# San Francisco World: Leveraging Structural Regularities of Slope for 3-DoF Visual Compass (Supplementary Material)

Jungil Ham[1], Minji Kim[1], Suyoung Kang[2], Kyungdon Joo[3], Haoang Li[4], and Pyojin Kim[1,†]

*Abstract*— **In this supplementary material, we provide detailed explanations on the utilization of the depth-based dominant plane tracking method and line segment detection, along with an illustrated analysis of the geometrical patterns of sloping line normals on the Gaussian sphere. Additionally, we present experimental results conducted in more large and diverse indoor and outdoor urban environments. These experiments demonstrate that the proposed SLOPe method enables 3D inter-floor navigation in urban settings, surpassing the limitations of 2D intra-floor navigation commonly associated with various robotics platforms, by leveraging consistent and repetitive slopes in large-scale environments. We further evaluate the robustness and sensitivity of the SLOPe method to noise using synthetic line data. Implementation details and runtime analyses of the key components are also reported. Finally, we provide a performance comparison between the proposed SLOPe method and existing approaches, including LIMAP with depth information and the mixture of Manhattan frames-based SLAM.**

## I. Detail on depth-based dominant plane tracking and LSD threshold.

When the tracking of the dominant plane sometimes fails, our *SLOPe* method finds and re-initializes a new dominant plane by analyzing the density distribution of the surface normal vectors from the depth camera (see Fig. 1), as proposed in our previous work [1]. In more detail, if the density distribution of the surface normal vectors around the currently tracked normal vector is too low, we re-initialize and detect a new dominant plane again with plane model-based RANSAC. We assign the normal vector of the new dominant plane to the closest axis in the SFW under the assumption that the SFW does not change too much between subsequent frames. Through the association with the previously tracked SFW frames, we can continuously and stably track the absolute 3-DoF orientation of the camera.

Since we have the SFW model (a kind of 3D orientation map for space) from the SFW detection step, our proposed method does not experience sudden and unexpected jumps in the prediction of dominant directions and of the 3D compass. Our *SLOPe* method also ensures smooth 3-DoF rotational motion tracking, preventing any potential jumps that can occur through the Kalman filter.

For the line segment detection, the horizontal and vertical lines that are visible to humans are sometimes not detected well by LSD [2] as shown in Fig. 2 because either the lines are too short or due to a uniform pattern and homogeneous features around the steps of the staircases. While adjusting the minimum line length threshold of LSD in Fig  2, we can detect more horizontal and vertical lines. However, short lines do not contribute to identifying structural patterns within the image, which is why we consistently set the LSD length threshold to 300.

## II. Geometrical Patterns of Sloping Line Normals on the Gaussian Sphere

Let us assume that we have non-Manhattan frame lines after the non-Manhattan frame lines filtering. We aggregate all the sloping line normals that follow distinct SDDs to maximize the number of inliers. We leverage the following two properties of SFW for the aggregation. For simplicity, we express the VDD, PHD, AHD, and the four SDDs in mathematical symbols as $v$, $h_p$, $h_a$, and $s_n$ as illustrated in Fig. 3(a).

***Symmetricity.*** In Fig. 3(c), the dominant planes of $s_1$ and $s_2$ are symmetric with respect to the $h_a$ and the $v$. The $s_3$ and $s_4$ dominant planes are symmetric with respect to the primary horizontal direction $h_p$ and $v$ in the same sense.

***Quarter-Turn Relation.*** In Fig. 3(d), the $s_1$ and $s_3$ dominant planes, and the $s_2$ and $s_4$ dominant planes, are related by a 90-degree rotation around $v$, respectively. By fully exploiting the geometric properties of SFW and aggregating sloping line normals onto a single dominant plane, we overcome the sparsity of line features, achieving effective and robust SFW detection.

## III. Demonstration of More Experiments in Large Indoor and Outdoor Environments, and Some Failure Cases

The ultimate goal of the proposed *SLOPe* method is to enable **3D inter-floor navigation** in urban areas rather than being limited to **2D intra-floor navigation** of various robotics platforms by effectively utilizing consistent and repetitive slopes in indoor/outdoor environments where slopes exist.

[1]School of Mechanical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea. {jungilham,minji0110,pjinkim}@gist.ac.kr

[2]Mannig College of Information and Computer Science, University of Massachusetts Amherst, Amherst, MA 01002, USA. {suyoungkang@umass.edu

[3]Artificial Intelligence Graduate School and the Department of Computer Science and Engineering, UNIST, Ulsan 44919, South Korea. kdjoo369@gmail.com, kyungdon@unist.ac.kr

[4]Thrust of Robotics and Autonomous Systems and the Thrust of Intelligent Transportation, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511458, China. {haoangli@hkust-gz.edu.cn


Fig. 1. **Dominant plane detection and tracking**. Clustered lines and tracked dominant planes are overlaid on the RGB images for each sequence, Half-Turn Stair and Quarter-Turn Stair, respectively. Our *SLOPe* method can adaptively change the dominant plane being tracked depending on the current situation.

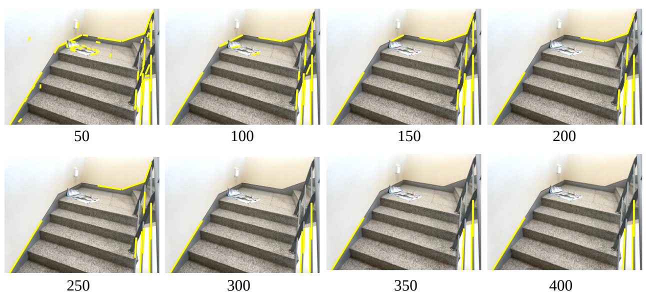



Fig. 2. **Comparison results for each LSD length threshold**. The number below indicates a minimum line length threshold used to filter line segments based on their squared line length. We have set this threshold to 300, and while lowering it results in more lines being detected, shorter lines provide less robust directional cues and are not useful at all.

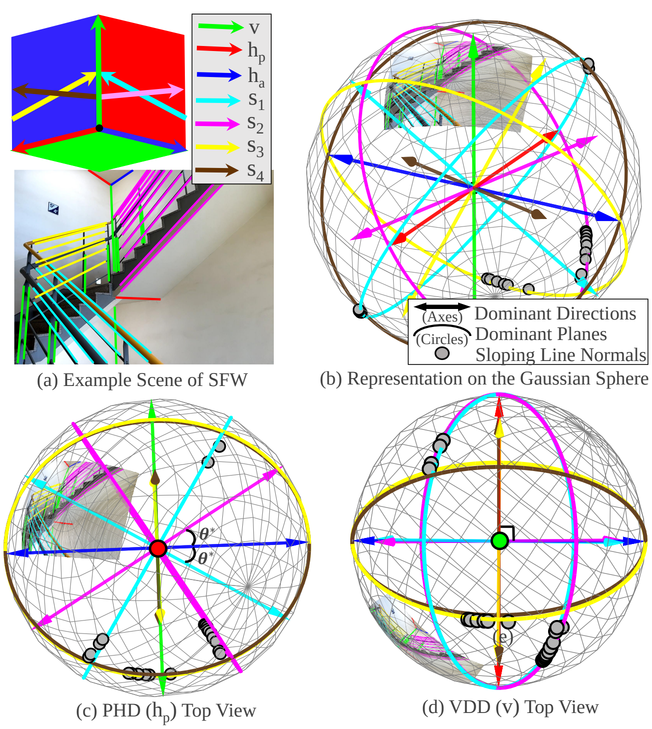



Fig. 3. **Geometrical patterns of sloping line normals**. Sloping line normals are aligned with four distinct sloping dominant planes, represented as great circles on the Gaussian sphere. The colors of the sloping dominant planes in (b),(c), and (d) indicate the respective color of vanishing direction for the sloping lines illustrated in (a).

TABLE I

ABSOLUTE ROTATION ERROR ON ADDITIONAL EXPERIMENTS

| **Sequence** | **ARE** |
|---|---|
| (a) Parking Building | 1.62° |
| (b) Indoor GIST S1 Building | 1.79° |
| (c) Outdoor Fire Escape Staircase | 2.09° |
| (d) Outdoor Pedestrian Bridge | 2.32° |

The proposed San Francisco world (SFW) is an optimal structure model, especially when consistent slopes and ramps are repeatedly observed, which are easy to find around us. In most ordinary buildings, the slope angle of stairs and ramps is quite regular and consistent. Our proposed *SLOPe* aims to improve the accuracy of positioning and rotational motion tracking in VO/SLAM by further simplifying it over Hong Kong world (HKW) [3].

In particular, the staircase environment within a building is very tricky, and it is challenging to apply SLAM skills like loop closure because similar images are continuously repeated. Staircases exist in all buildings, and they are spaces that must be passed through to move between floors. Without the help of loop closure in VO/SLAM in these staircase environments, position and rotation drift errors will accumulate over time [4], ultimately leading to overall positioning failure. Therefore, for various robotics platforms such as quadruped robots [5] and nano drones to freely navigate and explore between floors, it is essential to utilize the slope angles such as stairs and ramps, and we try to tackle this point with the proposed SFW and *SLOPe* method.

To demonstrate the generality and usefulness of our SFW model and the *SLOPe* method, we have performed additional experiments on more diverse and larger indoor/outdoor environments where slopes exist, satisfying the proposed SFW model as shown in Figs. 4 and 5. We utilize the same iPhone device and custom iOS app described in the manuscript to acquire various kinds of data, such as synchronized RGB/depth image sequences and Apple ARKit (VIO) camera poses. Since the accuracy and stability of Apple ARKit (VIO) are very high in a short period of time [6], we consider the 3-DoF rotational motion from Apple ARKit as the ground-truth, and evaluate our method quantitatively.

Our additional test scenes shown in Fig. 4 include various indoor/outdoor environments with slopes that we can easily find around us, such as (a) a parking building, (b) an indoor

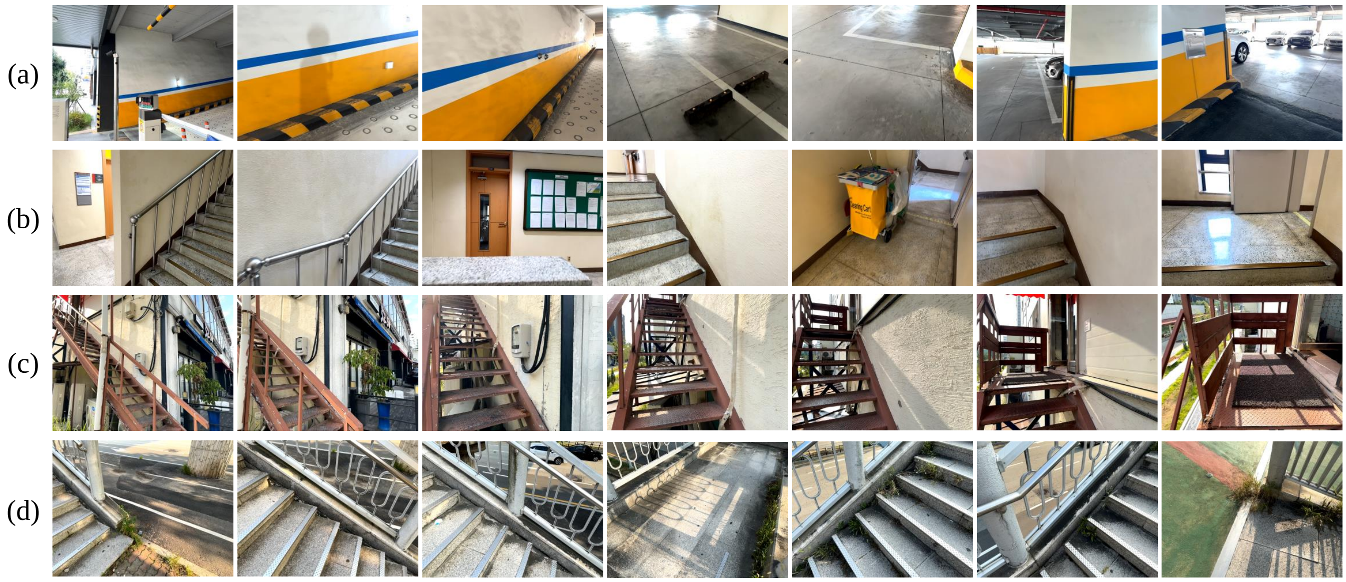



Fig. 4. **Various indoor and outdoor scenes with slopes around us**. Diverse indoor/outdoor environments with slopes that meet the proposed SFW model: (a) a parking building, (b) an indoor GIST S1 building, (c) an outdoor fire escape staircase, and (d) an outdoor pedestrian bridge. They involve complex camera translations, rotations, and numerous outlier line features.

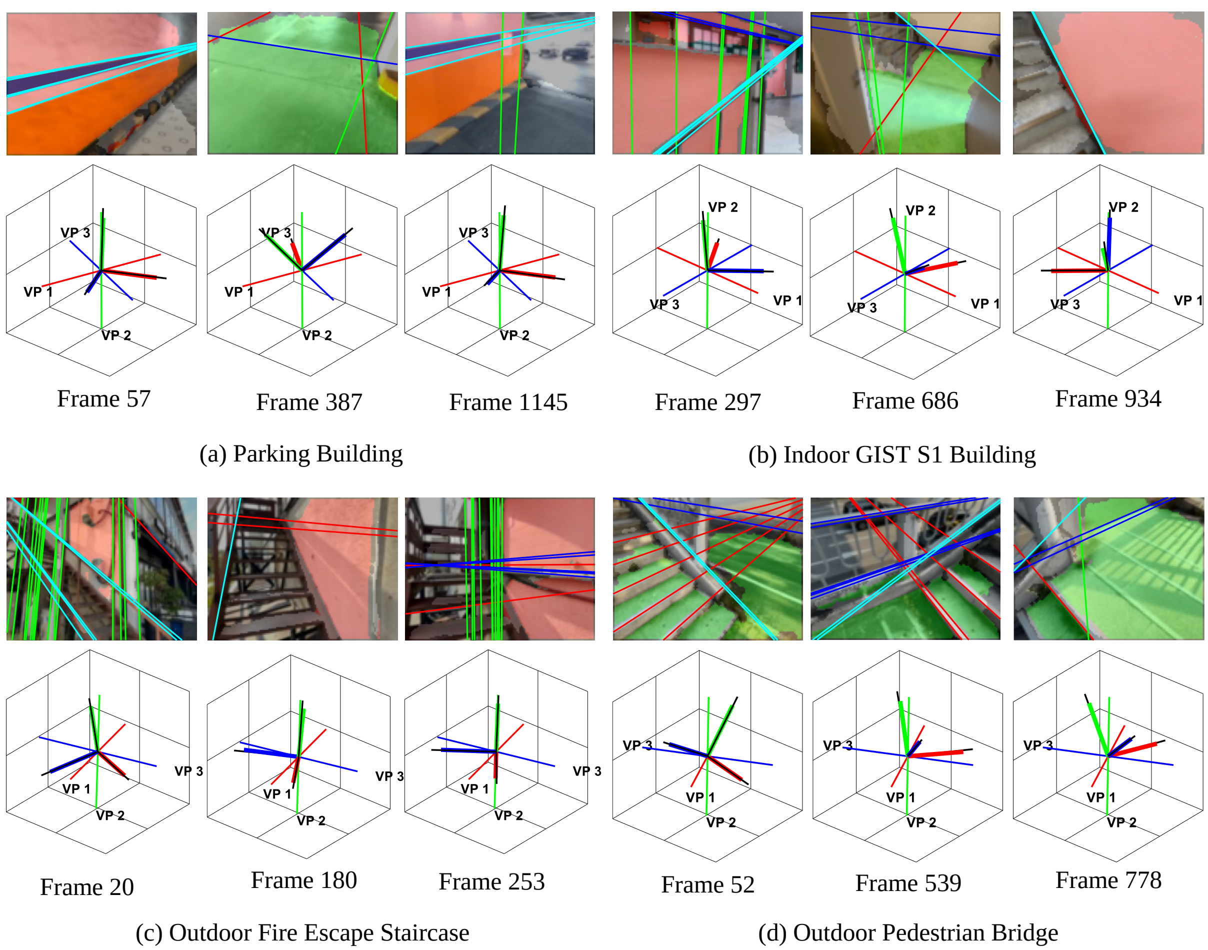



Fig. 5. **Our *SLOPe* results on additional experiments**. Clustered lines and tracked dominant plane with inferred SFW frame are overlaid on the RGB images (top). Colored thick and thin lines denote the estimated 3-DoF camera orientation and the MW (VPs), and the black lines represent the true pose of the camera from Apple ARKit (bottom).

GIST S1 building, (c) an outdoor fire escape staircase, and (d) an outdoor pedestrian bridge, involving complex camera translations and rotations, as well as numerous outlier line features. The proposed *SLOPe* method effectively leverages such consistent and repetitive slope information to achieve drift-free 3-DoF rotational motion tracking as shown in Fig. 5. Our *SLOPe* method can accurately track the 3-DoF rotational motion not only on the stairs but also on the parking lot ramp and the corridors between stairs. Table I shows the quantitative results of the absolute rotation error (ARE), which is about two degrees on average. Please refer to our project page at https://SanFranciscoWorld.github.io/ for supplementary video clips of each sequence in the additional experiments.

Although the proposed method performs highly robust rotational motion tracking in most urban areas with slopes, it does not always succeed in every environment. We present some failure cases in Fig. 6. For example, our method fails when no lines align with the initialized SFW frame or when there is a sizeable angular difference between the sloping lines.

We have also plotted boxplots of all the sequences in the GIST-SFW dataset in Fig. 7, showing similar trends to Fig. 13 (b) in the manuscript. Note that the TAMU dataset [7] does not provide a variety of SFW scenes and true 3-DoF rotational motion, making it unsuitable for precise rotational motion tracking evaluations. Recognizing the lack of appropriate datasets for evaluating VO/SLAM algorithms, we create and propose our own GIST-SFW dataset specifically designed for these purposes.

## IV. Study about sensitivity and robustness to noise for our *SLOPe* Method.

We have performed additional experiments to evaluate the sensitivity and robustness of the proposed method against noise with synthetic line data, especially for lines following the sloping directions. We synthesize several 3D lines aligned to the SFW and project them on the virtual image plane to generate lines satisfying the SFW model (red, green, blue, and magenta) as shown in Fig. 8. We perturb the endpoints of the lines following the sloping directions by a zero-mean Gaussian noise. Then, we synthesize the noisy lines (magenta) by randomizing their endpoints within the image, as shown in Fig. 8.

Figs. 9 and 10 show that the proposed SFW detection and tracking methods have an error of less than one degree even when a variance of about five degrees occurs.

## V. Implementation Details and Runtime Analysis of Key Components of the Proposed Method

We have implemented and tested the proposed *SLOPe* method in MATLAB R2023b on a desktop computer with an Intel Core i5-12400F (2.50 GHz) and 32 GB memory. It also has a graphic card with NVIDIA GeForce RTX 3060 to run only deep learning-based methods such as DROID-SLAM and LIMAP (DeepLSD [8]). No special MATLAB code optimization or parallelization was done.

TABLE II

Runtime Analysis of Proposed *SLOPe*

| Module | Runtime |
|---|---|
| Preprocessing (Line Detection) | 49.69 ms |
| Surface Normal & Mean Shift | 10.02 ms |
| Initial MW Detection | 6.67 ms |
| SFW Detection | 1.51 ms |
| SFW Tracking | 1.69 ms |

TABLE III

LIMAP ARE Results with/without Depth Map

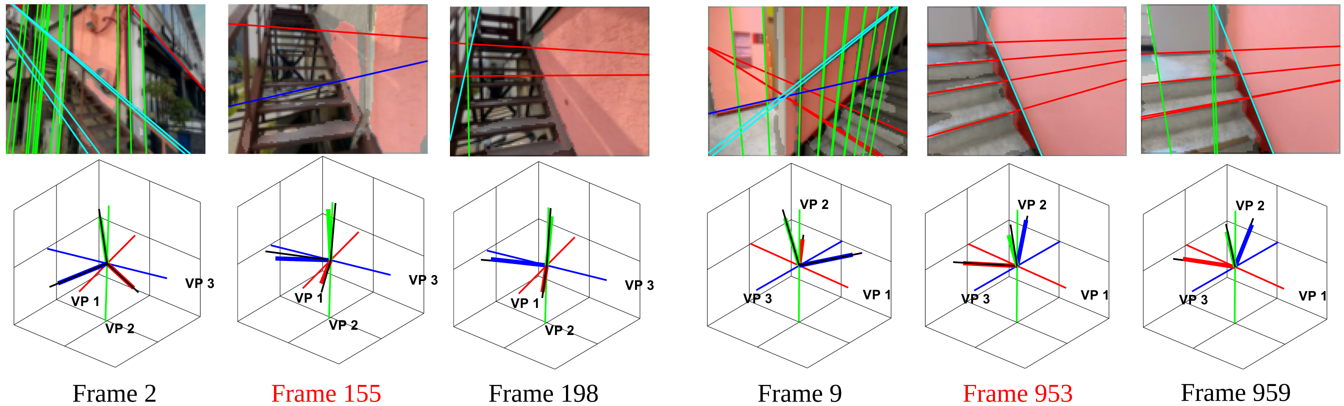
| Sequence | *SLOPe* | LIMAP w/o Depth | LIMAP w/ Depth |
|---|---|---|---|
| Half-Turn Stair 180 | 0.68 | 1.00 | 1.06 |
| Half-Turn Stair 360 | 1.19 | 1.92 | 1.38 |
| Quarter-Turn Stair 180 | 0.96 | 1.02 | 1.12 |
| Quarter-Turn Stair 360 | 1.21 | 1.79 | 1.41 |

We have analyzed the runtime of each module that constitutes the proposed method as shown in Table II. First, the image processing, such as line detection with LSD [2], takes the longest runtime in the proposed method, about $\sim 50$ ms for the 15 image lines. Note that the computational load of image processing such as line detection (LSD [2]) can vary significantly depending on various conditions such as input image size, the length of the detected line, etc. Then, we compute surface normals from the depth images and track the dominant plane, taking $\sim 10$ ms. Initial MW detection, finding the SFW model, and tracking the corresponding SFW frames take about $\sim 7, 2, 2$ ms, respectively. Overall, the total computation time of the proposed *SLOPe* method, excluding image processing such as line detection, is about $\sim 20$ ms per image frame, suggesting a computational improvement of the proposed method when implemented in C/C++.

## VI. Performance Comparison with LIMAP with depth information and mixture of Manhattan frames-based SLAM.

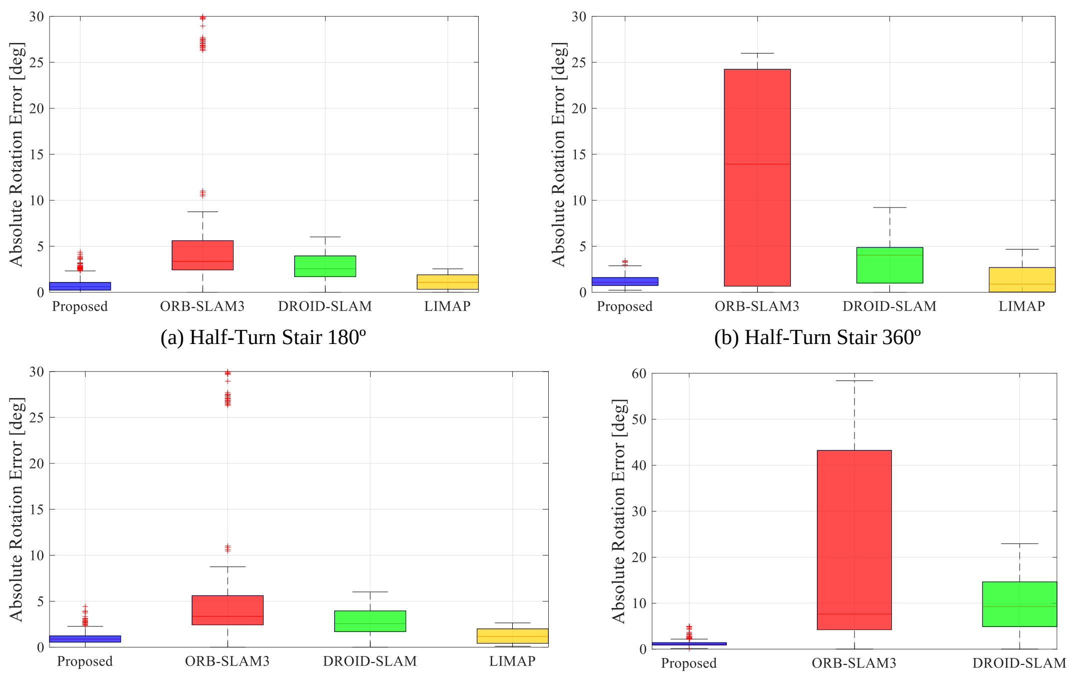
Table III shows the absolute rotation error (ARE) of the proposed *SLOPe* and the LIMAP depending on whether depth is used or not. Although the use of depth in LIMAP slightly improves performance by about 0.2 degrees on average, the proposed *SLOPe* method is still the most accurate. Note that LIMAP with depth maps sometimes fails when the depth map is sparse due to the limited range of RGB-D camera or when line feature matching fails as shown in Fig. 11. In these cases, we exclude such RGB/depth image frames and rerun LIMAP. On the other hand, the proposed method does not fail because we extract lines from RGB images and track the dominant plane from the depth maps complementarily.

We also have compared our *SLOPe* method to the ManhattanSLAM [9], which is a point, line, and plane-based SLAM technique leveraging a mixture of Manhattan frames, and the results are shown in Fig. 12. ManhattanSLAM is designed based on the ORB-SLAM [10], with additional detection and tracking of the Mixture of Manhattan Frames




(a) Outdoor Fire Escape Staircase (b) Indoor GIST S1 Building

Fig. 6. **Examples of some failure cases**. The frames marked in red indicate those with an error greater than five degrees. In (a) at frame 155, the red line contains redundant information for the same direction as PNV and is therefore useless. This results in the use of a blue line unrelated to the SFW frame, leading to a large error. In (b), the sloping angle at frame 9 differs from that at frame 953 by more than three degrees, resulting in a large error. In both cases, however, we can observe that the correct absolute camera orientation is tracked again in subsequent frames.




(a) Half-Turn Stair 180° (b) Half-Turn Stair 360°

(c) Quarter-Turn Stair 180° (d) In-Place Rotation

Fig. 7. **Boxplots of all the sequences in the GIST-SFW dataset**. Comparison of the proposed *SLOPe* versus other rotational tracking methods. The statistical distribution of the absolute rotation error (ARE) from other sequences in our own GIST-SFW dataset.

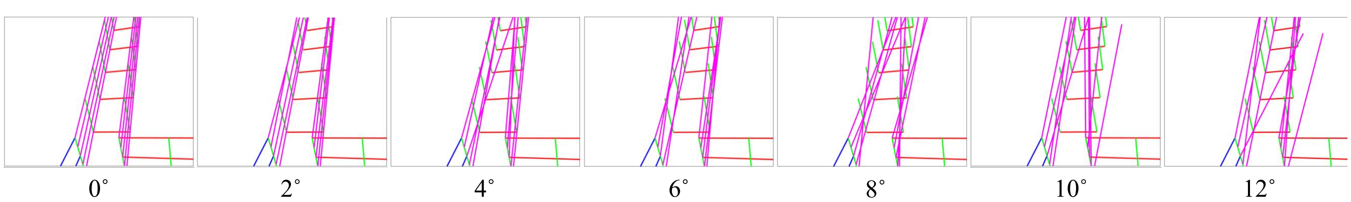
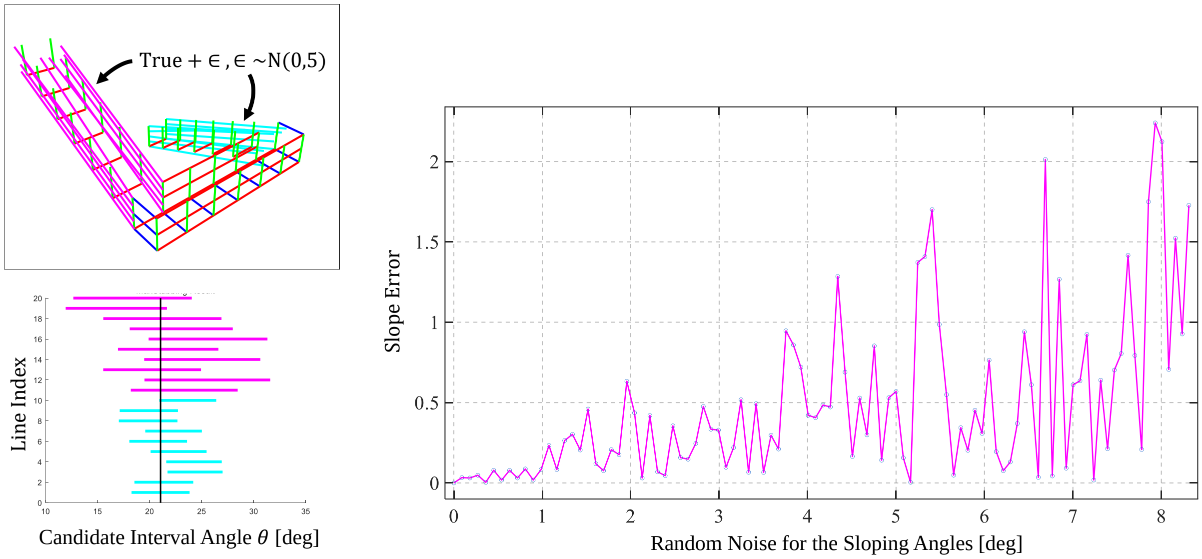



Fig. 8. **Synthetic line dataset with added random noise**. Synthetic line data for robustness analysis of our SFW detection and rotation tracking methods to noise. We incrementally add random noise to lines (magenta) following the dominant sloping direction. The numbers below images are the standard deviation of the added random noise from 0 degrees (true) to 12 degrees (very noisy).




(a) (b)

Fig. 9. **Study about sensitivity and robustness to noise for our SFW detection**. (a) shows the synthetic line data (top) when the standard deviation of the random noise applied to the sloping lines is about five degrees and the estimated sloping parameters found by the proposed MnS method (bottom). (b) illustrates the tendency of sloping parameter estimation error according to the added random noise from 0 degrees (true) to 8 degrees (very noisy).

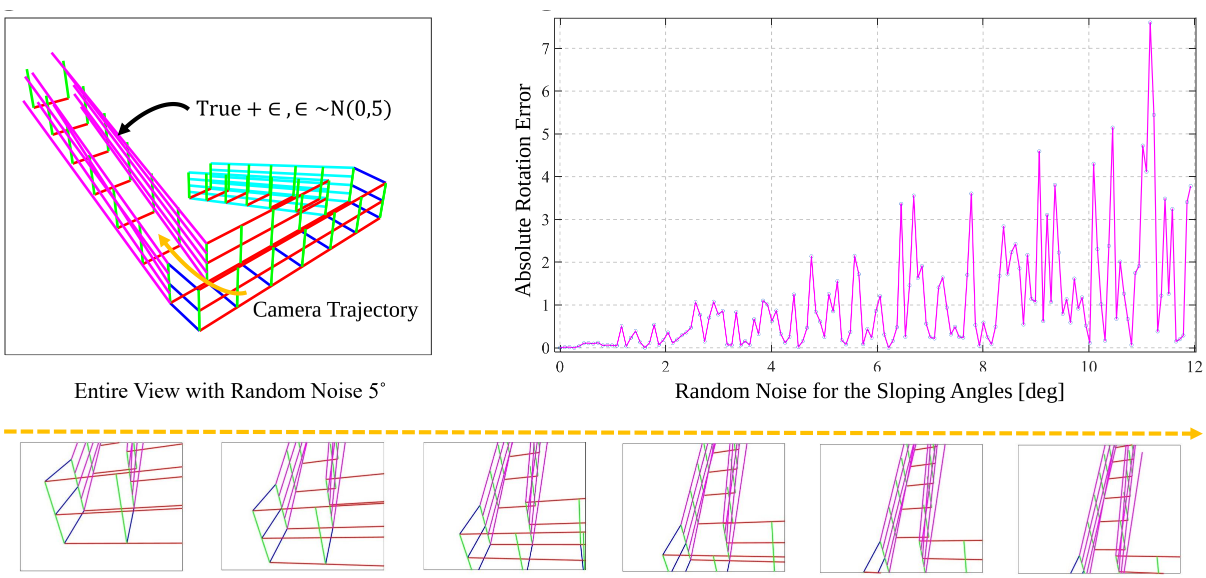



Fig. 10. **Study about sensitivity and robustness to noise for our SFW tracking**. Robustness analysis to noise for our 3-DoF rotational motion tracking. Tracking is performed using magenta-colored sloping lines along the yellow path.

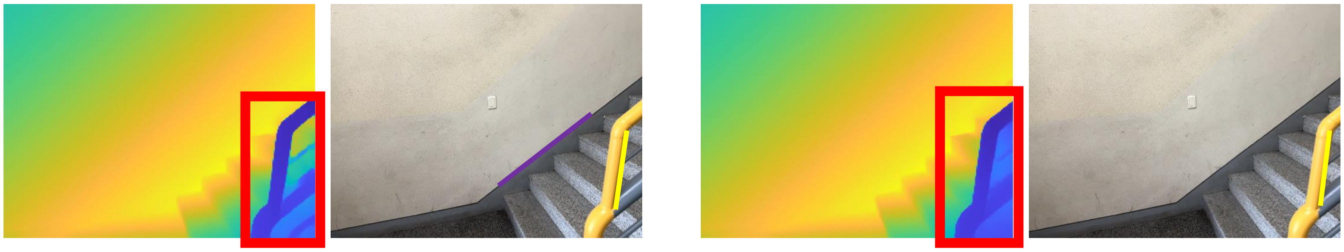
when sufficient orthogonality is observed between lines and planes. The biggest drawback of MMW compared to the proposed SFW is its high DoF. A new MW needs to be generated for each slope angle, meaning three new MWs should be initialized and tracked even for just climbing one floor. However, in texture-less environments like indoor/outdoor staircases and slopes, there are generally not enough directional features or lines in the slope direction (due to the lack of texture), making detecting, initializing, and tracking new MWs difficult. As a result, the slope angle direction cannot be effectively utilized. In Fig. 12, in texture-less environments such as climbing stairs, only point feature-based ORB-SLAM is operational, which results in overall performance similar to ORB-SLAM3.

## References

[1] P. Kim, B. Coltin, and H. J. Kim, "Indoor rgb-d compass from a single line and plane," in *IEEE CVPR*, 2018.

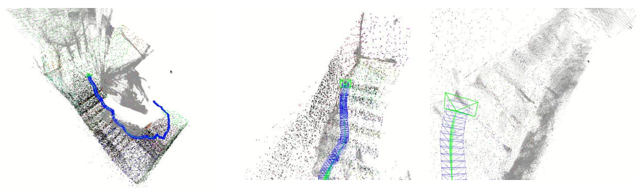
[2] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *IEEE T-PAMI*, 2008.

[3] H. Li, J. Zhao, J.-C. Bazin, P. Kim, K. Joo, Z. Zhao, and Y.-H. Liu, "Hong kong world: Leveraging structural regularity for line-based slam," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13 035–13 053, 2023.



(a) Frame 24 (b) Frame 25

Fig. 11. **Example of failure of LIMAP leveraging depth map**. In environments with few features, if the depth map becomes partially sparse, LIMAP sometimes fails due to the inability to match the surrounding 3D lines (highlighted in yellow) near the affected pixels.



(a) ManhattanSLAM on GIST-SFW Dataset (b) Texture-less Stair Scenes

Fig. 12. **A demonstration of Manhattan SLAM, a SLAM system based on Mixture of Manhattan World, running on GIST-SFW**. (b) In the texture-less stair scene, sloping lines are not effectively utilized, resulting in the failure to detect a new Manhattan frame; only point feature-based ORB-SLAM is used.


[4] Y. Zhou, L. Kneip, and H. Li, "Real-time rotation estimation for dense depth sensors in piece-wise planar environments," in *IEEE IROS*, 2016.

[5] J. Kim, S. Jung, S.-K. Kim, Y. Kim, and A.-a. Agha-mohammadi, "Staircase localization for autonomous exploration in urban environments," *arXiv preprint arXiv:2403.17330*, 2024.

[6] P. Kim, J. Kim, M. Song, Y. Lee, M. Jung, and H.-G. Kim, "A benchmark comparison of four off-the-shelf proprietary visual–inertial odometry systems," *Sensors*, vol. 22, no. 24, p. 9873, 2022.

[7] "TAMU Dataset," http://telerobot.cs.tamu.edu/MFG/rgbd/livo/data.html, accessed: March 21, 2024.

[8] R. Pautrat, D. Barath, V. Larsson, M. R. Oswald, and M. Pollefeys, "Deeplsd: Line segment detection and refinement with deep image gradients," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 327–17 336.

[9] R. Yunus, Y. Li, and F. Tombari, "Manhattanslam: Robust planar tracking and mapping leveraging mixture of manhattan frames," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6687–6693.

[10] C. Campos *et al.*, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.