processing at 100 Hz, enhancing feature matching during rapid movements. In contrast, the baseline

[2]In Figure 6 and Table 2, certain ORB-SLAM3 settings lack data due to frequent tracking failures in challenging environments.

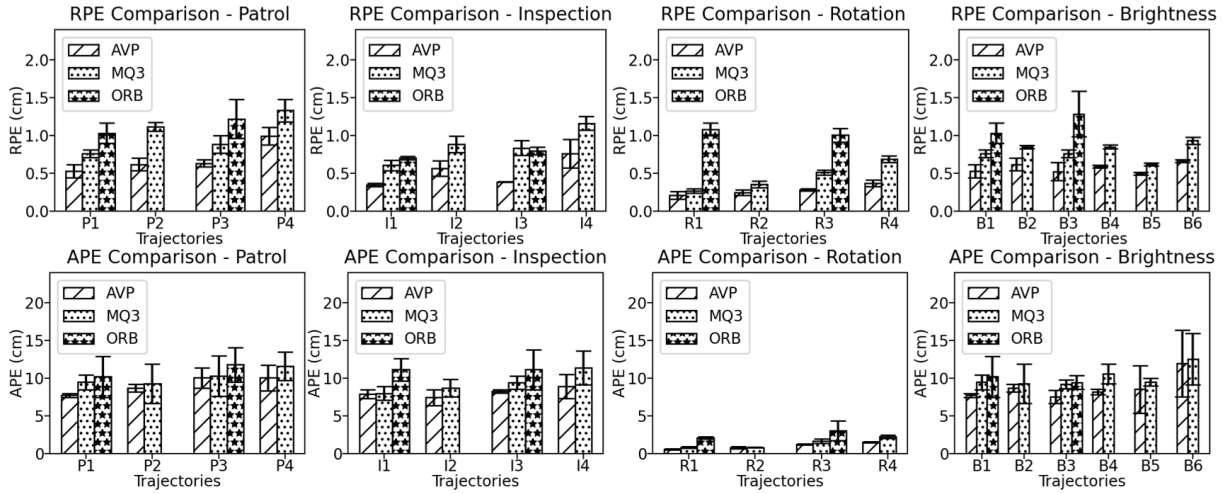


Figure 6: Evaluation results of XR headsets under different user movement and environment conditions with an open-source VI-SLAM baseline.

system, without hardware acceleration, processes at 30 Hz, making it more error-prone during fast movements.

5.2 User Movement

Our experiments show that both movement patterns and movement speeds significantly affect tracking performance.

Movement speed: Faster movement consistently increased tracking errors, especially in RPE. In P1 and P3, increasing speed from 0.75 m/s (slow walk) to 1.0 m/s (fast walk) raised RPE by 16.6% for the AVP and 19.9% for the MQ3. In featureless environments (P2 and P4), the same speed increase resulted in a 60.6% rise in RPE for the AVP.

Movement pattern: Head rotations yielded the best tracking accuracy, with RPEs and APEs in the millimeter range, suggesting that pose errors are primarily due to user movement. Patrol movements showed the worst performance, being more sensitive to environmental changes. Inspection movements outperformed patrol, indicating that consistent visual input, like focusing on the same object, improves tracking compared to scenes with frequent visual changes.

5.3 Environmental conditions

Our results reveal that environmental factors, such as feature level and brightness, greatly impact tracking performance.

Feature level: Environmental features are critical for accurate tracking. Both headsets showed reduced performance in RPE and APE when transitioning from feature-rich to featureless environments. Featureless settings frequently caused initialization failures or tracking loss, particularly for the baseline method. For instance, in high-brightness conditions, switching from a feature-rich (B1) to a featureless (B2) environment increased APE by 11.7% for the AVP. In low-brightness conditions, this transition (B5 to B6) led to a more pronounced 40.2% rise in APE.

Brightness level: Tracking performance remained stable as brightness decreased from 358 lux (B1, B2) to 133 lux (B3, B4). However, at 24 lux (B5, B6), APE increased significantly, with larger standard deviations, indicating greater drift and inconsistent tracking under low-light conditions.

6 CONCLUSIONS

In this work, we present a modular testbed for evaluating the tracking performance of the latest XR headsets under various user movements and environmental conditions, using an open-source VI-SLAM system as the baseline. Our approach facilitates the measurement and comparison of devices such as the AVP and MQ3, offering insights into performance data typically undisclosed by manufacturers. Results from our testbed show that the AVP outperforms the MQ3, with reductions in RPE and APE by 33.9% and 14.6%, respectively. While both headsets achieve sub-centimeter tracking accuracy in most cases, they exhibit considerable degradation in performance under certain conditions, with APEs exceeding 10 cm. This underscores the need for more robust and precise spatial tracking systems in XR headsets. Our methodology not only reveals otherwise safeguarded performance metrics but also provides valuable data for developers and users who rely on accurate and reliable tracking. As XR technology advances, our approach can be adapted to assess future headsets, ensuring continued transparency and improvement in tracking capabilities.

ACKNOWLEDGMENTS

We thank Prof. Boyuan Chen and Duke Robotics Lab for sharing their mocap facility. This work was supported in part by NSF grants CSR-2312760, CNS-2112562 and IIS-2231975, NSF CAREER Award IIS-2046072, NSF NAIAD Award 2332744, a CISCO Research Award, and a Meta Research Award.

REFERENCES

[1] Apple. 2024. Apple. https://www.apple.com/.

[2] Apple. 2024. Understanding World Tracking. https://developer.apple.com/documentation/arkit/arkit_in_ios/configuration_objects/understanding_world_tracking/

[3] Joris Boulo, Andréanne K Blanchette, Alexandra Cyr, and Bradford J McFadyen. 2024. Validity and reliability of the tracking measures extracted from the Oculus Quest 2 during locomotion. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 12, 1 (2024), 2274391.

[4] Mihai Bujanca, Xuesong Shi, Matthew Spear, Pengpeng Zhao, Barry Lennox, and Mikel Luján. 2021. Robust SLAM systems: Are we there yet?. In *Proceedings of IEEE/RSJ IROS*.

[5] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. 2021. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. *IEEE Transactions on Robotics* 37, 6 (2021), 1874–1890.

[6] Julian Chokkattu. 2024. Review: Apple Vision Pro. https://www.wired.com/review/apple-vision-pro/.

[7] Alvaro Collet and Tom Meyer. 2017. SLAM: Bringing art to life through technology. https://engineering.fb.com/2017/09/21/virtual-reality/slam-bringing-art-to-life-through-technology/

[8] Luca Ferranti, Xiaotian Li, Jani Boutellier, and Juho Kannala. 2021. Can you trust your pose? Confidence estimation in visual localization. In *Proceedings of IEEE ICPR*.

[9] Michael Grupp. 2017. EVO: Python package for the evaluation of odometry and SLAM. https://github.com/MichaelGrupp/evo.

[10] Joel Hesch, Anna Kozminski, and Oskar Linde. 2020. Measuring the accuracy of inside-out tracking in XR devices using a high-precision robotic arm. In *Proceedings of Springer HCI International*. Springer.

[11] Valentin Holzwarth, Joy Gisler, Christian Hirt, and Andreas Kunz. 2021. Comparing the accuracy and precision of SteamVR tracking 2.0 and Oculus Quest 2 in a room scale setup. In *Proceedings of ACM ICVARS*.

[12] Li Jinyu, Yang Bangbang, Chen Danpeng, Wang Nan, Zhang Guofeng, and Bao Hujun. 2019. Survey and evaluation of monocular visual-inertial SLAM algorithms for augmented reality. *Virtual Reality & Intelligent Hardware* 1, 4 (2019), 386–410.

[13] Tyler A Jost, Bradley Nelson, and Jonathan Rylander. 2021. Quantitative analysis of the Oculus Rift S in controlled movement. *Disability and Rehabilitation: Assistive Technology* 16, 6 (2021), 632–636.

[14] Peidong Liu, Xingxing Zuo, Viktor Larsson, and Marc Pollefeys. 2021. MBA-VO: Motion blur aware visual odometry. In *Proceedings of IEEE/CVF ICCV*.

[15] Lumafield. 2024. Apple Vision Pro and Meta Quest Non-Destructive Teardown. https://www.lumafield.com/article/apple-vision-pro-meta-quest-pro-3-non-destructive-teardown

[16] Meta. 2024. Meta. https://www.meta.com/.

[17] Riccardo Monica and Jacopo Aleotti. 2022. Evaluation of the Oculus Rift S tracking system in room scale virtual reality. *Virtual Reality* 26, 4 (2022), 1335–1345.

[18] Luigi Nardi, Bruno Bodin, M Zeeshan Zia, John Mawer, Andy Nisbet, Paul HJ Kelly, Andrew J Davison, Mikel Luján, Michael FP O'Boyle, Graham Riley, et al. 2015. Introducing SLAMBench, a performance and accuracy benchmarking methodology for SLAM. In *Proceedings of IEEE ICRA*.

[19] Ihsan Rabbi and Sehat Ullah. 2013. A survey on augmented reality challenges and tracking. *Acta graphica: znanstveni časopis za tiskarstvo i grafičke komunikacije* 24, 1-2 (2013), 29–46.

[20] Christoph Runde. 2021. Benchmarking of V/AR Components: Set-Up of a V/AR Measurement Lab and Technical Comparison of Headsets and Tracking. https://www.awexr.com/usa-2021/agenda/2265-benchmarking-of-xr-components-set-up-of-an-xr-meas

[21] Tim Scargill, Shreya Hurli, Jiasi Chen, and Maria Gorlatova. 2021. Will it move? Indoor scene characterization for hologram stability in mobile AR. In *Proceedings of ACM HotMobile*.

[22] Tim Scargill, Gopika Premsankar, Jiasi Chen, and Maria Gorlatova. 2022. Here to stay: A quantitative comparison of virtual object stability in markerless mobile AR. In *Proceedings of IEEE/ACM CPHS Workshop (co-located with CPS-IoT week 2022)*.

[23] Johannes L Schonberger and Jan-Michael Frahm. 2016. Structure-from-motion revisited. In *Proceedings of IEEE/CVF CVPR*.

[24] Xingdong Sheng, Shijie Mao, Yichao Yan, and Xiaokang Yang. 2024. Review on SLAM algorithms for Augmented Reality. *Displays* (2024), 102806.

[25] Carter Slocum, Xukan Ran, and Jiasi Chen. 2021. RealityCheck: A tool to evaluate spatial inconsistency in augmented reality. In *Proceedings of IEEE ISM*.

[26] Alan Truly. 2024. The most common Quest 3 problems and how to fix them. https://www.digitaltrends.com/computing/common-quest-3-problems-and-how-to-fix-them/.

[27] Shinji Umeyama. 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 13, 04 (1991), 376–380.

[28] Reid Vassallo, Adam Rankin, Elvis CS Chen, and Terry M Peters. 2017. Hologram stability evaluation for Microsoft HoloLens. In *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*, Vol. 10136. Spie, 295–300.

[29] Zichao Zhang and Davide Scaramuzza. 2018. A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry. In *Proceedings of IEEE/RSJ IROS*.

A APPENDIX

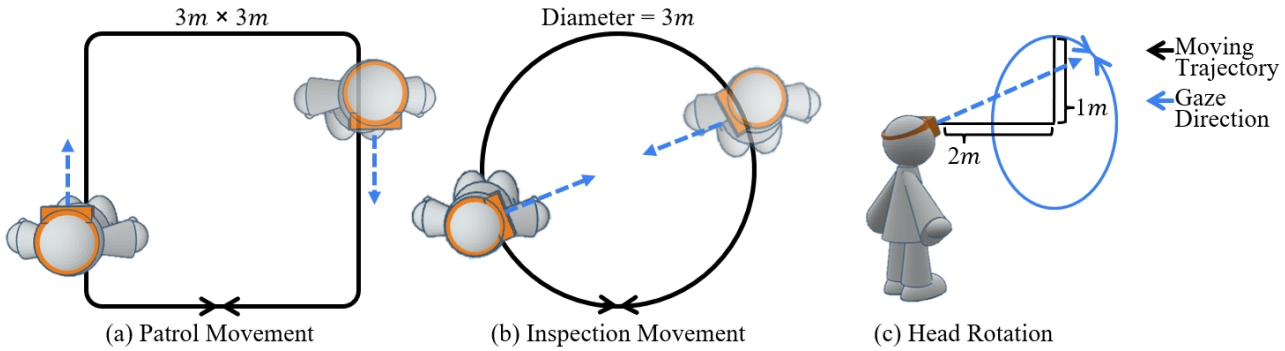


Figure 7: Our experiments encompass three movement patterns: (a) Patrol movement, where the user follows a square trajectory with their gaze aligned to the direction of movement; (b) Inspection movement, where the user follows a circular trajectory while keeping their gaze fixed at the center; (c) Head rotation movement, where the user's body remains stationary while rotating the head in circles.

Table 2: Tracking evaluation results of Apple Vision Pro, Meta Quest 3, and the open-source baseline, ORB-SLAM3 on Intel RealSense D435i.

Trial Group	AVP_RPE (m) (mean±SD)	AVP_APE (m) (mean±SD)	AVP_Speed (m/s) (mean±SD)	MQ3_RPE (m) (mean±SD)	MQ3_APE (m) (mean±SD)	MQ3_Speed (m/s) (mean±SD)	ORB_RPE (m) (mean±SD)	ORB_APE (m) (mean±SD)
B1	0.0053±0.0009	0.0772±0.0026	0.7681±0.0393	0.0076±0.0005	0.0944±0.0100	0.7644±0.0248	0.0103±0.0014	0.1017±0.0271
B2	0.0062±0.0008	0.0867±0.0051	0.7373±0.0294	0.0085±0.0002	0.0926±0.0258	0.7648±0.0309	NaN±NaN	NaN±NaN
B3	0.0052±0.0012	0.0752±0.0087	0.7621±0.0181	0.0076±0.0005	0.0918±0.0057	0.7739±0.0082	0.0128±0.0030	0.0943±0.0089
B4	0.0059±0.0002	0.0816±0.0036	0.7540±0.0252	0.0085±0.0002	0.1055±0.0129	0.7439±0.0158	NaN±NaN	NaN±NaN
B5	0.0049±0.0002	0.0850±0.0313	0.7848±0.0106	0.0062±0.0002	0.0947±0.0052	0.7890±0.0180	NaN±NaN	NaN±NaN
B6	0.0066±0.0002	0.1192±0.0442	0.7851±0.0237	0.0093±0.0005	0.1251±0.0340	0.7562±0.0220	NaN±NaN	NaN±NaN
I1	0.0034±0.0002	0.0789±0.0057	0.7472±0.0347	0.0060±0.0007	0.0797±0.0092	0.7596±0.0129	0.0070±0.0002	0.1112±0.0150
I2	0.0056±0.0010	0.0741±0.0104	0.7405±0.0344	0.0088±0.0011	0.0868±0.0118	0.7361±0.0182	NaN±NaN	NaN±NaN
I3	0.0038±0.0000	0.0823±0.0022	0.9830±0.0127	0.0083±0.0010	0.0941±0.0086	1.0317±0.0495	0.0080±0.0005	0.1111±0.0263
I4	0.0076±0.0019	0.0890±0.0158	0.9649±0.0331	0.0116±0.0009	0.1136±0.0223	1.0154±0.0412	NaN±NaN	NaN±NaN
P1	0.0053±0.0009	0.0772±0.0026	0.7681±0.0393	0.0076±0.0005	0.0944±0.0100	0.7644±0.0248	0.0103±0.0014	0.1017±0.0271
P2	0.0062±0.0008	0.0867±0.0051	0.7373±0.0294	0.0111±0.0006	0.0926±0.0258	0.7448±0.0309	NaN±NaN	NaN±NaN
P3	0.0063±0.0005	0.1003±0.0134	1.0304±0.0451	0.0088±0.0011	0.1029±0.0270	0.9858±0.1440	0.0122±0.0026	0.1178±0.0228
P4	0.0099±0.0012	0.1005±0.0170	1.0332±0.0480	0.0133±0.0015	0.1158±0.0187	1.0500±0.0290	NaN±NaN	NaN±NaN
R1	0.0021±0.0005	0.0056±0.0008	0.0779±0.0042	0.0027±0.0003	0.0081±0.0011	0.0810±0.0139	0.0108±0.0008	0.0207±0.0016
R2	0.0024±0.0003	0.0079±0.0013	0.0882±0.0084	0.0035±0.0004	0.0080±0.0006	0.0807±0.0017	NaN±NaN	NaN±NaN
R3	0.0028±0.0001	0.0121±0.0007	0.1993±0.0178	0.0051±0.0003	0.0163±0.0031	0.2145±0.0151	0.0100±0.0009	0.0304±0.0131
R4	0.0037±0.0004	0.0147±0.0008	0.2180±0.0118	0.0068±0.0004	0.0221±0.0021	0.2033±0.0204	NaN±NaN	NaN±NaN