Comparing Fiducial Marker Tracking Across Cameras in Virtual and Physical Environments

Felix Krumstroh [1], Andrés Eisenmann [1], Jannik Franssen [1], Moritz Rindermann [1], and Thies Pfeiffer [1]

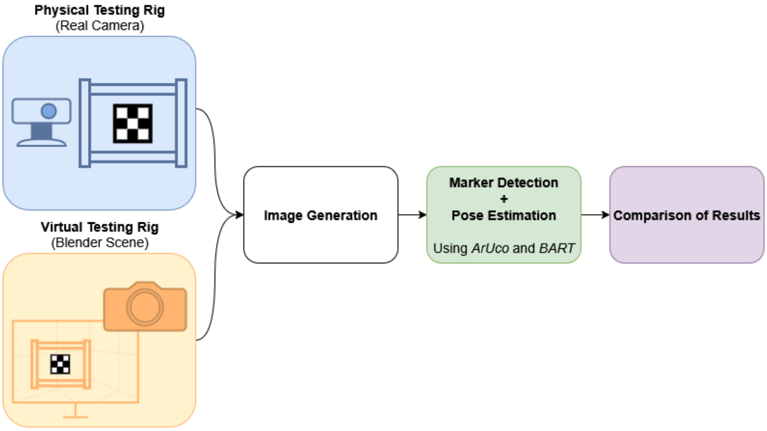

Fig. 1: Schematic overview of the hybrid evaluation pipeline for sim-to-real comparison and algorithm benchmarking

Abstract: Robust object tracking is a foundational technology for augmented and virtual reality (AR/VR) systems. While many benchmarking efforts rely solely on simulated environments, these often lack real-world fidelity, limiting practical insights. To bridge this gap, we propose a hybrid evaluation approach, combining a physical test rig—constructed from aluminum extrusion frames and ArUco markers—with a digital twin created in Blender. Our approach facilitates direct sim-to-real comparisons of fiducial marker tracking performance. We benchmark the widely used OpenCV ArUco (*A*ugmented *R*eality library from the *U*niversity of *C*órdoba) against a custom-optimized tracking approach, BART (*B*ielefeld *A*ugmented *R*eality *T*racker), a modern performance-optimized alternative, to evaluate whether tracking limitations stem from the ArUco implementation itself or are inherent to fiducial marker tracking in general. Our evaluation focuses on detection accuracy and pose estimation performance. The results offer practical insight into the differences between tracking performance in real and simulated environments.

Keywords: Object Tracking, Augmented Reality (AR), ArUco, Fiducial Marker Tracking, Benchmarking

[1] University of Applied Sciences Emden/Leer, Department of Technology, Constantiaplatz 4, 26723 Emden, Germany, felix.krumstroh@stud.hs-emden-leer.de, https://orcid.org/0009-0004-4480-0250; andres.antonio.eisenmann.franco@stud.hs-emden-leer.de, https://orcid.org/0009-0002-8612-9990; jannik.franssen@hs-emden-leer.de, https://orcid.org/009-0004-9690-8589; moritz.rindermann@stud.hs-emden-leer.de, https://orcid.org/0009-0000-9938-8698; thies.pfeiffer@hs-emden-leer.de, https://orcid.org/0000-0001-6619-749X

doi:10.18420/vrar2025_08

1 Introduction

Marker-based tracking is a foundational component in Extended Reality (XR) applications, particularly for spatial alignment between physical environments and their digital twins [Çö20; So23; Zh25]. It enables XR systems to anchor virtual content in real-world space, supporting use cases such as step-by-step maintenance assistance, part localization, and live system visualization. However, assessing the reliability of marker tracking remains challenging, as it typically requires extensive real-world testing under diverse and often unpredictable environmental conditions; ranging from variable lighting and reflective surfaces to differences in hardware like headsets and webcams [Me20; Su24].

In industrial contexts, recreating real-world conditions across different hardware configurations can be both time-consuming and costly, particularly for large-scale spaces such as factory floors or construction sites [Ma21]. This motivates the use of simulation-based testing as a low-risk, cost-effective pre-validation step. However, such simulations are often idealized—featuring simplified lighting, perfect camera parameters, and noise-free rendering, which may lead to overly optimistic assessments of tracking performance [Bl14; WSB23].

This paper presents a hybrid evaluation approach that combines a physical test rig with a digital twin created in Blender to assess both simulation fidelity and algorithmic performance in fiducial marker tracking. We benchmark the widely used OpenCV ArUco implementation against BART (Bielefeld Augmented Reality Tracker), a modern alternative, to distinguish between implementation-specific constraints and fundamental detection challenges. Fig. 1 illustrates our workflow, showing how data from both physical and virtual domains feed into a unified analysis process for direct comparison across multiple camera systems and tracking algorithms.

2 Related Work

Prior research of fiducial marker tracking has largely focused on algorithmic performance improvements and physical setup optimizations [Be24; KYW18; RMM18; RMM20]. Previous studies, such as one by Merino et al. [Me20], have evaluated marker detection under different lighting conditions, noise models, and distances, showing that detection accuracy is highly sensitive to external conditions such as glare, marker degradation, or motion blur. Additionally, several works explore aspects such as adaptive thresholding, sub-pixel corner refinement, and improved dictionary design to increase robustness [Ga14; Ga15].

However, few works have systematically compared real-world tracking performance with simulated equivalents [DRP15; SPS24]. While most benchmarking efforts focus on environmental factors affecting tracking performance, fewer studies systematically compare different algorithmic approaches under identical conditions. This gap is particularly relevant given the dominance of OpenCV's ArUco implementation in research and development.

2.1 Simulation-Based Evaluation

Previous evaluations focus either on physical testing or on fully synthetic data generation, without concern for how well the simulation reflects actual detection behavior. Synthetic testing approaches have explored tools such as Unity, Unreal Engine, and Blender for prototyping. Since these programs are developed primarily for real-time rendering, visual effects, and content creation, these environments tend to lack the ability to realistically simulate real-world optical effects such as lens distortion or sensor noise profiles, which are particularly relevant for devices like the *Meta Quest 3* [Ba24].

Diekmann; Renner; Pfeiffer [DRP15] presented a benchmarking framework based on synthetically generated video sequences to evaluate marker tracking algorithms under controlled, dynamic conditions. While effective for assessing algorithmic robustness, their approach did not prioritize high-fidelity environmental modeling, nor did it establish the external validity of simulated results with respect to real-world performance. Notably, their work included comparisons between tracking approaches such as BART and ArUco, which are also examined in our study. In a related line of research, Sivov; Poroykov; Shmatko [SPS24] evaluated pose estimation accuracy by comparing physical and Unity-based virtual environments. Although positional estimates aligned closely, significant depth discrepancies were observed in the real-world tests, primarily due to unmodeled sensor noise and optical distortions. These findings underscore the need to assess not only simulation fidelity but also the predictive validity of synthetic results across varying camera systems and acquisition conditions.

In contrast, our approach emphasizes simulation validity of simulation results how well outcomes from a virtual replica correspond to real-world tracking performance. Rather than treating synthetic data as generic input, we replicate physical measurements, lighting, and camera behavior within the virtual scene.

2.2 Detection Optimizations

Various enhancements to marker detection algorithms have been proposed over the years. The use of Kalman filters has been explored to stabilize detection over time, particularly under jittery or occluded conditions. An optimization also incorporated in approaches such as BART, a marker tracking system developed by the University of Bielefeld that prioritizes real-time processing for mobile and XR platforms [KYW18]. Recent work on dictionary optimization by Garrido-Jurado et al. [Ga15] introduced automated methods for generating marker dictionaries with maximal inter-marker distance, reducing detection ambiguity and improving robustness; particularly in challenging conditions. This has informed newer ArUco dictionaries' design such as `5x5_1000` used in this study.

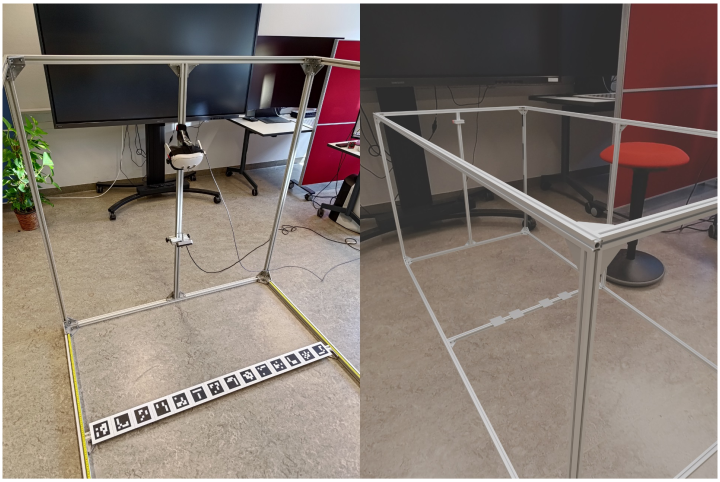

Fig. 2: Image of the physical (left) and simulated (right) setup.

3 Methodology

To compare the real and virtual experimental environments, we implemented the following approach:

3.1 Testing environments

- **Physical Testing Rig:** To ensure reliable and accurate real-world camera testing, we built a precise physical rig using aluminum extrusion profiles. This material provides a cost-effective balance of stability, modularity, and ease of assembly. The camera was mounted on a central vertical profile, with the ability to adjust its height and tilt angle. We conducted detections at 10 cm distance increments, both with the camera facing straight ahead and tilted 45° downward. These test configurations reflect a wide range of typical marker viewing conditions in real-world applications. All values were verified using digital protractors and measurement scales. Subsequent reviewing of the setup demonstrated an error tolerance below 1 mm.
- **Virtual recreation:** The virtual model replicates the physical geometry, including the camera position, marker layout, and environmental parameters such as lighting and resolution. This allows us to test identical scenarios under idealized and controlled conditions, ensuring consistency across multiple runs. To allow for realistic lighting and adjustable camera parameters, we used the program Blender. This software also allows for Python scripting, which we used to generate all digital counterparts of the real-world measurement images.

Both environments were calibrated using a ChArUco [Op] board, such that the camera data was consistent across all tests. We used 3D printed ArUco markers made of PLA filament in the physical testing rig due to the material's strength compared to paper, as well as its reflective properties [Do22].

Fig. 1 provides an overview of the hybrid evaluation workflow, illustrating the parallel structure of the real and simulated environments and their integration into a shared analysis pipeline. Fig. 2 shows both real and simulated testing environments, where the real setup (left) has the *Meta Quest 3* mounted and the virtual scene (right) is left blank.

3.2 Testing parameters

Cameras We selected three representative camera systems commonly used in consumer and XR development contexts to evaluate marker detection performance under both consumer-grade and industrial conditions. While the setup can accommodate a broader range of sensors, these devices were chosen based on practical relevance, image quality, and their distinct optical characteristics.

- ***Meta Quest 3*:** A widely available XR headset with integrated pass-through capabilities. Due to its unusual camera feed characteristics, such as non-standard aspect ratio, stitched image regions, distortion, and internal pre-processing, it represents a valuable test for mobile marker tracking systems.
  The camera records in a resolution of 2064 × 2208 with a field of view of ca. 100° FOV and a focal length of ca. $20mm$.
- ***Logitech MX Brio*:** A high-resolution RGB webcam capable of capturing 4K video, representing a more traditional desktop vision system. Unlike the *Quest*, it does not apply significant internal image processing, making it suitable as a baseline for high-fidelity detection and an interesting contrast to mobile XR devices.
  The camera records in a resolution of 3840 × 2160 with a field of view of ca. 90° FOV and a focal length of ca. $4.7mm$.
- ***HP 325 FHD*:** A low-end RGB webcam capable of capturing 1080p video. This represents a more affordable, flexible choice of webcam for industrial applications. With lower resolution, a reduced field of view, and lower image fidelity, it was selected as a contrast to the *Logitech Brio*.
  This camera records in a resolution of 1920 × 1080 with a field of view of ca. 66° FOV and a focal length of ca. $2.2mm$.

To replicate the cameras in the virtual scene, each device was modeled in Blender as a virtual camera object. The replication process included setting the focal length, sensor size, resolution, and aspect ratio according to the manufacturer specifications. Where available, calibration parameters were used to reproduce intrinsic properties. Optical distortion was visually approximated by applying Blender's compositing `Lens Distortion` node, allowing a close alignment between the simulated and real-world captures.

Markers To replicate realistic deployment conditions, marker selection was performed without cherry-picking. All markers were used as-is, regardless of individual detection performance, to reflect a typical application workflow. Markers were drawn from two ArUco dictionaries: `5x5_1000` and `ORIGINAL`. Both use a 5×5 bit grid but differ in optimization. The `5x5_1000` dictionary, generated using mixed-integer linear programming [Ga15], maximizes inter-marker Hamming distances to reduce false positives. In contrast, the `ORIGINAL` dictionary lacks such optimization, making it more prone to incorrect detections under challenging conditions. Including both allows comparison of legacy and modern dictionaries and ensures compatibility with detection pipelines limited to `ORIGINAL`.

Markers were printed at 10 cm × 10 cm, with a 7 cm × 7 cm coded area and a 1.5 cm white border for contrast. Each 5×5 bit grid had $1\,\mathrm{cm}^2$ bits, ensuring reliable detection over various distances and supporting sub-pixel corner refinement within the resolution limits of the cameras used.

Tracking algorithms To evaluate whether the ArUco detection pipeline represents a bottleneck in tracking performance, we benchmarked two different tracking systems: a standard ArUco-based implementation and a proprietary system named BART. This comparison assesses whether limitations in detection are inherent to ArUco or attributable to general marker tracking challenges.

BART Implementation To evaluate the proprietary BART tracking system against the current ArUco implementation, we updated its dependencies and processed the same generated dataset. BART employs a proprietary, time-based detection strategy designed to address occlusion, jitter, and dynamic detection challenges. Each frame is processed independently: once markers are detected, BART extracts their pose (position and orientation), records timing data, and visualizes marker IDs and body projections via OpenCV.

Performance Optimizations in BART BART outperforms standard OpenCV ArUco by incorporating several architectural optimizations that balance real-time efficiency with detection accuracy. Time-bounded processing ensures consistent frame rates by terminating detections exceeding preset thresholds. When markers persist across frames, the search space is constrained to ROIs near prior detections, reducing computational load.

A multi-scale detection strategy processes downsampled frames first, escalating to higher resolutions only if needed. Parallel full-frame searches are handled asynchronously via Boost threading, preserving the responsiveness of the main detection loop. Adaptive thresholding—using both mean and Otsu's methods [Ga22; Yo15]—enhances marker segmentation under varying lighting. Marker decoding uses lightweight bit matrix comparisons with minimal Hamming distance checks to reduce processing overhead.

Many aforementioned optimizations align conceptually with the acceleration strategies of Romero-Ramirez; Muñoz-Salinas; Medina-Carnicer [RMM18], which improve square fiducial marker detection without sacrificing accuracy.

4 Results

Our findings give insight into detection performances from all three cameras across varying camera and marker positions.

4.1 Marker Detection

Fig. 3 shows the marker detection results across different camera heights and marker distances for all three tested cameras. The left column presents detections from the **real-world environment**, while the right column shows detections in the **virtual simulation**. The heatmaps visualize the effective detection areas for each camera. In general, similar detection patterns are visible between real and simulated tests, indicating that the virtual environment provides a comparable—but not identical—representation of real-world behavior.

Tab. 1 summarizes the results. The key metric of **detection coverage** expresses the proportion of markers detected out of all available markers in the test setup. We calculate detection coverage using the formula Coverage = $\frac{\text{Detected Markers}}{\text{Total Number of Markers}} \times 100$.

This effectively reflects the system's overall ability to identify markers within our test environments, rather than the correctness of IDs. On average, both webcams achieved higher coverage than the Meta Quest 3, which suffered from reduced detection rates at larger distances and steep angles.

The other metric we used is **detection coverage**, which measures the number of overlap between the marker detection in the real world versus its counterpart in the simulated environment.

Let D_{real} denote the number of markers detected in the physical setup, and D_{sim} the number detected in the simulated environment under the same conditions.

$$\text{Precision} = \left| D_{\text{real}} - D_{\text{virtual}} \right| \times 100$$

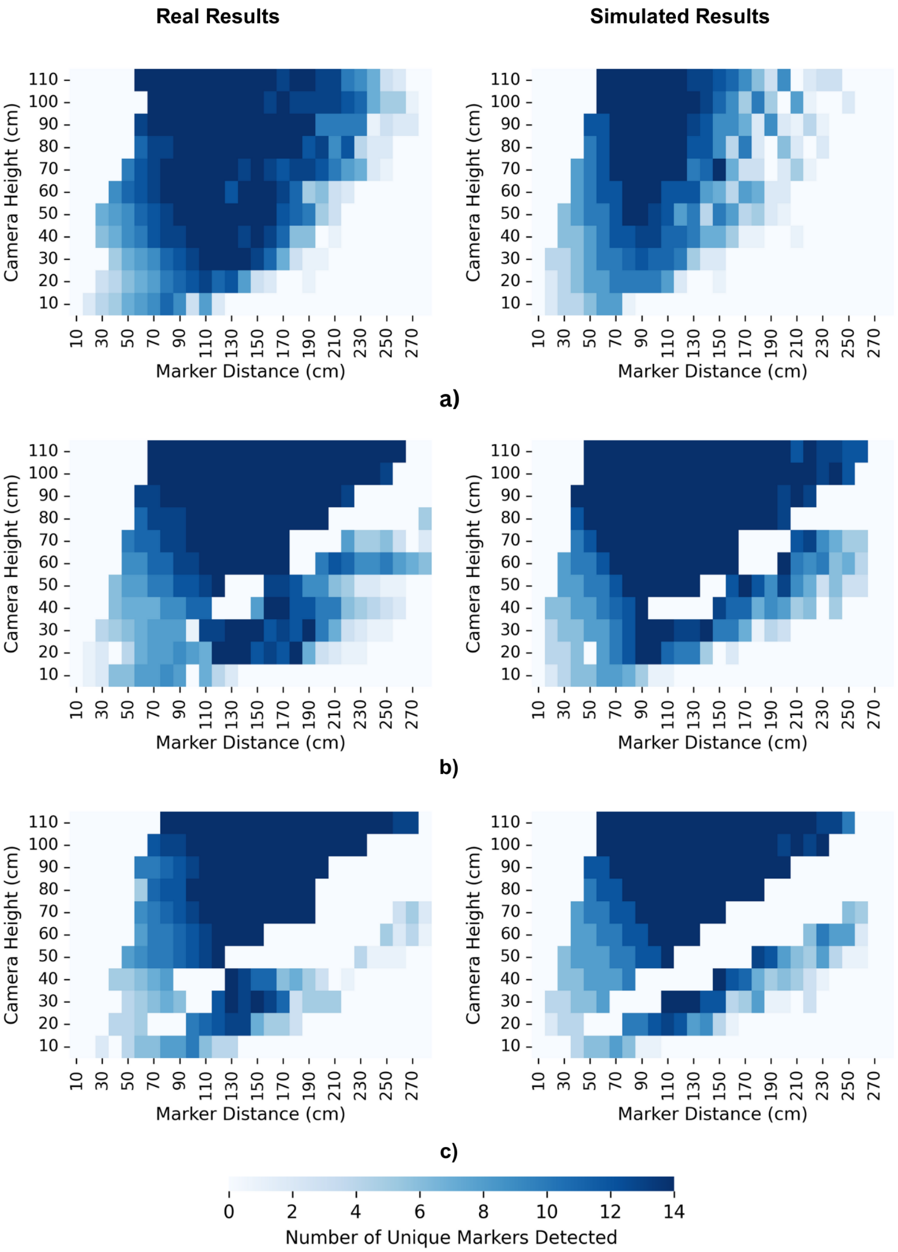

Fig. 3: Marker detection results across camera heights and marker distances (*Meta Quest 3* (**a**), *Logitech MX Brio* (**b**), *HP 325 FHD* (**c**))

Camera	Detection Coverage (%)	Precision (%)	Mean ΔDistance (cm)	Mean ΔRotation (deg)
HP 325	88.76	94	1.81	1.67
Logitech Brio	84.02	91	1.72	1.14
Meta Quest 3	72.41	97	3.43	4.41

Tab. 1: Comparison Summary Metrics for Each Camera.

The *Meta Quest 3* also showed the strongest discrepancies between real and simulated performance. Unlike webcams, where deviations can largely be traced to classical optical factors (e.g., field of view, resolution, or lens distortion), the Quest's issues seem to be linked to its internal passthrough image pipeline. Its video feed undergoes some processing, leading to spatially non-linear warping and stretched or stitched regions in the image. These effects, noted also in a study conducted by Bailenson et al. [Ba24], impact marker visibility and degrade pose estimation consistency (see Section 4.2). Because our simulation does not replicate such device-specific image generation steps, alignment between virtual and real results is inherently less reliable for the Quest.

By contrast, both webcams exhibited more consistent results between real and simulated conditions. Since they apply minimal image processing and have relatively low optical distortion, their detection behavior is primarily influenced by physical characteristics such as resolution and field of view. For example, the *HP* webcam's narrower 66º FOV limited its coverage at low heights and steep angles, whereas the *Logitech Brio* webcam benefited from its higher resolution and wider view. Minor discrepancies in the heatmaps are most plausibly attributable to small misalignments of the setup during measurement rather than the inherent behavior of the camera.

4.2 Pose estimation

While analyzing the pose estimation difference between real and virtual results, the *Meta Quest 3*'s rotation vector discrepancies, shown in Fig. 4, are particularly pronounced. The shaded regions in the plot indicate the variability of the measurements across trials, with wider bands reflecting less stable detection results. This often results from inverted marker orientations. These are instances where the correct marker ID is detected, but with one or more components of the rotation vector flipped, which occurs in cases where the pitch of the marker has a low pitch angle to the camera, which aligns with findings made by Rijlaarsdam; Zwick; Kuiper [RZK22].

This emphasizes the need for improved calibration and a more accurate simulation of the *Meta Quest*'s optical pipeline. The average rotation difference across all distances is also larger than the rotation vector differences with the webcams, which is likely due to the

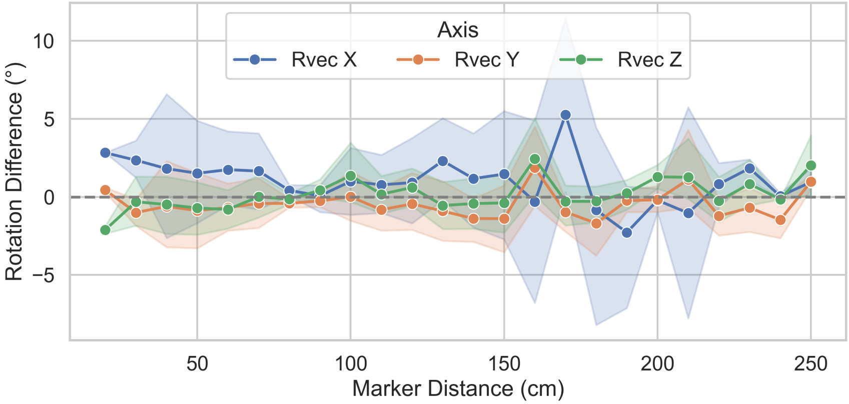

Fig. 4: Average detected rotation vector difference between real and simulated (*Meta Quest 3*)

unorthodox image generation pipeline and warping effect, which was only approximated in the virtual camera recreation. In contrast, the pose estimations were observed to be more similar between real and virtual environments with the two tested webcams. Both rotation and distance estimations were within small margins of ± 1−5 centimeters for the distance estimation and ± 4 degrees of rotation per axis, with similar occasional outliers as observed with the *Meta Quest 3*.

4.3 Marker Detection Approaches

Tab. 2 builds on the results presented in Tab. 1 by directly comparing the detection performance of ArUco and BART in all camera systems evaluated, using the same metrics introduced in Section 4.1. The comparison focuses solely on detection accuracy, as pose estimation differences remained within narrow margins for most devices, with the exception of the *Meta Quest 3*, which exhibited fewer inverted rotation vectors when using BART, as previously discussed in Section 4.1.

Camera	Algorithm	Coverage (%)	Precision (%)
HP 325	ArUco	88.76	94
HP 325	BART	90.11	98
Brio	ArUco	84.02	91
Brio	BART	87.30	97
Quest 3	ArUco	72.41	97
Quest 3	BART	80.15	99

Tab. 2: Comparison of ArUco vs BART across camera systems.

Although the overall detection coverage is largely comparable between the two algorithms, BART outperforms ArUco in terms of precision, indicating a notable reduction in false positives. These improvements are particularly significant on lower end hardware, where BART's adaptive thresholding and region-of-interest filtering more effectively suppress spurious detections. Even in high-quality sensors such as the *Logitech Brio*, BART maintains a precision advantage while achieving comparable coverage.

These findings support the view that while hardware characteristics largely constrain overall detection performance, algorithmic enhancements can meaningfully compensate for those limitations. As discussed in Section 6, remaining discrepancies between the real and simulated results are likely to be attributable to physical setup tolerances and human measurement error rather than algorithmic performance.

5 Discussion

Our results show that high-fidelity simulations can approximate general trends in marker visibility and detection zones but consistently overestimate detection performance when compared to real-world conditions. This overestimation is primarily caused by missing camera-specific effects in the simulation pipeline, such as lens distortion, sensor noise, and proprietary image processing steps. While the physical setup's geometric and lighting conditions were closely replicated in Blender, the absence of the aforementioned optical and computational effects limits the simulation's predictive accuracy.

Pose estimation fidelity further highlights this limitation. Although simulations produced reasonable geometric alignment for static configurations, rotation and distance estimations diverged significantly, particularly when device-specific distortions could not be modeled. Across all tested devices, pose estimation in simulation underperformed compared to real-world results, especially under off-angle or long-range conditions.

Overall, our digital twin approach proved to be effective for early-stage layout evaluation and marker placement testing, where relative visibility patterns are the primary concern. However, it remains insufficient for predicting precise detection rates or robust pose estimation performance across varied hardware. Our findings confirm that real-world validation remains essential when assessing practical marker tracking system performance.

In summary, virtual environments provide meaningful early-stage evaluation benefits, particularly for assessing geometric configurations and basic marker visibility. However, simulations are insufficient for final detection performance assessments, especially regarding device-specific distortions and pose accuracy.

6 Limitations

While our evaluation approach successfully facilitates early-stage testing, several key limitations constrain its ability to predict real-world tracking performance comprehensively:

- **Static Scenarios Only:** All tests were conducted under controlled, stationary conditions. Real-world XR applications involve dynamic motion, including head movement, object occlusion, and rapid viewpoint changes. The absence of motion artifacts, such as rolling shutter effects, motion blur, and temporal filtering, limits the realism of our simulation results.
- **Simplified Lighting and Materials:** Although virtual lighting was visually approximated, no photometric calibration or material reflectance modeling was performed. Missing optical effects—including specular reflections, shadows, and surface texture variations—reduce the visual realism of markers and surroundings, potentially influencing detection reliability.
- **Lack of Physical Camera Emulation:** The virtual camera models used in Blender omit critical real-world effects such as sensor noise, chromatic aberration, rolling shutter distortions, and proprietary in-camera processing. This is particularly impactful for devices like the *Meta Quest 3*, whose closed-source image pipeline introduces non-linear distortions not reproducible in standard 3D modeling software.
- **Approximate Real-World Calibration:** Physical setup parameters (camera positions, angles, and distances) were measured manually with tolerances under 5 mm. However, small calibration inaccuracies likely contributed to misalignment between real and simulated test cases, especially at large distances or shallow angles.
- **Limited Marker Types:** Only two ArUco dictionaries were evaluated (`5x5_1000` and `ORIGINAL`). Testing additional marker designs—including ChArUco boards and circular markers—could reveal broader insights regarding marker detectability under varying conditions. Additionally, the markers were selected from amongst the first 20 IDs across both tested dictionaries, which may lead to slightly idealized results, since lower-index markers typically feature higher inter-marker Hamming distances and less visual ambiguity than markers assigned higher indices [Ga14].
- **Limited Validation via Alternative Tracking Approaches:** This study only includes testing with two different fiducial marker tracking approaches (ArUco and BART), which leads to a lack of validating data for a true comparison between simulated and real environments. This limitation is extended to the fact that only static scenes were evaluated, which limits the amount of possible optimization approach comparisons.

In summary, while the virtual environment approximates static visibility performance, unmodeled optical effects, device-specific processing pipelines, and the exclusion of dynamic testing limit its predictive validity for real-world AR/VR tracking performance.

7 Conclusion

Our findings indicate that simulations are highly valuable for early design iterations and for identifying potential marker visibility issues. However, they cannot fully replace real-world validation when the goal is accurate detection or precise pose estimation.

In our experiments, simulations consistently tended to overestimate detection performance, particularly for devices that employ proprietary image processing, such as the Meta Quest 3. Simulations reliably predicted general marker visibility patterns. However, precise pose estimation and robust detection still depended on physical testing, since our simulations omitted real camera effects like distortion and sensor noise.

The central takeaway is clear: simulations are best used as a "negative filter." If a setup fails in simulation, it will almost certainly fail in practice, but success in simulation alone is not enough.

Looking ahead, future work should focus on integrating physically accurate camera models, including distortion profiles and sensor characteristics, and extending the evaluation framework to dynamic scenarios that involve motion blur, jitter, and occlusion. Another promising direction lies in the use of photogrammetry or LiDAR scans to enable higher-fidelity scene reconstruction, as well as testing under more diverse and realistic lighting conditions.

Acknowledgment

This research has been funded by the European Union - NextGenerationEU under the project TwinMaP (13IK028K). The views and opinions expressed are solely those of the author(s) and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

[Ba24] Bailenson, J. N. et al.: Seeing the World Through Digital Prisms: Psychological Implications of Passthrough Video Usage in Mixed Reality. Technology, Mind, and Behavior 5 (2: Summer 2024), https://tmb.apaopen.org/pub/ztfn3ubj, 2024.

[Be24] Berral-Soler, R. et al.: DeepArUco++: Improved detection of square fiducial markers in challenging lighting conditions. Image and Vision Computing 152, p. 105313, 2024, ISSN: 0262-8856, DOI: 10.1016/j.imavis.2024.105313.

[Bl14] Blech, J. O. et al.: Cyber-Virtual Systems: Simulation, Validation & Visualization, 2014, arXiv: 1410.1258 [cs.SE], https://arxiv.org/abs/1410.1258.

[Çö20] Çöltekin, A. et al.: Extended Reality in Spatial Sciences: A Review of Research Challenges and Future Directions. ISPRS International Journal of Geo-Information 9 (7), 2020, ISSN: 2220-9964, DOI: 10.3390/ijgi9070439.

[Do22] Dogan, M. D. et al.: InfraredTags: Embedding Invisible AR Markers and Barcodes Using Low-Cost, Infrared-Based 3D Printing and Imaging Tools. In: CHI Conference on Human Factors in Computing Systems. CHI ’22, ACM, pp. 1–12, 2022, DOI: 10.1145/3491102.3501951.

[DRP15] Diekmann, J.; Renner, P.; Pfeiffer, T.: Framework zur Evaluation von Trackingbibliotheken mittels gerenderter Videos von Tracking-Targets. In: 2015 12th Workshop GI VR/AR Bonn. 2015.

[Ga14] Garrido-Jurado, S. et al.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition 47 (6), pp. 2280–2292, 2014, DOI: 10.1016/j.patcog.2014.01.005.

[Ga15] Garrido-Jurado, S. et al.: Generation of fiducial marker dictionaries using mixed integer linear programming. Pattern Recognition 51, pp. 481–491, 2015, DOI: 10.1016/j.patcog.2015.09.023.

[Ga22] Garrido-Jurado, S. et al.: Reflection-Aware Generation and Identification of Square Marker Dictionaries. Sensors 22 (21), 2022, ISSN: 1424-8220, DOI: 10.3390/s22218548.

[KYW18] Kam, H. C.; Yu, Y.; Wong, K.-H.: An Improvement on ArUco Marker for Pose Tracking Using Kalman Filter. In: 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). IEEE, Busan, South Korea, pp. 65–69, 2018, DOI: 10.1109/SNPD.2018.8441049.

[Ma21] Malik, A. A.: Framework to model virtual factories: a digital twin view, 2021, arXiv: 2104.03034 [eess.SY], https://arxiv.org/abs/2104.03034.

[Me20] Merino, L. et al.: Evaluating Mixed and Augmented Reality: A Systematic Literature Review (2009-2019), 2020, arXiv: 2010.05988 [cs.HC], https://arxiv.org/abs/2010.05988.

[Op] OpenCV: Charuco Board Calibration, https://docs.opencv.org/4.x/da/d13/tutorial_aruco_calibration.html, Accessed on 01.07.2025, https://docs.opencv.org/4.x/da/d13/tutorial_aruco_calibration.html.

[RMM18] Romero-Ramirez, F.; Muñoz-Salinas, R.; Medina-Carnicer, R.: Speeded up detection of squared fiducial markers. Image and Vision Computing 76, pp. 38–47, 2018, DOI: 10.1016/j.imavis.2018.04.002.

[RMM20] Romero-Ramirez, F.; Muñoz-Salinas, R.; Medina-Carnicer, R.: Tracking fiducial markers with discriminative correlation filters. Image and Vision Computing 107, p. 104094, 2020, DOI: 10.1016/j.imavis.2020.104094.

[RZK22] Rijlaarsdam, D. D. W.; Zwick, M.; Kuiper, J. (: A novel encoding element for robust pose estimation using planar fiducials. Frontiers in Robotics and AI Volume 9 - 2022, 2022, ISSN: 2296-9144, DOI: 10.3389/frobt.2022.838128.

[So23] Song, J. et al.: Augmented Reality-Based BIM Data Compatibility Verification Method for FAB Digital Twin implementation. Buildings 13 (11), 2023, ISSN: 2075-5309, DOI: 10.3390/buildings13112683.

[SPS24] Sivov, N. Y.; Poroykov, A. Y.; Shmatko, E. V.: Estimating the Error in Locating Fiducial Markers in Space Using Physical Simulation. In: 2024 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). Pp. 1–5, 2024, DOI: 10.1109/WECONF61770.2024.10564651.

[Su24] Sulistiyono, M. et al.: Comparative study of marker-based and markerless tracking in augmented reality under variable environmental conditions. Journal of Soft Computing Exploration 5 (4), pp. 413–422, 2024.

[WSB23] Wooley, A.; Silva, D. F.; Bitencourt, J.: When is a simulation a digital twin? A systematic literature review. Manufacturing Letters 35, 51st SME North American Manufacturing Research Conference (NAMRC 51), pp. 940–951, 2023, ISSN: 2213-8463, DOI: https://doi.org/10.1016/j.mfglet.2023.08.014.

[Yo15] Yousefi, J.: Image Binarization using Otsu Thresholding Algorithm, 2015, DOI: 10.13140/RG.2.1.4758.9284.

[Zh25] Zhang, J. et al.: Digital twin embodied interactions design: Synchronized and aligned physical sensation in location-based social VR. Frontiers in Virtual Reality Volume 6 - 2025, 2025, ISSN: 2673-4192, DOI: 10.3389/frvir.2025.1499845.