Kalpa Publications in Computing

Volume 22, 2025, Pages 356–367

Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

3D Indoor Reconstruction Based on Spatial Layout Estimation using Panoramic Inspection Video from Building Sites

Shuo Wang[1], Yujie Lu[1,2], Tao Zhong[1], Lijian Zhong[1], Yufan Chen[1]

[1] College of Civil Engineering, Tongji University, Shanghai, China.
[2] Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai, China.
wangshuo2207@tongji.edu.cn, lu6@tongji.edu.cn

Abstract

As the demand for digital delivery of construction projects in the Architectural Engineering and Construction (AEC) industry continues to increase, the importance of worksite inspection and supervision is emphasized. Digital twin modeling of construction processes can reflect real-time site conditions, aiding refined management and project delivery. This paper explores the task of 3D layout reconstruction of interior construction sites through inspection by employing a portable 360-degree panoramic camera. The method uses visual simultaneous localization and Mapping (vSLAM) technology to precisely estimate camera poses during inspections, generating a motion trajectory and selecting key panoramic frames through an optimal capture point searching algorithm. Before reconstruction, the system integrates Inertial Measurement Unit (IMU) data to determine positional relationships between panoramic camera viewpoints, aligning multiple panoramic images into a unified coordinate system for accurate spatial reconstruction. Three-dimensional indoor layouts are reconstructed from panoramic images using a deep learning-based algorithm to automatically detect vertices through panoramic geometry calculations from a single panorama. An experiment with the existing floor plan is conducted to demonstrate the validity of the proposed method. This research introduces a novel approach that enhances the real-time capabilities and automation of spatial layout modeling for construction sites, laying the groundwork for intelligent inspections and holding significant engineering potential for rapid spatial layout recovery and object space mapping in future applications.

Keywords: Intelligent construction, Layout estimation, 3D reconstruction, Panoramic inspection, Simultaneous localization and mapping

I. Hkust, J. Cheng and Y. Yantao (eds.), ICCBEI 2025 (Kalpa Publications in Computing, vol. 22), pp. 356–367

1 Introduction

Digital delivery in civil and infrastructure projects is vital for ensuring project success, quality assurance, owner satisfaction, and long-term operations and maintenance (Guo et al. 2017). A fundamental concept of the digital delivery model is the transition from two-dimensional to three-dimensional management frameworks, enabling more comprehensive and dynamic oversight of project processes. The digital model of existing construction scenes generated through 3D reconstruction can visually demonstrate the differences between the construction site conditions and the design model, enabling project managers to assess conditions efficiently, make informed decisions, and conduct post-project verification and documentation, serving it as a critical technology in the development of an effective project delivery management system.

3D layout reconstruction extracts object profiles, including dimensions, positions, and range data, which is crucial for creating digital twins with detailed geometrical information and enabling human interaction and predictive analysis (Lu et al. 2024). Data for 3D layout reconstruction can be obtained using structured light sensors, laser scanners, X-ray scanners, or imaging devices (Verykokou and Ioannidis 2023). In recent years, digital cameras have become a preferred choice due to their quality, efficiency, and cost-effectiveness. Thus, image-based 3D layout reconstruction techniques have gained increased attention from both theoretical and practical perspectives in the field of construction engineering and management.

Current mainstream methods for 3D reconstruction of indoor layouts fall into two categories. The first method involves using multiple consecutive images with Structure from Motion (SfM) (Ullman 1979) or Simultaneous Localization and Mapping (SLAM) (Durrant-Whyte and Bailey 2006) techniques for depth estimation, resulting in a sparse point cloud model. However, in indoor environments with limited texture and features, image-based 3D reconstruction requires capturing a large number of images from various angles, leading to time-consuming data processing and insufficient model accuracy. The second method employs a single panoramic image to estimate depth or directly infer layout parameters for indoor reconstruction, which is more efficient in data acquisition and processing, providing higher accuracy specifically for layout reconstruction of interior space.

This study examines the digital delivery of 3D construction modeling through site inspection following the sequence of inspection trajectory mapping, capture point determination, and 3D layout reconstruction, as shown in Figure 1. Panoramic video serves as the data source to develop a 3D digital site layout information model. The method employs a portable 360-degree panoramic camera to capture indoor panoramic images and addresses two main challenges in the modeling process: (1) To determine the optimal panoramic capture points during dynamic inspections, an OpenVSLAM-based path trajectory mapping and capture point retrieval algorithm is proposed; (2) To address the issue of missing the actual size of the room layout model, the HorizonNet algorithm is applied for single panoramic image-based 3D layout reconstruction of interior spaces, enabling the recovery of accurate room boundary dimensions. The following sections will focus on related work, proposed methods, experimental analysis, and conclusions regarding 3D layout reconstruction of interior construction based on panoramic site inspection.

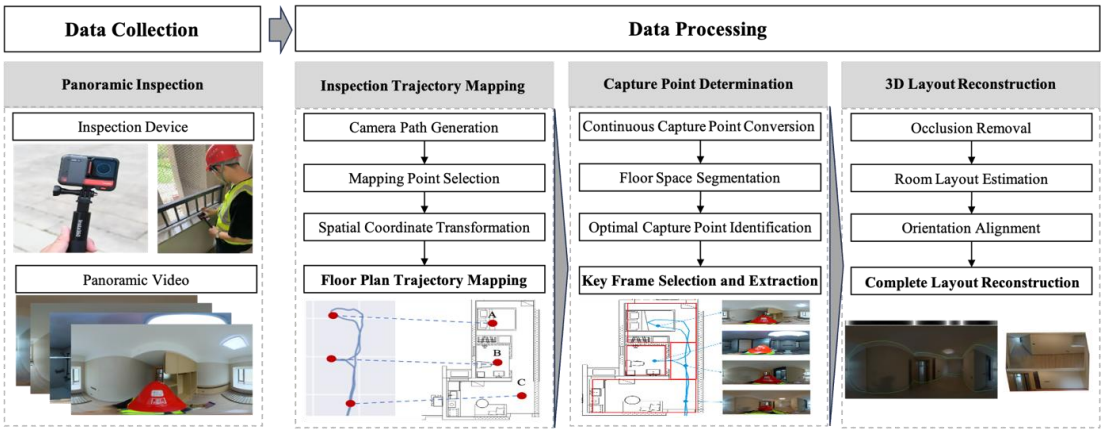

Figure 1: Framework of panoramic image-based 3D layout reconstruction for interior construction space

2 Related Work

2.1 Vision-based 3D Reconstruction

The generation of vision-based 3D reconstruction can be achieved using active or passive methods, distinguished by how depth information is extracted from images. Active methods use RGB-Depth cameras directly to acquire depth data, with Structured Light (Geng 2011) and Time-of-Flight (ToF) (Cui et al. 2010) being common techniques. In contrast, passive methods rely on environmental reflections, such as natural light, to estimate geometric data and depth, which is then used to generate a dense point cloud. Due to limitations of active methods, including environmental and equipment constraints, recent research has increasingly focused on passive vision approaches. Generally, the passive methods can be categorized based on the number of cameras: monocular, binocular, and multi-view systems (Khan et al. 2020).

Monocular vision uses a single camera for 3D reconstruction and is a simple, flexible, and cost-effective method with fast processing times. Its versatility makes it widely applicable in areas such as 3D object measurement and site inspection. Monocular vision extracts features from images like brightness, depth, texture, contours, geometric shapes, and key points. The data processing includes feature extraction, matching, camera motion estimation, bundle adjustment, and depth recovery (Lu et al. 2024). The first four steps form the Structure from Motion (SfM) process, which estimates camera movement and generates sparse point clouds. This sparse reconstruction is then enhanced with Multi-View Stereo (MVS) to achieve spatial consistency, cluster points, and fill in missing data, resulting in a dense point cloud (Furukawa and Hernández 2015). However, the narrow field of view of a standard monocular camera requires capturing multiple images to obtain information about the entire scene, making it challenging to cover all angles. Additionally, since feature point matching is necessary, the images used for reconstruction must have sufficient overlap, increasing the data acquisition complexity.

2.2 Single Panoramic Image-Based 3D Layout Reconstruction

Single-image reconstruction simplifies the image-based 3D reconstruction problem by relying on a single input image rather than multiple images, providing parallax for depth estimation. This approach benefits from easy data acquisition and a simplified model. With the growing prevalence of 360-degree cameras and panoramic techniques applications in engineering fields, layout reconstruction based on

panoramic images has become a popular research focus in computer vision (Cinnamon and Jahiu 2023). Single panoramic image-based 3D layout reconstruction methods are categorized into two main types: constraint-based and deep learning-based approaches.

The constraint-based 3D layout reconstruction approach extracts structural features from images using geometric constraints to estimate 3D layout. These constraints draw from prior knowledge and include vanishing points, parallelism, coplanarity, orthogonality, and perspective relationships (Hoiem et al. 2007). Among these, the Manhattan world assumption is frequently applied in single-image 3D layout reconstruction algorithms. This assumption posits that most artificial environments align object lines and planes with one of the three axes of the Cartesian coordinate system (Coughlan and Yuille 1999). Researchers have widely utilized this assumption for single-image 3D layout reconstruction, particularly in architectural applications, where it allows the segmentation of indoor scenes into floors, walls, and ceilings, thereby reconstructing the 3D model of the scene. Additionally, several scholars have expanded the Manhattan World model to improve its applicability in real-world environments, enhancing its accuracy and effectiveness in diverse scenarios (Straub et al. 2018).

The deep learning-based 3D layout reconstruction approach constructs a target training dataset to map semantic labels to scene geometry using large volumes of data. This method effectively predicts a room's geometric structure through end-to-end learning techniques. LayoutNet (Zou et al. 2018) inputs a pre-processed Manhattan line map into a U-Net structure consisting of an encoder and decoder, generating a probability map for wall corners and boundaries between walls, floors, and ceilings, marking the first deep learning approach to perform 3D layout reconstruction as an end-to-end learning task. Dula-Net (Yang et al. 2019) introduces a deep learning framework for predicting Manhattan-world 3D room layouts from a single RGB panorama, using a dual-branch architecture with equirectangular and perspective ceiling views connected through a feature fusion scheme to enhance accuracy in 2D-floor plans and layout height predictions. LED2-Net (Wang et al. 2021) conducts a differentiable layout-to-depth transformation, which reformulates 360-degree room layout estimation as a 360-degree depth estimation problem, enabling end-to-end training and improving model generalizability by integrating geometric information. HorizonNet (Sun et al. 2019) introduces a 1D representation for 3D room layout estimation from panoramic images, using end-to-end learning and Pano Stretch Data Augmentation to reduce computational complexity and accurately recover complex room shapes. Its advantages over existing methods include lower computational cost and more straightforward implementation, improving performance and generalizability across various indoor environments.

2.3 3D Layout Reconstruction for Construction Management

Information gathering on construction sites traditionally relies on manual inspections and record-keeping, resulting in low levels of digitization and making it difficult for managers to grasp comprehensive site information effectively. The advent of 360-degree capture devices, which can document an entire scene in a single shot, has enabled the effective use of panoramic images for recording construction site conditions.

Research has conducted a comprehensive review of 360° panoramic visualization technologies in the AEC industry, highlighting key applications and benefits for enhancing construction education, monitoring, visualization, and safety training (Shinde et al. 2023). Panoramic image-based 3D layout reconstruction has been applied in construction management scenarios, including automatic progress assessment of interior construction sites (Fang et al. 2023), detection of indoor functional elements like dome lights and outlets (Pintore et al. 2018), partial construction space reconstruction virtual reality applications (Feng et al. 2018). However, existing researches have focused on generating 3D models from panoramic images captured in stationary positions, with limited exploration of mobile technology for capturing panoramic images to reconstruct complete interior construction spaces. This paper explores 3D layout reconstruction of interior construction sites through mobile inspections using a portable 360-degree panoramic camera, ensuring minimal disruption to daily operations.

3 Methodology

3.1 Inspection Trajectory Mapping

This paper utilizes the OpenVSLAM (Sumikura et al. 2019) algorithm to analyze and process panoramic inspection videos to obtain inspection trajectories, recording the spatial positions and location information at each moment during the inspection process. As an indirect method that extracts ORB feature points from target scenes, OpenVSLAM operates through three modules: tracking, mapping, and global optimization.

To obtain the position information of the inspection trajectory from camera to floor plan, an affine transformation is applied to associate the pixel coordinate system C_p with the camera trajectory coordinate system C_t. This process involves fundamental operations, including scaling, rotation, and shearing. By abstracting these fundamental operations, we simplify the task of matching the path to the floor plan into an affine transformation process. Specifically, from the path generated by OpenVSLAM, three trajectory coordinates requiring relocation are recorded as (x_i^t , y_i^t). Given the existing construction floor plan, three reference points are selected on the plan and recorded as (x_i^p, y_i^p). Next, the affine matrix M calculates the pixel coordinates of the inspection trajectory points on the floor plan, as shown in Equation 1, where a is the Linear Transformation Matrix and t is the Translation Vector. Finally, the pixel coordinates of all trajectory points are plotted on the floor plan, illustrating the actual inspection path and trajectory within the plan.

$$C_p = C_t M = \begin{bmatrix} x_i^t & y_i^t & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_i^t & y_i^t & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ t_1 \\ a_{21} \\ a_{22} \\ t_2 \end{bmatrix} \tag{1}$$

3.2 Capture Point Determination

After mapping the inspection trajectory onto the floor plan, the system divides it into continuous capture points based on the building's layout. The target interior space is segmented into multiple rectangles, with each rectangle's center assumed to be the best capture point. The system calculates and calibrates the center coordinates for each segment. The optimal capture point is the one closest to the center of each rectangle, ensuring a comprehensive view of the interior space.

The system then constructs a KD Tree using the camera's capture points along the trajectory for an efficient nearest neighbor search (Bentley 1975). The global optimal capture point search algorithm traverses the KD Tree, as shown in Equation (2), and identifies the closest capture point to each calibrated center based on the shortest distance. This method locates the point along the trajectory nearest the target space's center, ensuring the optimal viewpoint for capturing the panoramic image and extracting the corresponding video frame.

$$Nearest(q) = \arg\min_{p_i \in P} d(q, p_i) \quad \text{where } d(q, p_i) = \sqrt{(x_q - x_i)^2 + (y_q - y_i)^2} \tag{2}$$

3.3 3D Layout Reconstruction

Since the inspection recording process involves handheld shooting, the resulting panoramic images in the form of equirectangular projection contain large areas of occlusion (e.g., the body and recording device), leading to the loss of indoor layout corner points and boundary details. This section uses the Large Mask Inpainting (LaMa) (Suvorov et al. 2021) algorithm to remove occlusions and restore boundary information. Using the original RGB image and the mask image from the panoramic inspection video as inputs, the algorithm, based on Fourier Convolutions, processes the image by leveraging the global structure and contextual information to generate natural and coherent inpainting results. This method retrieves the complete background and restores the full boundary contours of the interior room space.

After Recovering scene boundaries and corner points from indoor panoramic images, HorizonNet reconstructs the room layout into a point cloud format. The panoramic image captured by the camera is an equidistant rectangular image projected from a square relative to the camera's coordinate system. To ensure that the final layout aligns with the world coordinate system, the panoramic image must be re-projected before being input into the HorizonNet neural network. The camera's pose is determined using its 6-axis IMU data and the Mahony algorithm. The py360convert tool then uses the Euler angles derived from the pose calculation to re-project the panoramic image into the world coordinate system.

The image data processing based on HorizonNet can be divided into three stages: pre-processing, processing, and post-processing. In the pre-processing stage, the algorithm starts with panoramic images as input, where the data is aligned to ensure consistency under the equirectangular projection. This alignment helps simplify the process of detecting vertical wall boundaries, a critical step for accurately estimating the room layout. In the Processing stage, ResNet-50 is used to extract relevant multi-scale features. HorizonNet employs a 1D representation that predicts floor-wall, ceiling-wall, and wall-wall boundaries for each image column. A bidirectional LSTM processes these columns sequentially, capturing long-range dependencies and geometric patterns across the layout. In the Post-processing stage, the predicted 1D boundaries are projected into 3D space using the assumption of a 1.8-meter camera height for real-world scaling. HorizonNet enhances accuracy through Pano Stretch Data Augmentation, which generates diverse training samples. The final result is a detailed point cloud model that captures the room's geometric structure, suitable for both cuboid and non-cuboid layouts. The flowchart of 3D layout reconstruction for interior space is shown in Figure 2.

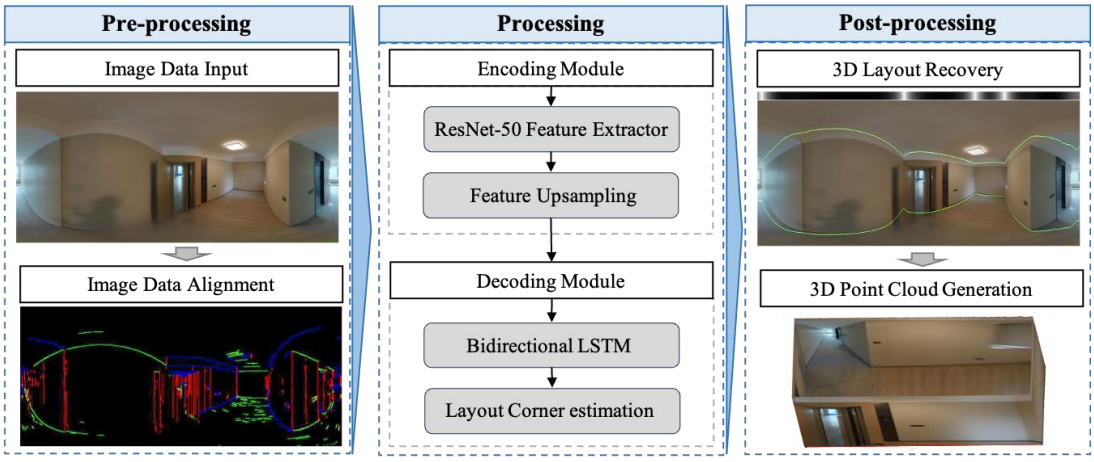

Figure 2: Flowchart of HorizonNet

4 Experiment

The proposed method was validated on a concrete Modular Integrated Construction (MiC) project during the final interior finishing phase. A handheld Insta360 One RS was used for inspection, recording an 87-second panoramic video. After equidistant cylindrical projection, the OpenVSLAM algorithm reconstructed the inspection trajectory, which was projected onto known construction blueprints using affine transformation, as shown in Figure 3. The trajectory was discretized into inspection points, and the KD-tree search identified the optimal points nearest to the center of each space, as shown in Figure 4. Frames from these points were extracted for further image processing and layout reconstruction.

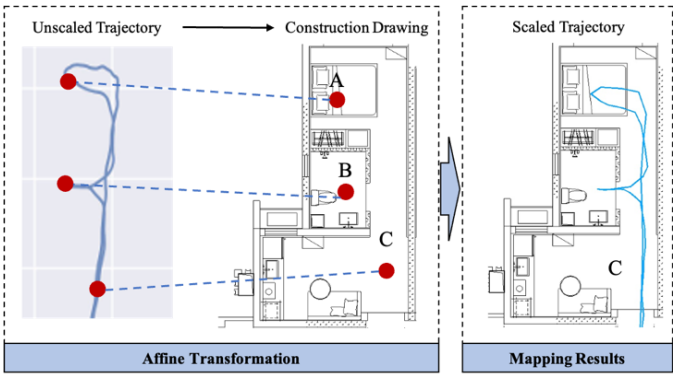

Figure 3: Matching process of inspection path trajectory with the floor plan

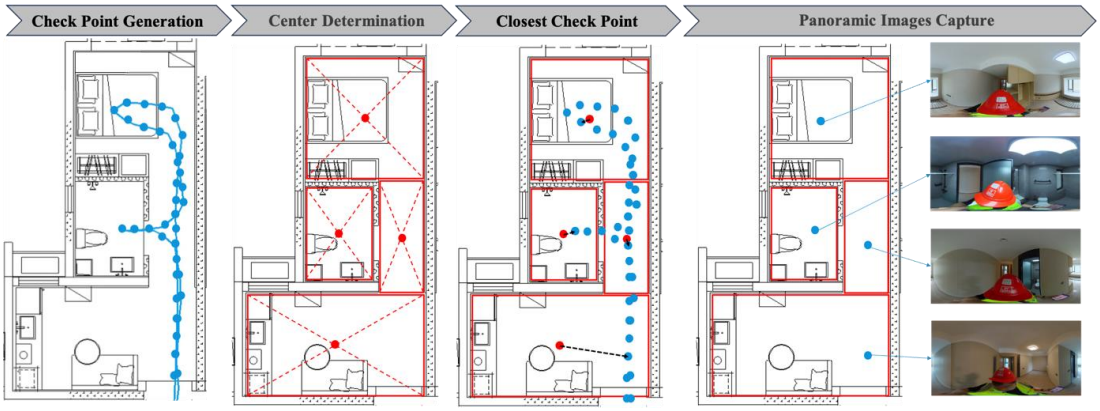

Figure 4: Determination of optimal capture point from inspection trajectory

Before image processing, IMU data calibrated the camera's pose to ensure that all panoramic images used for layout reconstruction had a unified orientation. This calibration step was essential for aligning the images correctly in 3D space. Using the LAMA algorithm for key frame occlusion removal, the process begins by generating a mask for the occluded regions in the original image. The original image and its corresponding mask are then input into the LAMA algorithm, which inpaints the occluded areas and restores corner points along the boundaries of the panoramic image, such as where the floor meets the walls. The steps for removing occlusions in panoramic images are shown in Figure 5.

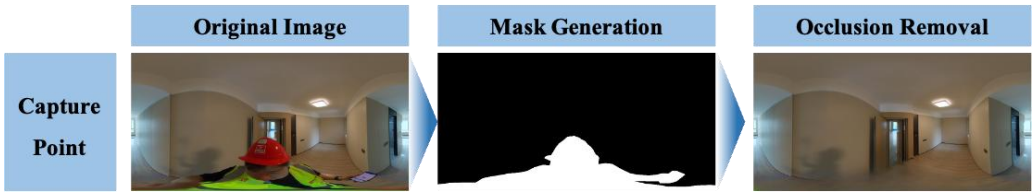

Figure 5: Process of occlusion removal for panoramic image

Our research employed HorizonNet to reconstruct the target room layouts from panoramic images. The model was trained on a computing system equipped with an Intel® Xeon® CPU E5-2683 v4 @ 2.10GHz and two Nvidia GeForce RTX 3090 GPUs, running on the Ubuntu 20.04.06 operating system. The training dataset consisted of a total of 335 panoramic images collected from three different construction sites. Of these, 88 images contained occlusions, while 247 images were free from occlusions. The dataset encompassed various room types, including living rooms, dining rooms, bedrooms, restrooms, kitchens, and hallways. The dataset was partitioned into a training set (268 images, 80%) and a testing set (67 images, 20%) to ensure robust model training and evaluation. The model was fine-tuned using pre-trained weights from the Structure3D model, with a learning rate of 0.0003, a batch size of 16, and training conducted over 100 epochs. The fine-tuning process took approximately 2.5 hours to complete. To evaluate the performance of the trained model, we used two key metrics: 3D Intersection over Union (IoU) and Corner Error. The 3D IoU metric was employed to assess the overlap between the 3D room layout predicted by HorizonNet and the corresponding ground truth, achieving an IoU score of 77.2%. Additionally, the Corner Error, calculated as 0.82%, quantifies the Euclidean distance between the predicted and actual corner points of the room layout, normalized by the diagonal length of the image. These two metrics provide valuable insights into the model's accuracy and precision in predicting room layouts.

The complete process of room layout reconstruction, including all the steps involved, is visually summarized in Figure 6. However, it is important to note that dynamic elements present on construction sites—such as moving objects, workers, or temporary structures—can significantly influence layout accuracy. These elements may cause occlusion of key layout features, such as corners, during the image capture process, thereby introducing inaccuracies in the reconstruction. To mitigate this issue, images with occluded areas larger than one-third of the image resolution should be excluded from the dataset, as they may lead to unreliable predictions and degrade model performance. This consideration is crucial for maintaining the integrity of the dataset and ensuring the quality of the room layout reconstruction.

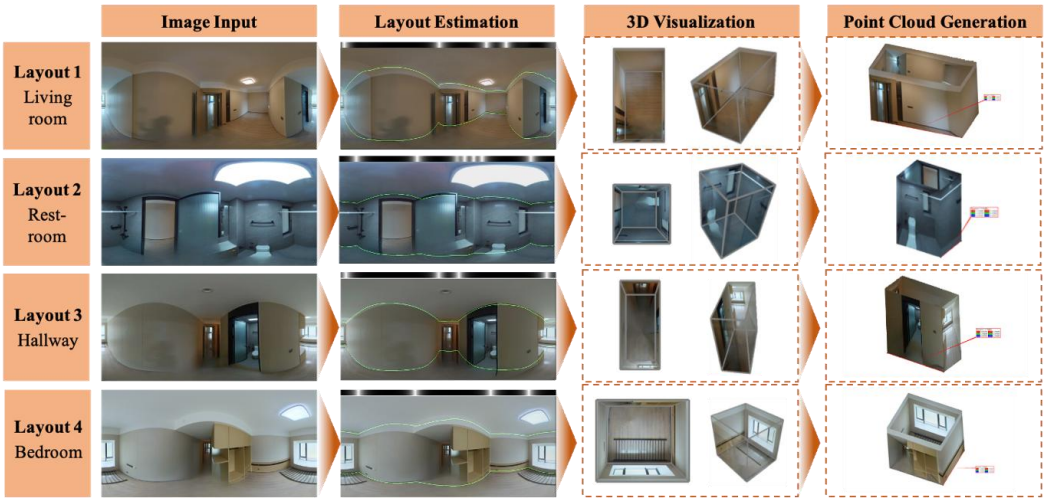

Figure 6: The process of room layout reconstruction

In HorizonNet, a camera height of 1.8 meters is assumed for projecting the predicted 2D layout boundaries (floor-wall and ceiling-wall) into a 3D space. This assumption allows for the accurate calculation of floor-to-ceiling distances and wall positions relative to the camera, facilitating realistic room reconstruction from panoramic image inputs. Following the 3D layout reconstruction, one boundary from each reconstructed layout was selected and compared with the as-designed floor plan. The average discrepancy between the two was approximately 0.3 meters, as depicted in Figure 7. However, the study is subject to two limitations regarding layout estimation and real-scale recovery. In terms of layout estimation, the average prediction accuracy is constrained by the limited number of panoramic images available from the construction site. Regarding real-scale recovery, measurement errors may arise from fluctuations in camera height (±5 cm) during inspections, which limit the precision of the 3D reconstruction. Future research could focus on augmenting panoramic image data through image stretching techniques. Additionally, the adoption of a fixed-height camera system may enhance the accuracy of future studies.

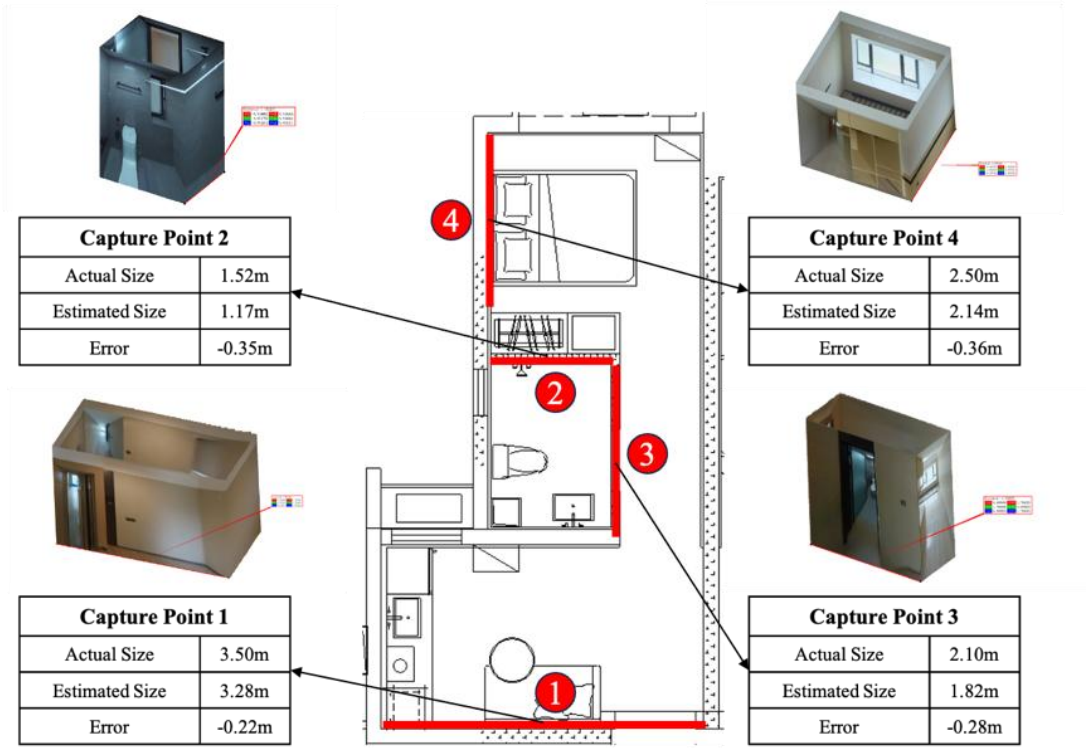

Figure 7: Size evaluation of proposed method for room layour reconstruction

5 Conclusions

This paper introduces a new strategy for 3D layout reconstruction of interior construction sites using a portable 360-degree panoramic camera. The three key contributions are inspection trajectory mapping, capture point determination, and 3D layout reconstruction. It addresses challenges in capturing optimal panoramic data through an OpenVSLAM-based path mapping and capture point retrieval algorithm. The KD-tree is applied to locate points along the trajectory nearest to the center of each target space, ensuring the best viewpoint for capturing panoramic images and extracting key video frames. Using these key images, HorizonNet reconstructs 3D layouts, achieving 77.2% 3D IoU and 0.82% Corner Error, with a 0.3-meter boundary error compared to as-designed floor plans. This method serves as an alternative to stationary layout estimation, offering valuable insights into intelligent site inspections and precise spatial modeling.

This research can stand out as a flexible tool for future construction site monitoring, enhancing accuracy and real-time capability in dynamic environments. By incorporating interior layout reconstruction with higher accuracy, construction professionals can improve both the efficiency and completeness of site inspections, ensuring more reliable data collection and spatial analysis throughout the construction process.

Acknowledgments

The study was supported by the National Key Research & Development Program of China (2022YFC3801700), The Science Foundation for the Science and Technology Commission of Shanghai Municipality (22dz1207100 and 22dz1207800), Fundamental Research Funds for the Central Universities (2024-1-ZD-02).

References

Bentley, J. L. 1975. "Multidimensional binary search trees used for associative searching." *Commun. ACM*, 18 (9): 509–517. https://doi.org/10.1145/361002.361007.

Cinnamon, J., and L. Jahiu. 2023. "360-degree video for virtual place-based research: A review and research agenda." *Computers, Environment and Urban Systems*, 106: 102044. https://doi.org/10.1016/j.compenvurbsys.2023.102044.

Coughlan, J. M., and A. L. Yuille. 1999. "Manhattan World: compass direction from a single image by Bayesian inference." *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 941–947 vol.2. Kerkyra, Greece: IEEE.

Cui, Y., S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 2010. "3D shape scanning with a time-of-flight camera." *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1173–1180. San Francisco, CA, USA: IEEE.

Durrant-Whyte, H., and T. Bailey. 2006. "Simultaneous localization and mapping: part I." *IEEE Robot. Automat. Mag.*, 13 (2): 99–110. https://doi.org/10.1109/MRA.2006.1638022.

Fang, X., H. Li, H. Wu, L. Fan, T. Kong, and Y. Wu. 2023. "A fast end-to-end method for automatic interior progress evaluation using panoramic images." *Engineering Applications of Artificial Intelligence*, 126: 106733. https://doi.org/10.1016/j.engappai.2023.106733.

Feng, Z., V. A. González, L. Ma, M. M. A. Al-Adhami, and C. Mourgues. 2018. "Rapid 3D Reconstruction of Indoor Environments to Generate Virtual Reality Serious Games Scenarios." arXiv. https://doi.org/10.48550/ARXIV.1812.01706.

Furukawa, Y., and C. Hernández. 2015. "Multi-View Stereo: A Tutorial." *FNT in Computer Graphics and Vision*, 9 (1–2): 1–148. https://doi.org/10.1561/0600000052.

Geng, J. 2011. "Structured-light 3D surface imaging: a tutorial." *Adv. Opt. Photon.*, 3 (2): 128. https://doi.org/10.1364/AOP.3.000128.

Guo, F., C. T. Jahren, Y. Turkan, and H. David Jeong. 2017. "Civil Integrated Management: An Emerging Paradigm for Civil Infrastructure Project Delivery and Management." *J. Manage. Eng.*, 33 (2): 04016044. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000491.

Hoiem, D., A. A. Efros, and M. Hebert. 2007. "Recovering Surface Layout from an Image." *Int J Comput Vis*, 75 (1): 151–172. https://doi.org/10.1007/s11263-006-0031-y.

Khan, F., S. Salahuddin, and H. Javidnia. 2020. "Deep Learning-Based Monocular Depth Estimation Methods—A State-of-the-Art Review." *Sensors*, 20 (8): 2272. https://doi.org/10.3390/s20082272.

Lu, Y., S. Wang, S. Fan, J. Lu, P. Li, and P. Tang. 2024. "Image-based 3D reconstruction for Multi-Scale civil and infrastructure Projects: A review from 2012 to 2022 with new perspective from deep learning methods." *Advanced Engineering Informatics*, 59: 102268. https://doi.org/10.1016/j.aei.2023.102268.

Pintore, G., R. Pintus, F. Ganovelli, R. Scopigno, and E. Gobbetti. 2018. "Recovering 3D existing-conditions of indoor structures from spherical images." *Computers & Graphics*, 77: 16–29. https://doi.org/10.1016/j.cag.2018.09.013.

Shinde, Y., K. Lee, B. Kiper, M. Simpson, and S. Hasanzadeh. 2023. "A Systematic Literature Review on 360° Panoramic Applications in Architecture, Engineering, and Construction (AEC) Industry." *ITcon*, 28: 405–437. https://doi.org/10.36680/j.itcon.2023.021.

Straub, J., O. Freifeld, G. Rosman, J. J. Leonard, and J. W. Fisher. 2018. "The Manhattan Frame Model—Manhattan World Inference in the Space of Surface Normals." *IEEE Trans. Pattern Anal. Mach. Intell.*, 40 (1): 235–249. https://doi.org/10.1109/TPAMI.2017.2662686.

Sumikura, S., M. Shibuya, and K. Sakurada. 2019. "OpenVSLAM: A Versatile Visual SLAM Framework." *Proceedings of the 27th ACM International Conference on Multimedia*, 2292–2295. Nice France: ACM.

Sun, C., C.-W. Hsiao, M. Sun, and H.-T. Chen. 2019. "HorizonNet: Learning Room Layout With 1D Representation and Pano Stretch Data Augmentation." *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1047–1056. Long Beach, CA, USA: IEEE.

Suvorov, R., E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky. 2021. "Resolution-robust Large Mask Inpainting with Fourier Convolutions." arXiv.

Ullman, S. 1979. "The interpretation of structure from motion." *Proc. R. Soc. Lond. B.*, 203 (1153): 405–426. https://doi.org/10.1098/rspb.1979.0006.

Verykokou, S., and C. Ioannidis. 2023. "An Overview on Image-Based and Scanner-Based 3D Modeling Technologies." *Sensors*, 23 (2): 596. https://doi.org/10.3390/s23020596.

Wang, F.-E., Y.-H. Yeh, M. Sun, W.-C. Chiu, and Y.-H. Tsai. 2021. "LED ^{2} -Net: Monocular 360° Layout Estimation via Differentiable Depth Rendering." *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12951–12960. Nashville, TN, USA: IEEE.

Yang, S.-T., F.-E. Wang, C.-H. Peng, P. Wonka, M. Sun, and H.-K. Chu. 2019. "DuLa-Net: A Dual-Projection Network for Estimating Room Layouts From a Single RGB Panorama." *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3358–3367. Long Beach, CA, USA: IEEE.

Zou, C., A. Colburn, Q. Shan, and D. Hoiem. 2018. "LayoutNet: Reconstructing the 3D Room Layout from a Single RGB Image." *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2051–2059. Salt Lake City, UT: IEEE.