

Traversability-aware Consistent Situational Graphs for Indoor Localization and Mapping

Jeewon Kim¹, Minho Oh¹, and Hyun Myung^{1,*}

* Corresponding author

School of Electrical Engineering, KAIST (Korea Advanced Institute of Science and Technology), Daejeon, 34141, Republic of Korea,
{ddarong2000, minho.oh, hmyung}@kaist.ac.kr
<http://urobot.kaist.ac.kr>

Abstract. Scene graphs enhance 3D mapping capabilities in robotics by understanding the relationships between different spatial elements, such as rooms and objects. Recent research extends scene graphs to hierarchical layers, adding and leveraging constraints across these levels. This approach is tightly integrated with pose-graph optimization, improving both localization and mapping accuracy simultaneously. However, when segmenting spatial characteristics, consistently recognizing rooms becomes challenging due to variations in viewpoints and limited field of view (FOV) of sensors. For example, existing real-time approaches often over-segment large rooms into smaller, non-functional spaces that are not useful for localization and mapping due to the time-dependent method. Conversely, their voxel-based room segmentation method often under-segment in complex cases like not fully enclosed 3D space that are non-traversable for ground robots or humans, leading to false constraints in pose-graph optimization. We propose a traversability-aware room segmentation method that considers the interaction between robots and surroundings, with consistent feasibility of traversability information. This enhances both the semantic coherence and computational efficiency of pose-graph optimization. Improved performance is demonstrated through the re-detection frequency of the same rooms in a dataset involving repeated traversals of the same space along the same path, as well as the optimization time consumption.

Keywords: Scene graphs, Traversable terrain, Room segmentation, SLAM

1 Introduction

Scene graphs offer a structured representation of 3D environments, enhancing semantic-rich mapping in robotics. Unlike traditional methods that focus solely on geometric data, this approach captures both objects and their spatial relationships, thereby improving navigation and understanding of surroundings for robots and humans alike. However, constructing accurate scene graphs heavily depends on precise robot localization, which relies on detailed elements like

This work was supported in part by the Technology Innovation Program(or Industrial Strategic Technology Development Program-Robot Industry Technology Development)(00427719, Dexterous and Agile Humanoid Robots for Industrial Applications) funded By the Ministry of Trade Industry Energy(MOTIE, Korea) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(RS-2024-00348461).

keyframes or point clouds. Recent improvements in SLAM methodologies conducted by Bavle, H. *et al.* [1, 2, 7] include pose-graph optimization (PGO) to use geometric data across different layers for better localization and mapping.

However, the existing approaches for online room segmentation that utilize the Euclidean Signed Distance Field (ESDF) map [9] often produce results lacking an understanding of the interaction between the robot and its surroundings. Additionally, when entering a new space and segmenting spatial characteristics, consistently recognizing rooms becomes challenging due to voxel-based free space detection method and the time dependency of room segmentation method. For example, while some real-time approaches over-segment large rooms into smaller, non-functional spaces that are not useful for localization and mapping, others tend to under-segment in more complex cases, such as open spaces that are non-traversable for ground robots or humans but estimated as free space because the 3D space is not fully enclosed. This under-segmentation issue arises from the use of a voxel-based method, leading to false constraints in PGO.

To address these limitations, we propose a traversability-aware room segmentation method that aligns with conditions such as the field of view (FOV), making room segmentation using traversability information consistently feasible. This method demonstrates the following contributions:

- Our proposed traversability-aware free-space clustering method, which considers the traversability of the terrain, improves the consistency of room segmentation.
- Consistent room segmentation by adding lower-layer data flushing rule, not using timer dependency.
- The effectiveness of our method is shown by evaluating repetitive operations along the same path, specifically focusing on improvements in room re-detection rates and reductions in PGO processing time.

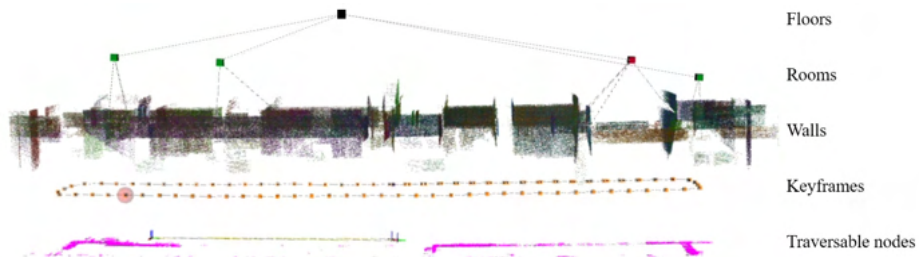


Fig. 1. Our proposed hierarchical graph layers. From the bottom, the hierarchical graph consists of traversable nodes, keyframes, walls, rooms, and floors layers.

2 Related Works

2.1 SLAM and Scene Graphs

Many recent SLAM methods use geometric features from point clouds collected by LiDAR or depth camera [17, 11], which show robustness to complex environments and notable SLAM performance. These methods, however, are unable to handle tasks such as utilizing maps without semantic information or performing data association, such as loop closure, as they require processing entire point sets. Accordingly, semantic SLAM which assigns meanings to terrain-level or object-level points and enhances human-robot interactions is getting interests [10, 3]. Still, challenges remain because the semantic segmentation accuracy suffers from sensor noise or localization error. Recent efforts aim to combine the strengths of geometric and semantic approaches by leveraging the geometric relationships between semantic elements [18].

The concept of a scene graph for SLAM has emerged to capitalize on structural constraints across different layers, using a hierarchical approach. Studies like [4, 14] employ visual sensors for object segmentation and map semantic labels, such as assigning a dining table to a kitchen, enhancing the level of detail in the graph. On the other hand, [1, 2, 7] only use point clouds to create layered scene graphs representing walls, rooms, and floors, with a foundational layer of keyframes. This hierarchical graph optimization results in more precise robot trajectories and better geometric reconstruction of indoor spaces. However, they were only tested on small and simple-shaped rooms, while they actually failed in larger or complex spaces like long corridors. Especially, in the most progressed work among them, Bavle, H. *et al.* [7] introduces a Graph Neural Network (GNN) based approach for efficient semantic segmentation in SLAM systems, inferring relationships between wall surfaces to construct enriched graphs.

2.2 Room Segmentation

Accurate room segmentation is crucial because the room triggers the creation of new branch of the graph which affects the PGO process. In the approach described by S-graphs+ [2], room segmentation is conducted in two main steps: ESDF map-based free-space clustering and room extraction using wall data. An ESDF map calculates the shortest distance from a free-space voxel to the nearest obstacle, typically a wall. By setting a distance threshold, edges and vertices close to walls or doorways are removed, allowing the graph to break apart near these boundaries, which helps in distinguishing between different rooms. The wall data from lower-layer and the free-space clusters are stacked for the fixed time intervals, and then the wall set lying near the clusters is chosen to define the rooms.

Furthermore, GNN-based approach [7] used a proximity graph and inferred relationships between walls for efficient room detection. The method identifies room structures quickly and improves detection time over S-graphs+.

3 Traversability-aware Room Segmentation

The overall architecture of our method is illustrated in Fig. 2. The situational graph layers are similar with *S-graphs+* [2], but with additional traversable nodes layer and the strategy. We enhance room segmentation by implementing a traversability-aware strategy, which ensures the extraction of meaningful and consistent rooms. Here, meaningful room indicates traversable room for ground robots and humans, including interaction between them and terrain.

This method consists of three steps: extracting traversable ground nodes, clustering these nodes, and extracting rooms by utilizing the clustered traversable spaces.

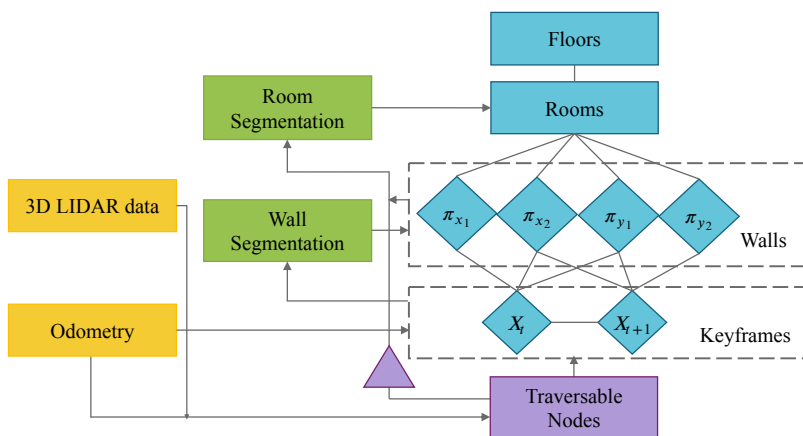


Fig. 2. Overview of our method. Inputs are the 3D LiDAR measurements, which is pre-filtered while calculating traversability, and robot odometry. Front-end is to extract wall planes, rooms, floors and traversable nodes. and the graphs are jointly optimized in the back-end. We added traversability-aware room segmentation strategy for consistency.

3.1 Traversable Ground Points Extraction

In robotics, the demand for accurate detecting and representing the surrounding terrain is growing, especially for unmanned ground vehicles. Research in this area has primarily focused on identifying drivable regions and object identification [16, 15], odometry estimation [13], and global localization techniques [6].

We employ the ground segmentation method described in [8], which identifies likely ground points based on their minimal height. To determine traversable areas, [8] calculates the local convexity and vertical alignment of each region and incorporates a Bayesian generalized kernel (BGK) to consider the surrounding points together. Additionally, we added an outlier points rejection step for better

initial ground extraction and global update process to ensure global consistency in the map representation. Thereby, we obtained global consistent traversable terrain nodes.

3.2 Free-space Clustering

ESDF map-based free-space causes under-segmentation of room in complex indoor environments, such as corridors with central hollows as illustrated in Fig. 3, which appears frequently in many types of buildings. Segmentation inaccurately spans from wall to opposite one, as shown in Fig. 4(a), instead of more logical boundaries such as from wall to handrail. Also, another baseline Bavle, H. *et al.* [7] demonstrating faster and better room segmentation, showed the same limitation as the network was trained not considering the interaction between the robot and the surroundings.

In this paper, we have developed a method that utilizes traversable ground points for free-space clustering. This process begins by constructing a node graph of ground region centers connected by edges, then selectively disconnected based on their proximity to occupied point clouds, with a threshold set at λ_{th} . Several clusters are extracted near the obstacles or doorways as a result, thereby we only consider a cluster that includes the robot’s current position to avoid including areas outside doorways.

3.3 Room Extraction Strategy

Furthermore, while *S-graphs+* traditionally extracts rooms at every set time interval, which is effective in small or simple environments, it struggles with larger spaces or extended corridors. In such scenarios, room definitions can become unclear, and inconsistent room identifications can occur at different times. To address the challenges in room segmentation, we introduce a novel strategy that utilizes changes in aisle width or direction as indicators. This method involves flushing the lower-layer data that are typically accumulated for room extraction. For instance, if a large bookshelf limits the examination of an entire room, the space between the wall and the bookshelf should be considered a distinct room. Doing so, it enhances the relevance and accuracy of room segmentation, ensuring that each defined room is meaningful and directly related to the navigational and operational context of the robot.

4 Experimental Results

In this section, we evaluate our proposed traversability-aware approach in terms of efficiency and consistency, comparing it with *S-graphs+* [2] and Bavle, H. *et al.* [7]. We utilize the TIERS dataset [12], which includes a scenario where the starting point is revisited at the end. The odometry was calculated with scan-matching based SLAM algorithm [5]. The experiment was run repetitively to

demonstrate the efficiency and consistency of our approach. We conducted the experiment 10 times and averaged the results.

This experimental setup not only highlights the consistency and reliability of our approach but also demonstrates significant improvements in room segmentation and SLAM efficiency, achieving the objectives outlined in our methodological contributions.

4.1 Consistent Room Segmentation

As shown in Fig. 3, it is an open space in three dimensions, so it is shared spatially, but it is actually a different layer and a different area, so it should be distinguished. However, in the case of existing baselines, when faced with that environment, spatial under-segmentation occurs, as shown in Fig. 4(a).

In contrast, our method accounts for both robot and human traversability, resulting in reasonable free-space clustering as depicted in Fig. 4(d-f), from the wall to the handrail.

The consistency of scene graphs was evaluated using the following metrics; frequency of room re-detection, and geometric precision metrics. First, we measured the ability to extract rooms and re-detect the same rooms in successive iterations. The re-detection ability was quantified by the ratio of the number of re-detected rooms to the number of rooms initially extracted in the first iteration, aiming for a high re-detection rate to confirm consistency.

Second, we assessed the precision of our segmentation with two further measures: Dice Coefficient Score (DCS) and distance of room center for the most overlapped rooms from the first and second traverse. The Dice coefficient measures the similarity between two sets. For two square-shaped rooms A and B , it is defined as:

$$DCS(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

Here, $A \cap B$ is the intersection area, while $|A|$ and $|B|$ are the areas of room A and B respectively. DCS indicates strong overlap between the rooms segmented from the initial and second traverse in this paper, and the distance of room centers reflects minimal spatial deviation between the positions of the most overlapped room pairs. The average result values of the 10 iterative evaluations are shown in Table 1.

The baseline methods may produce slightly different predictions depending on how message-passing evolves across layers, and learned embeddings especially for GNN-based method [7]. The variability in these predictions could also contribute to inconsistent room segmentation results. Whereas they extract over 10 rooms in the first traverse and re-detect only less than half among them, our method extracts a smaller number of rooms but is meaningful with traversability information, so the re-detection frequency was significantly high.

Also, higher DCS and smaller room center distance than S-graphs+ demonstrated consistent room segmentation in our method. In the case of [7], room

width was not provided in the room information but only room center position, so the DCS calculation requires an additional process. However, we considered it unnecessary to calculate the DCS for this method due to the inconsistency shown with f_{re} .

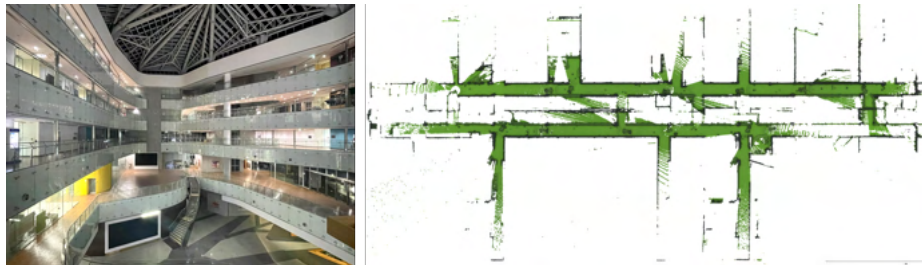


Fig. 3. (L-R) Examples of complex corridor environments: (L) An open-corridor structure that shares 3D free space, where ground robots or humans cannot traverse. (R) Bird's-eye view of the global point cloud map of the 3rd floor data from the Indoor sequence of the TIERS dataset, which traverses through open-corridor environments [12].

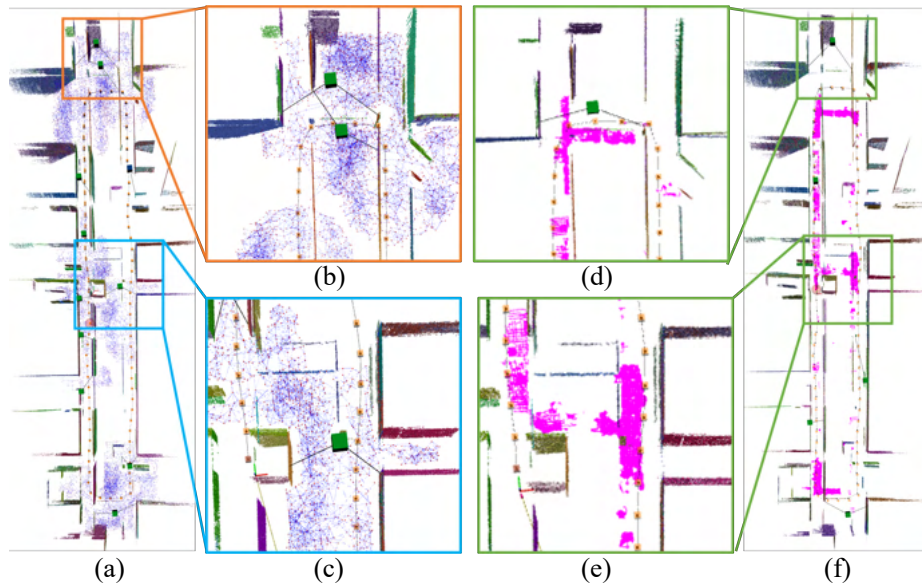


Fig. 4. Qualitative results on the TIERS dataset [12]: (a-c) The results of *S-graphs+* [2] demonstrate (b) over-segmentation and (c) under-segmentation in complex corridor environments due to the use of ESDF-based 3D free-space clustering (blue points). (d-f) The results of our proposed method show that by leveraging traversable nodes (magenta points) in free-space clustering, room segmentation is consistently maintained.

4.2 Efficient and Robust Pose-graph Optimization

To demonstrate that the proposed method improves not only the pose estimate performance but also the efficiency, we evaluate the absolute trajectory error (ATE) of the distance between the start and end point position on *Indoor* sequence of the Tiers dataset.

The ATE of the distance between the start and end point position was used to evaluate the navigational accuracy of SLAM. As the dataset returns to the start point at the end, the ground truth is zero, and the smaller value would indicate better performance. We measured and compared the time efficiency of the hierarchical PGO process across the data to show the enhancements in SLAM operational speed facilitated by our method. We tracked each time consumption spent for PGO performed while the same rooms were detected and calculated the average value and overall summation during the whole data time.

Table 1. Quantitative comparison for room segmentation consistency and pose-graph optimization efficiency. Room segmentation consistency was compared in terms of the frequency of room re-detection (f_{re}) of two iterations in TIERS LiDAR dataset, calculated by N_{re} over N^{1st} , where N^{1st} and N^{2nd} indicate the numbers of rooms extracted from the first and second iteration and N_{re} indicates the number of rooms detected again that are the same as the rooms in the first iteration among the rooms in the second iteration. Also, room consistency was compared with the similarity between the pair of rooms segmented from two successive traverses, using Dice Coefficient Score (DCS) and distance of room center (d_{center}) measured in $[m]$. Also, pose-graph optimization efficiency was compared with ATE measured in $[m]$ and PGO computation time during the whole process (T_{PGO}) measured in $[s]$.

| Metrics | Room Segmentation Consistency | | | | | | PGO Efficiency | |
|-----------------------------|-------------------------------|-----------|----------|-------------------|----------------|-------------------------|------------------|----------------------|
| | N^{1st} | N^{2nd} | N_{re} | $f_{re} \uparrow$ | DCS \uparrow | $d_{center} \downarrow$ | ATE \downarrow | $T_{PGO} \downarrow$ |
| S-graphs+ [2] | 10.5 | 7.6 | 5.3 | 0.50 | 0.61 | 0.97 | 0.24 | 1.68 |
| Bavle, H. <i>et al.</i> [7] | 9.25 | 7.5 | 1.5 | 0.16 | - | 3.67 | 0.48 | 2.66 |
| Ours | 4.7 | 4.5 | 4.0 | 0.85 | 0.74 | 0.57 | 0.22 | 1.62 |

Table 1 tells GNN-based method [7] brings even lower re-detection rate of the rooms than non-learning based method, as a result strongly depends on the prior learned model and not robust to various situations. This over-segmentation leads to high ATE and increased PGO time consumption.

Our method showed the most promising performance in ATE and PGO time consumption compared to *S-graphs+*. *S-graphs+* tends to extract the duplicated room in successive timings, giving ambiguity and additional useless edges to the situational graphs, which brings a negative effect on the PGO and also takes more time. Our method addressed this by extracting room with one characteristic just once, improving efficiency of PGO.

5 Conclusion

In conclusion, our proposed traversability-aware strategy substantially improves both the semantic meaningfulness including robot-terrain interaction information, and consistency of the situational graphs. The results demonstrate that our method achieves a higher frequency of room re-detection and improved DCS compared to *S-graphs+* and [7]. This suggests a more reliable and consistent construction of situational graphs, and accordingly, this brings accurate SLAM performance and reduced time consumption for PGO. These results confirm the effectiveness of our methodology in advancing scene graph-based SLAM systems.

For future work, we plan to explore how free-space clustering with an ESDF map and traversable nodes may vary depending on sensor orientation and height. We also aim to assess segmentation consistency across different robot platforms to offer cooperative scene graph construction in multi-platform environments.

References

1. Bavle, H., Sanchez-Lopez, J.L., Shaheer, M., Civera, J., Voos, H.: Situational graphs for robot navigation in structured indoor environments. *IEEE Robotics and Automation Letters* **7**(4), 9107–9114 (2022)
2. Bavle, H., Sanchez-Lopez, J.L., Shaheer, M., Civera, J., Voos, H.: S-graphs+: Real-time localization and mapping leveraging hierarchical representations. *IEEE Robotics and Automation Letters* **8**(8), 4927–4934 (2023)
3. Chen, X., Milioto, A., Palazzolo, E., Giguere, P., Behley, J., Stachniss, C.: SUMA++: Efficient LiDAR-based semantic SLAM. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2019), 4530–4537
4. Hughes, N., Chang, Y., Carlone, L.: Hydra: A real-time spatial perception system for 3D scene graph construction and optimization. *arXiv preprint arXiv:2201.13360* (2022)
5. Koide, K., Miura, J., Menegatti, E.: A portable three-dimensional LiDAR-based system for long-term and wide-area people behavior measurement. *International Journal of Advanced Robotic Systems* **16**(2) (2019)
6. Lim, H., Kim, B., Kim, D., Mason Lee, E., Myung, H.: Quatro++: Robust global registration exploiting ground segmentation for loop closing in LiDAR SLAM. *The International Journal of Robotics Research* **43**(5), 685–715 (2024)
7. Millan-Romera, J.A., Bavle, H., Shaheer, M., Oswald, M.R., Voos, H., Sanchez-Lopez, J.L.: Better situational graphs by inferring high-level semantic-relational concepts. *arXiv preprint arXiv:2310.00401* (2023)
8. Oh, M., Jung, E., Lim, H., Song, W., Hu, S., Lee, E.M., Park, J., Kim, J., Lee, J., Myung, H.: TRAVEL: Traversable ground and above-ground object segmentation using graph representation of 3D LiDAR scans. *IEEE Robotics and Automation Letters (RA-L)* **7**(3), 7255–7262 (2022)
9. Oleynikova, H., Taylor, Z., Fehr, M., Siegwart, R., Nieto, J.: Voxblox: Incremental 3D Euclidean signed distance fields for on-board MAV planning. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2017), 1366–1373
10. Shan, T., Englot, B.: LeGO-LOAM: Lightweight and ground-optimized LiDAR odometry and mapping on variable terrain. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2018), 4758–4765
11. Shan, T., Englot, B.J., Meyers, D., Wang, W., Ratti, C., Rus, D.: LIO-SAM: tightly-coupled LiDAR inertial odometry via smoothing and mapping. *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 5135–5142 (2020)

12. Sier, H., Li, Q., Yu, X., Peña Queraltá, J., Zou, Z., Westerlund, T.: A benchmark for multi-modal LiDAR SLAM with ground truth in GNSS-denied environments. *Remote Sensing* **15**(13), 3314 (2023)
13. Song, S., Yu, B., Oh, M., Myung, H.: BIG-STEP: Better-initialized state estimator for legged robots with fast and robust ground segmentation. In: *Proc. International Conference on Control, Automation Systems (ICCAS)* (2023), 184–188
14. Werby, A., Huang, C., Büchner, M., Valada, A., Burgard, W.: Hierarchical open-vocabulary 3D scene graphs for language-grounded robot navigation. In: *Proc. First Workshop on Vision-Language Models for Navigation and Manipulation at ICRA 2024* (2024)
15. Xue, H., Fu, H., Ren, R., Zhang, J., Liu, B., Fan, Y., Dai, B.: LiDAR-based drivable region detection for autonomous driving. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2021), 1110–1116
16. Zermas, D., Izzat, I., Papanikolopoulos, N.: Fast segmentation of 3D point clouds: A paradigm on LiDAR data for autonomous vehicle applications. In: *Proc. IEEE International Conference on Robotics and Automation (ICRA)* (2017), 5067–5073
17. Zhang, J., Singh, S.: LOAM: LiDAR odometry and mapping in real-time. In: *Proc. Robotics: Science and Systems (RSS)*. Volume 2, Berkeley, CA (2014), 1–9
18. Zhou, L., Wang, S., Kaess, M.: π -lsam: LiDAR smoothing and mapping with planes. In: *Proc. IEEE International Conference on Robotics and Automation (ICRA)* (2021), 5751–5757