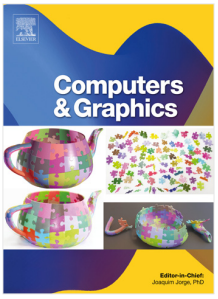
Journal Pre-proof



Bridging BIM and reality: A hardware-optimized registration pipeline for mixed reality in indoor construction environments

Marcos Arroyo-Ruiz, Gonzalo Gómez-Nogales, José Antonio Gómez-Fernández, Carlos Andújar, Marc Comino-Trinidad

PII: S0097-8493(26)00099-3
DOI: https://doi.org/10.1016/j.cag.2026.104628
Reference: CAG 104628

To appear in: *Computers & Graphics*

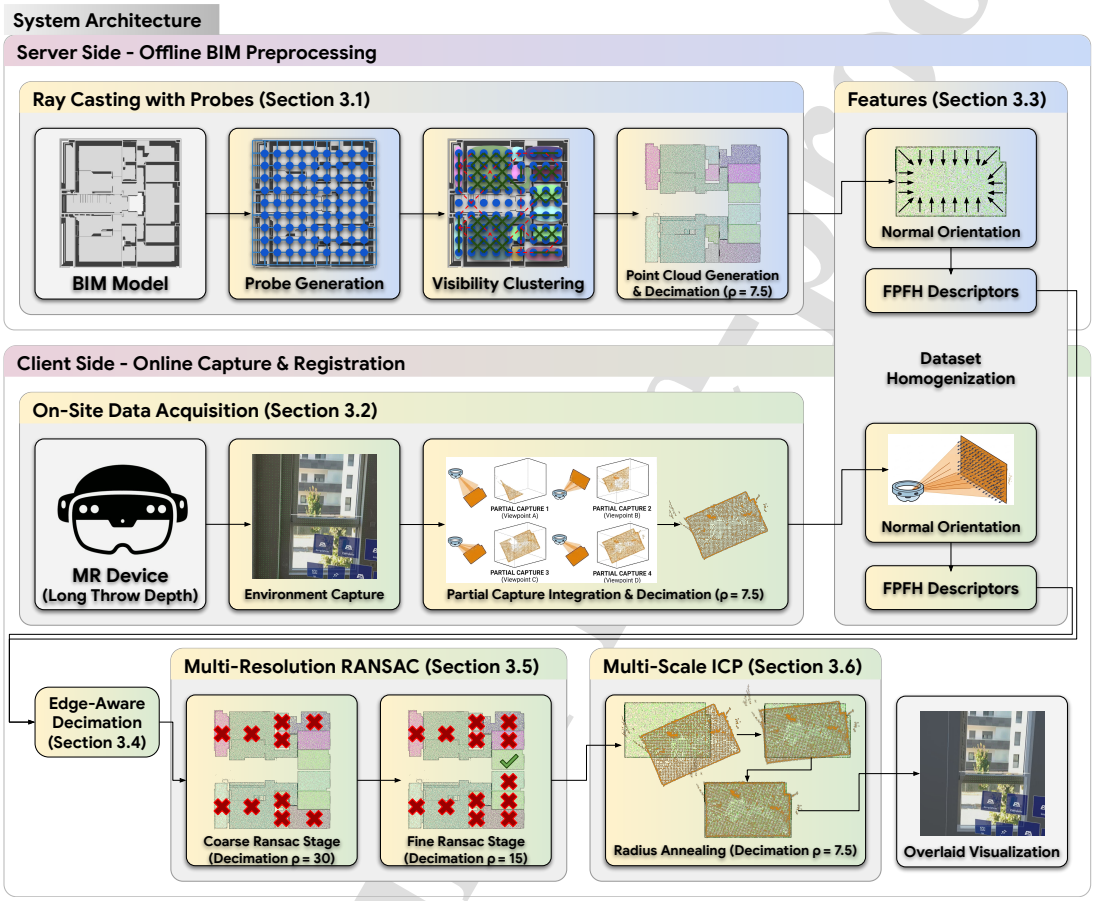
Received date : 8 April 2026
Revised date : 1 May 2026
Accepted date : 13 May 2026

Please cite this article as: M. Arroyo-Ruiz, G. Gómez-Nogales, J.A. Gómez-Fernández et al., Bridging BIM and reality: A hardware-optimized registration pipeline for mixed reality in indoor construction environments. *Computers & Graphics* (2026), doi: https://doi.org/10.1016/j.cag.2026.104628.

Graphical Abstract

Bridging BIM and Reality: A Hardware-Optimized Registration Pipeline for Mixed Reality in Indoor Construction Environments

Anonymous



Highlights

Bridging BIM and Reality: A Hardware-Optimized Registration Pipeline for Mixed Reality in Indoor Construction Environments

Anonymous

- A standalone geometric registration pipeline for HoloLens 2 achieves centimeter-level accuracy without cloud offloading.
- Automated alignment significantly outperforms manual and QR-based methods in lateral precision and on-site usability.
- Registration degradation directly compromises architectural spatial judgment and safety-critical hazard assessments.

Bridging BIM and Reality: A Hardware-Optimized Registration Pipeline for Mixed Reality in Indoor Construction Environments★

Anonymous[a,*,1]

[a], , , , ,

ARTICLE INFO

Keywords:
Mixed Reality
Building Information Modeling
Point Cloud Registration
Construction Inspection
Spatial Perception

ABSTRACT

Building Information Modeling (BIM) has transformed the Architecture, Engineering, and Construction (AEC) industry by digitizing project data, yet its full potential remains unrealized due to persistent gaps between virtual models and physical sites. These gaps contribute to inefficiencies, with studies reporting substantial waste in labor and coordination. Extended Reality (XR) technologies offer a promising solution by enabling immersive, real-scale visualization of BIM models on-site.

This article introduces a Mixed Reality (MR) application for Microsoft HoloLens 2 that superimposes BIM representations onto construction environments at a 1:1 scale, supporting real-time detection of differences between the as-designed BIM model and the as-built construction on site. We present a robust registration pipeline that integrates commercial XR hardware with advanced algorithms to achieve precise alignment under challenging conditions. To validate the system, we conducted a controlled user study comparing three registration paradigms (manual gesture-based, QR-assisted, and fully automatic) and analyzing their impact on alignment accuracy and user experience (UX) in AEC-related tasks. Results show that our automatic approach provides advantages over state-of-the-art alternatives and significantly improves registration precision and usability ratings over the baseline methods. Furthermore, the study demonstrates that alignment errors strongly influence spatial perception and decision-making, highlighting the necessity of high-fidelity registration for effective MR integration in construction workflows.

1. Introduction

The Architecture, Engineering, and Construction (AEC) industry has experienced a profound transformation with the adoption of Building Information Modeling (BIM). BIM represents a paradigm shift from static drawings to interoperable digital models that support collaboration and automation throughout the project lifecycle [27]. Architects and engineers can now create parametric, object-oriented n-D models enriched with detailed construction data [4].

Despite these advances, BIM's potential remains underutilized during the execution phase. A key challenge is the cognitive gap that arises when rich 3D data is consumed through 2D screens, limiting spatial understanding and real-scale perception [22]. Extended Reality (XR), including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), offers a promising solution by overlaying digital twins onto physical environments. Devices such as Microsoft HoloLens 2 [15] enable immersive interaction with 3D elements through natural gestures and voice commands.

However, the effectiveness of XR in construction depends critically on the accurate alignment between virtual and real-world models. In this work, we primarily focus on indoor construction environments. In contrast to outdoor open spaces, where usually geolocating technologies are available and can provide a robust cold-start position for localization algorithms, indoor settings such as tunnels or unfinished shells pose unique challenges: GPS signals are unavailable, connectivity is limited, textures are sparse, and lighting conditions are often poor. While recent computer vision literature proposes advanced registration methods such as Teaser++ [30], Fast Global Registration (FGR) [31], or 4-Point Congruent Sets (4PCS) [1, 14], our analysis shows that these approaches are unsuitable for standalone MR head-mounted displays (HMDs) in AEC contexts due to either high computational overhead or extreme sensitivity to environmental noise and parameter configuration.

To address this gap, we developed a registration pipeline optimized for the HoloLens 2 and integrated it into an XR application that allows on-site users to visually compare BIM models with the built environment and identify deviations in real time. Our approach employs a customized RANSAC-based algorithm [8] to estimate the user's 6-DOF pose from point clouds, achieving low-latency performance while maintaining resilience to lighting variations and geometric ambiguities.

However, algorithmic optimization is insufficient if it does not translate to improved operator performance, as the true utility of Mixed Reality in construction is fundamentally dictated by how registration errors affect human users. Misalignments can lead to incorrect spatial judgments, significantly increasing the risk of costly mistakes during inspection or assembly. Therefore, we posit that designing a robust algorithmic pipeline and evaluating its perceptual impact on user experience (UX) and task performance are two inseparable pillars of effective MR integration. Additionally, we investigate how these registration discrepancies affect the

★This work is part of the project *removed for anonymity*, funded by *removed for anonymity*.

*Corresponding author
anon@anonymous.com (Anonymous)
https://anonymous.github.io/ (Anonymous)
ORCID(s):
[1]This author has been anonymized.

utility of virtual avatars when employed as relatable references to assess accessibility and depth perception within the digital-physical overlay.

In this article, we present a holistic framework that bridges technical geometric alignment with human-centric validation. We compare our automatic approach with manual gesture-based alignment and QR-code alignment, and we analyze how varying levels of registration accuracy influence spatial perception and usability. The main contributions of this work are:

- **Hardware-Optimized Registration Framework:** A robust algorithm tailored for compute-constrained MR devices, demonstrating that a tuned RANSAC approach outperforms heavier alternatives in AEC scenarios.
- **Comparative Analysis:** Evaluation of three registration paradigms (automatic, manual, marker-based) to quantify trade-offs in accuracy, setup time, and computational load.
- **User-Centric Validation:** Empirical evidence on how registration quality impacts user confidence and task performance, providing guidelines for MR adoption in construction workflows.

2. Related Work

In this section, we review the state of the art in Mixed Reality applications for the AEC industry. We organize our analysis into two main areas. First, in Section 2.1, we examine the evolution of registration algorithms, contrasting visual Simultaneous Location and Mapping (SLAM) approaches with geometric methods. We highlight the specific challenge of aligning *as-built* reality with *as-planned* BIM data in computationally constrained environments. Second, in Section 2.2, we discuss the impact of registration quality on user experience, establishing the theoretical link between alignment accuracy and spatial understanding in construction tasks.

2.1. Registration Methodologies in AEC

The concept of Augmented Reality relies fundamentally on the accurate estimation of the camera's pose relative to the real world, as defined in the seminal survey by Azuma [2]. However, in the specific context of the AEC industry, superimposing BIM models onto physical sites remains a persistent technical bottleneck due to the scale and complexity of construction environments [29]. Prior studies indicate that even small alignment errors can significantly affect task performance and trust in MR systems [7].

Early solutions utilized marker-based systems (e.g., QR codes) to provide reliable anchors [28]. While more recent highly automated inspection frameworks still rely on physical markers to solve the initial global registration cold-start problem [5], we note that their deployment on active job sites is logistically complex; persistent occlusion from machinery, dust accumulation, and the constant transformation of construction surfaces make marker maintenance inefficient for continuous 1:1 scale tracking [21]. To address this dependency, established methods focus on marker-less tracking, specifically Visual SLAM systems like ORB-SLAM [16]. While these algorithms excel at mapping the current reality and tracking the user's movement within it, they depend critically on the richness and stability of environmental visual features. Registering this internal map to the BIM coordinate system is non-trivial. Modern learning-based feature matchers like LoFTR [23] have improved correspondence detection, but as noted by Bosché et al. [6], the process is complicated by physical discrepancies between the design (BIM) and the construction state (reality), where deviations or unfinished elements can cause alignment failures in difficult environments.

Given these visual challenges (and the frequent lack of texture in concrete environments), geometric registration using depth or LiDAR data is often preferred. The standard approach involves coarse alignment followed by fine refinement using Iterative Closest Point (ICP) [3]. The challenge lies in the initialization. Fast Global Registration (FGR) [31] offers speed, but it exhibits high instability when faced with the massive noise levels typical of mobile depth sensors. Conversely, robust global solvers like Teaser++ [30], or 4PCS [1] utilize truncated least squares to handle extreme outliers.

However, our benchmarks indicate that these methods are computationally demanding for the ARM architecture of standalone HMDs. While offloading these heavy computations to edge or cloud servers could theoretically alleviate hardware constraints [26, 12, 5, 18], such architectures introduce a dependency on high-bandwidth network connectivity or local workstations. In many AEC contexts, particularly in deep infrastructure, tunnels, or early-phase construction sites, connectivity is often intermittent or nonexistent, and installing local hardware is unfeasible. Consequently, a fully standalone, on-device solution is not merely a preference but an indispensable operational requirement to ensure system reliability under real working conditions.

It is also important to distinguish our proposed geometric registration from recent advances in UAV-based and robotic mapping. Studies such as BIM-SLAM [26], D-PC2BIM [12], dense multi-sensor indoor SLAM [19], and the Pose Hough Transform [18] demonstrate impressive robustness for global alignment. However, while these systems leverage high-end mechanical or drone-mounted LiDARs and workstation-class processing, our work is specifically optimized for the wearable form factor. These external pipelines rely on sensor payloads and memory footprints that exceed the computational limits of standalone HMDs. Similarly, works focused on Construction Progress Monitoring [25] utilize 4D BIM to automate the verification of schedule completion. These systems generally prioritize the binary detection of built elements (existence) and often

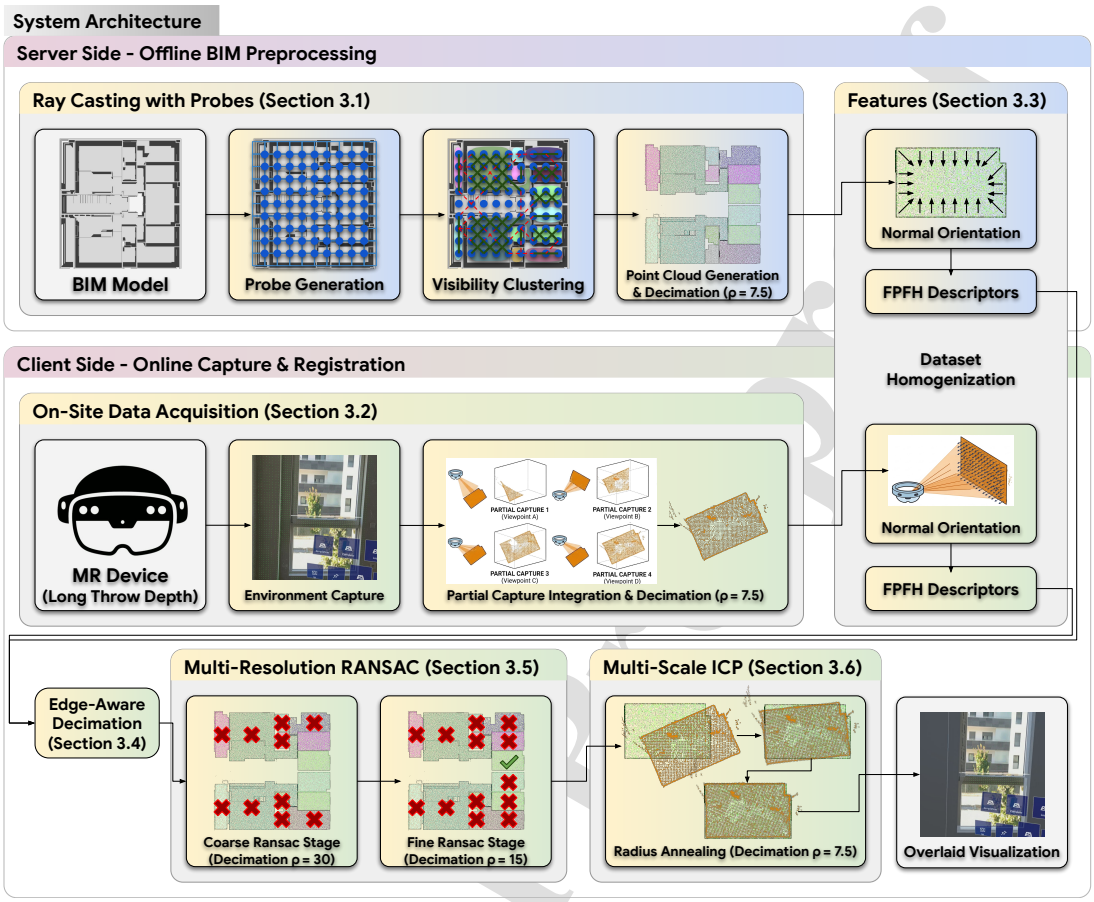


Figure 1: Comprehensive system architecture of the proposed Mixed Reality registration framework. The diagram illustrates the functional division between the server-side offline preprocessing and the client-side online execution. (Top) The offline BIM preprocessing stage begins with Ray Casting with Probes (Section 3.1), where a parametric BIM model is partitioned into distinct workspaces through automated probe generation and visibility logic. The resulting point cloud is decimated using a uniform grid with an edge length of ρ_{base} = 7.5 cm. Identical parameters for normal orientation and FPFH descriptor extraction are then applied (Features (Section 3.3)) to ensure dataset homogeneity. (Bottom) The online capture and registration stage is performed on-site by a Microsoft HoloLens 2 using its Long Throw depth stream (Section 3.2). After an initialization sweep, the captured data undergoes Edge-Aware Decimation (Section 3.4) to preserve structural anchors. The core of the alignment is a multi-resolution geometric matching sequence: a coarse, high-speed Multi-Resolution RANSAC (Section 3.5) is followed by a precise local refinement using Multi-Scale ICP (Section 3.6) with radius annealing. Finally, the registered BIM model is rendered as a 1:1 scale, real-time Overlaid Visualization for interactive on-site inspection.

operate asynchronously. In contrast, our work targets the human operator's need for interactive, high-fidelity geometric inspection entirely on a standalone device.

Our proposed system addresses these challenges by occupying a middle ground between efficiency and robustness. We implement a customized RANSAC-based pipeline [8] that utilizes Fast Point Feature Histograms (FPFH) [20] for robust correspondence matching. This architecture provides greater resilience to noise than FGR while exhibiting superior generalization capabilities and a significantly lower dependency on parameter fine-tuning than that of global solvers like Super4PCS. These optimizations render the pipeline suitable for real-time, offline operation within the constrained computational envelope of the HoloLens 2.

2.2. The Impact of Registration on Spatial Perception

The utility of MR in construction is not solely defined by algorithmic precision, but by the user's ability to interpret the augmented information. Misalignments between the digital twin and physical reality can lead to cognitive tunneling or errors in spatial judgment. Swan et al. [24]

established that depth perception in Optical See-Through displays is inherently compromised by hardware factors, a problem exacerbated when the virtual overlay drifts from its physical anchor.

Recent work focuses on mitigating these perceptual issues through better interaction design. Lazaro and Kim [11] highlight that multimodal interaction (gaze, gesture, voice) reduces the cognitive load required to manipulate misaligned or complex models. Furthermore, in safety-critical applications such as evacuation drills [13] or indoor navigation [10], the user's trust in the system is directly correlated with registration stability. Our study extends this body of knowledge by quantifying the relationship between registration error magnitudes and the user's subjective confidence.

3. Method

In this section, we detail the technical architecture of the proposed MR application, optimized for the computational constraints of the Microsoft HoloLens 2 (Figure 1). We describe a two-stage registration pipeline comprising offline preprocessing, which converts BIM models into optimized point clouds via a probe-based visibility approach, and online alignment utilizing a custom, ARM-optimized RANSAC algorithm and ICP refinement. We also discuss the choices made to ensure low-latency performance and alternative fallback mechanisms, including manual and QR-based options.

3.1. Offline BIM Preprocessing: Ray Casting with Probes

To bridge the gap between heavy parametric BIM data and the limited computational resources of standalone XR HMDs, we developed a server-side preprocessing pipeline termed Ray Casting with Probes. This stage segments the building into navigable workspaces by extracting visible geometry for registration while discarding internal structural elements. By offloading these computations to a remote workstation, the system manages high-polygon models and provides the HoloLens 2 with a lightweight model that prevents memory overflow.

The process begins by calculating the Axis-Aligned Bounding Box of the BIM model and subdividing it into a coarse $\delta_{grid} = 5\,m^3$ grid. A geometry-presence filter discards empty cells, after which virtual probes are instantiated at the vertices of a secondary $\delta_{probe} = 0.5\,m^3$ grid. A reciprocal visibility algorithm then groups these probes into architectural clusters (workspaces) based on unobstructed lines-of-sight, as depicted in Figure 2.

For each cluster, we generate a synthetic reference point cloud using the Embree ray-tracing engine (Figure 3). From the center of each probe, $N_{rays} = 200$ rays are cast in a uniform spherical distribution. By capturing only the first intersection point, we reconstruct the inner shell of the environment, effectively eliminating redundant geometry such as back-faces and occluded installations that could lead to false correspondences.

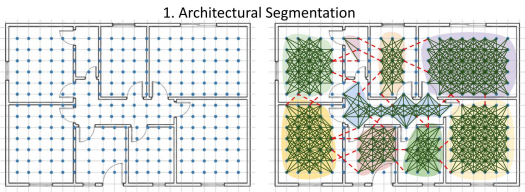


Figure 2: Architectural segmentation through probe visibility logic. The left image shows the initial uniform probe distribution within the BIM model's navigable space. The right image illustrates the visibility graph and resulting clusters, where solid lines indicate reciprocal visibility and dashed red lines represent structural obstructions.

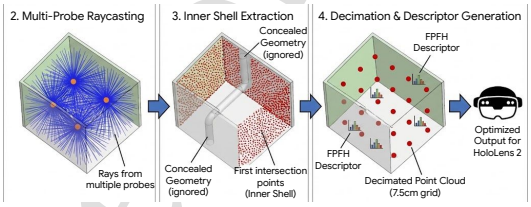


Figure 3: Point cloud generation from the segmented BIM model. Initially, multi-probe ray casting extracts workspaces' inner shells by isolating first-intersection points and discarding concealed geometry. Subsequently, uniform 7.5 cm decimation and FPFH descriptor pre-computation are applied to facilitate efficient streaming to the Microsoft HoloLens 2.

Finally, the resulting point cloud is decimated to a base resolution using a uniform grid with an edge length of $\rho_{base} = 7.5\,cm$ and pre-computed with geometric descriptors to ensure that only contextually relevant data is transmitted to the HMD.

Figure 4 shows different segmented models.

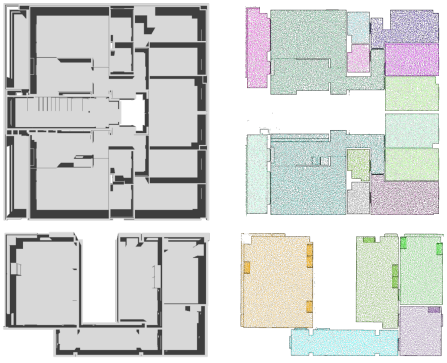


Figure 4: Examples of BIM models and generated synthetic point clouds segmented into rooms. The top row illustrates the apartment model used for numerical experiments and ablation studies. The bottom row displays the office model used for the user studies. The left column shows the original mesh, while the right column shows the resulting inner-shell point cloud.

3.2. On-Site Data Acquisition and Scene Reconstruction

To capture the physical environment, we utilize the Microsoft HoloLens 2 Research Mode via the Long Throw depth stream, which prioritizes mid-range geometry over standard high-frequency near-field tracking. While official specifications suggest a 5.46 m range, our empirical testing in construction environments established a reliable data threshold of 3.5 to 4.0 m, used here to calibrate capture volumes. Data acquisition is managed through a custom C++ plugin for Unity that interacts with raw data streams to perform temporal synchronization between depth frames and spatial odometry.

During initialization, the user performs a 360° sweep to integrate captures into a unified local point cloud, with real-time visual feedback indicating regions requiring further scanning. To ensure resilience against the volatile lighting and sparse textures of unfinished sites, the pipeline ignores RGB data and relies exclusively on geometry. Once complete, the captured cloud is consolidated and decimated to match the BIM reference resolution (ρ_{base}), ensuring geometric homogeneity for the registration stages.

3.3. Geometric Feature Characterization

To achieve homogeneity between the feature descriptors of the digital model and the captured environment, we implement a consistent geometric characterization pipeline for both datasets.

A significant challenge in BIM-to-Reality registration is data asymmetry: BIM models have perfect, discrete, and sharp normals, while sensor-derived clouds are inherently noisy and smoothed. To mitigate this, we ignore the original BIM mesh normals and instead estimate normals for both datasets [9]. By applying identical estimation parameters to both clouds, we achieve a homogeneous response in the subsequent feature descriptors.

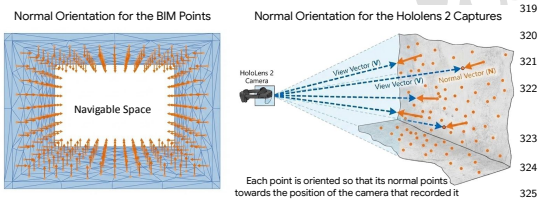


Figure 5: Normal orientation strategy for dataset homogenization. BIM normals are oriented toward the navigable interior (left) while captured points are oriented toward the sensor (right). This ensures a consistent front-facing orientation across both datasets to prevent feature descriptor inversion.

Consistency in orientation is maintained through a dual-logic approach to homogenize the datasets (Figure 5): for the BIM point cloud, normals are oriented deterministically to match the source surface topology, ensuring they point toward the navigable interior. In contrast, normals for captured points are oriented dynamically toward the optical center of the HoloLens depth sensor. Specifically, we enforce a negative dot product between the view vector and the surface normal to ensure a consistent front-facing orientation across both datasets and prevent feature descriptor inversion.

These oriented normals are utilized to compute Fast Point Feature Histograms (FPFH) [20], encoding local geometry into 33-dimensional vectors. To minimize on-device latency and thermal load, BIM FPFH descriptors are precalculated on the server, allowing the HMD to only compute descriptors for the locally captured cloud after decimation.

3.4. Edge-Aware Decimation

To further optimize the registration process, we implement an Edge-Aware Decimation filter that reduces informational redundancy in large planar areas while preserving critical structural features. While the initial uniform grid decimation stabilizes point density, this second stage identifies points defining the environment's topology, such as wall junctions, corners, and door frames.

The algorithm is based on the surface variation analysis proposed by Pauly et al. [17]. For each point p_i in the captured cloud, we analyze a local neighborhood $N(p_i)$ and construct a covariance matrix to examine the spatial distribution of its neighbors. Through eigenvalue decomposition ($\lambda_0 < \lambda_1 < \lambda_2$), we evaluate the surface curvature c_i as the ratio of the smallest eigenvalue to the sum of the total variance:

$$c_i = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \tag{1}$$

In perfectly planar surfaces, such as long concrete walls, the value of λ_0 tends toward zero, resulting in negligible curvature. Conversely, in structural intersections, the point distribution becomes more heterogeneous, increasing c_i. We establish a curvature threshold of $\tau_{curve} = 0.075$; only points exceeding this value are retained for the final registration cloud. This selective filtering eliminates the geometric ambiguity and sliding errors common in ICP when dealing with featureless planes, concentrating the computational budget on the most descriptive structural anchors.

3.5. Global Registration: Multi-Resolution RANSAC

To resolve the primary alignment between the BIM model and the captured cloud, we use a customized Multi-Resolution RANSAC pipeline. This approach addresses the kidnapped robot problem (e.g., identifying the user's location within the global BIM model without prior pose information). By utilizing the FPFH descriptors computed in the previous stage, the algorithm performs an iterative, probabilistic search to find the transformation matrix that maximizes the fitness score between the captured and reference clouds.

To ensure real-time performance on the HoloLens 2's ARM architecture, we implement a two-stage hierarchical search:

1. **Coarse Stage ($\rho_{coarse} = 30\,cm$ grid)**: This phase acts as a high-speed structural filter. It processes a heavily decimated version of the room clusters with a limited budget of $I_{coarse} = 10,000$ iterations. This allows the system to rapidly discard room candidates that lack basic structural correlation with the user's current perception.
2. **Precision Stage ($\rho_{fine} = 15\,cm$ grid)**: This stage is executed exclusively on the top 10 candidates identified in the coarse phase. We increase the stochastic effort to $I_{fine} = 100,000$ iterations over a more dense point set. This resolves fine-grained geometric ambiguities and ensures the resulting global pose is sufficiently accurate to serve as a seed for local refinement, preventing the subsequent ICP from converging into local minima.

This hierarchical approach balances construction site entropy with the need for low-latency feedback. By operating on feature histograms rather than raw coordinates, the RANSAC process remains robust against dynamic noise, such as machinery or personnel present on-site but absent from the static BIM model.

3.6. Local Refinement: Multi-Scale Point-to-Plane ICP

To reach centimeter-level precision and ensure long-range stability, the coarse RANSAC pose is refined using a Multi-Scale Point-to-Plane ICP variant. Standard ICP often converges to local minima in AEC environments due to geometric ambiguities from structural symmetries. We mitigate this using a radius annealing strategy that modulates the search radius and iteration budget across four phases.

The first two phases utilize a large search radius (up to $3.0\times\rho_{base}$) to maximize structural overlap and correct global registration displacements. The third phase reduces this to $1.2 \times \rho_{base}$ to prioritize edges and junctions, eliminating subtle rotational errors. Finally, RMSE minimization with a restrictive $0.5 \times \rho_{base}$ radius acts as an outlier filter against sensor noise and temporary objects. This hierarchical locking ensures the 6-DOF pose is anchored to the rigid building structure.

3.7. Sensory Fusion and Persistence

To maintain registration stability as the user navigates through the site, the result of the ICP refinement is handed off to the device's native Scene Tracking framework. Spatial Anchors bind the optimized transformation matrix to the local spatial map of the Hololens 2. While high-frequency Visual-Inertial Odometry manages relative movement, our geometric registration provides the absolute metric anchor. This architecture prevents typical long-term drift and allows for manual re-calibration if the environment undergoes significant structural changes.

3.8. Redundancy and Fallback Mechanisms

The first level of redundancy is a marker-based registration using QR codes. These fiducial markers are placed at surveyed coordinates on the job site and serve as deterministic anchors. When scanned via the HoloLens 2 RGB camera, the system resolves the user's pose through a Perspective-n-Point solver, bypassing the need for geometric matching in featureless areas like long, unfinished tunnels. The second level is a user-assisted manual alignment, where the operator can directly manipulate the BIM hologram using natural hand gestures. This mode includes precision controls for fine-tuning the 6-DOF pose, allowing the human-in-the-loop to resolve ambiguities that automated computer vision algorithms might struggle with in highly cluttered scenes.

4. Algorithmic Performance Evaluation and Ablation

In this section, we quantitatively evaluate the registration pipeline's stability within the computational limits of the Microsoft HoloLens 2. We present ablation studies isolating the effects of hierarchical RANSAC and edge-aware decimation on robustness and latency. Furthermore, we analyze the convergence of our radius annealing strategy and benchmark the framework against state-of-the-art global registration methods. Conducted across varying architectural complexities, these tests validate the system's ability to maintain high-fidelity alignment in challenging conditions of construction environments.

4.1. Ablation Study of RANSAC Design Choices

To provide a rigorous evaluation of the proposed system, we conducted a series of ablation studies designed to isolate the impact of our architectural optimizations on both mathematical robustness and computational efficiency. The experiments were performed using two distinct test environments: a synthetic laboratory model (Figure 6) composed of 16 rooms and a high-complexity setting comprising a complete floor of an apartment building (partially illustrated in Figure 4) with 94 segmented rooms.

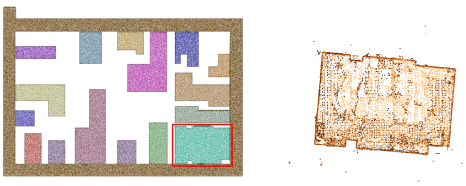


Figure 6: Experimental test environments for the ablation study. (Left) The segmented BIM reference for the laboratory model. The red bounding box highlights the target workspace for registration. (Right) The raw point cloud captured by the Microsoft HoloLens 2.

All tests were executed directly on the Microsoft HoloLens 2 to ensure the results reflect real-world performance on standalone hardware. The metrics reported in Table 1 represent mean values obtained across 10 independent runs per configuration. These baseline comparisons validate the

necessity of the hierarchical search and edge-aware filtering modules.

Registration Method	Apartment		Laboratory	
	Time (ms)	Fitness	Time (ms)	Fitness
Ours	1861	0.853	591	0.617
Global BIM Cloud	34640	0.000	133	0.000
Single-Pass RANSAC	14413	0.822	960	0.426
Regular Decimation	1912	0.799	390	0.387

Table 1
Consolidated Ablation Study of RANSAC Performance Metrics across Apartment and Laboratory Environments.

The **Global BIM Cloud** configuration represents an attempt to register against the entire unsegmented building model. As shown in the results, this approach failed to converge in 100% of the trials, yielding a fitness score of 0.000. This failure is mathematically consistent with the stochastic nature of RANSAC, where the required number of iterations increases exponentially as the inlier ratio drops. In large-scale infrastructure, the volume of a single room relative to the entire building results in an inlier ratio often below 2%. Stress tests further emphasize this limitation; even after ten minutes of uninterrupted computation, the system failed to reach a satisfactory alignment. Furthermore, attempting to process a global building cloud risks exceeding the device's RAM capacity, leading to system saturation.

Our divide-and-conquer strategy enables the system to implement adaptive filters. In contrast, although the **Single-Pass RANSAC** configuration eventually achieved a valid fitness score (0.822 in the apartment), it incurred a latency of over 14 seconds. By implementing a high-speed coarse stage with a 30 cm grid (10,000 iterations), the system can rapidly filter room candidates, allowing the final 100,000-iteration precision stage to concentrate exclusively on high-probability candidates. This optimization reduced total registration time by 87% in the apartment scenario.

Finally, the **Regular Decimation** variant demonstrates that geometric quality is significantly compromised when edge-aware filtering is omitted. Although a purely regular grid results in faster execution in the laboratory (390 ms), the fitness score drops to 0.387. This is due to the high density of points on large, featureless planar surfaces, such as walls and floors, which induce sliding errors during the stochastic matching process. Data indicates that the initial Principal Component Analysis overhead is offset by the reduced complexity of the resulting cloud, providing a robust kinematic lock and superior fitness without a significant time penalty.

4.2. Analysis of Radius Annealing in ICP Refinement

To validate the performance of the proposed radius annealing strategy, we performed a comparative convergence analysis designed to assess local refinement stability against conventional static-radius ICP variants. The evaluation utilized three distinct levels of initial registration error (Figure 7) comprising Low, Medium, and High offsets to measure the robustness of the Coarse, Medium, Fine, and Multi-Scale ICP radius configurations.

In scenarios characterized by low initial error, the multi-scale approach achieves a high-precision alignment equivalent to the Fine ICP variant, successfully reaching the centimeter-level threshold. However, the limitations of static methodologies become apparent as the initial displacement increases. For moderate and high initial errors, the Fine ICP variant remains trapped in local minima and provides non-feasible registrations. While the Coarse and Medium variants achieve basic structural alignment in these challenging cases, they inevitably reach a convergence plateau because they lack the sensitivity required for precise metric locking.

The data demonstrates that only the multi-scale strategy successfully navigates the error landscape across all test cases, consistently reducing RMSE to the target precision required for long-range stability. As illustrated in Figure 8, qualitative comparisons confirm the algorithm's ability to correct substantial spatial deviations and maintain structural alignment better than fixed-radius alternatives across varying distances. For an extended iteration-by-iteration analysis, we provide comprehensive visualizations in Supplementary Material Section A of the Appendix.

4.3. Comparative Performance Analysis and Benchmarking

To provide a rigorous justification for our algorithmic selection, we evaluate our multi-resolution RANSAC pipeline against representative state-of-the-art global registration methods. A viable methodology within the XR framework must demonstrate metric precision against the noise and multi-path interference characteristic of Time-of-Flight sensors while remaining compatible with the computational constraints of the HoloLens 2. Although Teaser++ is recognized for its robustness to high outlier ratios, it was excluded from this comparison due to resource demands that exceed mobile hardware limits and the significant complexity of porting its libraries to the Universal Windows Platform environment. Furthermore, cloud-based offloading was dismissed to ensure system reliability in subterranean or remote sites where offline operation is an indispensable requirement. Consequently, our comparative analysis focuses on Fast Global Registration (FGR) [31] and 4-Point Congruent Sets (4PCS) [14].

Experimental trials across the apartment and laboratory environments showed that deterministic alternatives like FGR and Super4PCS exhibited extreme sensitivity to parameter configuration. Parameters optimized for a specific dataset frequently failed when applied to another, limiting their utility in the unpredictable conditions of an active job site. In contrast, our RANSAC-based method provided superior generalization across diverse geometric scales without requiring extensive manual tuning.

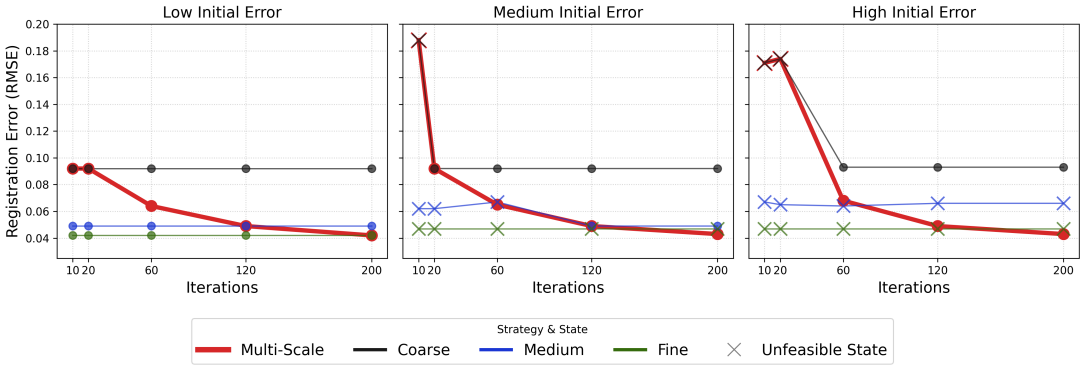


Figure 7: RMSE convergence analysis across varying initial registration error levels for the Multi-Scale radius annealing strategy and static-radius ICP variants (Coarse, Medium, and Fine). Circular markers (•) denote feasible, convergent registrations, while cross markers (×) indicate instances where the algorithm remained trapped in local minima.

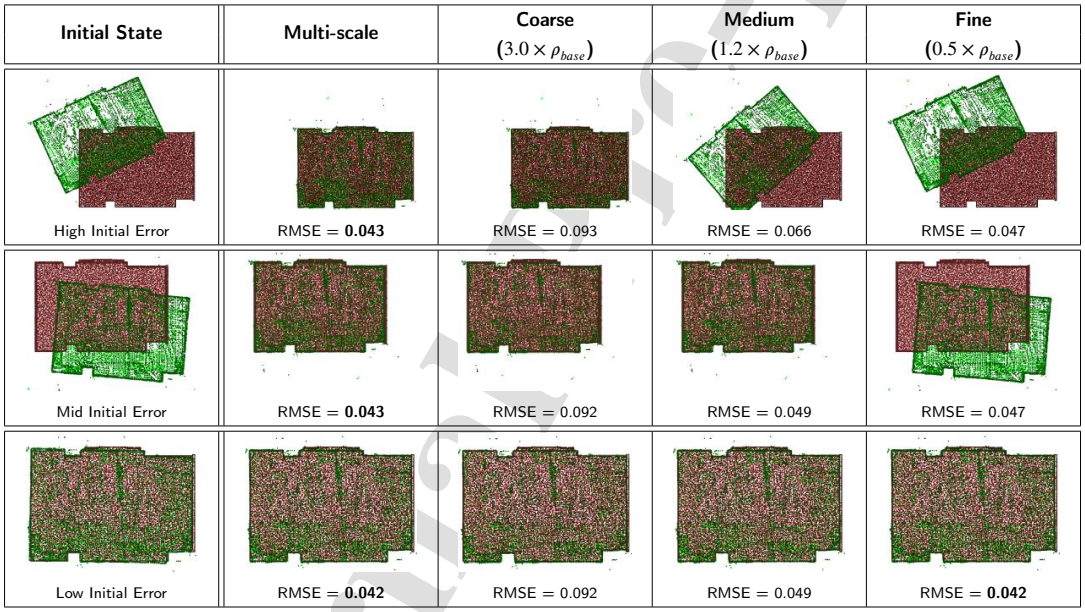


Figure 8: Visual comparison of convergence between the multi-scale ICP version and fixed-radius versions across scenarios with high, medium, and low initial registration errors. The target point cloud (BIM) is shown in red, and the captured point cloud (HoloLens) is shown in green.

As detailed in Table 2, our proposed system achieved significantly higher fitness scores than the competing baselines. In the apartment environment, our method reached a fitness of 0.853, whereas FGR only achieved 0.426. This substantial reliability margin ensures that the initial alignment is robust enough to prevent subsequent ICP refinement from diverging, effectively securing the centimeter-level precision required for professional AEC applications.

The scalability of our implementation provides a critical advantage through the dynamic adjustment of the computational budget. At a reduced budget of 2,000 iterations, the system maintained a fitness score of 0.501, outperforming Super4PCS while matching its execution speed. Although high-precision registration introduces a latency of 0.5 to 2 seconds, this remains well within the acceptable threshold for on-site inspection tasks where reliability and total hardware autonomy are prioritized over pure real-time refresh rates.

Registration Method	Apartment		Laboratory	
	Time (ms)	Fitness	Time (ms)	Fitness
Ours (100,000 iter.)	1861	0.853	591	0.617
Ours (10,000 iter.)	1659	0.775	425	0.616
Ours (5,000 iter.)	871	0.740	373	0.617
Ours (2,000 iter.)	426	0.501	430	0.544
FGR	254	0.426	320	0.149
Super4PCS	480	0.481	380	0.253

Table 2
Comparative analysis of performance (ms) and registration quality (fitness) between the proposed RANSAC-based algorithm and state-of-the-art methods FGR and Super4PCS.

5. Human-Centric Validation: Performance and Perceptual Analysis

In this section, we present an empirical evaluation focused on the user experience and perceptual performance of the proposed Mixed Reality registration pipeline. The evaluation is structured into two distinct phases, where Study I defines a performance benchmark by comparing the precision and robustness of our automatic alignment against standard interaction paradigms. Transitioning from technical precision to human-centric outcomes, Study II investigates the specific thresholds at which registration degradation compromises spatial perception, interaction efficiency, and user confidence. By synthesizing objective performance data with subjective usability metrics, we identify the operational requirements for ensuring fluid and reliable interaction within complex physical environments. Detailed normality test results for all user study data are provided in Supplementary Material Section B.

5.1. Study I: Evaluation of Registration Paradigms

The primary objective of this study was to quantify the alignment accuracy of the Automatic registration pipeline relative to existing methods. Specifically, we sought to measure the placement error that occurs when a user registers the model in one location and traverses to a secondary site, a key requirement for on-site inspection.

5.1.1. Participants

A total of $N = 16$ participants were recruited for this specific study. The cohort was demographically diverse, with ages ranging from 18 to over 45. Participants were screened to ensure no significant visual or motor impairments that would hinder the use of stereoscopic HMDs or the execution of precise gestures.

5.1.2. Experimental Design

We employed a within-subjects design with the Registration Method as the single independent variable. To mitigate learning effects and order bias, the sequence of the three conditions was counterbalanced using a randomized block design. The three experimental conditions were:

1. **Automatic (Proposed):** The user scans the room geometry using the HoloLens 2. The system utilizes our RANSAC-based pipeline to match the environmental point cloud with the BIM reference. Captured depth samples are shown as green points to provide feedback on coverage sufficiency.
2. **Marker-Based (QR):** The user scans a printed QR code placed at a pre-calibrated physical location. The system aligns the virtual coordinates based on the marker's relative pose.
3. **Manual (Gesture-Based):** Serving as the baseline, users select their approximate location on a 2D floor plan and then fine-tune the 3D model alignment using standard manipulation gestures (two-handed pinch to drag, rotation constrained to the Y-axis).

5.1.3. Procedure

The experimental protocol commenced with an induction phase where participants reviewed a technical briefing and a video tutorial of the Automatic registration interface. Following this, users executed the alignment in the starting location (Workspace A) using their assigned paradigm. To evaluate the stability of the holographic overlay during movement, participants were then required to traverse approximately 10 meters to a secondary location (Workspace B). In this second environment, the system rendered four virtual electrical plugs on physical walls; participants were tasked with marking the top-left corner of each virtual element in the real world. The procedure concluded by measuring the placement error, defined as the absolute distance between the virtual target and the physical mark, decomposed into lateral (xz) and vertical (y) deviations. Immediately following each condition, participants provided a subjective difficulty rating via the Single Ease Question (SEQ) on a 7-point Likert scale (1=Very Difficult, 7=Very Easy). Because post-registration tracking was delegated to the standard HoloLens 2 SLAM pipeline in all conditions, any drift-related effects were comparable across methods and did not bias the relative comparison of registration performance.

5.1.4. Results: Registration Accuracy

We analyzed the placement accuracy across the 16 participants. To provide a robust metric of deviation, we calculated the Root Mean Square Error (RMSE) for each modality and participant. Due to the distinct nature of the errors, we separated the analysis into Lateral and Vertical components.

Lateral Accuracy: The registration method had a significant effect on lateral accuracy (Friedman $\chi^2(2) = 16.625$, $p < 0.001$). As detailed in Table 3, the Automatic method achieved superior precision with an RMSE of 2.37 cm. In contrast, the Manual method resulted in the highest error of 23.16 cm. The QR method performed moderately, with a mean error of 8.00 cm. Post-hoc analysis confirmed that the Automatic method significantly outperformed both Manual ($p < 0.001$) and QR ($p < 0.01$) approaches in the horizontal plane. Additionally, the QR method significantly outperformed the Manual approach ($p = 0.019$) for lateral alignment.

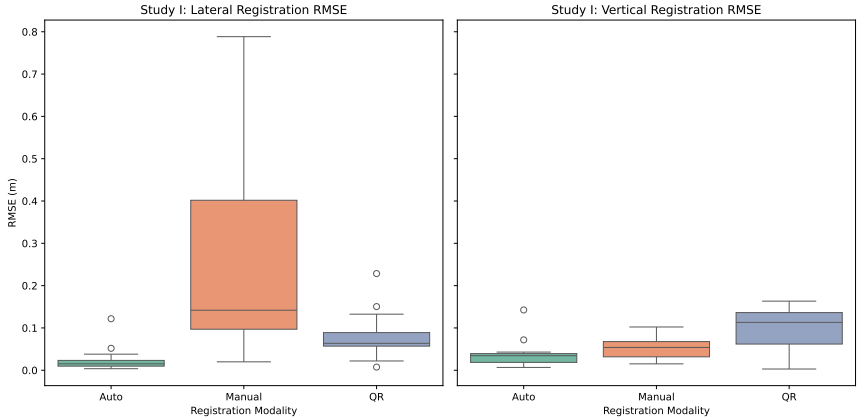


Figure 9: Boxplots representing the distribution of registration errors in Study I ($N = 16$). Left: Absolute Lateral Error, showing high variance for the Manual modality. Right: Absolute Vertical Error, highlighting the increased deviation in the QR-based method. The Automatic approach (green) demonstrates consistent stability across both axes.

Modality	Lateral Error (cm)		Vertical Error (cm)		Perceived Difficulty (SEQ)		
	Mean	SD	Mean	SD	Mean	Median	SD
Automatic	**2.37**	2.90	**3.77**	3.20	6.1	6.0	0.8
QR-Code	8.00	5.32	9.98	4.81	**6.7**	7.0	0.6
Manual	23.16	21.06	5.10	2.50	3.0	2.0	1.9

Table 3
Study I Results: Quantitative alignment errors and subjective difficulty scores (SEQ) across the three registration paradigms.

Vertical Accuracy: Analysis of vertical error revealed a distinct pattern (Friedman $\chi^2(2) = 10.125$, $p < 0.01$). The Automatic method (RMSE = 3.77 cm) and the Manual method (RMSE = 5.10 cm) showed no statistically significant difference ($p = 0.1329$). This performance parity is likely attributable to the specific constraints of the manual interaction, which restricted rotation to the vertical Y-axis. By eliminating complex pitch and roll adjustments, the task was effectively simplified, allowing users to easily match the virtual floor plane with the physical ground. Conversely, the QR method performed significantly worse vertically (RMSE = 9.98 cm) compared to both Automatic ($p = 0.0064$) and Manual ($p = 0.033$) methods. This error typically stems from pitch estimation inaccuracies when scanning markers at oblique angles, which translates into larger vertical offsets when projected over distance. To better illustrate the reliability of these paradigms beyond central tendencies, Figure 9 visualizes the error distribution for both axes. The boxplots highlight the severe dispersion of the manual baseline compared to the consistent stability of the proposed automatic approach.

Perceived Difficulty: Table 3 summarizes the descriptive statistics for the perceived difficulty (SEQ) scores. The results indicate a clear dichotomy in user experience. The Manual method was rated as the most difficult, with a mean score of 3.0 ($SD = 1.9$) and a median of 2.0. Conversely, the semi-automated methods were rated highly: the QR-Code method achieved the highest usability scores ($Mean = 6.7, SD = 0.6$), followed closely by the Automatic method ($Mean = 6.1, SD = 0.8$). A Friedman test revealed a statistically significant difference in perceived difficulty between methods, $\chi^2(2) = 19.32, p < 0.001$. Post-hoc analysis confirmed that the Manual method was significantly more burdensome than both the QR ($Z = 3.35, p < 0.001$) and Automatic ($Z = 3.26, p = 0.001$) methods. No significant difference was found between the QR and Automatic approaches ($Z = 1.89, p = 0.058$).

User Preference: Subjective rankings mirrored these findings. The QR-Code method was the primary choice for 81% of participants ($N = 13$), favored for its speed. The Automatic method was the preferred alternative for the remaining participants, while the Manual method was consistently ranked last ($N = 14$), with users citing the frustration of perceived drift.

5.2. Study II: Evaluation of the Impact of Registration Quality on UX and Spatial Perception

The primary objective of Study II is to evaluate the impact of registration degradation on User Experience (UX) and spatial decision-making. While Study I focused on quantifying technical precision across different registration paradigms, this second phase investigates the threshold at which alignment errors compromise human performance. By manipulating alignment parameters to introduce controlled errors, we scientifically measure their effect on user operability, architectural perception, and safety judgments.

5.2.1. Participants

A total of $N = 10$ participants were recruited for this specific study. The cohort was demographically diverse,

with ages ranging from 18 to over 45. Participants were screened to ensure no significant visual or motor impairments that would hinder the use of stereoscopic HMDs or the execution of precise gestures, matching the baseline requirements of the previous study.

5.2.2. Experimental Design

We employed a within-subjects design with Registration Quality as the single independent variable. To mitigate learning effects and order bias, the exposure sequence was organized using a Balanced Latin Square (4×4), randomly assigning each participant to one of the four sequential groups.

The independent variable, Registration Quality, was defined through four experimental conditions designed to simulate misalignment levels commonly encountered in construction scenarios. To ensure ecological validity, the offset magnitudes were empirically derived from the second and third quartiles of the baseline errors recorded in Study I (utilizing the Manual mode for horizontal error and QR-based mode for vertical error):

- **C1 (Control):** Ideal visualization with no added distortion, establishing the performance baseline.
- **C2 (Moderate Offset):** Introduction of a moderate deviation (vertical offset $\approx$ 11 cm, horizontal offset $\approx$ 14 cm), corresponding to the second quartile of the baseline errors.
- **C3 (Severe Offset):** Introduction of a severe deviation (vertical offset $\approx$ 13 cm, horizontal offset $\approx$ 40 cm), corresponding to the third quartile of the baseline errors.
- **C4 (Noise):** Introduction of jitter-type instability (rotational noise) without translational displacement.

The dependent variables encompassed both objective and subjective measures. Objective metrics included absolute error (m), 2D Euclidean distance, binary success rates, and task execution times. Subjective metrics focused on perceived difficulty and operational impact of visual noise.

5.2.3. Procedure

The experimental protocol proceeded through four distinct blocks (C1–C4) according to the randomized sequence defined by the Balanced Latin Square. At the start of each block, the system applied the specific registration distortion condition. Participants then performed a battery of six tasks categorized into two phases:

Phase A: Precision and Interaction (T1–T2) Participants executed tasks requiring direct interaction with the environment:

- **T1 Measurement:** Users estimated the length of virtual bars using a virtual measuring tool to evaluate metric consistency.
- **T2 Positioning:** Participants placed a physical marker on a wall to match a virtual stimulus, allowing for the measurement of hand-eye coordination drift.

Phase B: Spatial Perception with Avatars (T3–T5) Subsequently, virtual avatars were introduced as human-scale references to evaluate architectural judgment:

- **T3 Spatial Reference:** Participants identified physical objects indicated by an avatar's gesture to determine if registration errors compromised the perceived Line of Sight.
- **T4 Accessibility:** Users judged whether an avatar in a wheelchair could navigate specific trajectories without colliding with real-world physical elements (affordance judgment).
- **T5 Hazard Assessment:** Participants determined if a walking avatar was in physical danger regarding real obstacles.

Immediately following the completion of each individual task, participants provided a subjective difficulty rating via the Single Ease Question (SEQ) on a 7-point Likert scale (1=Very Difficult, 7=Very Easy). The procedure concluded with a final questionnaire specifically evaluating the operational impact of the noise condition, rating the extent to which visual instability impeded their ability to perform the assigned tasks.

5.2.4. Results: Task T1 – Measurement

In this task, participants were presented with four groups of four virtual bars (one group per condition) and were required to provide an estimate of their length using a virtual measuring tool. In instances where registration conditions prevented a full measurement of the bar, users were asked to provide a visual estimate of the total length. The order of the bars remained constant across subjects, while the sequence of conditions (C1 to C4) was organized using a Balanced Latin Square design.

Metric Precision Analysis: Registration Quality had a significant impact on metric precision (Friedman $\chi^2(3) = 24.36, p < 0.0001$). As detailed in Table 4, the Control condition (C1) exhibited superior precision with an estimated RMSE of 1.71 cm. Conversely, the Severe Offset (C3) yielded the lowest precision with an error of 24.72 cm, followed by C2 (6.01 cm) and C4 (3.84 cm). Post-hoc analysis confirmed that C1 achieved a significantly lower error compared to C2 ($p = 0.0234$), C3 ($p = 0.0117$), and C4 ($p = 0.0234$). Additionally, C2 ($p = 0.0234$) and C4 ($p = 0.0117$) both performed significantly better than C3. No significant differences were found between C2 and C4 ($p = 0.1641$).

Perceived Difficulty: Table 4 summarizes the descriptive statistics for the perceived difficulty (SEQ) scores. Results indicate a clear preference for the Control condition (C1), rated as the easiest with a mean score of 6.2 ($SD = 0.6$) and a median of 6.0. In contrast, the Severe Offset (C3)

Condition	Measurement Error (RMSE - cm)		Perceived Difficulty (SEQ)		
	Mean	SD	Mean	Median	SD
C1	**1.71**	1.00	**6.2**	6.0	0.6
C2	6.01	3.51	4.3	4.5	1.5
C3	24.72	31.48	2.9	2.5	1.5
C4	3.84	1.91	4.7	5.0	1.3

Table 4
Task T1 Results: Comparative statistics of measurement error and subjective difficulty across registration conditions.

represented the greatest difficulty, obtaining the lowest score in the study ($Mean = 2.9, SD = 1.5$). Conditions C2 and C4 showed an intermediate performance, both with a median of 5.0. A Friedman test revealed a statistically significant difference in perceived difficulty between conditions, $\chi^2(3) = 25.30, p < 0.001$. Post-hoc analysis confirmed that C1 was preferred over C2 ($p = 0.0234$), C3 ($p = 0.0117$), and C4 ($p = 0.0234$). Furthermore, the analysis confirms a significant preference for conditions C2 ($p = 0.0469$) and C4 ($p = 0.0234$) over C3, with no significant differences found between C2 and C4.

5.2.5. Results: Task T2 – Positioning

In this task, participants were presented with four groups of four virtual elements (one group per condition) positioned over a virtual workspace. Users were asked to mark the intended physical coordinates (as if no registration error was present) as accurately as possible on the physical workspace. The order of the elements remained constant across subjects, while the sequence of conditions (C1 to C4) was organized using a Balanced Latin Square design.

Metric Precision Analysis: Registration Quality had a significant impact on metric precision (Friedman $\chi^2(3) = 19.80, p = 0.00018$). As detailed in Table 5, the Control condition (C1) exhibited superior precision with an estimated RMSE of 1.80 cm. In contrast, the Moderate (C2), Severe (C3), and Noise (C4) conditions yielded significantly lower precision with errors of 9.32 cm, 10.35 cm, and 5.95 cm, respectively. Post-hoc analysis confirmed that C1 achieved a significantly lower error compared to C2 ($p = 0.0117$), C3 ($p = 0.0117$), and C4 ($p = 0.0234$). Interestingly, no significant differences were found between the degraded conditions: C2 and C3 ($p = 1.0$), C2 and C4 ($p = 0.1172$), or C3 and C4 ($p = 0.1641$).

Condition	Positioning Error (RMSE - cm)		Perceived Difficulty (SEQ)		
	Mean	SD	Mean	Median	SD
C1	**1.80**	0.94	**6.9**	7.0	0.3
C2	9.32	3.94	4.3	4.0	1.2
C3	10.35	5.94	3.4	3.0	1.4
C4	5.95	2.60	4.0	3.0	1.8

Table 5
Task T2 Results: Comparative statistics of positioning error and subjective difficulty across registration conditions.

Perceived Difficulty: Table 5 summarizes the descriptive statistics for the perceived difficulty (SEQ) scores. Results indicate a strong preference for the Control condition (C1), which was rated as the easiest with a mean score of 6.9 ($SD = 0.32$) and a median of 7.0, reflecting consistent user satisfaction. Conversely, the remaining conditions represented a much higher difficulty: Severe Offset (C3) was rated the lowest ($Mean = 3.4, SD = 1.4$), followed by C4 ($Mean = 4.0, SD = 1.8$) and C2 ($Mean = 4.3, SD = 1.2$). A Friedman test revealed a statistically significant difference in perceived difficulty between conditions, $\chi^2(3) = 22.38, p < 0.0001$. Post-hoc analysis with Wilcoxon signed-rank tests confirmed that C1 was significantly preferred over C2 ($p = 0.0117$), C3 ($p = 0.0117$), and C4 ($p = 0.0117$). No significant differences in perceived difficulty were found between conditions C2, C3, and C4.

5.2.6. Results: Task T3 – Spatial Reference

In this task, participants observed four groups of four virtual avatars (one per condition) pointing toward distinct real-world objects. Users identified the specific physical object referenced by each avatar. Performance was measured using a discrete success score (0 to 4) per condition to account for the clustered data structure. This approach avoids the independence assumptions of traditional categorical tests like the McNemar, which are violated by individual learning curves. While the sequence of avatars remained constant across participants, the presentation order of experimental conditions (C1 to C4) followed a Balanced Latin Square design.

Identification Accuracy Analysis: Registration Quality had a significant impact on identification accuracy (Friedman $\chi^2(3) = 19.80, p = 0.000187$). As detailed in Table 6, the Control condition (C1) achieved absolute success with a 100% identification rate. Performance significantly declined as registration quality degraded: the Moderate (C2) condition yielded a 65% success rate, while the Severe (C3) and Noise (C4) conditions resulted in notably lower success rates of 40% and 37.5%, respectively. Post-hoc analysis confirmed that C1 outperformed C2 ($p = 0.0469$), C3 ($p = 0.0117$), and C4 ($p = 0.0234$). No significant differences were found between the degraded conditions (C2, C3, and C4).

Condition	Success Score		Success	Perceived Difficulty		
	Mean	SD	**Rate**	Mean	Median	SD
C1	**4.0**	0.00	**100%**	**6.9**	7.0	0.3
C2	2.6	1.08	65%	4.7	4.5	1.6
C3	1.6	1.17	40%	3.7	3.5	1.8
C4	1.5	1.18	37.5%	3.9	4.5	2.1

Table 6
Task T3 Results: Comparative statistics of identification success and subjective difficulty across registration conditions.

Perceived Difficulty: Table 6 summarizes the descriptive statistics for the perceived difficulty (SEQ) scores. Results indicate a clear preference for the Control condition (C1), rated as the easiest with a mean score of 6.9 ($SD =$

0.3) and a median of 7.0. Conversely, the remaining conditions represented a much higher cognitive burden: the Severe Offset (C3) condition was rated the lowest ($Mean = 3.7, SD = 1.8$), followed by C4 ($Mean = 3.9, SD = 2.1$) and C2 ($Mean = 4.7, SD = 1.6$). A Friedman test revealed a statistically significant difference in perceived difficulty between conditions, $\chi^2(3) = 19.88, p = 0.00018$. Post-hoc analysis with Wilcoxon signed-rank tests confirmed that C1 was significantly preferred over C2 ($p = 0.0234$), C3 ($p = 0.0117$), and C4 ($p = 0.0234$). No significant differences in perceived difficulty were found between the degraded conditions C2, C3, and C4.

5.2.7. Results: Task T4 – Accessibility

In this task, participants observed a virtual wheelchair avatar performing four distinct trajectories (one per condition). Users judged whether the avatar could complete each path without colliding with real-world physical elements. To evaluate how registration quality impacts passability affordances, we introduced alignment offsets to distort the avatar's perceived position (despite all paths being technically clear). While the sequence of trajectories remained constant, the order of experimental conditions (C1 to C4) was randomized using a Balanced Latin Square design.

Affordance Judgment Analysis: Both the Cochran Q test ($p = 0.0719$) and the Friedman test ($\chi^2(3) = 7, p = 0.0719$) indicated that no statistically significant differences were found at the standard confidence level.

Condition	Accuracy (Binary)		Success	Perceived Difficulty		
	Mean	SD	Rate	Mean	Median	SD
C1	**1.0**	0.00	**100%**	**6.7**	7.0	0.7
C2	0.7	0.48	70%	5.2	6.0	1.4
C3	0.5	0.53	50%	4.3	4.0	1.7
C4	0.6	0.52	60%	4.9	6.0	2.0

Table 7
Task T4 Results: Comparative statistics of success rate and subjective difficulty across registration conditions.

Perceived Difficulty: Table 7 summarizes the descriptive statistics for the perceived difficulty (SEQ) scores. The results indicate a clear preference for the Control condition (C1), which was rated as the easiest with a mean score of 6.7 ($SD = 0.7$) and a median of 7.0. Conversely, C3 represented the greatest difficulty, obtaining a mean score of 4.3 ($SD = 1.7$). A Friedman test revealed a statistically significant difference in perceived difficulty between conditions ($\chi^2(3) = 18.31, p = 0.000379$). Post-hoc analysis with Wilcoxon signed-rank tests confirmed that C1 was significantly preferred over C2 ($p = 0.0234$), C3 ($p = 0.0234$), and C4 ($p = 0.0469$). No significant differences in perceived difficulty were found between the degraded conditions C2, C3, and C4. These results reveal a dissociation between objective efficacy and subjective experience. While users compensated for registration errors to reach correct judgments, perceived difficulty scores demonstrate that any perturbation increases the cognitive load required to process spatial information.

5.2.8. Results: Task T5 – Hazard Assessment

In this final task, participants observed a virtual avatar performing four distinct trajectories (one per condition). Users judged whether the avatar traversed a pre-identified hazard zone during its path. To evaluate how registration impacts safety perception, we introduced alignment offsets to distort the avatar's perceived position (despite all paths technically intersecting the zone). While the sequence of trajectories remained constant, the order of experimental conditions (C1 to C4) was randomized using a Balanced Latin Square design.

Hazard Judgment Analysis: A notable ceiling effect was observed in this task, with a 100% success rate recorded across all conditions. This indicates that the task had limited sensitivity for discriminating performance differences between registration conditions. Accordingly, T5 should not be interpreted as strong evidence of condition-dependent differences in objective hazard detection accuracy.

Condition	Accuracy (Binary)		Success	Perceived Difficulty		
	Mean	SD	Rate	Mean	Median	SD
C1	1.0	0.00	100%	**6.8**	7.0	0.4
C2	1.0	0.00	100%	5.1	5.5	1.1
C3	1.0	0.00	100%	4.5	5.0	1.6
C4	1.0	0.00	100%	5.3	6.0	1.7

Table 8
Task T5 Results: Comparative statistics of success rate and subjective difficulty across registration conditions.

Perceived Difficulty: Table 8 summarizes the descriptive statistics for the perceived difficulty (SEQ) scores. The results indicate a clear preference for the Control condition (C1), which was rated as the easiest with a mean score of 6.8 ($SD = 0.4$) and a median of 7.0. In contrast, the degraded conditions (C2, C3, and C4) exhibited lower usability ratings, with the Severe Offset (C3) receiving the lowest score ($Mean = 4.5, SD = 1.6$). A Friedman test revealed a statistically significant difference in perceived difficulty between conditions, $\chi^2(3) = 22.21, p = 0.00005$. Post-hoc analysis with Wilcoxon signed-rank tests confirmed that C1 was significantly preferred over C2 ($p = 0.0117$) and C3 ($p = 0.0117$). No significant differences were found between C1 and C4, nor among the three degraded conditions (C2, C3, and C4). These findings reveal a dissociation between objective performance and subjective perception. While users identified hazards despite registration errors, degraded SEQ scores indicate this occurred at the cost of higher cognitive load and reduced system confidence.

6. Limitations and Future Work

While the proposed registration pipeline demonstrates robust performance in AEC environments, it is bound by the computational constraints of the Microsoft HoloLens 2. The decision to restrict the scope to an on-device computational setup was driven by the reality of active construction sites, which frequently lack the reliable, high-bandwidth network connectivity required for cloud-based processing. However,

acknowledging these limitations, future research could evaluate a hybrid computational model. This approach would allow the two ecosystems to coexist seamlessly by preserving the fully standalone pipeline as a fallback mechanism for disconnected sites, with the possibility of dynamically offloading computations to heavy solvers on edge or cloud servers when adequate network coverage is detected.

Additionally, the HoloLens 2 relies on an infrared Time-of-Flight depth sensor, which produces reliable samples up to 3.5 to 4.0 meters in the Long Throw setting. In open outdoor environments, this range may be limiting, and the capture may also suffer from severe infrared interference from ambient sunlight, which can make geometric scanning unreliable.

Furthermore, it is important to note that HoloLens 2 production has ended ("last time to buy"), with critical servicing continuing only for a limited window. This constrains the long-term applicability of the current hardware framework. To address this, future work should focus on extending the portability of the proposed pipeline to modern video see-through Mixed Reality devices, such as the Meta Quest 3. Transitioning to a video see-through architecture offers significant perceptual advantages over optical see-through displays like the HoloLens, which often struggle to render opaque holograms in the bright-light conditions typical of outdoor or well-lit construction sites. Devices like the Meta Quest 3 resolve this by blending high-resolution color passthrough with digital elements, allowing virtual BIM models to be seamlessly and solidly integrated into the user's view, albeit with a potential latency trade-off.

Adapting this registration framework to such devices will require a paradigm shift in how environmental scanning is handled. Instead of relying solely on explicit depth sweeps, future implementations could interface with the platform's native spatial understanding stack. For instance, querying persistent scene models (planes, volumes, meshes) across sessions and leveraging shared spatial anchors with colocation could enable robust multi-user alignment. Subsequent research should investigate how to interface BIM preprocessing and alignment algorithms directly with these spatial mapping capabilities, moving away from manual scanning phases in favor of built-in environmental understanding tools.

Finally, to explicitly address the challenges of symmetric or feature-poor spaces, subsequent iterations of the framework could formalize and evaluate layered intra-pipeline fallback mechanisms. These might include semantic seeding from detected planes, QR or marker-based initialization, minimal manual nudges, re-binding to fresh spatial anchors over long traversals, and hybrid offloading when available, thereby increasing the overall robustness of the system.

7. Conclusion

This article presented a comprehensive MR registration framework for the AEC industry, tailored for the HoloLens 2. We leverage several optimization strategies, specifically offline probe-based ray casting for BIM inner-shell extraction, edge-aware point cloud decimation, and a multi-resolution RANSAC pipeline coupled with multi-scale ICP refinement, to achieve robust, centimeter-level precision (RMSE = 2.37 cm) without relying on external markers or cloud-based offloading. While the current implementation is tailored to the HoloLens 2, the core methodology is highly generalizable to the broader AEC sector and modern video see-through architectures.

Because technical precision in Mixed Reality is insufficient if it does not translate to improved human decision-making, we conducted comparative evaluations to prove our automatic pipeline superior to manual and QR-code alternatives. In the lateral plane, our method's 2.37 cm precision significantly outperformed QR markers (8.00 cm) and manual alignment (23.16 cm). Although users preferred QR codes for simplicity, pitch estimation errors at oblique angles caused vertical offsets approximately three times greater than those achieved by our automated system. Participants ranked the manual method least favorably, citing frustration with perceived drift and high cognitive load. These results confirm our alignment method as essential for reliable 1:1 scale visualization on-site.

User study findings highlight how registration quality affects architectural judgment and safety. A drop to a 50% success rate in wheelchair affordance judgments (Task T4) under severe offsets proves that registration errors can cause incorrect site management decisions. Similarly, significantly degraded usability scores during hazard assessments indicate that misalignments force a fatiguing mental recalibration, proving that algorithmic robustness directly dictates user trust and task efficacy. These findings reinforce the core idea that securing a high-fidelity initial alignment yields measurable and lasting benefits. A superior initial registration establishes a stable cognitive baseline that enhances spatial perception and decision-making throughout the entire inspection session, even if the subsequent tracking relies on the standard device spatial mapping pipeline.

In summary, this research establishes that reliable, on-device geometric registration is a fundamental prerequisite for the effective integration of Mixed Reality in construction workflows. By successfully balancing computational efficiency with algorithmic robustness, our framework provides a viable pathway for 1:1 scale on-site inspection. As the AEC industry transitions toward more advanced immersive technologies, the adaptable, multi-layered alignment strategies and hardware considerations discussed in this work will be crucial for the continued evolution of spatial computing in the built environment.

CRediT authorship contribution statement

Anonymous: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing..

References

[1] Aiger, D., Mitra, N.J., Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface registration, in: ACM SIGGRAPH 2008 Papers, Association for Computing Machinery, New York, NY, USA. pp. 1–10. doi:https://doi.org/10.1145/1399504.1360684.

[2] Azuma, R.T., 1997. A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6, 355–385. doi:https://doi.org/10.1162/pres.1997.6.4.355.

[3] Besl, P.J., McKay, N.D., 1992. Method for registration of 3-d shapes, in: Schenker, P.S. (Ed.), Sensor Fusion IV: Control Paradigms and Data Structures, International Society for Optics and Photonics. SPIE, pp. 586–606. doi:https://doi.org/10.1117/12.57955.

[4] Bianchini, C., Inglese, C., Ippolito, A., Maiorino, D., Senatore, L.J., 2017. Building information modeling (bim): Great misunderstanding or potential opportunities for the design disciplines?, in: Handbook of research on emerging Technologies for Digital Preservation and Information Modeling. IGI Global, pp. 67–90.

[5] Boan, T., Jiajun, L., Bosché, F., 2025. Autonomous mixed reality framework for real-time construction inspection. Journal of Information Technology in Construction (ITcon) 30, 852–874. doi:https://doi.org/10.36680/j.itcon.2025.035.

[6] Bosché, F., 2010. Automated recognition of 3d cad model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. Advanced Engineering Informatics 24, 107–118. doi:https://doi.org/10.1016/j.aei.2009.08.006. informatics for cognitive robots.

[7] Chi, H.L., Kang, S.C., Wang, X., 2013. Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. Automation in Construction 33, 116–122. doi:https://doi.org/10.1016/j.autcon.2012.12.017. augmented Reality in Architecture, Engineering, and Construction.

[8] Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395. doi:https://doi.org/10.1145/358669.358692.

[9] Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W., 1992. Surface reconstruction from unorganized points, in: Proceedings of the 19th annual conference on computer graphics and interactive techniques, pp. 71–78.

[10] Khan, D., Cheng, Z., Uchiyama, H., Ali, S., Asshad, M., Kiyokawa, K., 2022. Recent advances in vision-based indoor navigation: A systematic literature review. Computers & Graphics 104, 24–45. doi:https://doi.org/10.1016/j.cag.2022.03.005.

[11] Lazaro, M.J., Kim, S., 2025. Review of multimodal interaction in optical see-through augmented reality. International Journal of Human–Computer Interaction 41, 11227–11243. doi:https://doi.org/10.1080/10447318.2024.2442128.

[12] Liu, H., Liu, D., Chen, J., 2024. Depth-informed point cloud-to-bim registration for construction inspection using augmented reality. Advanced Engineering Informatics 62, 102867. doi:https://doi.org/10.1016/j.aei.2024.102867.

[13] Manfredi, G., Capece, N., Carlo, R.P.D., Erra, U., 2024. A mixed reality application for multi-floor building evacuation drills using real-time pathfinding and dynamic 3d modeling, in: Caputo, A., Garro, V., Giachetti, A., Castellani, U., Dulecha, T.G. (Eds.), Smart Tools and Applications in Graphics - Eurographics Italian Chapter Conference, The Eurographics Association. pp. 1–8. doi:https://doi.org/10.2312/stag.20241331.

[14] Mellado, N., Aiger, D., Mitra, N.J., 2014. Super 4pcs fast global pointcloud registration via smart indexing. Computer Graphics Forum 33, 205–215. doi:https://doi.org/10.1111/cgf.12446.

[15] Microsoft, 2025. Official microsoft hololens website. https://www.microsoft.com/hololens. Accessed: Apr 2025.

[16] Mur-Artal, R., Montiel, J.M.M., Tardós, J.D., 2015. Orb-slam: A versatile and accurate monocular slam system. IEEE Transactions on Robotics 31, 1147–1163. doi:https://doi.org/10.1109/TRO.2015.2463671.

[17] Pauly, M., Kobbelt, L.P., Gross, M., 2002. Efficient simplification of point-sampled surfaces, in: Visualization Conference, IEEE, IEEE Computer Society, Los Alamitos, CA, USA. pp. 163–170. doi:https://doi.org/10.1109/VISUAL.2002.1183771.

[18] Qiao, Z., Huang, H., Liu, C., Yu, Z., Shen, S., Zhang, F., Yin, H., 2025. Speak the same language: Global lidar registration on bim using pose hough transform. IEEE Transactions on Automation Science and Engineering 22, 22282–22295. doi:https://doi.org/10.1109/TASE.2025.3549176.

[19] Ress, V., Zhang, W., Skuddis, D., Haala, N., Soergel, U., 2024. Slam for indoor mapping of wide area construction environments. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences X-2-2024, 209–216. doi:https://doi.org/10.5194/isprs-annals-X-2-2024-209-2024.

[20] Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration, in: 2009 IEEE International Conference on Robotics and Automation, pp. 3212–3217. doi:https://doi.org/10.1109/ROBOT.2009.5152473.

[21] Sidani, A., Matoseiro Dinis, F., Duarte, J., Sanhudo, L., Calvetti, D., Santos Baptista, J., Poças Martins, J., Soeiro, A., 2021. Recent tools and techniques of bim-based augmented reality: A systematic review. Journal of Building Engineering 42, 102500. doi:https://doi.org/10.1016/j.jobe.2021.102500.

[22] Sun, C., Jiang, S., Skibniewski, M.J., Man, Q., Shen, L., 2017. A literature review of the factors limiting the application of bim in the construction industry. Technological and Economic Development of Economy 23, 764–779. doi:https://doi.org/10.3846/20294913.2015.1087071.

[23] Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X., 2021. Loftr: Detector-free local feature matching with transformers, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA. pp. 8918–8927. doi:https://doi.org/10.1109/CVPR46437.2021.00881.

[24] Swan, J.E., Jones, A., Kolstad, E., Livingston, M.A., Smallman, H.S., 2007. Egocentric depth judgments in optical, see-through augmented reality. IEEE Transactions on Visualization & Computer Graphics 13, 429–442. doi:https://doi.org/10.1109/TVCG.2007.1035.

[25] Vassena, G.P.M., Perfetti, L., Comai, S., Mastrolembo Ventura, S., Ciribini, A.L.C., 2023. Construction progress monitoring through the integration of 4d bim and slam-based mapping devices. Buildings 13. doi:https://doi.org/10.3390/buildings13102488.

[26] Vega Torres, M.A., Braun, A., Borrmann, A., 2023. Bim-slam: Integrating bim models in multi-session slam for lifelong mapping using 3d lidar, in: García de Soto, B., Gonzalez-Moret, V., Brilakis, I. (Eds.), Proceedings of the 40th International Symposium on Automation and Robotics in Construction, International Association for Automation and Robotics in Construction (IAARC), Chennai, India. pp. 521–528. doi:https://doi.org/10.22260/ISARC2023/0070.

[27] Volk, R., Stengel, J., Schultmann, F., 2014. Building information modeling (bim) for existing buildings — literature review and future needs. Automation in Construction 38, 109–127. doi:https://doi.org/10.1016/j.autcon.2013.10.023.

[28] Wagner, D., Schmalstieg, D., 2007. Artoolkitplus for pose tracking on mobile devices, in: Proceedings of the 12th Computer Vision Winter Workshop (CVWW'07), Sankt Lambrecht, Austria. pp. 139–146. URL: https://api.semanticscholar.org/CorpusID:14966142.

[29] Wang, J., Wang, X., Shou, W., Xu, B., 2014. Integrating bim and augmented reality for interactive architectural visualisation. Construction Innovation 14, 453–476. doi:https://doi.org/10.1108/CI-03-2014-0019.

[30] Yang, H., Shi, J., Carlone, L., 2021. Teaser: Fast and certifiable point cloud registration. IEEE Transactions on Robotics 37, 314–333. doi:https://doi.org/10.1109/TRO.2020.3033695.

[31] Zhou, Q.Y., Park, J., Koltun, V., 2016. Fast global registration, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision - ECCV 2016, Springer International Publishing, Cham. pp. 766–782. doi:https://doi.org/10.1007/978-3-319-46475-6_47.

Bridging BIM and Reality: A Hardware-Optimized Registration Pipeline for Mixed Reality in Indoor Construction Environments⋆

Marcos Arroyo-Ruiz[a], Gonzalo Gómez-Nogales[a], José Antonio Gómez-Fernández[b], Carlos Andújar[c] and Marc Comino-Trinidad[a,*]

[a] *Universidad Rey Juan Carlos, C/Tulipán s/n, Móstoles, 28933, Madrid, Spain*

[b] *Asociación para el fomento de la innovación y la sostenibilidad en la Arquitectura, ingeniería y construcción (I2Con), Plaza Mayor 51, Piso 9, Puerta 18, Alzira, 46600, Valencia, Spain*

[c] *Universitat Politècnica de Catalunya, Jordi Girona, 31, Barcelona, 08034, Catalunya, Spain*

ARTICLE INFO

Keywords:
Mixed Reality
Building Information Modeling
Point Cloud Registration
Construction Inspection
Spatial Perception

ABSTRACT

Building Information Modeling (BIM) has transformed the Architecture, Engineering, and Construction (AEC) industry by digitizing project data, yet its full potential remains unrealized due to persistent gaps between virtual models and physical sites. These gaps contribute to inefficiencies, with studies reporting substantial waste in labor and coordination. Extended Reality (XR) technologies offer a promising solution by enabling immersive, real-scale visualization of BIM models on-site.

This article introduces a Mixed Reality (MR) application for Microsoft HoloLens 2 that superimposes BIM representations onto construction environments at a 1:1 scale, supporting real-time detection of differences between the as-designed BIM model and the as-built construction on site. We present a robust registration pipeline that integrates commercial XR hardware with advanced algorithms to achieve precise alignment under challenging conditions. To validate the system, we conducted a controlled user study comparing three registration paradigms (manual gesture-based, QR-assisted, and fully automatic) and analyzing their impact on alignment accuracy and user experience (UX) in AEC-related tasks. Results show that our automatic approach provides advantages over state-of-the-art alternatives and significantly improves registration precision and usability ratings over the baseline methods. Furthermore, the study demonstrates that alignment errors strongly influence spatial perception and decision-making, highlighting the necessity of high-fidelity registration for effective MR integration in construction workflows.

CRediT authorship contribution statement

Marcos Arroyo-Ruiz: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. **Gonzalo Gómez-Nogales:** Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing. **José Antonio Gómez-Fernández:** Data curation, Investigation, Resources, Software, Validation, Visualization. **Carlos Andújar:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing. **Marc Comino-Trinidad:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

Acknowledgments

This work is part of the project CPP2021-008847 "Dispositivo de Realidad Mixta para la visualización de modelos BIM a escala real en obras de edificación" (Mixed Reality Device for Full-Scale Visualization of BIM Models in Building Construction Projects), funded by MCIN/AEI/10.13039/501100011033 and by the European Union 'NextGenerationEU'/PRTR. The authors also gratefully acknowledge the collaboration and support of the rest of the project consortium members, Tesicnor and I2CON.

⋆This work is part of the project CPP2021-008847 "Dispositivo de Realidad Mixta para la visualización de modelos BIM a escala real en obras de edificación" (Mixed Reality Device for Full-Scale Visualization of BIM Models in Building Construction Projects), funded by MCIN/AEI/10.13039/501100011033 and by the European Union 'NextGenerationEU'/PRTR.

*Corresponding author

marcos.arroyo@urjc.es (M. Arroyo-Ruiz); gonzalo.gomez@urjc.es (G. Gómez-Nogales); ja.gomez@i2con.org (J.A. Gómez-Fernández); carlos.andujar@upc.edu (C. Andújar); marc.comino@urjc.es (M. Comino-Trinidad)

https://gonzalogn.com/ (G. Gómez-Nogales); https://www.cs.upc.edu/~virtual/home/index.html (C. Andújar); http://marccomino.github.io/ (M. Comino-Trinidad)

ORCID(s):

- A standalone geometric registration pipeline for HoloLens 2 achieves centimeter-level accuracy without cloud offloading.

- Automated alignment significantly outperforms manual and QR-based methods in lateral precision and on-site usability.

- Registration degradation directly compromises architectural spatial judgment and safety-critical hazard assessments.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: