Adaptive Voxel-Based Weighted Multisensor Point Cloud Fusion with Color Consistency for Indoor Digital Twin Construction

Zhouyan Qiu[1], Arshia Ghasemlou[2], Rabia Rashdi[3]

[1] School of Robotics, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University,
215400 Taicang, Suzhou, China - zhouyan.qiu@xjtlu.edu.cn
[2] CINTECX, Universidade de Vigo, GeoTECH Group, Campus Universitario de Vigo,
As Lagoas, Marcosende, 36310 Vigo, Spain - arshia.ghasemlou@uvigo.gal
[3] University of Minho, Institute for Sustainability and Innovation in Structural Engineering (ISISE),
Associate Laboratory Advanced Production and Intelligent Systems (ARISE),
Department of Civil Engineering, Guimarães, Portugal - rabiarashdi@civil.uminho.pt

Keywords: Laser Scanning, Point Cloud Registration, Sensor Fusion, Data Integration, Indoor Digital Twin, Smart Infrastructure.

Abstract

Building accurate indoor digital twins is essential for smart-building services such as asset tracking, space planning, and AR/VR navigation. Yet a single LiDAR sensor cannot supply both millimeter-level accuracy and complete coverage: fixed terrestrial laser scanners (TLSs) leave occlusion holes, whereas handheld mobile laser scanners (HMLSs) suffer from lower geometric stability and color drift. We propose an adaptive voxel-based fusion pipeline that combines a FARO Focus3D X330 TLS with a CHCNAV RS10 HMLS to overcome these limits. First, the handheld point cloud is rigidly registered to the TLS reference using Iterative Closest Point. TLS voxels with sparse points are flagged as holes; for each hole we admit only handheld points whose point-to-plane distance and normal deviation fall below strict thresholds, ensuring geometric consistency. Next, we correct color bias by learning a global linear RGB mapping from overlapping scans and refining it locally with weighted regression. Finally, we blend colors across the TLS–handheld boundary to remove visible seams. Experiments on classroom scenes from a smart-campus testbed show that our method recovers 85.7 % of missing surfaces, lowers the global point-to-plane RMSE by 14.8 %, and improves mean color difference by 22.2%. The resulting high-fidelity, color-consistent indoor models give facility managers and planners reliable data for maintenance scheduling, occupancy analysis, and long-term space optimization.

1. Introduction

1.1 Background

Indoor environments play a crucial role in a wide range of applications, including smart building management, robotic navigation, facility maintenance, and emergency responses. Recently, digital twins are often used for accurate and up-to-date digital representations of indoor environments enabling decisions in real time (Shaharuddin et al., 2022). Digital twins can be created using a range of technologies, including the Internet of Things (IoT), LiDAR, photogrammetry, and machine learning, which enable the automated generation of 3D models. These models not only capture the geometric properties of indoor scenes but also embed semantic information such as object labels, spatial relationships, and functional attributes.

Despite the growing deployment of these devices, the integration of heterogeneous data continues to pose significant challenges (Rashdi et al., 2022). Differences in scanning angles, lighting conditions, and point density make the registration process more complex. For example, no single LiDAR device can capture both the high-precision geometry and holistic coverage required: fixed terrestrial laser scanners deliver millimeter-level accuracy (Rashdi et al., 2024) but suffer from occlusion-induced holes, while handheld scanners boast flexible viewpoints yet exhibit lower structural stability and inconsistent point density (Shin and Kwak, 2024; Balado et al., 2025). Equally problematic is the visual seam introduced by sensor-specific color bias: different camera systems and lighting conditions yield systematic RGB offsets between scans, undermining model realism.

In this paper, we introduce a unified fusion pipeline that co-registers a mobile SLAM LiDAR scan with a TLS reference and then automatically fills in holes and color-matches the data. In this regard, we present a unified registration and color-harmonization framework that seamlessly fuses FARO and handheld point clouds into a complete, consistently colored indoor twin (Yoon and Koo, 2023; González-Collazo et al., 2024). We demonstrate the approach on real indoor scenes and report significant gains in completeness, accuracy, and visual quality over using TLS alone.

1.2 Related Work

LiDAR for Indoor Mapping. Indoor LiDAR mapping systems broadly fall into static terrestrial laser scanners (TLS) and mobile/mobile-mapping LiDAR platforms, each with distinct advantages and drawbacks. TLS devices deliver millimeter-level accuracy and dense point measurements but are susceptible to occlusion-induced gaps in cluttered environments and require time-consuming multi-station setups for full coverage (Luhmann et al., 2020; Qiu et al., 2023). In contrast, handheld and SLAM-based systems provide flexible viewpoints and rapid area coverage, though they introduce cumulative drift, uneven point density, and color inconsistencies; comparative evaluations of LiDAR SLAM algorithms (ICP variants, graph optimization, particle filters) highlight varied performance in loop closure and stability across feature-poor indoor scenes (Zou et

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W15-2025-127-2025 | © Author(s) 2025. CC BY 4.0 License.

al., 2021). To address these challenges, recent reviews of mobile mapping systems survey the integration of LiDAR with cameras and IMUs, demonstrating how sensor fusion enhances pose estimation, enables direct georeferencing, and enriches semantic content for autonomous navigation and digital twin generation (Elhashash et al., 2022). Beyond algorithmic improvements, systematic analyses of low-cost 3D mapping solutions emphasize the trade-offs between accuracy, coverage, and operational efficiency across handheld, trolley, and vehicle-mounted platforms (Balado et al., 2025). Finally, voxel-based fusion of heterogeneous scans—including aerial, terrestrial, and handheld data—has shown promise in recovering occluded surfaces and harmonizing color biases, achieving centimeter-level completeness and visual consistency in complex indoor and heritage environments (Roggero and Diara, 2024).

Point Cloud Registration. Robust alignment of overlapping point clouds is critical for building accurate 3D models of indoor environments. The Iterative Closest Point (ICP) algorithm, introduced by Besl and McKay, is still the standard for fine registration: it iteratively minimizes point-to-point distances between two clouds, but it often converges slowly and is sensitive to the initial alignment (Besl and McKay, 1992). To address these limitations, Das and Waslander (2014) propose a Segmented Region Growing NDT (SRG-NDT) that first removes ground points and then clusters remaining points into Gaussian distributions, yielding a smooth cost function and reducing runtime by over 90 % compared to voxel-based NDT and ICP variants. In large-scale SLAM settings, Zou et al. (2021) provide a comparative analysis of LiDAR-SLAM registration methods—scan-to-scan and scan-to-map ICP, Generalized-ICP, NDT, and graph-based optimizations—highlighting trade-offs in robustness, loop-closure performance, and computational efficiency in feature-poor indoor scenarios. Together, these works form a comprehensive foundation for selecting and extending registration techniques in our adaptive TLS–HMLS fusion pipeline.

Color Calibration Color calibration in LiDAR–RGB point clouds refers to the process of adjusting the raw RGB measurements—typically captured by a co-registered camera rigidly mounted to the LiDAR or by an integrated imaging sensor—so that colors are consistent and accurate across multiple scans or viewpoints. Giacomini et al. (2024) introduce Ca²Lib, which uses small planar markers (e.g., A3 chessboards) to detect correspondences between LiDAR returns and image pixels, and then solves a joint non-linear optimization to estimate both the extrinsic alignment and per-channel photometric parameters. Li et al. (2024) propose LVBA, a LiDAR–Visual bundle adjustment framework: it first performs a global LiDAR-only bundle adjustment to refine sensor poses, then incorporates planar 3D features into a photometric bundle adjustment that optimizes camera poses and exposures for globally consistent RGB mapping. Xing et al. (2022) tackle illumination variation directly on RGB–D scans with Point Cloud Color Constancy (PCCC), a deep-learning approach based on PointNet that estimates the scene's illumination chromaticity per point and applies a global correction to achieve color constancy under varying lighting conditions.

Accurate color calibration is critical for applications that rely on both geometric and photometric fidelity. In heritage documentation and virtual tourism, it ensures that textured 3D reconstructions faithfully reproduce material appearance under varying lighting. In robotics and autonomous navigation, consistent color in point clouds enhances semantic segmentation, object recognition, and change detection. For AR/VR and digital-twin platforms, seamless color integration across scans improves visual realism and user immersion, reducing artifacts at sensor boundaries and enabling reliable photometric measurements for downstream analytics.

In Luhmann et al. (2020), terrestrial laser scans, UAV photogrammetric point clouds, and close-range imagery are co-registered by first aligning each dataset to a common control-point network, refining the result via target detection (retro-reflective spheres or ground control points), and applying a global ICP adjustment to achieve millimeter-level residuals. Roggero and Diara (2024) segment TLS, airborne LiDAR, and UAV data into voxels based on point-density and roughness, perform feature-based coarse alignment followed by ICP for each sensor modality, and merge the datasets voxel-wise by selecting the most accurate source per region. Elhashash et al. (2022) survey mobile mapping systems that tightly fuse LiDAR, cameras, and IMUs through extrinsic calibration (using planar targets), time synchronization, SLAM-based pose estimation (e.g., MSCKF or bundle adjustment), and back-projection of LiDAR points into image frames for colorization and semantic labeling.

In contrast to existing mixed-sensor workflows that rely on external targets, photogrammetric imagery, or manual control networks for alignment and color calibration, our fully automated, voxel-based fusion pipeline operates entirely within the LiDAR domain. By extracting ceiling and wall primitives for robust coarse registration, selectively admitting only geometrically consistent handheld points to fill occlusions in terrestrial laser scans, and learning both global and local color mappings directly from overlapping scan regions, our approach delivers precise registration, comprehensive surface reconstruction, and seamless color integration in purely indoor environments—promising a fast, turnkey solution for high-fidelity 3D modeling without the overhead of mixed-sensor calibration or image processing.

2. Sensor Specifications

We experiment with two devices (Table 1). The CHCNAV RS10 is a SLAM-based handheld scanner with an integrated 4th-generation GNSS RTK antenna (CHCNAV, 2024). It provides approximately 1–5 cm accuracy (absolute horizontal/vertical RMS < 5 cm, relative < 1 cm) and covers a $360^\circ \times 270^\circ$ field of view (FoV). In 16-channel mode, it can achieve a range of approximately 120 m; using a higher 32-channel mode extends its range to around 300 m. The RS10 captures data at up to 320,000 points per second (in 16-channel mode) and operates on a swappable battery lasting approximately 1 hour. Its portability and real-time SLAM capabilities facilitate quick scanning of cluttered indoor environments; however, its angular resolution (approximately 0.18° per step) and precision remain moderate compared to TLS devices.

In contrast, the FARO Focus3D X330 is a tripod-mounted, phase-shift TLS offering exceptionally high precision with a ranging error of approximately ± 2 mm and a fine angular step resolution of 0.009° (around 40,960 samples per full 360° sweep) (FARO, 2025). It has an effective range of up to 330 m under optimal conditions, coupled with low measurement noise. Additionally, the X330 features a battery life of roughly 4.5 hours and captures up to 70 MP color imagery. However, its vertical FoV is $300^\circ \times 360^\circ$, resulting in a 60° blind zone vertically (up/down), and each scan is comparatively slow. In practice, certain occluded regions (e.g., beneath tables and behind pillars) remained unobserved. Table 1 summarizes these differences.

Feature	**CHCNAV RS10**	**Faro Focus3D X330**
Resolution	0.18°	0.009°
Accuracy	10 mm	2 mm
Range	0.5 m – 120 m	0.6 m – 130 m
Battery duration	1 h	4.5 h
Field of view	360° vertical / 270° horizontal	300° vertical / 360° horizontal
Measurement rate	320,000 points/s	976,000 points/s

Table 1. Comparison of CHCNAV RS10 (Handheld SLAM LiDAR) vs. FARO Focus3D X330 (TLS) capabilities.

3. Methodology

Our pipeline consists of three main stages: (1) Point Cloud Registration, (2) Hole Detection and Filling, and (3) color Harmonization.

3.1 Point Cloud Registration

To enhance the robustness of the point cloud registration method against initial rotational variations, this study employs the approach proposed by Qiu et al. (2025). This method establishes a clearly defined spatial coordinate framework by extracting floor, ceiling, and wall boundaries from indoor point cloud data, effectively resolving issues arising from arbitrary initial orientations.

Specifically, the ceiling plane of the indoor scene is initially estimated using the RANSAC algorithm, resulting in a plane equation defined as:

$$ax + by + cz + d = 0 \tag{1}$$

where (a, b, c) represents the normal vector of the plane, and d is the plane offset. Subsequently, points from the original cloud within a distance δ from this plane are removed according to:

$$|ax + by + cz + d| < \delta \tag{2}$$

The remaining point cloud data is then segmented vertically into multiple horizontal thin layers, each projected onto a reference plane parallel to the ceiling. Given the original point coordinates (x, y, z), the projected two-dimensional coordinates become (x', y'). The frequency $f(i, j)$ of occupancy for each grid cell in the two-dimensional projected plane is computed to identify stable wall structures:

$$f(i,j) = \sum_{k=1}^{K} v_k(i,j) \tag{3}$$

where K represents the total number of layers, and $v_k(i, j)$ indicates the occupancy state of grid cell (i, j) in the k-th layer.

Subsequently, RANSAC-based line fitting is applied to these stable structures to derive a set of linear segments representing wall boundaries. These segments are further merged and extended to clearly define the room layout.

Following the establishment of the standardized spatial coordinate framework described above, the classical Iterative Closest Point (ICP) algorithm is employed for accurate registration of the point clouds. The ICP algorithm iteratively matches pairs of points to minimize the distance error between the source point cloud $P = \{p_i\}$ and the target point cloud $Q = \{q_i\}$, optimizing the following objective function (Sun et al., 2024):

$$\min_{R,t} \sum_i \|Rp_i + t - q_{\sigma(i)}\|^2 \tag{4}$$

where R is the rotation matrix, t is the translation vector, and $\sigma(i)$ indexes the corresponding matching points. This comprehensive approach significantly enhances the robustness of the registration method, ensuring accuracy and reliability in subsequent fusion and analysis processes.

3.2 Hole Detection and Filling

To accurately identify and fill gaps within the Faro TLS data, we first discretize the point cloud space into cubic voxels of size r. For each point p in the Faro cloud $\mathcal{P}_F$, we define its voxel coordinates $v(p)$ as:

$$v(p) = \left\lfloor \frac{p}{r} \right\rfloor. \tag{5}$$

The set of occupied voxels by Faro points $\mathcal{V}_F$ is then identified as:

$$\mathcal{V}_F = \{v \mid \text{count}(\{p \in \mathcal{P}_F : v(p) = v\}) \geq \tau\}, \tag{6}$$

where τ is the minimum occupancy threshold.

Next, we find voxel occupancy for the handheld cloud $\mathcal{P}_H$:

$$\mathcal{V}_H = \{v(p) \mid p \in \mathcal{P}_H\}. \tag{7}$$

To efficiently identify and compare voxel occupancy, we define a simple but effective hashing function $h : \mathbb{Z}^3 \to \mathbb{Z}$. This function maps 3D voxel coordinates to a unique integer, enabling fast set operations:

$$h(v) = 73856093 \cdot v_x + 19349663 \cdot v_y + 83492791 \cdot v_z, \tag{8}$$

where $v = (v_x, v_y, v_z) \in \mathbb{Z}^3$. This type of spatial hashing is commonly used in graphics and geometry processing to accelerate lookup operations.

The set of hole voxels $\mathcal{V}_{\text{hole}}$ is computed by subtracting occupied Faro voxels from handheld voxel occupancy:

$$\mathcal{V}_{\text{hole}} = \{v \mid h(v) \in h(\mathcal{V}_H) \setminus h(\mathcal{V}_F)\}. \tag{9}$$

Handheld points corresponding to these hole voxels form the gap-filling point set $\mathcal{P}_{H,gap}$:

$$\mathcal{P}_{H,gap} = \{p \in \mathcal{P}_H \mid h(v(p)) \in h(\mathcal{V}_{\text{hole}})\}. \tag{10}$$

To maintain consistency in point density with the Faro cloud, we apply density-based upsampling. Let d_F be the median density of Faro-occupied voxels. For each voxel $v \in \mathcal{V}_{\text{hole}}$, the number of points to be added $a(v)$ is:

$$a(v) = \max(0, d_F - |\{p \in \mathcal{P}_{H,gap} : v(p) = v\}|). \tag{11}$$

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W15-2025-127-2025 | © Author(s) 2025. CC BY 4.0 License.

New points are interpolated by randomly selecting point pairs (p_1, p_2) within the same voxel and creating midpoint interpolations:

$$p_{\text{new}} = \frac{p_1 + p_2}{2} + \delta, \quad \delta \sim U\left(-\frac{r}{2}, \frac{r}{2}\right)^3, \tag{12}$$

3.3 Color Harmonization

After geometric alignment and hole filling, we harmonize the color information of the fused point cloud. Since the Faro TLS point cloud typically provides more stable and reliable color measurements, we use the RGB values from the TLS data as the global color reference. To improve computational efficiency, only the handheld points used for hole filling are color-calibrated, while the original TLS points retain their measured colors.

For each handheld point that fills a gap in the TLS data, we first establish color correspondences (c_H, c_F) within the overlapping region. A global affine color mapping is estimated via least squares regression:

$$c'_H = A\, c_H + b \tag{13}$$

where A is a 3×3 color transformation matrix and b is an offset vector. The parameters (A, b) are computed using only the paired colors from the overlap and are applied solely to the RS points in the hole-filled regions, rapidly aligning their color appearance to the global TLS standard.

To further account for local color variation caused by lighting or material differences, we perform a local color adjustment for each filled handheld point. Specifically, for each handheld point, we select its k nearest color correspondences in the overlap and fit a weighted local affine model:

$$c'_H(p) = A_p\, c_H(p) + b_p \tag{14}$$

where (A_p, b_p) are locally adapted parameters. This ensures that each filled handheld point is corrected according to its specific local context, improving visual coherence.

Finally, to avoid visible seams at the boundary between TLS and handheld regions, we perform color blending within a narrow transition band using distance-weighted interpolation. By restricting the color correction to the newly added handheld points and transition areas, our approach achieves both high computational efficiency and seamless visual integration in the final indoor model.

4. Results and Discussion

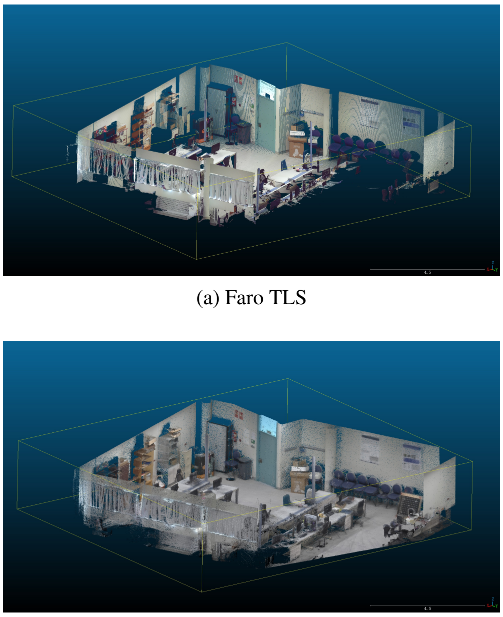
4.1 Experimental Setup

We evaluated our method in a classroom characterized by a complex and cluttered environment. Table 2 summarizes the key characteristics of the point clouds captured by the TLS (Faro), the handheld device (RS10), and their fusion, including acquisition durations and sensor measurement rates.

Despite the Faro TLS possessing a significantly higher measurement rate (976,000 points/s) compared to the RS10 handheld scanner (320,000/s), the total number of points captured by the handheld device is much greater. This is primarily because the RS10 scan was performed over a longer period and along a continuous, flexible trajectory, allowing for dense sampling throughout the scene and greater coverage, especially in occluded areas. In contrast, the Faro TLS acquired data from a fixed tripod position with a shorter scan duration, which, while offering high precision and measurement reliability, resulted in a lower overall point count and density due to limited spatial coverage and potential occlusions.

The similar room sizes and ceiling areas measured from each point cloud highlight the repeatability and reliability of both acquisition methods. Furthermore, both datasets were acquired within a one-hour window with the environment kept stable (no personnel entering or exiting), ensuring a consistent and controlled setting for comparison. By integrating the complementary advantages of each approach, the fused point cloud achieves both high geometric fidelity and comprehensive spatial completeness.

As can be seen from Figure 1a, the original Faro TLS point cloud exhibits high color fidelity but contains evident holes and missing surfaces in occluded regions. In contrast, the handheld RS10 point cloud (Figure 1b) provides denser coverage and effectively captures regions that are not visible from static TLS positions, but its color consistency and geometric precision are relatively lower.


(a) Faro TLS

(b) RS10 handheld

Figure 1. Original RGB point clouds captured by (a) the Faro TLS and (b) the RS10 handheld scanner.

4.2 Registration Results

The registration pipeline proceeds through four sequential stages. First, the ceiling is extracted using RANSAC plane fitting with a 0.02m distance threshold, requiring 11.35 s to robustly isolate the dominant horizontal surface. The resulting ceiling slice, as illustrated in Figure 2, serves as a stable reference for subsequent processing. Next, the scene is segmented vertically and each layer is projected onto the ceiling plane, enabling the detection of major wall axes via RANSAC line fitting; this wall detection step takes 23.71s. The extracted orthogonal wall-line segments, shown in Figure 3, provide critical

Source	Room Size	Ceiling Area	XY Density	Total Points	Acquisition Time	Measurement Rate
	(L×W×H), m	(m^2)	(points/m^2)		(s)	(points/s)
Faro TLS	12.02×9.84×3.71	118.34	52,794.33	6,247,838	309	976,000
RS10	12.07×9.88×3.48	119.20	178,927.67	21,327,571	473	320,000
Fused	12.07×10.01×3.71	120.88	147,337.78	17,809,660	–	–

Table 2. Summary of point cloud characteristics and acquisition details for the classroom scene

geometric cues for initial alignment. Coarse alignment is then performed, achieving an initial pose error below 2° with negligible runtime compared to other stages. Finally, fine registration is accomplished by point-to-plane ICP refinement, which converges in just 3.44s.

Overall, the entire coarse-to-fine pipeline completes in approximately 38s. By leveraging ceiling and wall primitives for initial alignment, this workflow not only accelerates convergence but also avoids the inefficiency and local minima pitfalls commonly encountered with direct global ICP, ensuring robust and rapid registration well-suited for indoor digital-twin applications.

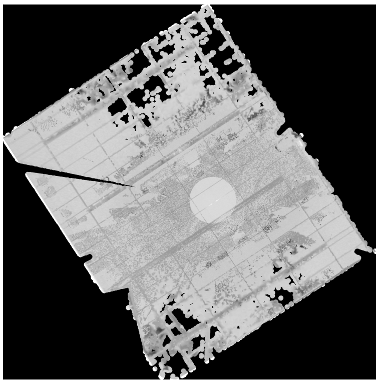


Figure 2. Ceiling slice extracted by RANSAC plane fitting.

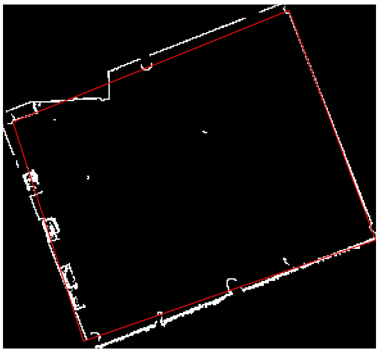


Figure 3. Orthogonal wall-line segments detected from the layered slices.

4.3 Quantitative Evaluation

To evaluate the performance of our proposed fusion and color harmonization framework for creating high-fidelity indoor digital twins, we employ three quantitative metrics: surface recovery rate, geometric accuracy via point-to-plane Root Mean Square Error (RMSE), and color consistency using the mean color difference. These metrics assess the method's ability to reconstruct missing surfaces, improve geometric alignment, and achieve visual coherence, respectively.

The surface recovery rate quantifies the extent to which our method reconstructs missing surfaces in the TLS point cloud using data from the HMLS. To compute this, we voxelize the point clouds with a voxel size r and identify the unique voxels occupied by each point cloud. Let $\mathcal{V}_F$, $\mathcal{V}_H$, and $\mathcal{V}_M$ denote the sets of unique voxels in the TLS point cloud ($\mathcal{P}_F$), the handheld point cloud ($\mathcal{P}_H$), and the fused point cloud ($\mathcal{P}_M$), respectively. The hole voxels, representing missing surfaces, are defined as:

$$\mathcal{V}_{\text{hole}} = \mathcal{V}_H \setminus \mathcal{V}_F \tag{15}$$

The new voxels, indicating areas filled by the fusion process, are:

$$\mathcal{V}_{\text{new}} = \mathcal{V}_M \setminus \mathcal{V}_F \tag{16}$$

The surface recovery rate is then calculated as:

$$\text{Recovery Rate} = \left(\frac{|\mathcal{V}_{\text{new}}|}{|\mathcal{V}_{\text{hole}}|}\right) \times 100\% \tag{17}$$

where $|\cdot|$ denotes the number of voxels in the set. In our experiments, we achieved a surface recovery rate of 85.7 %, indicating that 85.7 % of the missing surfaces in the TLS point cloud were successfully reconstructed using handheld data. This high recovery rate demonstrates the method's effectiveness in addressing occlusion-induced gaps, resulting in a more complete digital twin suitable for applications such as facilities management.

Geometric accuracy is evaluated using the point-to-plane Root Mean Square Error (RMSE), which measures how closely the points in the fused point cloud align with the high-precision TLS point cloud, used as the reference due to its superior accuracy (2 mm as per Table 1). For each point p_i in the fused point cloud $\mathcal{P}_M$, we find its nearest neighbor q_i in the TLS point cloud $\mathcal{P}_F$ using a k-d tree search. The point-to-plane distance is computed as:

$$d(p_i) = |(p_i - q_i) \cdot n_i| \tag{18}$$

where n_i is the normal vector at q_i, estimated using principal component analysis on the local neighborhood of q_i. The RMSE is then:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d(p_i)^2} \tag{19}$$

where N is the number of points in the point cloud being evaluated. We compute the RMSE for both the original handheld point cloud ($\mathcal{P}_H$) and the fused point cloud ($\mathcal{P}_M$). The RMSE for the handheld point cloud was 0.168 m, while the fused point cloud achieved an RMSE of 0.143 m, representing a 14.8 % improvement.

This reduction in RMSE indicates that the fusion process, which includes Iterative Closest Point (ICP) alignment and error-driven point selection, significantly enhances the geometric alignment of the handheld data with the TLS reference, improving the structural accuracy of the indoor model.

Color consistency is assessed by comparing the colors of the handheld points to their nearest neighbors in the TLS point cloud, using the CIELAB color space, which is designed to approximate human perception of color differences. For each point in the handheld point cloud $\mathcal{P}_H$, we identify its nearest neighbor in $\mathcal{P}_F$ and compute the ΔE color difference:

$$\Delta E = \sqrt{(L_H - L_F)^2 + (a_H - a_F)^2 + (b_H - b_F)^2} \quad (20)$$

where (L_H, a_H, b_H) and (L_F, a_F, b_F) are the CIELAB color coordinates of the handheld and TLS points, respectively. The mean ΔE is the average over all points:

$$\text{Mean } \Delta E = \frac{1}{N} \sum_{i=1}^{N} \Delta E_i \quad (21)$$

We evaluate the mean ΔE for both the uncalibrated and calibrated handheld colors. Before harmonization, the mean ΔE was 15.71, indicating noticeable color discrepancies due to sensor-specific biases and lighting variations. After applying our color harmonization method, which includes global and local regression to adjust handheld colors, the mean ΔE decreased to 12.22, representing a 22.2 % improvement.

This improvement reflects a more visually coherent model, with reduced color seams at the TLS-handheld boundaries, making it suitable for immersive applications such as AR/VR navigation.

In conclusion, the high surface recovery rate of 85.7 % ensures that most occluded areas in the TLS scans are effectively filled, resulting in a more complete digital twin. The 14.8 % reduction in point-to-plane RMSE confirms improved geometric alignment, crucial for applications requiring precise measurements, such as facility management. The 22.2 % improvement in mean ΔE enhances the visual realism of the model, making it ideal for applications where aesthetic quality is paramount. These results collectively demonstrate that our adaptive fusion pipeline significantly enhances the structural completeness, geometric accuracy, and visual quality of 3D indoor models, contributing to more reliable digital representations for smart building management.

4.4 Qualitative Discussion

As shown in Figure 4, the fused point cloud achieves color characteristics closely matching the high-fidelity Faro scan, while simultaneously filling numerous voids and gaps caused by occlusion in the original TLS data. This demonstrates that our fusion approach is highly effective, yielding a complete and visually coherent indoor point cloud model.

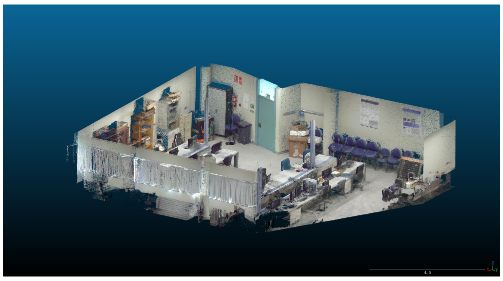


Figure 4. Fused RGB point cloud integrating Faro TLS and handheld RS10 data.

4.5 Compare to FARO Scene

Faro Scene is a commercial software suite that performs automated point-cloud registration and color-consistent fusion (FARO Technologies Inc., 2025), making it a suitable benchmark against which we compare our method.

Metric	Ours (Proposed)	Scene
Surface Recovery (%)	85.7 %	120.8 %
Geometric Accuracy (RMSE, m)	0.143 m (↓14.8 %)	0.155 m (↓7.7 %)
color Consistency (ΔE)	12.22 (↓22.2 %)	13.43 (↓14.5 %)

Table 3. Comparison of surface recovery, geometric accuracy, and color consistency for Ours and Faro Scene software.

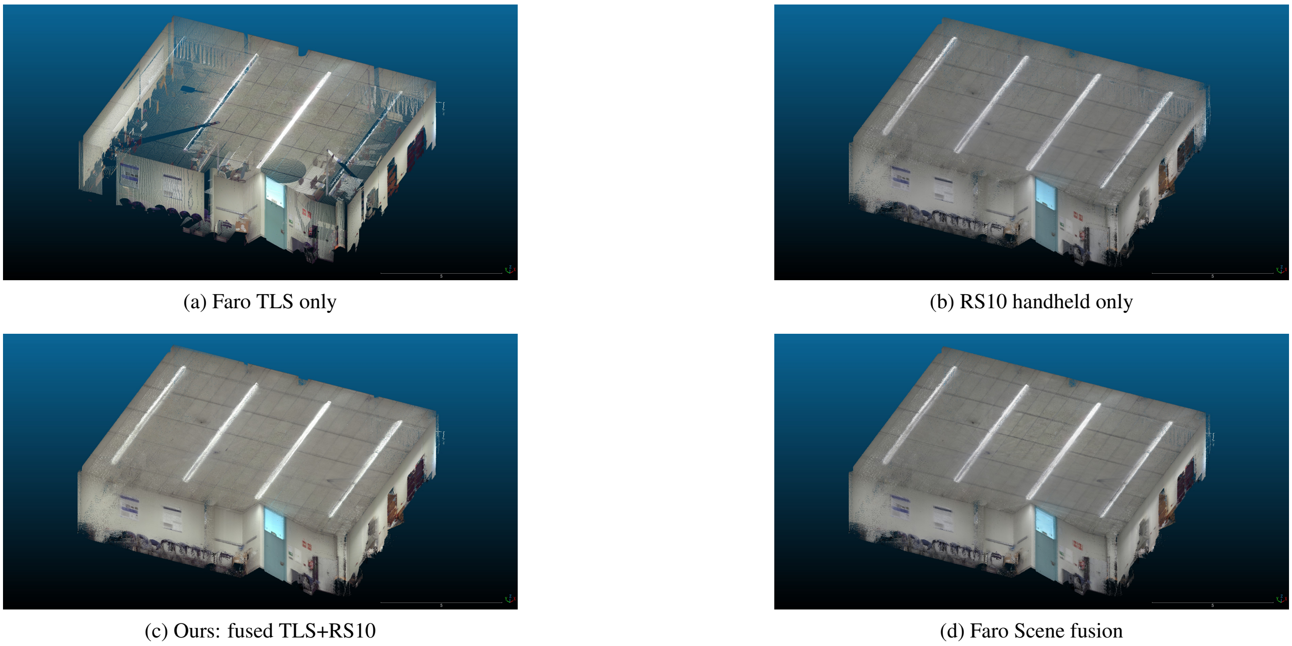
As shown in Table 3, our method achieves better overall performance across all evaluation metrics compared to the Scene software. Specifically, it reconstructs 85.7 % of the missing surfaces without overfilling, whereas Scene exceeds 100 % recovery, suggesting potential over-completion or noise. In terms of geometric accuracy, Ours reduces the global point-to-plane RMSE by 14.8 %, achieving a final error of 0.143m, lower than Scene's 0.155m. For color consistency, Ours also outperforms Scene, with a 22.2 % improvement in mean ΔE.

As illustrated in Figure 5(a), the Faro TLS delivers the sharpest geometry but leaves large wedge-shaped gaps around the door and ceiling edges. Figure 5(b) shows that the RS10 handheld scan fills these gaps thanks to its dense sampling, yet the colors are darker because the camera operated under changing illumination. Our fusion result in Figure 5(c) first rectifies geometry by selectively adding RS10 points only where the TLS is occluded, then applies a global–local color correction anchored to the reliable TLS RGB values; therefore appear neutral and crisp, with no "dragging" artefacts.

In contrast, Figure 5(d) (Faro Scene) blends colors purely by point-count weighting. Because the handheld scan contains many more points than the TLS, its noisier RGB values dominate: the overall hue shifts toward RS10's cast, and high-contrast edges (e.g. lamp strips, wall–ceiling junctions) exhibit visible streaks. This visual evidence supports the quantitative findings. Scene's surface-recovery exceeds 100 % (over-filling) while its RMSE and mean ΔE are higher—illustrating that more points do not necessarily translate into more reliable information.

5. Conclusion

We introduced an adaptive voxel-based fusion pipeline that combines a Faro Focus TLS with a CHCNAV RS10 handheld scanner to generate complete, color-consistent indoor digital twins. The workflow integrates a ceiling- and wall-guided



(a) Faro TLS only

(b) RS10 handheld only

(c) Ours: fused TLS+RS10

(d) Faro Scene fusion

Figure 5. Visual comparison of the classroom ceiling and walls obtained from four data sources.

coarse alignment followed by point-to-plane ICP, geometry-based hole detection with selective handheld infill and density equalization, and a two-stage color harmonization that applies global RGB mapping, local regression, and seam blending.

On a cluttered classroom scene, the proposed method reconstructs 85.7 % of TLS occlusion holes without over-filling, reduces the global point-to-plane RMSE from 0.168 m to 0.143 m, and lowers the mean color difference ΔE from 15.71 to 12.22. Compared with the commercial Faro Scene software, our pipeline achieves lower RMSE and better color consistency. These results confirm that the fused models deliver the geometric fidelity and visual realism needed for smart-building applications such as asset tracking, maintenance scheduling, and AR/VR navigation.

Future work will extend the method to larger indoor datasets and leverage deep learning for feature extraction, semantic segmentation, and color correction, aiming for greater automation, robustness, and near-real-time performance.

Acknowledgment

We sincerely thank CHCNAV for providing the instruments and support for this research. This work was supported by the Research Development Fund of Xi'an Jiaotong–Liverpool University (RDF-24-02-025). This work also has been supported by the Spanish Ministry of Science, Innovation and Universities by projectCPP2021-008374 funded by MICIU/AEI/10.13039/501100011033 and by European Union NextGenerationEU/PRTR; and by grant ED431F 2024/02 funded by Xunta de Galicia, Spain-GAIN.

References

Balado, J., Garozzo, R., Winiwarter, L., Tilon, S., 2025. A Systematic Literature Review of Low-Cost 3D Mapping Solutions. *Information Fusion*, 114, 102656.

Besl, P. J., McKay, N. D., 1992. Method for registration of 3-d shapes. *Sensor fusion IV: control paradigms and data structures*, 1611, Spie, 586–606.

CHCNAV, 2024. RS10: Handheld SLAM 3D Laser Scanner + GNSS RTK. `https://www.chcnav.com/product-detail/rs10`. Accessed: 2025-06-30.

Das, A., Waslander, S. L., 2014. Scan registration using segmented region growing NDT. *The International Journal of Robotics Research*, 33(13), 1645–1663.

Elhashash, M., Albanwan, H., Qin, R., 2022. A review of mobile mapping systems: From sensors to applications. *Sensors*, 22(11), 4262.

FARO, 2025. User Manuals and Quick Start Guides for the Focus Laser Scanner. `https://knowledge.faro.com/Hardware/Focus/Focus/User_Manuals_and_Quick_Start_Guides_for_the_Focus_Laser_Scanner`. Accessed: 2025-06-30.

FARO Technologies Inc., 2025. SCENE User Manual. FARO Technologies. Accessed: 2025-06-30.

Giacomini, E., Brizi, L., Di Giammarino, L., Salem, O., Perugini, P., Grisetti, G., 2024. Ca2Lib: Simple and Accurate LiDAR-RGB Calibration Using Small Common Markers. *Sensors*, 24(3).

González-Collazo, S. M., Balado, J. et al., 2024. Santiago Urban Dataset (SUD): Combination of Handheld and Mobile Laser-Scanning Point Clouds. *Expert Systems with Applications*, 238, 121842.

Li, R., Liu, X., Li, H., Liu, Z., Lin, J., Cai, Y., Zhang, F., 2024. LVBA: LiDAR-Visual Bundle Adjustment for RGB Point Cloud Mapping. *arXiv preprint arXiv:2409.10868*.

Luhmann, T., Chizhova, M., Gorkovchuk, D., 2020. Fusion of UAV and terrestrial photogrammetry with laser scanning for 3D reconstruction of historic churches in georgia. *Drones*, 4(3), 53.

Qiu, Z., Martínez-Sánchez, J., Arias, P., 2025. Fusion of thermal images and point clouds for enhanced wall temperature uniformity analysis in building environments. *Energy and Buildings*, 115781.

Qiu, Z., Martínez-Sánchez, J., Arias-Sánchez, P., Rashdi, R., 2023. External multi-modal imaging sensor calibration for sensor fusion: A review. *Information Fusion*, 97, 101806.

Rashdi, R., Garrido, I., Balado, J., Del Río-Barral, P., Rodríguez-Somoza, J. L., Martínez-Sánchez, J., 2024. Comparative Evaluation of LiDAR systems for transport infrastructure: case studies and performance analysis. *European Journal of Remote Sensing*, 57(1), 2316304.

Rashdi, R., Martínez-Sánchez, J., Arias, P., Qiu, Z., 2022. Scanning Technologies to Building Information Modelling: A Review. *Infrastructures*, 7(4). https://www.mdpi.com/2412-3811/7/4/49.

Roggero, M., Diara, F., 2024. Multi-Sensor 3D survey: Aerial and terrestrial data fusion and 3D modeling applied to a complex historic architecture at risk. *Drones*, 8(4), 162.

Shaharuddin, S., Abdul Maulud, K. N., Syed Abdul Rahman, S. A. F., Che Ani, A. I., 2022. DIGITAL TWIN FOR INDOOR DISASTER IN SMART CITY: A SYSTEMATIC REVIEW. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-4/W3-2021, 315–322. https://isprs-archives.copernicus.org/articles/XLVI-4-W3-2021/315/2022/.

Shin, H., Kwak, Y., 2024. Enhancing Digital Twin Efficiency in Indoor Environments: Virtual Sensor-Driven Optimization of Physical Sensor Combinations. *Automation in Construction*, 161, 105326.

Sun, W., Qu, X., Wang, J., Jin, F., Li, Z., 2024. Multi-Platform Point Cloud Registration Method Based on the Coarse-To-Fine Strategy for an Underground Mine. *Applied Sciences*, 14(22), 10620.

Xing, X., Qian, Y., Feng, S., Dong, Y., Matas, J., 2022. Point cloud color constancy. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 19750–19759.

Yoon, S., Koo, J., 2023. In-situ Model Fusion for Building Digital Twinning. *Building and Environment*, 243, 110652.

Zou, Q., Sun, Q., Chen, L., Nie, B., Li, Q., 2021. A comparative analysis of LiDAR SLAM-based indoor navigation for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 6907–6921.