DFBI 2025 aims to encourage the international exchange of innovative ideas between researchers from academia and industry. In addition to knowledge dissemination, the conference offers a valuable platform for professional networking, particularly benefiting university professors, graduate students, and postdoctoral researchers.

Research Article

Construction Quality Inspection by Local Point Cloud to BIM Registration and Component Association Based on Vision Foundation Model

Boyu WANG[1], Mingkai LI[1], Fangzhou LIN[1], Zhenyu LIANG[1], Jack CHENG[1]

Affiliation 1: Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

Correspondence: bwangbb@connect.ust.hk

Abstract

Construction quality inspection is essential for verifying conformity between physical installations and design specifications. Traditional inspection methods rely heavily on manual comparison between construction drawings and onsite conditions, making the process labor-intensive, subjective, and lacking in digital traceability. Although laser scanning technologies have been widely adopted for capturing as-built data, the automated registration of local LiDAR scans to BIM models and the association of onsite components with their designed counterparts remain underdeveloped. To address these challenges, this study proposes a novel framework for local point cloud to BIM registration and component association, leveraging vision foundation models to enable component-level inspection directly on construction sites. The proposed framework comprises four main stages. First, raw LiDAR scans are preprocessed and converted into binary visibility maps centered on the LiDAR sensor. Second, synthetic visibility maps are generated by virtually scanning the BIM model from corresponding viewpoints. Third, both real and synthetic visibility maps are encoded using a vision foundation model to extract high-dimensional features. These features are compared to generate similarity maps for coarse localization, which is further refined using the Iterative Closest Point (ICP) algorithm to achieve accurate registration. Finally, onsite components are identified using the vision foundation model and matched to their corresponding BIM components through bipartite graph matching, solved via the Hungarian algorithm. Experiments conducted on real-world construction sites demonstrate the robustness and accuracy of the proposed method, validating its effectiveness for practical applications in construction quality inspection and digital twin development.

Keywords: Construction Digital Twin; Point Clouds; Cross-domain Registration; Vision Foundation Model; As-built As-designed Association

Highlights

- LiDAR-centric visibility map features are designed and encoded using vision foundation models to capture spatial context effectively.
- Local point cloud to BIM registration is achieved through feature comparison between real LiDAR scans and virtually generated scans from the BIM model.
- Component association between onsite installations and BIM elements is formulated as a bipartite graph matching problem and solved using the Hungarian algorithm.

Cite this article:
Lastname, F. M., & Lastname, F. M. (Year). Title of the article. Title of the Journal, Volume 2025, Page range: 193 -202

1 Section 1- Introduction

Systematic monitoring of construction quality and progress is critical in the Architecture, Engineering, and Construction (AEC) industry, as it enables the timely detection of deviations between design intent and actual construction, thereby reducing delays and controlling costs. Accurate onsite alignment between construction status and design specifications facilitates essential tasks such as as-built verification and progress assessment. However, conventional inspection practices remain predominantly manual, relying on two-dimensional drawings and printed documentation. These methods are not only time-consuming and cognitively demanding but also prone to errors due to limited spatial perception and human memory constraints.

To overcome these limitations, construction digital twins have gained increasing attention. On the digital side, Building Information Modeling (BIM) provides structured representations of design intent and lifecycle information, often integrated with mobile-accessible platforms such as Autodesk BIM 360 Field. Simultaneously, LiDAR-based laser scanning technologies offer high-resolution spatial data, enabling precise documentation of as-built conditions and accurate digital archiving. A key requirement for realizing the potential of digital twins lies in the accurate registration of LiDAR-derived point clouds with BIM models, which is foundational for advanced applications such as augmented reality-based inspections and automated progress monitoring. While various localization technologies—such as QR codes, GPS, RFID, and Wi-Fi fingerprinting—have been proposed, they often suffer from practical deployment challenges, limited accuracy, and signal instability. In contrast, LiDAR-based approaches inherently provide rich spatial data suitable for localization, eliminating the need for additional infrastructure and simplifying implementation.

To address the challenges of accurate alignment and component association, this study proposes a robust point cloud-to-BIM registration and component association framework leveraging vision foundation models. The framework enables effective onsite, component-level inspection and comprises four key stages: (1) preprocessing the as-built point cloud into LiDAR-centric visibility maps; (2) generating synthetic visibility maps from BIM through virtual scanning; (3) extracting high-dimensional visual features using a vision foundation model (e.g., DINOv2) to compute similarity scores between real and synthetic views; and (4) associating as-built elements with their corresponding BIM components via bipartite graph matching, solved using the Hungarian algorithm.

2 Section 2 – Literature Review

2.1 Point Cloud Registration

Point cloud registration is the process of aligning overlapping point cloud datasets by estimating spatial transformations, which is fundamental for applications such as SLAM, mapping, and aligning as-built data with digital models. Registration methods are typically categorized into optimization-based, feature-based, and learning-based approaches, with feature-based methods being the most widely adopted. These approaches generally involve four steps: local feature extraction, correspondence matching, outlier filtering, and refinement using the Iterative Closest Point (ICP) algorithm.

Local descriptors have evolved significantly—from early handcrafted methods such as Fast Point Feature Histograms (FPFH)(Rusu *et al.*, 2009), to more robust learning-based descriptors such as sparse convolutional networks FCGF (Choy *et al.*, 2019), and fully convolutional techniques D3Feat (Bai *et al.*, 2020). Effective outlier rejection is crucial to registration accuracy, with traditional methods like

RANSAC and more advanced techniques such as Fast Global Registration (FGR)(Zhou *et al.*, 2016) and Teaser++ (Yang *et al.*, 2020) significantly improving robustness.

In the Architecture, Engineering, and Construction (AEC) domain, registration often involves aligning point clouds acquired via terrestrial laser scanning, SLAM, or Structure-from-Motion (SfM) with Building Information Models (BIM). This alignment supports tasks such as construction monitoring, quality inspection, and model enrichment. Current BIM-to-point cloud registration approaches typically follow a coarse-to-fine strategy, beginning with manual correspondences or commercial tools (e.g., Autodesk ReCap, Leica Cyclone), followed by refinement using ICP. Recent research has focused on automating this process by leveraging the geometric regularity of built environments. Notable geometric approaches include Bosché's RANSAC-based plane matching (Bosché, 2012), Bueno's Four-Plane Congruent Set method (Bueno *et al.*, 2018).

Local-to-global registration addresses a more complex scenario in which locally captured, partial point clouds must be aligned with global models without access to a complete global scan. This typically involves coarse localization using global descriptors, followed by fine registration. Scan Context (Kim and Kim, 2018), which uses polar representations, is a widely used global descriptor in LiDAR-based localization. Other notable methods include 3D-BBS (Aoki *et al.*, 2024) and PointNetVLAD (Uy and Lee, 2018), which leverage optimized voxel maps and deep-learned descriptors, respectively. More recently, semantic-driven registration approaches have emerged to address cross-domain challenges. For instance, SPVLoc (Gard *et al.*, 2024) performs semantic panoramic viewport matching, while Zhao et al. utilize CNN-derived semantic features to bridge the visual-to-BIM domain gap, enhancing robustness in complex construction environments (Zhao and Cheah, 2023).

2.2 Quality Inspection Based on Scan-vs-BIM

Laser scanning has significantly advanced construction quality assurance by enabling the efficient capture of precise geometric data, reducing inspection time by over 60% compared to traditional visual methods. It supports infrastructure health monitoring by analyzing geometric indicators such as deformations, surface flatness, cracks, and seepage. Central to quality verification is the Scan-vs-BIM process, which compares as-built point clouds against as-designed BIM models to assess conformance with design specifications.

At the point-level, the Scan-vs-BIM process begins with rough registration, followed by ICP-based fine alignment to unify the coordinate systems of the scan and the BIM model. Once aligned, deviation maps are generated to quantify geometric discrepancies, which are then visualized using threshold-based binary classification to identify compliant and non-compliant areas. Notable applications include Bosché's steel element classification via BIM matching (Bosché, 2010) and Nahangi and Haas's automated defect detection in MEP pipe spools using neighborhood-based ICP (Nahangi and Haas, 2014).

At the object-level, Scan-vs-BIM involves establishing correspondences between as-built and as-designed components to enable more detailed dimensional analysis. For instance, Kim et al. introduced a dimensional quality control method for shipbuilding blocks by comparing measurement points with design-quality points (Kim *et al.*, 2024). Fang et al. assessed tunnel excavation profiles against Dynamo parametric models (Fang *et al.*, 2024). Mirzaei et al. employed neural networks for dimensional quality assessment of steel structures by segmenting structural elements and comparing lengths and spacings with design specifications (Mirzaei *et al.*, 2023). Truong-Hong and Lindenbergh proposed a region-growing method tailored for extracting concrete structural surfaces based on contextual geometry (Truong-Hong and Lindenbergh, 2022). Hu and Brilakis extended shape detection to match planar, curved, and linear structural elements using descriptor-based methods (Hu and Brilakis, 2024). Additional applications include rebar inspection and cylindrical MEP component verification using Hough Transform techniques.

Despite these advances, existing methods often suffer from over-sensitivity to noise or depend heavily on custom semantic extraction rules, which limit their applicability in dynamic and complex construction environments. More critically, robust identification and precise matching of components remain challenging, making it difficult to systematically detect and quantify differences between as-built and as-designed conditions.

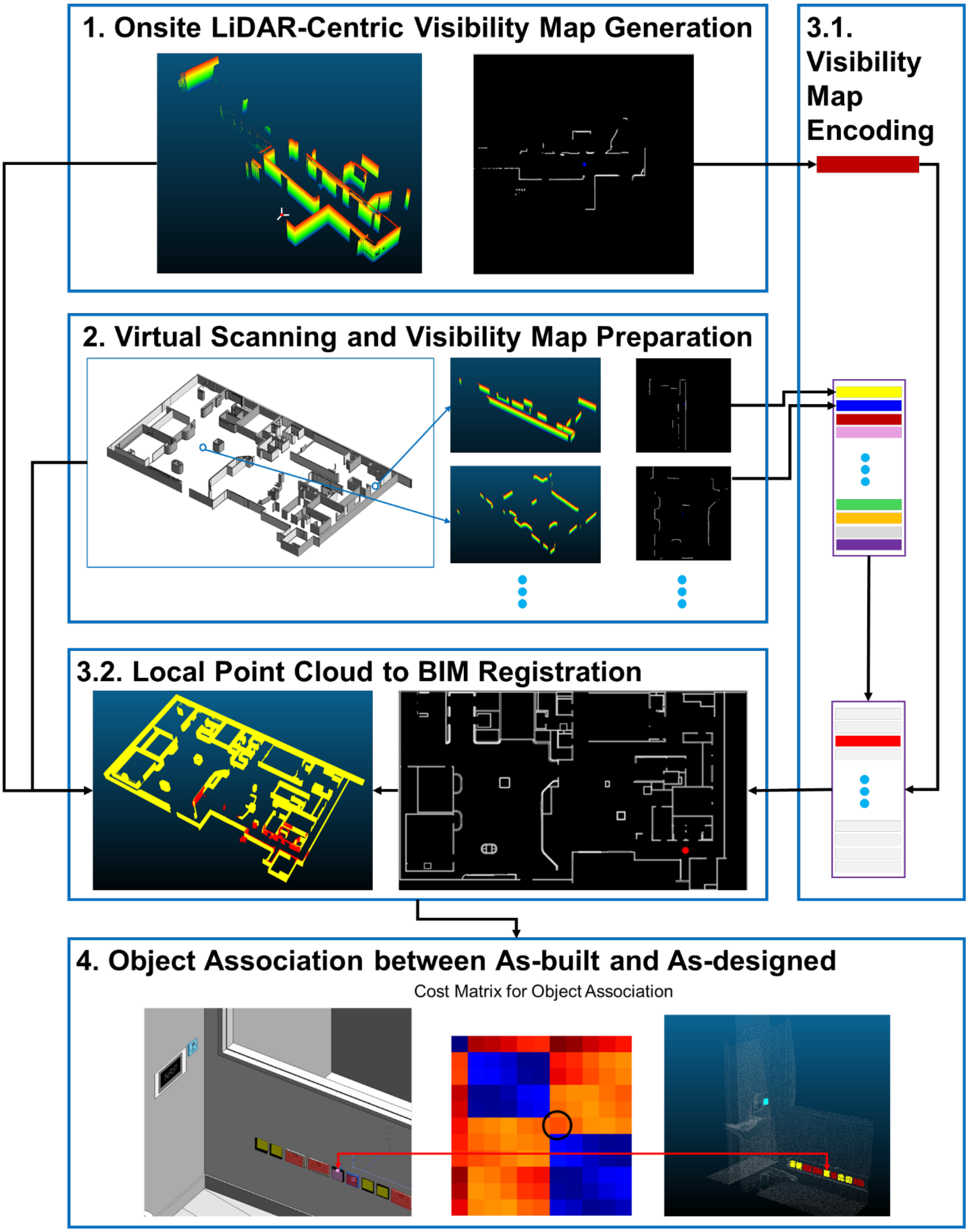
3 Methodology

his study presents a novel construction quality inspection framework based on local point cloud to BIM registration and component association, leveraging a vision foundation model for robust feature representation. As illustrated in Figure 1, the proposed framework comprises four core modules: (1) generation of LiDAR-centric visibility maps from onsite point clouds, (2) virtual scanning of BIM models to generate synthetic reference visibility maps, (3) extraction and comparison of learned visual features using a vision foundation model, and (4) component association through bipartite graph matching. The detailed methodology for each module is presented in Sections 3.1 to 3.4.

3.1 LiDAR-Centric Visibility Map Generation

Accurate registration between onsite point clouds and as-designed digital models depends on obtaining a reliable initial estimate of the sensor's location and orientation. Achieving this initial alignment requires the use of modality-invariant features that are robust to differences in data acquisition methods. LiDAR-centric visibility maps are particularly well suited for this purpose, as they maintain geometric consistency across various scanning modalities. To generate these visibility maps, the as-built point cloud undergoes a sequence of preprocessing steps. The process begins with the detection of the ground plane to establish a consistent reference level. A horizontal slice is then extracted at a height strategically chosen to lie below typical architectural elements such as doors and windows, while remaining above common sources of clutter. This slicing operation yields a simplified yet informative geometric representation of the local environment. The resulting slice is projected into a two-dimensional binary grid-based map. This is achieved by determining the horizontal spatial extent of the point cloud and assigning each three-dimensional point to a corresponding grid cell using an efficient nearest neighbour search algorithm. Cells containing points are marked as occupied, forming a binary occupancy map. To improve the continuity and connectivity of structural features, morphological dilation is applied to the occupancy map, resulting in a refined LiDAR-centric visibility map that supports robust initial registration. To ensure consistency in feature extraction across different visibility maps, a standardization procedure is introduced. Specifically, a fixed spatial region within a ten-meter radius cantered at the LiDAR sensor position is extracted from the visibility map, providing a normalized input for subsequent processing stages.

Figure 1. Overview of the proposed local point cloud to BIM registration and component association framework based on vision foundation model



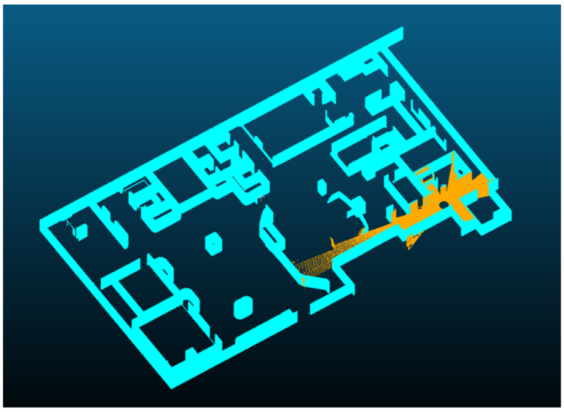
3.2 Virtual Scanning and Visibility Map Preparation

In typical building environments, where primary structural and architectural elements adhere to their digital representations. This observation provides a reliable foundation for feature-based alignment between as-built and as-designed data. To facilitate accurate estimation of initial sensor positions and orientations, LiDAR-centric visibility maps are generated from BIM through a virtual scanning process that simulates real-world sensor viewpoints and occlusion patterns. The process begins by isolating primary static structural components, such as walls and columns, from the BIM model to ensure consistency with elements typically captured during onsite scanning. A binary two-dimensional floor plan is then generated from the model, and a uniform grid of candidate viewpoints is sampled across this plan. At each candidate viewpoint, virtual LiDAR scans are simulated using ray-casting techniques that replicate key sensor parameters, including field of view, angular resolution, and maximum sensing range. This generates noise-free synthetic point clouds that approximate real sensor outputs. These virtual scans are then processed using the same slicing and projection procedures applied to real-world LiDAR data. Specifically, each scan is sliced at a fixed height and projected into a two-dimensional binary occupancy map, resulting in standardized visibility maps. These synthetic maps are directly comparable to those derived from onsite scans, enabling robust cross-domain registration.

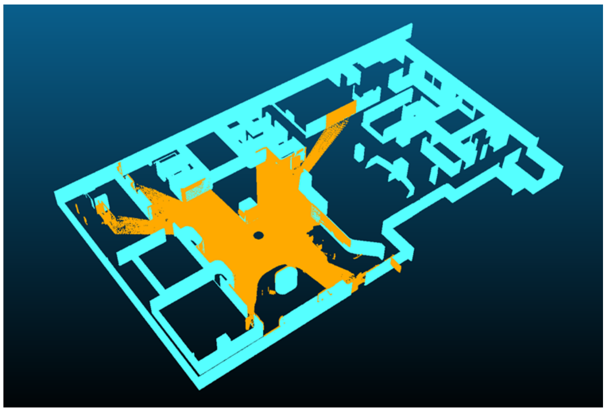
3.3 Feature Encoding and Comparison for Local Point Cloud to BIM Registration

To enable robust and semantically meaningful comparison between real and synthetic visibility maps, this study employs the vision foundation model DINOv2 for feature extraction (Oquab *et al.*, 2023). DINOv2 is a state-of-the-art self-supervised model that has demonstrated exceptional performance in visual representation learning across large-scale datasets such as ImageNet and COCO. Its learned features exhibit strong semantic clustering properties, effectively grouping visually similar patterns while maintaining clear distinctions between dissimilar ones. This makes it particularly suitable for encoding visibility maps, which require both geometric and semantic discriminability.

Figure 2. Local point cloud to BIM registration visualization



Registration Results for Station A



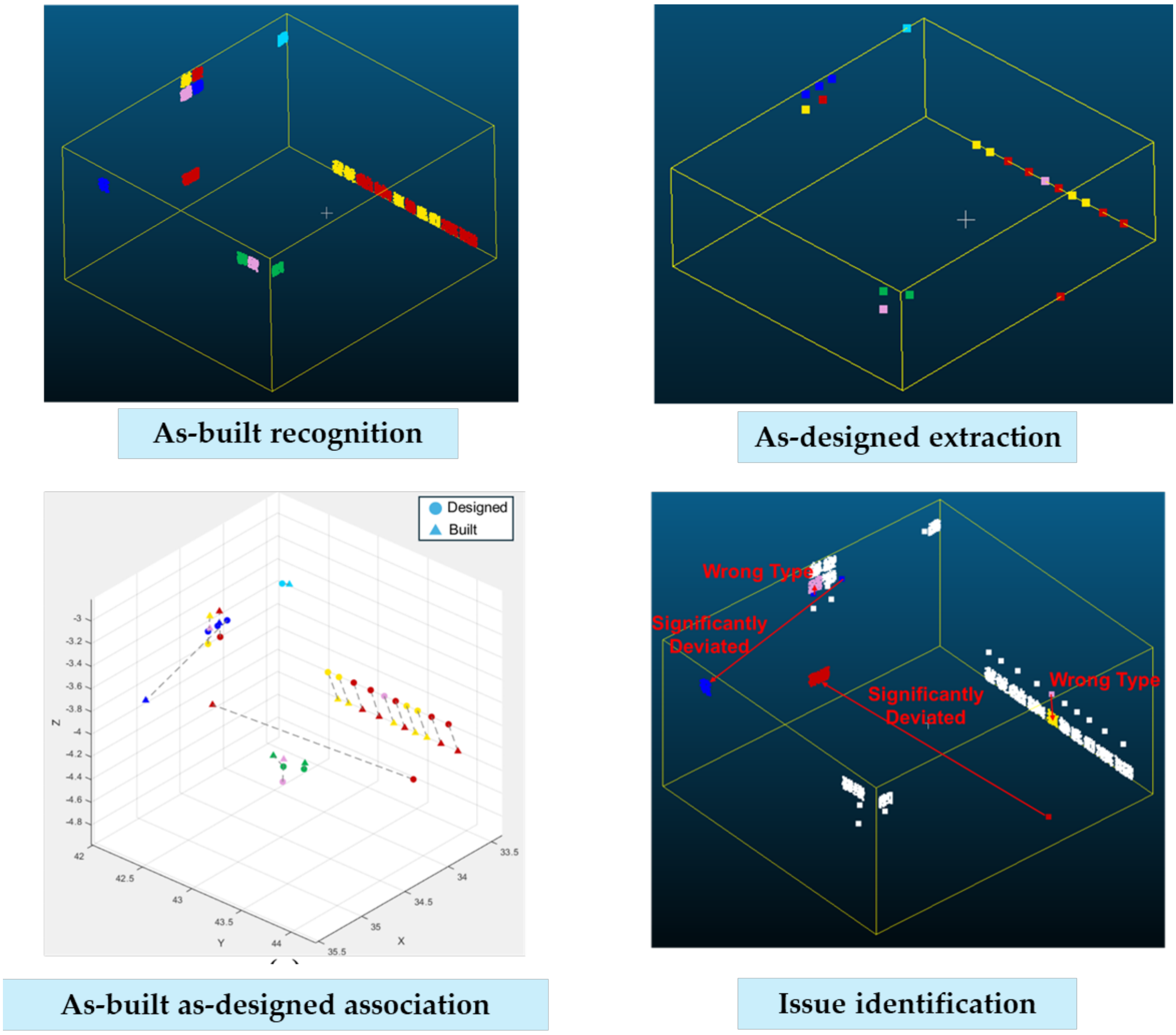
Registration Results for Station B

Following feature extraction, the similarity between visual feature vectors from onsite and synthetic visibility maps is computed using cosine similarity, a widely used metric for high-dimensional vector comparison. The candidate viewpoint yielding the highest similarity score is selected as the most probable estimate of the sensor's location. The corresponding position and orientation are then adopted as the initial alignment parameters. To further refine the registration, the Iterative Closest Point (ICP) algorithm is applied, producing an accurate transformation that aligns the onsite point cloud with the BIM-derived model as shown in Figure 2.

3.4 Component Association through Bipartite Graph Matching

Following the registration of the as-built point cloud to the as-designed BIM, it is essential to establish accurate component-level associations to enable effective quality inspection. Object recognition can be performed using advanced techniques such as Omni-Scan2BIM (Wang *et al.*, 2024) or GroundDINO (Liu *et al.*, 2024), which provide semantic segmentation and classification of structural and architectural elements from point cloud data. Once recognized, the inspection process focuses on identifying four primary types of discrepancies between the as-built and as-designed components: (1) spatial deviations (i.e., components installed in the wrong location), (2) incorrect family types (i.e., mismatches in component specifications), (3) missing components (i.e., modeled elements not present in the physical environment), and (4) redundant components (i.e., installed elements not represented in the design model). To support this analysis, essential object-level metadata—such as component identifiers, three-dimensional coordinates, categories, family types are extracted from the as-designed BIM.

Figure 3. Component association between the as-built installations and the as-designed counterparts



For robust and accurate component matching, this study adopts a bipartite graph matching framework solved using the Hungarian algorithm, rather than relying on traditional sequential matching techniques. Sequential approaches are often sensitive to installation order and may fail under conditions of large spatial deviation or incomplete data. In the proposed framework, a weighted bipartite graph is constructed, where one set of nodes represents components from the BIM model and the other represents components identified in the as-built point cloud. The edges between nodes are weighted based on a composite metric incorporating spatial proximity, family type similarity, and geometric alignment, including orientation consistency for linear components. A cost matrix is derived from these edge weights and used by the Hungarian algorithm to determine the optimal one-to-one assignments between design and as-built components. The outcomes of this association process are interpreted as follows: matched components with correct types but positional discrepancies are classified as spatially deviating. Matched components with type mismatches are labelled as wrong type. Unmatched BIM components are considered missing, while unmatched as-built components are identified as redundant. The process has been illustrated in Figure 3.

4 Validation

To evaluate the effectiveness of the proposed local point cloud to digital model registration approach, point cloud data were acquired from 15 scanning stations across an active hospital construction site in Hong Kong. The data were collected using the LEICA RTC360 laser scanner, which offers high-precision measurements with an accuracy of approximately 2 mm, a resolution of 5 mm at a 10-meter range, a horizontal field of view of 360 degrees, and a vertical field of view of 300 degrees. The system is also equipped with high dynamic range (HDR) cameras capturing imagery at a resolution of 2048 × 2048 pixels, enabling comprehensive visual documentation of the site.

Registration performance was quantitatively assessed using registration recall, defined by a root-mean-square error (RMSE)-based criterion. A registration was considered successful if the RMSE between the estimated and ground-truth scan alignment was below a 0.2-meter threshold. For comparative evaluation, a traditional local feature-based registration pipeline composed of Fast Point Feature Histograms (FPFH) and RANSAC was employed as the baseline method.

To assess the accuracy of component-level association, a representative room containing 20 distinct components was selected. The association accuracy was computed as the success rate in correctly matching as-built components to their as-designed counterparts, based on human-defined ground truth. The proposed method was compared with a greedy matching algorithm, which sequentially associates components based on proximity or similarity in a first-come, first-served manner.

Table I. Registration accuracy analysis of the proposed registration method and local feature based methods.

Methods	Hospital
ICP Directly	0/15
FPFH + RANSAC	1/15
Proposed Method	14/15

The results in Table I demonstrate that conventional local feature-based registration methods frequently struggle or fail outright in environments with repetitive architectural patterns, which introduce significant ambiguity into the matching process. In contrast, the proposed approach, which leverages deep visual features, consistently identifies the correct sensor location even under highly challenging conditions. These findings highlight the robustness and superior generalization capability of deep visual representations in complex real-world construction scenarios.

Also, as shown in Table I I, the proposed method successfully matched all 20 components in the test area with complete accuracy. In contrast, sequential matching methods based on sequential method exhibited inconsistent results due to their sensitivity to component processing order. This limitation was evident across three randomized trials, where the outcome varied significantly depending on the order of processing. The proposed bipartite graph matching approach consistently outperforms sequential methods, offering stable and accurate results even in the presence of substantial discrepancies between the BIM model and the built environment. These findings underscore the effectiveness of the proposed association strategy in supporting reliable and scalable construction quality inspection.

Table I I. Component association accuracy analysis of the proposed method and the sequential method.

Methods	First Run	Second Run	Third Run	Fourth Run	Average
Sequential Method	15/20	12/20	14/20	13/20	13.5/20
Proposed Method	20/20	20/20	20/20	20/20	20/20

5 Conclusions

This study proposes a visibility-based approach for registering as-built point clouds to digital models, differing from traditional geometry-based methods that require manual feature design. Instead, it utilizes a pretrained vision foundation model for feature encoding, offering a more intuitive and holistic understanding aligned with human perception. The method addresses two key challenges in indoor registration: differing data modalities between point clouds and digital models, and the high self-similarity of features in indoor environments. It also demonstrates robustness to discrepancies between the site and the model. After aligning the data, the correspondence between as-designed and as-built elements is framed as a bipartite graph matching problem, solved using the Hungarian algorithm. This enables automated quality checks by categorizing issues as wrong type, deviating, redundant, or missing. The method remains effective even for components with substantial installation deviations.

References

Aoki, K., Koide, K., Oishi, S., Yokozuka, M., Banno, A., & Meguro, J. 3d-bbs: Global localization for 3d point cloud scan matching using branch-and-bound algorithm. 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024. IEEE, 1796-1802.

Bai, X., Luo, Z., Zhou, L., Fu, H., Quan, L., & Tai, C.-L. D3feat: Joint learning of dense detection and description of 3d local features. Proceedings of the Ieee/Cvf Conference on Computer Vision and Pattern Recognition, 2020. 6359-6367. doi:http://doi.org/10.1109/cvpr42600.2020.00639

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*(1), 107-118. doi:https://doi.org/10.1016/j.aei.2009.08.006

Bosché, F. (2012). Plane-based registration of construction laser scans with 3D/4D building models. *Advanced Engineering Informatics, 26*(1), 90-102.

Bueno, M., Bosché, F., González-Jorge, H., Martínez-Sánchez, J., & Arias, P. (2018). 4-Plane congruent sets for automatic registration of as-is 3D point clouds with 3D BIM models. *Automation in Construction, 89*, 120-134.

Choy, C., Park, J., & Koltun, V. Fully convolutional geometric features. Proceedings of the Ieee/Cvf International Conference on Computer Vision, 2019. 8958-8966. doi:http://doi.org/10.1109/ICCV.2019.00905

Fang, W., Chen, W., Love, P. E., Luo, H., Zhu, H., & Liu, J. (2024). A status digital twin approach for physically monitoring over-and-under excavation in large tunnels. *Advanced Engineering Informatics, 62*, 102648. doi:https://doi.org/10.1016/j.aei.2024.102648

Gard, N., Hilsmann, A., & Eisert, P. SPVLoc: Semantic Panoramic Viewport Matching for 6D Camera Localization in Unseen Environments. European Conference on Computer Vision, 2024. Springer, 398-415.

Hu, Z., & Brilakis, I. (2024). Matching design-intent planar, curved, and linear structural instances in point clouds. *Automation in Construction, 158*, 105219. doi:https://doi.org/10.1016/j.autcon.2023.105219

Kim, G., & Kim, A. Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018. IEEE, 4802-4809.

Kim, S., Yoon, Y. U., & Kim, B. C. (2024). Method for estimating management points for dimensional quality management in shipbuilding and offshore plant construction using terrestrial laser scanning and a computer-aided design model. *Advanced Engineering Informatics, 62*, 102591. doi:https://doi.org/10.1016/j.aei.2024.102591

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., & Su, H. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. European Conference on Computer Vision, 2024. Springer, 38-55.

Mirzaei, K., Arashpour, M., Asadi, E., Feng, H., Mohandes, S. R., & Bazli, M. (2023). Automatic compliance inspection and monitoring of building structural members using multi-temporal point clouds. *Journal of Building Engineering, 72*, 106570. doi:https://doi.org/10.1016/j.jobe.2023.106570

Nahangi, M., & Haas, C. T. (2014). Automated 3D compliance checking in pipe spool fabrication. *Advanced Engineering Informatics, 28*(4), 360-369. doi:https://doi.org/10.1016/j.aei.2014.04.001

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., & El-Nouby, A. (2023). DINOv2: Learning robust visual features without supervision. *Arxiv Preprint*. doi:https://doi.org/10.48550/arXiv.2304.07193

Rusu, R. B., Blodow, N., & Beetz, M. Fast point feature histograms (FPFH) for 3D registration. 2009 Ieee International Conference on Robotics and Automation, 2009. IEEE, 3212-3217. doi:http://doi.org/10.1109/ROBOT.2009.5152473

Truong-Hong, L., & Lindenbergh, R. (2022). Extracting structural components of concrete buildings from laser scanning point clouds from construction sites. *Advanced Engineering Informatics, 51*, 101490. doi:https://doi.org/10.1016/j.aei.2021.101490

Uy, M. A., & Lee, G. H. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. 4470-4479.

Wang, B., Chen, Z., Li, M., Wang, Q., Yin, C., & Cheng, J. C. (2024). Omni-Scan2BIM: A ready-to-use Scan2BIM approach based on vision foundation models for MEP scenes. *Automation in Construction, 162*, 105384. doi:https://doi.org/10.1016/j.autcon.2024.105384

Yang, H., Shi, J., & Carlone, L. (2020). Teaser: Fast and certifiable point cloud registration. *IEEE Transactions on Robotics, 37*(2), 314-333.

Zhao, X., & Cheah, C. C. J. A. i. C. (2023). BIM-based indoor mobile robot initialization for construction automation using object detection. *146*, 104647.

Zhou, Q.-Y., Park, J., & Koltun, V. Fast global registration. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, 2016. Springer, 766-782.