

## MR-MD: Multi-Robot Mapping with Manhattan Descriptor

Haiyang Wu<sup>1,2,3</sup>, Ruofei Zhong<sup>1,2,3</sup>, \*, Donghai Xie<sup>1,2,3</sup>, Chi Chen<sup>4,5,6</sup>, Jie Tang<sup>1,2,3</sup>, Chaohong Wu<sup>1,2,3</sup>, Xingyu Qi<sup>1,2,3</sup>

<sup>1</sup> Key Laboratory of 3D Information Acquisition and Application, MOE, Capital Normal University, Beijing 100048, China –  
2210902160@cnu.edu.cn(H.W.), zrf@cnu.edu.cn(R.Z.), 6121@cnu.edu.cn(D.X.), 2210902147@cnu.edu.cn(J.T.),  
2220902148@cnu.edu.cn(C.W.), 2220902187@cnu.edu.cn(X.Q.)

<sup>2</sup> Base of the State Key Laboratory of Urban Environmental Process and Digital Modeling, Capital Normal University, Beijing  
100048, China

<sup>3</sup> College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China

<sup>4</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072,  
China - chichen@whu.edu.cn

<sup>5</sup> Engineering Research Center of Space-Time Data Capturing and Smart Application, the Ministry of Education of P.R.C., Wuhan  
430072, China

<sup>6</sup> Institute of Geospatial Intelligence, Wuhan University, Wuhan 430072, China

**KEY WORDS:** Multi-robot, Manhattan World Assumption, Indoor Scenes, Planar Constraint, LiDAR SLAM.

### ABSTRACT:

Simultaneous Localization and Mapping (SLAM) technology, utilizing Light Detection and Ranging (LiDAR) sensors, is crucial for 3D environment perception and mapping. However, the absence of absolute observations and the inefficiency of single-robot perception present challenges for LiDAR SLAM in indoor environments. In this paper, we propose a multi-robot (MR) collaborative mapping method based on the Manhattan descriptor (MD) named MR-MD to overcome these limitations and improve the perception accuracy of LiDAR SLAM in indoor environments. The proposed method consists of two modules: MD generation and MD optimization. In the first module, each robot builds a local submap and constructs MD by parameterizing the planes in the submap. In the second module, the global main direction is updated using the historical MD of each robot, and constraints are built for each robot's horizontal and vertical directions according to their current MD and optimized. We conducted extensive comparisons with other multi-robot and single-robot LiDAR SLAM methods using real indoor data, and the results show that our method achieved higher mapping accuracy.

## 1. INTRODUCTION

### 1.1 Problem Statement

Simultaneous Localization and Mapping (SLAM) technology, employing Light Detection and Ranging (LiDAR) sensors, is a general approach for generating 3D maps of indoor environments. With the development of LiDAR SLAM technology, some LiDAR SLAM algorithms (Xu et al., 2022; He et al., 2023; Chen et al., 2023) can obtain high-precision point cloud maps in small-scale indoor environments with rich geometric features. However, in indoor scenes with unclear geometric features or complex environments such as long corridors and staircases, there are still challenges, such as scene degradation (Zhang et al., 2016) and difficulty eliminating accumulated errors (Liu et al., 2021). Moreover, the low efficiency of single-robot perception impedes the efficient construction of high-quality maps of large-scale indoor 3D environments. Existing multi-robot collaborative LiDAR SLAM algorithms (Zhong et al., 2022; Cramariuc et al., 2022; Huang et al., 2021) primarily address inter-robot collaboration issues, yet they continue to encounter challenges in the abovementioned complex indoor environments.

### 1.2 Proposed Solution

In this paper, we apply the Manhattan World assumption (Coughlan et al., 1999), a priori assumption of artificial building geometry relationships, to structured indoor environments. In visual SLAM, some algorithms (Li et al., 2020; Yunus et al.,

2021) employ the Manhattan World assumption to enhance the robustness and accuracy of the algorithm. In LiDAR SLAM, (Wu et al., 2023) apply the Manhattan World assumption to provide constraints for single-robot SLAM. However, the application of this assumption remains limited. (Dai et al., 2022) used the Manhattan World assumption to construct maps based on navigation needs but found a slight improvement in mapping accuracy. Based on this, we propose a multi-robot collaborative mapping method named MR-MD that uses Manhattan descriptors (MD) to overcome these limitations and improve the perception accuracy of LiDAR SLAM in indoor environments. The proposed method consists of two modules: MD generation and MD optimization. In the first module, each robot builds a local submap and constructs MD by parameterizing the planes in the submap. In the second module, the global main direction is updated using the historical MD of each robot, and constraints are built for each robot's horizontal and vertical directions according to their current MD and optimized. We conducted extensive comparisons with other multi-robot and single-robot LiDAR SLAM methods using real indoor data, and the results show that our method achieved higher mapping accuracy.

The remainder of this paper is organized as follows. Section 2 introduces the principles of the methodology proposed in this study. Section 3 presents the experiments conducted in a real-world environment and analyzes the results. Finally, in Section 4, we provide a summary of the paper.

\* Corresponding author

## 2. METHODOLOGY

The overall process of the proposed algorithm is shown in Figure 1. We accumulate point cloud submaps using a fixed-size sliding window comprising 50 or 30 frames. Input data can originate from one or multiple robots and includes raw LiDAR scan point cloud data and IMU data. Upon obtaining a submap, we perform preprocessing operations such as point cloud downsampling, passthrough filtering, and noise removal to obtain point cloud data better suited for subsequent processing.

A two-stage MD generation module is employed to extract potential planes from the input point cloud and parameterize them to calculate their normals. Simultaneously, the module computes the necessary values for Manhattan descriptors and generates complete Manhattan descriptors. Subsequently, a collaborative update of MDs from all robots is performed in the MD optimization module. Then, plane constraints are constructed based on MDs to facilitate loop closure matching. Finally, a global factor graph is constructed, incorporating odometry factors, loop closure factors, and plane constraint factors, enabling global optimization.

This process results in optimal poses in environments conforming to the Manhattan World assumption, ultimately generating a globally consistent point cloud map.

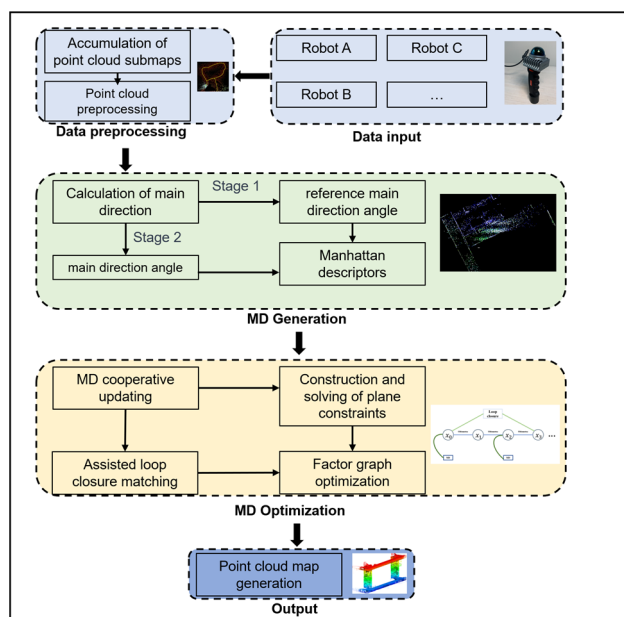


Figure 1. The overall process of the algorithm.

### 2.1 MD Generation

According to the Manhattan World assumption, artificial buildings have three mutually orthogonal main directions, including a horizontal main direction (HMD) approximately parallel to the ground and two vertical main directions (VMD) perpendicular to the HMD and each other. The two VMDs are transformed into one VMD through their orthogonal relationship to facilitate the construction of plane constraints. Due to the presence of specific accumulated errors in the currently acquired point cloud data that have not been processed, the main direction angle is extracted from the optimized historical point cloud submaps, which is the vertical reference main direction angle

(VRMD) and the horizontal reference main direction angle (HRMD).

We employ a two-stage strategy when generating the MD for each robot at a fixed frequency. Firstly, in the first stage, we select a point cloud submap from 200 frames before the current frame as the reference point cloud. We assume this point cloud has undergone partial optimization and has a more minor cumulative error. From the reference point cloud, we extract the vertical and horizontal planes it contains and calculate their respective normal vectors. A point cloud submap may contain multiple planes, and we use statistical methods to eliminate incorrectly extracted planes, obtaining a statistically best estimate of the orientation for the scene.

In the second stage, we obtain a point cloud submap consisting of the last 50 frames as the current point cloud submap. We extract this point cloud's vertical and horizontal planes while calculating their normal vectors. We then use the same statistical methods as in the first stage to obtain the current best orientation estimate.

In indoor environments such as long corridors, the current environment often cannot provide effective constraints for a certain direction during loop closure matching (including loop closure within a single robot and between multiple robots), resulting in unreliable results. However, according to the Manhattan World assumption, planes in the environment can be divided into three orthogonal categories based on their normal directions. Therefore, the area ratio of each category of planes to the total plane area can be calculated as the reliability score for the current direction of loop closure, which can provide constraints for possible subsequent loop closure matching. Based on this, the MD of each robot in the current multi-robot collaborative 3D LiDAR SLAM system is constructed. The MD is defined as the following equation:

$$MD(pre, fra) = [D_h, D_v, D_{hr}, D_{vr}, C_0, C_1, C_2, F], \quad (1)$$

where  $D_h, D_v$  = HMD, VMD  
 $D_{hr}, D_{vr}$  = HRMD, VRMD  
 $C_0, C_1, C_2$  = Confidence level in three directions  
 $F$  = initialization flag

### 2.2 MD Optimization

After each robot in the multi-robot system generates its own MD, the HRMD and VRMD are collaboratively updated for the MD that has completed coordinate alignment. Next, vertical and horizontal plane constraints are constructed based on the differences between VMD and VRMD, as well as between HMD and HRMD. The constructed plane constraints are then added as edges in the factor graph (Kaess et al., 2012) between the current variable nodes, and optimization is performed using incremental smoothing and mapping with Bayesian trees.

In addition, assistance can be provided for the loop closure matching based on the MD when a loop closure is detected. When a loop closure is detected within a single robot, the MD with the closest frame number to the current and historical frames that form the loop is first found, and the confidence level information in each direction is obtained for the loop closure matching based on the confidence level information in both directions of the two MDs. At the same time, the difference between the main directions of the two MDs can provide a good initial value for this loop closure matching, thereby improving the accuracy and efficiency of the matching. When a loop closure is detected

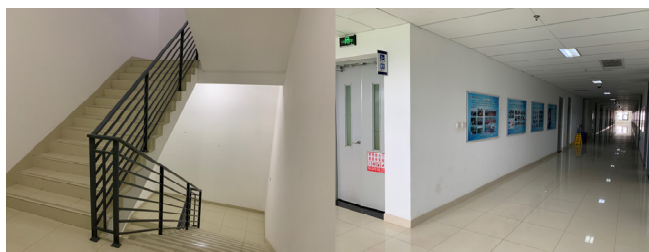
within multiple robots, the MD closest to the current frame number of each robot is found based on the frame numbers and robot IDs of the two robots that form the loop. Then, the initialization flag is checked. If any robot has not completed initialization, only the confidence level information of the MD is used, and no initial value is provided for this matching. The result of the loop closure matching is added as a constraint between two variable nodes in the overall factor graph, and the confidence level in each direction calculated at this time is used as the confidence level of this constraint.

### 3. EXPERIMENTS AND ANALYSIS

To verify the effectiveness of the proposed method in this paper, we applied the proposed algorithm to the DCL-SLAM algorithm as a plug-in. We used a hand-held device equipped with Livox MID-360 LiDAR as the experimental dataset acquisition device, as shown in Figure 2. The experimental dataset was collected from real indoor environments. Specifically, the dataset shown in Figure 3(a) contains long corridors and long stairs, which were used to verify the effectiveness of our algorithm for single-robot systems. The dataset shown in Figure 3(b) contains data collected from two different areas of an underground parking lot simultaneously, which were used to verify the effectiveness of our algorithm for multi-robot systems.



Figure 2. Handheld device equipped with MID-360 LiDAR.



(a) Indoor corridor and staircase dataset.



(b) Indoor parking lot dataset.

Figure 3. Real-world images from two experimental datasets.

#### 3.1 Single-Robot Mapping Results

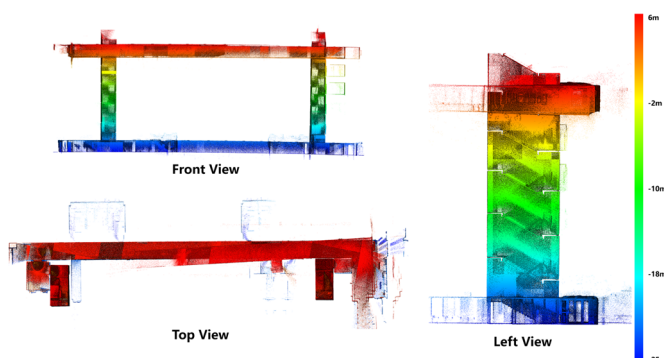
We compared our proposed algorithm with the SOTA (state of the art) single-robot LiDAR SLAM algorithm, FAST-LIO2, and FAST-LIO2+SC algorithm, which incorporates loop closure using ScanContext descriptors, using the indoor corridor and staircase dataset. The mapping results of the three methods are shown in Figure 4, where the point clouds are colored based on their elevation values.

For the indoor corridor and staircase dataset, the data collection path starts from the sixth floor of the teaching building, goes through the left staircase to the first floor, and returns to the sixth floor using the right staircase.

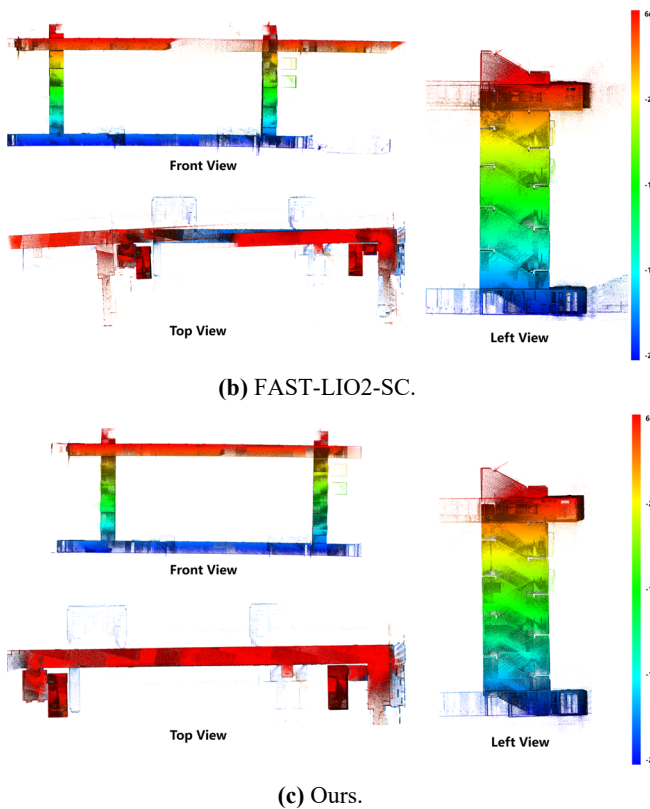
By comparing the left views of the three algorithms in Figure 4, it can be observed that there is no significant variation in the Z-direction among the three methods. Part of this is attributed to the larger vertical field of view (FOV) of the Livox MID-360, allowing it to capture vertical point cloud features better.

When comparing the front and top views of the three methods in Figure 4, it is evident that FAST-LIO2-SC exhibits significant errors in corridor areas, with the displayed corridor length far exceeding the actual length. This error is attributed to incorrect loop closure matching when compared to FAST-LIO2.

Furthermore, FAST-LIO2's front view in Figure 4 is largely consistent with our method, but the top view needs to be more accurate. This misalignment is due to FAST-LIO2's inability to mitigate cumulative errors. Our method demonstrates good mapping quality and global consistency.

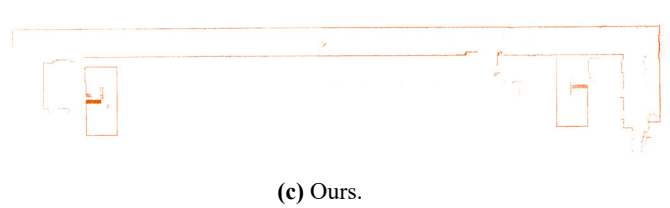
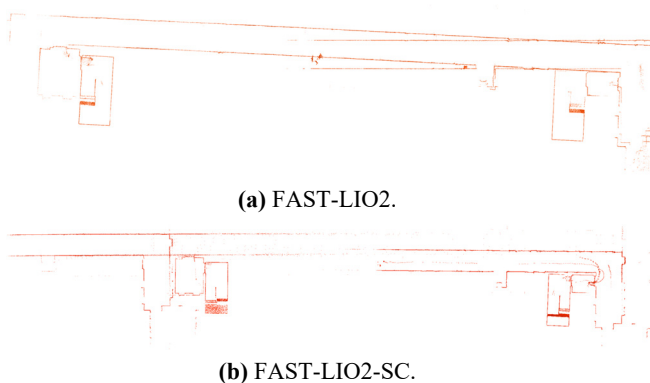


(a) FAST-LIO2.



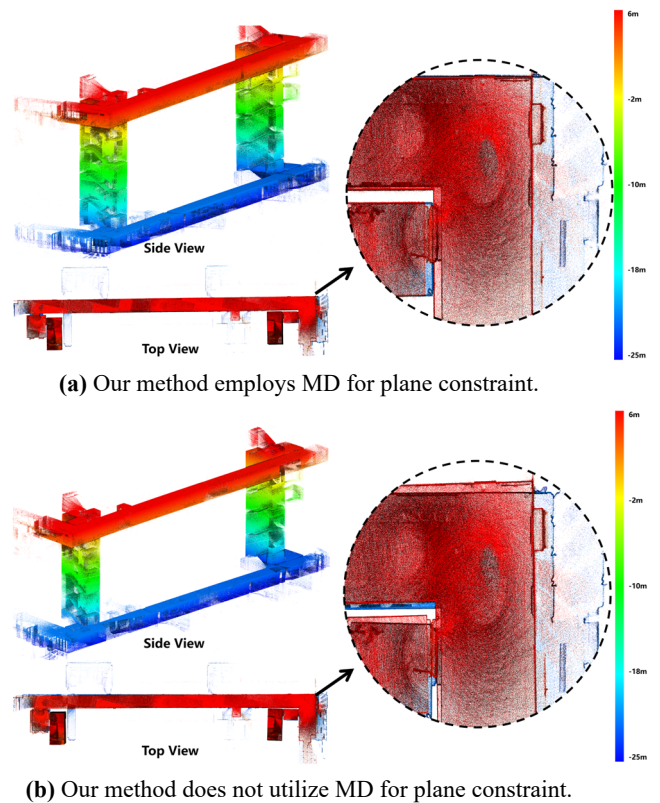
**Figure 4.** Comparison of three methods on indoor corridor and staircase dataset.

To better compare the mapping quality of the three methods, we obtained six cross-sectional profiles of the point cloud maps at a height of 5 meters with a thickness of 0.2 meters, as shown in Figure 5. From the mapping results of the three methods, it can be observed that the FAST-LIO2 algorithm exhibits significant drift when returning to the sixth floor due to the absence of a loop closure module to mitigate cumulative errors. Although the FAST-LIO2-SC algorithm incorporates a loop closure module, the overall mapping quality is poorer due to inadequate point cloud registration when detecting loops on the sixth floor, resulting in significant degradation in the long corridor environment. On the other hand, our method shows no significant degradation or drift and outperforms the other two methods regarding mapping quality.



**Figure 5.** Cross-sectional profiles of the three methods.

To verify the effectiveness of using MD for plane constraint, we conducted ablation experiments while keeping other components of the algorithm, such as MD generation and the assistance of MD for loop closure matching, unchanged. The sole purpose was to evaluate the impact of the MD plane constraint on the overall mapping quality by turning the MD plane constraint on or off. The experimental results are shown in Figure 6. When MD plane constraint was not enabled, noticeable stratification was observed at the corners of the corridor. However, a globally consistent point cloud map without evident stratification was obtained when the MD plane constraint was enabled.



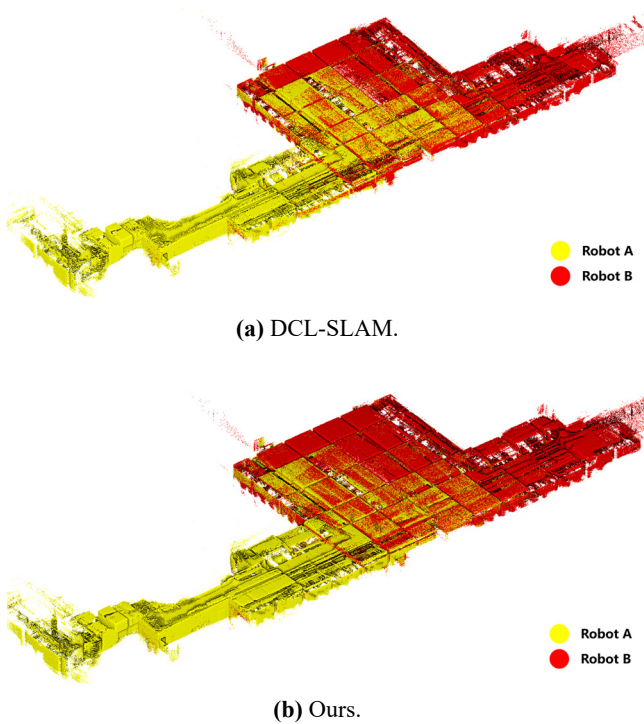
**Figure 6.** Ablation experiments of MD for plane constraint.

### 3.2 Multi-Robot Mapping Results

In the indoor parking lot dataset, the parking area was divided into two parts and data collection was performed using the handheld device shown in Figure 2 for both parts. There is a partial overlap between the two data collection sessions. Robot A and Robot B represent the two data collection sessions, respectively.

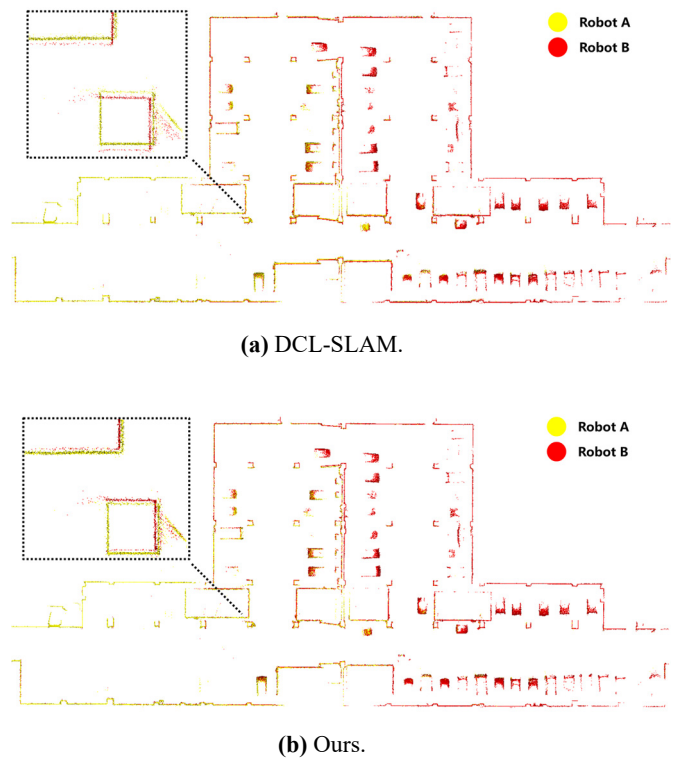
Using the indoor parking lot dataset, we compared our proposed algorithm with the multi-robot LiDAR SLAM algorithm, DCL-SLAM. The mapping results of the two methods are shown in Figure 7, where the point clouds are colored based on their elevation values.

By comparing the mapping results of Figure 7 on the indoor parking lot dataset, there are noticeable differences between DCL-SLAM and our method in the overlapping area (where the red and yellow overlap).



**Figure 7.** Comparison of two methods on indoor parking lot dataset.

We obtained cross-sectional profiles of the mapping results of the two methods at a height of 1.3 meters with a thickness of 0.2 meters, as shown in Figure 8. The cross-sectional profiles provide a clearer visualization of the differences in mapping quality between the two methods. In Figure 8(a), the DCL-SLAM algorithm exhibits significant point cloud overlap in the overlapping region between the two robots, indicating the method's inability to fuse point cloud maps acquired by different robots effectively. In Figure 8(b), our method demonstrates smaller point cloud overlap than DCL-SLAM, reflecting our method's ability to achieve higher-quality mapping results.



**Figure 8.** Cross-sectional profiles of the two methods.

In Table 1, we selected eight uniformly distributed building corners in the overlapping region of the two robots' data collection. The distance of point cloud overlap at these corner points was manually measured to describe the mapping quality of the two algorithms quantitatively. We use bold font to indicate more minor overlap errors and use the average overlap error to evaluate mapping quality. Our algorithm achieves more minor overlap errors at all locations, with an average overlap error reduction of 64.4% compared to DCL-SLAM.

Point Number	DCL-SLAM	Ours
1	0.113	<b>0.034</b>
2	0.107	<b>0.016</b>
3	0.084	<b>0.041</b>
4	0.092	<b>0.015</b>
5	0.092	<b>0.060</b>
6	0.119	<b>0.024</b>
7	0.094	<b>0.043</b>
8	0.129	<b>0.064</b>
Mean	0.104	<b>0.037</b>

**Table 1.** Multi-robot Error (m) in Parking Lot Dataset.

#### 4. CONCLUSIONS

In our study, we conducted experiments involving single-robot and multi-robot scenarios to assess the effectiveness of our proposed multi-robot collaborative mapping method, MR-MD, in indoor environments. Given the absence of ground truth data in indoor settings, we employed both qualitative and quantitative evaluations to validate the efficacy of our approach. We assessed mapping quality and quantified the distance of point cloud overlap at corner points as evaluation metrics.

Our results demonstrate that our method significantly enhances mapping quality compared to the SOTA single-robot algorithm, FAST-LIO2, in the scenarios above. Furthermore, compared to the SOTA multi-robot algorithm, DCL-SLAM, our method reduces the distance of point cloud overlap at corner points by 64.4%.

It is worth noting that our proposed method does not currently exhibit real-time performance. This limitation primarily stems from the method's reliance on plane extraction and plane normal vector computation, necessitating a high density of point cloud submaps for accurate extraction and calculation. As a result, maintaining a high-frequency real-time extraction and computation process is challenging.

Additionally, our method's effectiveness is constrained by the Manhattan World assumption upon which it is based. The accuracy and mapping quality improvements are limited in atypical environments that do not conform to the Manhattan World assumption, such as high-dynamic or outdoor settings.

In the future, our research will continue to address the limitations of the method presented in this paper. On the one hand, we will explore the feasibility of real-time extraction of reliable planes and computation of their normals from point clouds. On the other hand, we will endeavor to extend the applicability of our method to a broader range of scenarios, such as outdoor environments, to enhance its overall versatility.

#### ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant U22A20568 and 42071444, in part by the National Key Technologies Research and Development Program of China under Grant 2022YFB3904101.

#### REFERENCES

- Chen, K., Nemiroff, R., & Lopez, B. T., 2023. Direct LiDAR-Inertial Odometry and Mapping: Perceptive and Connective SLAM. *arXiv preprint arXiv:2305.01843*.
- Coughlan, J. M., Yuille, A. L., 1999. Manhattan World: Compass direction from a single image by Bayesian inference. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 941–947.
- Cramariuc, A., Bernreiter, L., Tschopp, F., Fehr, M., Reijgwart, V., Nieto, J., ... & Cadena, C., 2022. maplab 2.0—a modular and multi-modal mapping framework. *IEEE Robotics and Automation Letters*, 8(2), 520-527.
- Dai, A., Lund, G., & Gao, G., 2022. PlaneSLAM: Plane-based LiDAR SLAM for Motion Planning in Structured 3D Environments. *arXiv preprint arXiv:2209.08248*.
- He, D., Xu, W., Chen, N., Kong, F., Yuan, C., & Zhang, F., 2023. Point - LIO: Robust High - Bandwidth Light Detection and Ranging Inertial Odometry. *Advanced Intelligent Systems*, 2200459.
- Huang, Y., Shan, T., Chen, F., & Englot, B., 2021. DiSCo-SLAM: Distributed scan context-enabled multi-robot lidar slam with two-stage global-local graph optimization. *IEEE Robotics and Automation Letters*, 7(2), 1150-1157.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J. J., & Dellaert, F., 2012. iSAM2: Incremental smoothing and mapping using the bayes tree. *The International Journal of Robotics Research* 31(2), 216-235.

Li, Y., Brasch, N., Wang, Y., Navab, N., & Tombari, F., 2020. Structure-slam: Low-drift monocular slam in indoor environments. *IEEE Robotics and Automation Letters*, 5(4), 6583-6590.

Liu, Z., & Zhang, F., 2021. Balm: Bundle adjustment for lidar mapping. *IEEE Robotics and Automation Letters*, 6(2), 3184-3191.

Wu, H., Wu, W., Qi, X., Wu, C., An, L., & Zhong, R., 2023. Planar Constraint Assisted LiDAR SLAM Algorithm Based on Manhattan World Assumption. *Remote Sensing*, 15(1), 15.

Xu, W., Cai, Y., He, D., Lin, J., & Zhang, F., 2022. Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics* 38(4), 2053-2073.

Yunus, R., Li, Y., & Tombari, F., 2021. Manhattanslam: Robust planar tracking and mapping leveraging mixture of manhattan frames. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6687-6693.

Zhang, J., Kaess, M., Singh, S., 2016. On degeneracy of optimization-based state estimation problems. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 809-816.

Zhong, S., Qi, Y., Chen, Z., Wu, J., Chen, H., & Liu, M., 2022. DCL-SLAM: A Distributed Collaborative LiDAR SLAM Framework for a Robotic Swarm. *arXiv preprint arXiv:221011978*.