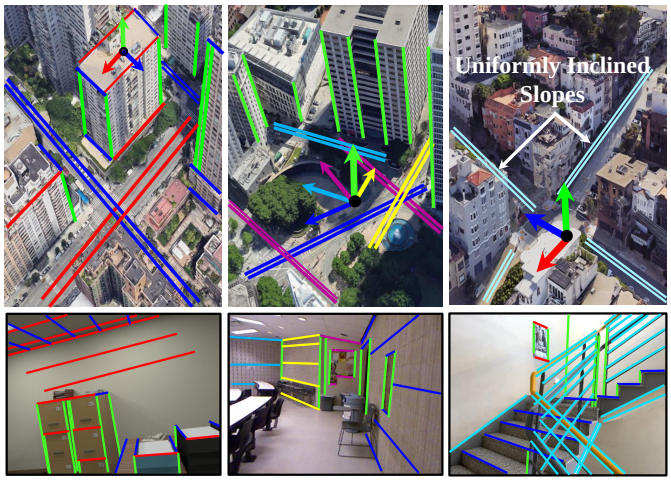
San Francisco World: Leveraging Structural Regularities of Slope for 3-DoF Visual Compass

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

***Abstract*—We propose the San Francisco world (SFW) model, a novel structural model inspired by San Francisco's hilly terrain, enabling 3D inter-floor navigation in urban areas rather than being limited to 2D intra-floor navigation of various robotics platforms. Our SFW consists of a single vertical dominant direction (VDD), two horizontal dominant directions (HDDs), and four sloping dominant directions (SDDs) sharing a common inclination angle. Although SFW is a more general model than the Manhattan world (MW), it is a more compact model than the mixture of Manhattan world (MMW). Leveraging the structural regularities of SFW, such as uniform inclination angle and geometric patterns of the four SDDs, we design an efficient and robust DD/vanishing point estimation method by aggregating sloping line normals on the Gaussian sphere. We further utilize the structural patterns of SFW for the 3-DoF visual compass, the rotational motion tracking from a single line and plane, which corresponds to the theoretical minimal sampling for 3-DoF rotation estimation. Our method demonstrates enhanced adaptability in more challenging inter-floor scenes in urban areas and the highest rotational tracking accuracy compared to state-of-the-art methods. We release the first dataset of sequential RGB-D images captured in San Francisco world (SFW) and open source codes at: https://SanFranciscoWorld.github.io/.**

***Index Terms*—Mapping, Vision-Based Navigation, SLAM, Data Sets for SLAM, RGB-D Perception.**



(a) Manhattan (b) Atlanta (c) San Francisco

Fig. 1. Representative city views for each structural model (top) and image samples of off-the-shelf and author-collected datasets (bottom). (a) Manhattan world with three mutually orthogonal DDs, (b) Atlanta world with multiple horizontal DDs and one vertical DD, and (c) San Francisco with multiple sloping DDs (cyan) sharing uniform inclination angle and three mutually orthogonal DDs (red, blue, and green axis). Each SDD is mutually orthogonal to one of the HDDs in SFW.

I. Introduction

STRUCTURED environments have been well-studied in computer vision [1]–[4] and robotics fields [5]–[8]. The Manhattan world (MW) [1] represents the scenes with three dominant directions (DDs) that are mutually orthogonal (see Figs. 1(a) and 2(a)). The Atlanta world (AW) [2] holds for the scenes with multiple horizontal DDs (HDDs) and a single vertical DD (VDD) (see Figs. 1(b) and 2(b)). The mixture of Manhattan worlds (MMW) [3] assumes multiple Manhattan worlds that are independent (see Fig. 2(d)), describing more complex man-made environments than MW and AW. The Hong Kong world (HKW) [8] has multiple sloping DDs (SDDs) with distinct slopes and HDDs sharing a common VDD (see Fig. 2(e)).

The main shortcoming of MW and AW is their limited generality, describing scenes on flat lands but not environments with slopes. Although MMW and HKW models can describe more general structured environments, they are unstable for use in VO/SLAM and rotational motion tracking and sensitive to noise owing to their high degree of freedom (DoF).

To address these issues, we propose a novel structural model named the *San Francisco world* (SFW) inspired by San Francisco's hilly terrain (see Fig. 1(c)). Our proposed SFW model is 4-DoF, which consists of a single VDD, two HDDs, and four SDDs sharing a common inclination angle. Our SFW is more general than MW because it can represent environments with slopes, e.g., structured environments with hilly terrain or a staircase. It is more compact and accurate than MMW and HKW since DDs are tightly coupled based on uniform inclination angle and orthogonality constraints. Although our method estimates a single SDD, leveraging the

Manuscript received: June 7, 2024; Revised: September 20, 2024; Accepted: October 20, 2024. This letter was recommended for publication by Editor Pascal Vasseur upon evaluation of the Associate Editor and Reviewers' comments. The work of Pyojin Kim was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2024-00358374). The work of Kyungdon Joo was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2024-00457065). The work of Haoang Li was supported by Guangzhou Municipal Science and Technology Project 2023A03J0011. *(Corresponding author: Pyojin Kim.)*

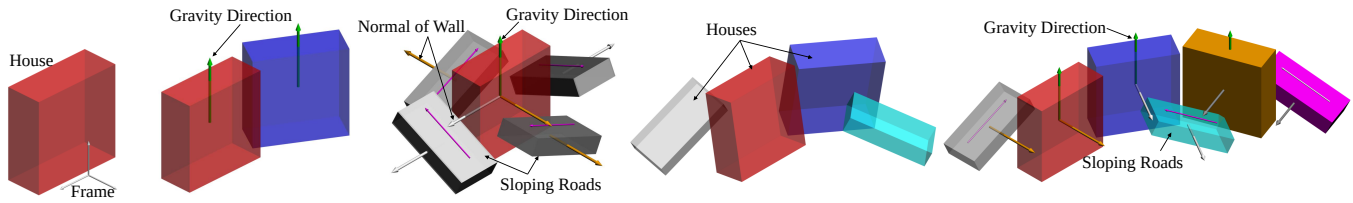
[1]School of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea. {jungilham,minji0110,pjinkim}@gist.ac.kr

[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

Digital Object Identifier (DOI): see top of this page.



(a) Manhattan World (b) Atlanta World (c) San Francisco World (d) Mixture of Manhattan World (e) Hong Kong World

Fig. 2. Illustration of various structural models. (a) Manhattan world [1] corresponds to a single block or frame. (b) Atlanta world [2] corresponds to multiple blocks sharing a common vertical DD, e.g., gravity direction. (c) In our San Francisco world, sloping blocks (gray boxes) share a uniform inclination angle. The sloping lines are orthogonal to the normal of the walls (silver and golden axes). (d) Mixture of Manhattan world [3] with independent blocks. (e) Hong Kong world [8] with multiple sloping roads and blocks that share a common vertical DD.

structural regularities of SFW, our proposed method can utilize line segments from four SDDs (see Fig. 2(c) magenta axes on the gray boxes).

Our SFW has a large variety of potential application fields, such as scene understanding, visual odometry (VO), and 3-DoF rotation estimation. We focus on applying it to a 3-DoF rotation estimation. Our proposed method, *SLOPe* (Single Line and Plane-based Absolute Orientation Perception), includes efficient and robust SFW detection and a visual 3-DoF compass from only a single line and a plane. Our *SLOPe* enables 3D inter-floor navigation in urban areas rather than being limited to 2D intra-floor navigation by effectively utilizing consistent and repetitive slopes in indoor/outdoor environments. Extensive evaluations show that our DD detection and rotational motion tracking method produces the highest accuracy compared to state-of-the-art methods. Our main contributions are as follows:

- We propose a general and compact structural model called San Francisco World (SFW) for structured environments with slopes, leveraging uniformly inclined slopes and the mutual orthogonality between HDDs and SDDs.
- We present a novel approach named *SLOPe*, a highly accurate, drift-free rotational motion tracking method in SFW that requires only a single line and plane.
- We establish the first dataset of sequential RGB-D images collected in SFW and evaluate our method on the author-collected and the TAMU datasets, showing robust, stable, and accurate performance.

II. Related Work

Recent approaches in vision applications such as single-view DD estimation [9]–[11] and DD tracking over time for camera orientation estimation [5], [6], [8], [12]–[14] have exploited structural models as precursors. The accuracy of DD/rotation estimation has been improved dramatically in [5], [9], [12] under MW assumption. Joo et al. [6], Li et al. [13], and Zou et al. [15] have exploited AW for more complex environments with multiple HDDs. MW and AW lack generality and can only describe the structured scenes on flat lands. Antunes and Barreto [16], and Yunus et al. [14] infer the full MMW, the mixture of independent MWs. The main limitation of MMW lies in unsatisfactory accuracy and overwhelming computation owing to its high DoF. Li et al. [8] have proposed HKW for a more strict assumption than MMW, with multiple HDDs sharing a common VDD. However, HKW lacks compactness in describing most structured environments with less complexity because it assumes an infinite number of HDDs and SDDs with distinct slopes, resulting in over-clustering DDs and unsatisfactory accuracy.

Numerous studies have exploited structural prior for estimating DDs utilizing line and plane normals. Kim et al. [5] incorporate the 3-DoF rotational motion tracking from a single line and plane into the model estimation step of the RANSAC, utilizing the structural regularities of MW. Li et al. [17] utilize line normals and a parameter search-based Mine-and-Stab (MnS) method to guarantee quasi-global optimality in HDDs estimation in AW. They reduce the search space effectively by utilizing the orthogonality of VDD and HDDs. Methods under MW and AW assumption [5], [17] fail to estimate SDDs. Li et al. [8] propose a line-based SLAM under HKW assumption, leveraging the orthogonality of HDDs and SDDs. The shortcomings of their approach are the over-clustering DDs and unsatisfactory accuracy owing to the HKW's high DoF. Their method also often fails because it necessitates a substantial number of inlier lines.

Existing approaches have neglected to utilize the unique geometric properties of structured environments with slopes and fail in accurate DD/rotation estimation. Our method overcomes these limitations thanks to the structural regularities of SFW, such as the uniform inclination angle of SDDs and the orthogonality between HDDs and SDDs. We design a new and efficient MnS approach for optimal sloping parameter search in SFW. We further leverage the structural regularities for designing drift-free rotational motion tracking from a single line and plane.

III. Preliminaries

A. Gaussian Sphere

A Gaussian sphere is a unit sphere centered on the camera's center of projection. Fig. 3 illustrates how we represent geometric elements such as lines and plane normal vectors (PNV) on the Gaussian sphere. A line observed in the image is projected onto the Gaussian sphere as a great circle. We express the line in the image as a normal vector of the great circle. Great circles representing parallel lines in the image intersect at two antipodal points on the Gaussian sphere. A unit vector from the center of projection to the intersection point is a vanishing direction, calculated as the cross-product of the normal vectors of two great circles representing parallel lines in the image. The normal vectors from parallel lines lie

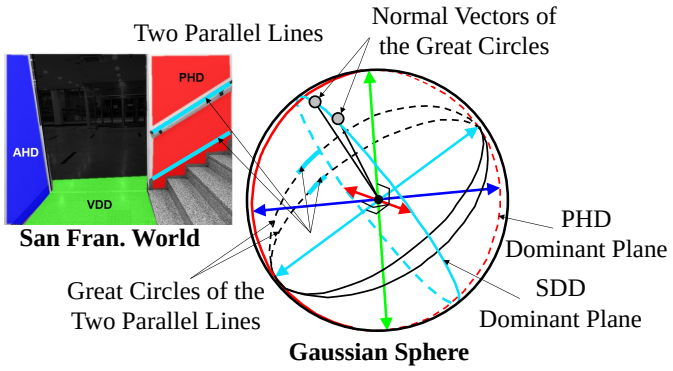


Fig. 3. The 3D geometric relationship between the lines and plane normals on the Gaussian sphere. The colors of the DDs on the Gaussian sphere indicate the respective color of the plane normal vector or lines sharing a common vanishing direction. A DD is defined by at least two parallel line segments (see the sloping lines (cyan) projected on the Gaussian sphere). We express a line segment on the image as a normal vector of the great circle.

on the "dominant plane". We express the estimated vanishing direction and a plane normal vector as the "dominant direction (DD)" and the normal vector of the great circle from the image line as the "line normal".

In Section IV-A, we aggregate sloping lines aligned with the four distinct sloping dominant planes to be aligned with a single sloping dominant plane for an effective sloping parameter search. The primary horizontal direction (PHD) in Fig. 3 indicates an HDD mutually orthogonal to that first initialized sloping dominant direction in the aggregation step, and the auxiliary horizontal direction (AHD) is the remaining HDD. We summarize the abbreviations in Table I.

B. Mine-and-Stab (MnS) Algorithm

The Mine-and-Stab (MnS) method guarantees global optimality in terms of maximizing the number of inliers. We briefly summarize the basic idea of the MnS approach in terms of our sloping parameter θ^* search (for full details of the MnS method, refer to [17]). In Fig. 4(a), a set of sphere points p_i of the i-th noise-free sloping lines normal vector n_i should lie on the same sloping dominant plane. In practice, however, the sphere points p_i cannot strictly lie on the sloping dominant plane due to the noise of the 2D line position in the image.

TABLE I
ACRONYMS WITH COMPLETE WORDS

Acronym	Meaning
MW	Manhattan World
AW	Atlanta World
SFW	San Francisco World
HKW	Hong Kong World
MMW	Mixture of Manhattan World
DD	Dominant Direction
SDD	Sloping Dominant Direction
PHD	Primary Horizontal Direction
AHD	Arbitrary Horizontal Direction
PNV	Plane Normal Vector
MF	Manhattan Frame
MnS	Mine-and-Stab

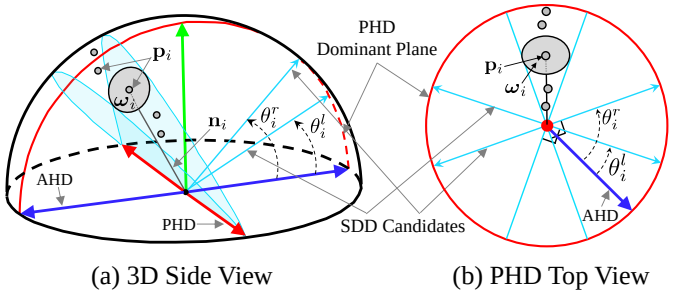


Fig. 4. The Gaussian sphere with the normal vector (n_i) of projected i-th image line. The sphere point p_i (gray dot) is aligned with the sloping dominant plane. We expand p_i into the spherical cap ω_i, the candidate region to obtain the candidate interval $[\theta_i^l, \theta_i^r]$.

To consider this error, the MnS method employs a strategy involving the extension of points p_i to circles w_i. We first mine candidate intervals $[\theta_i^r\ \theta_i^l]$ for each circle w_i. The candidate intervals for circles w_i are calculated as angles between AHD (blue axis) and SDD candidates $s(\theta_i^r)$, $s(\theta_i^l)$ (cyan axes), with the PHD (red axis) as the rotation axis. Given the candidate intervals, our goal is to find an optimal sloping parameter θ^* that stabs the maximum number of the candidate intervals.

IV. Proposed Method

We propose a new and efficient method for SFW detection and rotational motion tracking and name our method *SLOPe* (Single Line and Plane-based Absolute Orientation Perception), a novel approach for estimating a drift-free rotational motion in SFW. Our method is robust in sparsely textured environments because we effectively utilize geometric patterns and structural regularities to overcome sparsity. An overview of the proposed method is shown in Figs. 5 and 8.

A. San Francisco World Detection

1) Initial MW Detection: We detect line segments using LSD [18] and calculate their corresponding unit normal vectors of great circles on the Gaussian sphere. We detect a PNV from a depth image's 3D point cloud with a RANSAC algorithm [19]. We assume at least one Manhattan frame (MF) line exists. To define the second MF axis, we take the cross product between PNV and the normal vector of the MF line. The third MF axis is automatically determined. For full details of the MW detection, refer to [5].

2) Non-Manhattan Frame (Non-MF) Lines Filtering: We filter the non-MF lines by eliminating the line normals aligned with the three dominant planes of the MF.

3) Sloping Parameter θ^* Search: Given the Non-MF lines, we aggregate all the sloping line normals that follow distinct SDDs to maximize the number of inliers. We leverage the following 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. 6(a).

Property 1. In Fig. 6(a), the s_1 and s_2 are symmetric with respect to the h_a and the v. The s_3 and s_4 are symmetric with respect to the h_p and v in the same sense.

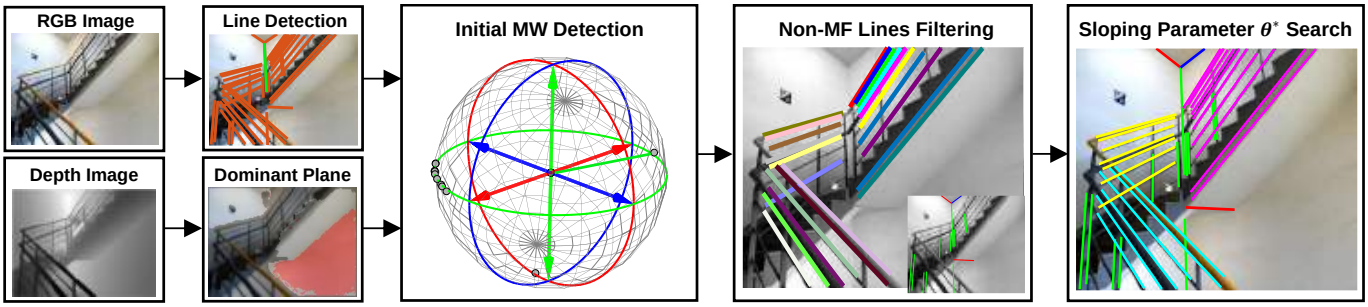
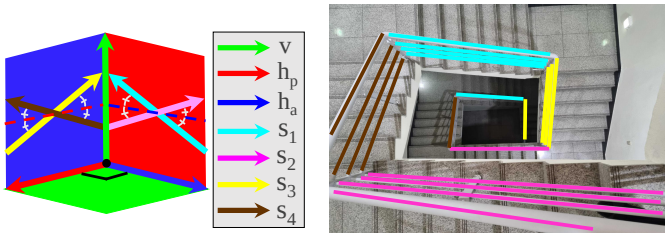


Fig. 5. Overview of our San Francisco world detection. We first detect the initial Manhattan world in the current image to filter out the non-Manhattan frame (non-MF) lines among all detected lines. Our novel sloping parameter θ^* search algorithm finds the optimal 1-DoF slope by effectively utilizing the geometric properties of the SFW, achieving quasi-global optimality.



(a) Illustration of SFW (b) Example of an SFW Scene

Fig. 6. The symmetric and quarter-turn relationships of four SDDs in SFW. The colors of the SDDs in (a) indicate the respective color of the vanishing direction for the sloping lines illustrated in (b).

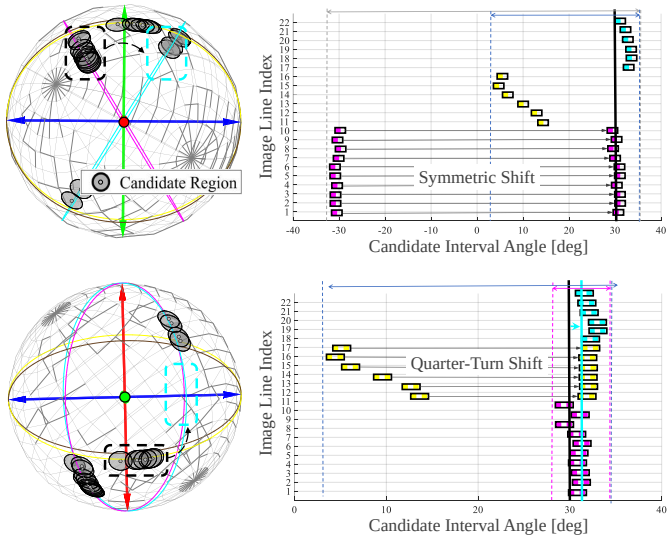


Fig. 7. The sloping line normals on the Gaussian sphere in PHD top view (top left) and VDD top view (bottom left) and corresponding candidate intervals $[\theta_l, \theta_r]$ (right). By utilizing the symmetry and the quarter-turn relationship of the sloping line normals, we can aggregate all the sloping line normals to be aligned on a single dominant plane, effectively finding the optimal probe that maximizes the number of stabbed intervals.

Property 2. The s_1 and s_3, and the s_2 and s_4, are related by a 90-degree rotation around v, respectively. We call this relationship as a quarter-turn relation.

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. Specifically, by utilizing ***Property 1***, we aggregate line normals following s_2 onto s_1 dominant plane with a symmetric shift (see Fig. 7 (top)), and ***Property 2***, we aggregate line normals following s_3 and s_4 into the same s_1 dominant plane with a quarter-turn shift (see Fig. 7 (bottom)). Our efficient MnS approach reduces the search space of $[-\frac{\pi}{2}, \frac{\pi}{2}]$ to $[0, \frac{\pi}{2}]$ and obtains an accurate sloping parameter in a "quasi-globally" optimal manner. Specifically, it guarantees the retrieval of the maximum number of inliers under the condition that 3-DoF (the initial Manhattan frame) is constrained.

B. Rotational Motion Tracking

After the SFW detection, we estimate drift-free rotational motion from a single line and plane, which corresponds to the theoretical minimal sampling for 3-DoF rotation estimation. Given the PNV (the first DD, DD_1) tracked with a mean shift algorithm [12], our *SLOPe* first aims to determine DD_2. The remaining DDs of the SFW at the N^{th} frame can be automatically determined utilizing the structural regularities of the SFW. Algorithm 1 outlines the procedure for the proposed *SLOPe* method.

Fig. 9 depicts the geometric relationships between each DD. The tracked PNV and an unknown-but-sought DD_2 from a sampled line can be either mutually orthogonal (Fig. 9 (a)) or related by the sloping parameter θ^* (Fig. 9 (b) and (c)). To operate for every plane and line pair existing in SFW, an algorithm that works differently for each case is necessary. We propose two approaches of DD_2 estimation, the *Orthogonal Method/Non-Orthogonal Method* for the implementation of *SLOPe*.

1) The Orthogonal Method: If the PNV and the unknown-but-sought DD_2 are mutually orthogonal, DD_2 is determined by the orthogonal method, taking a cross product between the PNV and the line normal. Given the DD_2 estimated from the orthogonal method, we compute the angle between the estimated DD_2 and each axis in the $(N-1)^{th}$ San Francisco frame (SF frame). If the angle satisfies the angle threshold, we skip the non-orthogonal method and recover N^{th} SF frame.

2) The Non-Orthogonal Method: If no axis satisfies the angle threshold check, we assume that the PNV and the unknown-but-sought DD_2 are in a relationship of Fig. 9 (b) or (c), and the sampled line is a sloping line. We propose a novel *Non-Orthogonal method* for determining DD_2, leveraging the

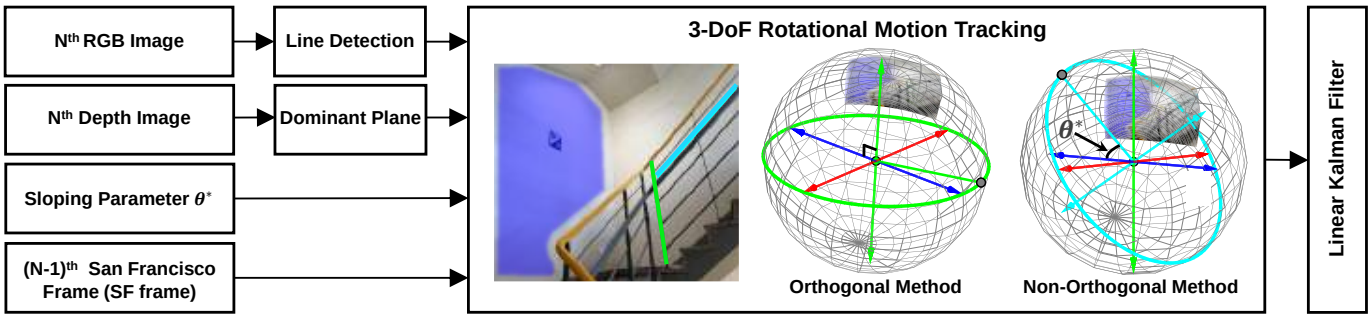


Fig. 8. Overview of our visual compass in San Francisco world. The proposed method estimates the drift-free rotational motion by using only a single line and plane in the RANSAC framework. We utilize the sloping parameter θ^* for the non-orthogonal method and the $(N-1)^{th}$ SF frame for verifying the estimated DD. We apply a Linear Kalman Filter for smoothing.

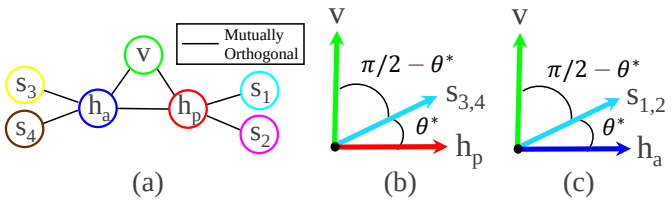


Fig. 9. Illustration of the mutual relationships between DDs in SFW. v, h_p, h_a, s_n, θ^* denote VDD, PHD, AHD, the four SDDs, and the sloping parameter, respectively. The two distinct DDs can be either (a) mutually orthogonal or (b, c) related by θ^* or $\frac{\pi}{2} - \theta^*$.

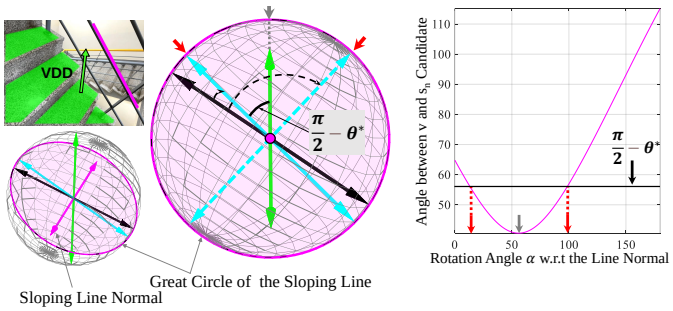


Fig. 10. Non-orthogonal method for SDD detection. The second column depicts the top view of the sloping line normal. Utilizing the property that a sloping line normal and the unknown-but-sought SDD are orthogonal, we parameterize SDD as $s_n(\alpha)$ (cyan axis), which is the arbitrary vector u (black axis), lying on the great circle of the sloping line, rotated by an angle α. We search for the α where the rotated vector u makes an angle $\frac{\pi}{2} - \theta^*$ with the VDD.

geometrical relationship between the sloping line normal and the three potential PNVs.

In Fig. 10, the tracked PNV is VDD v, and a single line follows s_1. The angle between v and s_1 is $\frac{\pi}{2} - \theta^*$ (the known sloping parameter) as Fig. 9 (c) describes. To parameterize the unknown-but-sought SDD s_n, we utilize the orthogonality between the SDD and the sloping line normal. We rotate an arbitrary vector u, perpendicular to the known $(n)^{th}$ sloping line normal $normal_n$, by an unknown-but-sought angle $\alpha \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ around the sloping line normal $normal_n$ as

$$s_n(\alpha) = R_{\langle normal_n, \alpha \rangle} u, \tag{1}$$

where $R_{\langle axis, angle \rangle}$ denotes a rotation based on axis-angle representation [20]. $normal_n$ and $s_n(\alpha)$ indicate the sloping line normal and the unknown-but-sought s_n, respectively. We aim to find α that makes the angle between s_n and v as close to the known sloping parameter $\frac{\pi}{2} - \theta^*$. Mathematically,

$$\arg\min_{\alpha} \left| \cos^{-1}\left(s_n(\alpha) \cdot h_a\right) - \left(\frac{\pi}{2} - \theta^*\right) \right|^2 \tag{2}$$

The problem in Eq. 2 has two distinct real solutions, as shown in Fig. 10, marked by the red arrows. We determine the real root by recovering two SF frames based on each solution and comparing the clustered inliers of the SF frames.

Algorithm 1 *SLOPe*

Input: N line normals, tracked PNV (DD_1)
Output: $(N)^{th}$ SF frame
while RANSAC break condition **do**
  Sample $(n)^{th}$ line normal $normal_n$
  DD_2 = PNV $(DD_1) \times normal_n$
  for each axis a_i of the $(N-1)^{th}$ SF frame **do**
    Compute angle θ_i between DD_2 and a_i
    if $\theta_i \leq$ angle threshold **then**
      DD_2 matches with a_i
      break
    end if
  end for
  if no $\theta_i \leq$ angle threshold **then**
    DD_2 = NonOrthogonal(PNV, $normal_n$)
  end if
  Recover SF frame
end while

V. Experiments

We provide our dataset of sequential images in SFW and evaluate the proposed method on the author-collected and the TAMU datasets. We compare the proposed method against other state-of-the-art approaches.

A. Datasets

1) Our GIST-SFW Dataset: Existing datasets [21], [22] are only suitable for experiments in Manhattan and Atlanta worlds. To evaluate the proposed algorithms, we establish the first dataset of sequential images in the SFW. We collect data on the Gwangju Institute of Science and Technology (GIST) campus using an iPhone 15 Pro Max, which provides RGB, depth, and confidence images, and name our dataset the "GIST-SFW dataset". Fig. 11 shows that our dataset includes image sequences recorded on two different types of staircases. One is

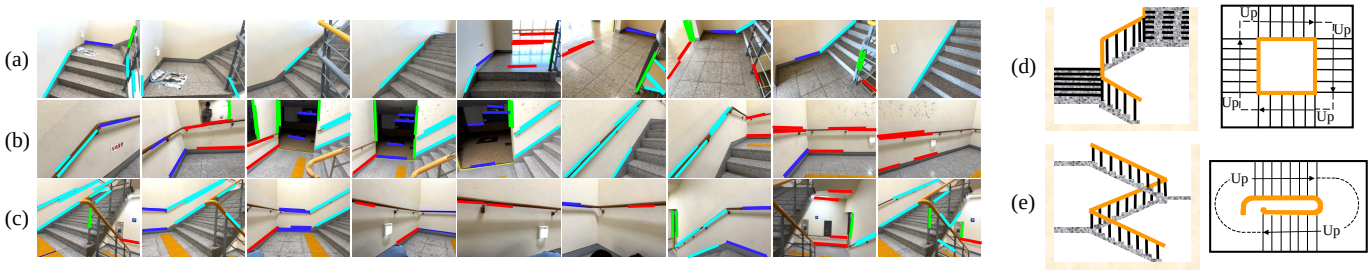


Fig. 11. Sample image sequences from our GIST-SFW dataset, the first indoor RGB-D dataset in SFW, with line clustering results from the proposed method. We capture sequence (a) on a quarter-turn staircase illustrated in (d). We capture sequences (b) and (c) on a half-turn staircase illustrated in (e). We establish and release the GIST-SFW dataset for the evaluation of the proposed *SLOPe* and other methods in SFW.

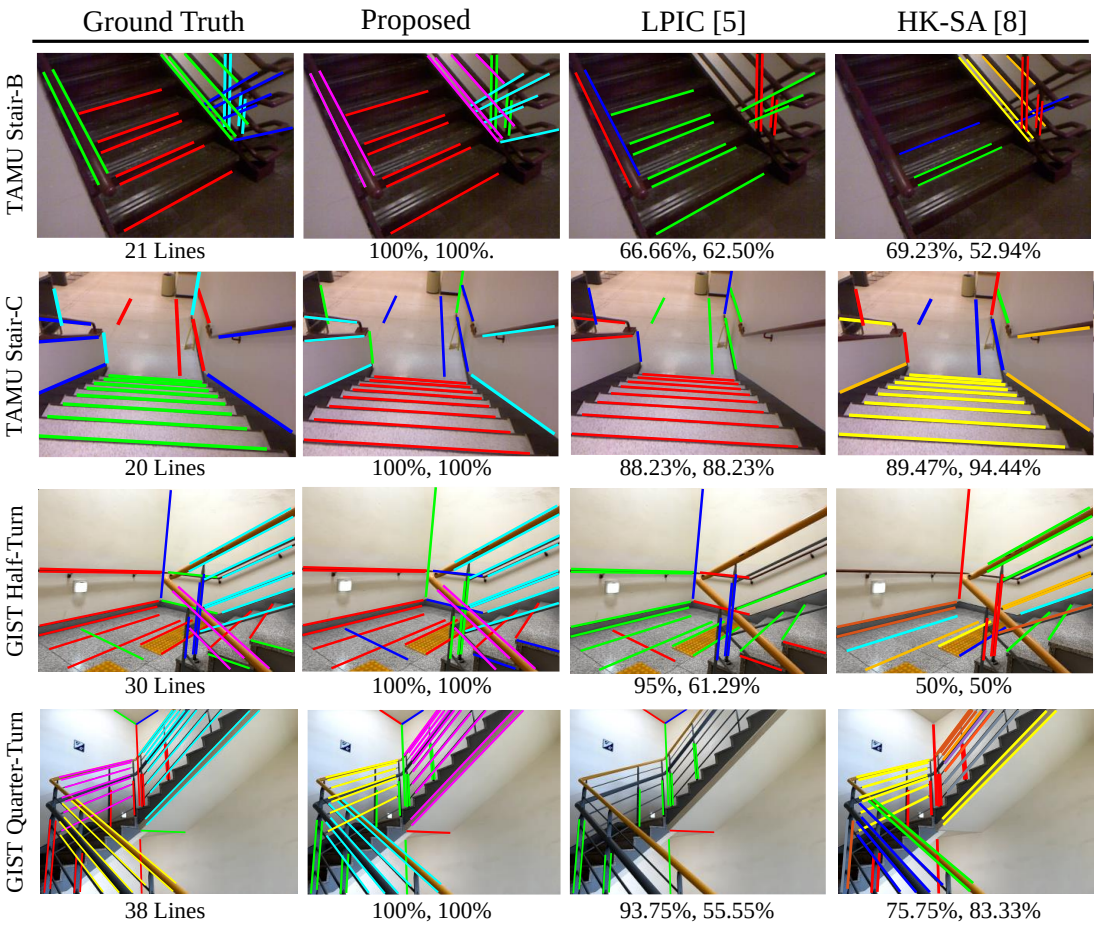


Fig. 12. Representative evaluations on TAMU [25] and our GIST-SFW datasets. Each row represents a tested dataset, and each column denotes an evaluated algorithm. We utilize the manually extracted lines as the ground truth. The numbers below the images are the precision and recall.

a quarter-turn staircase (Fig. 11 (a), and (d)), and the other is a half-turn staircase (Fig. 11 (b), (c) and (e)). This is the first indoor RGB-D dataset in SFW. Our dataset contains Apple ARKit's odometry to validate the proposed *SLOPe* and other methods in SFW. [23], [24] have evaluated the accuracy of 6-DoF pose estimation in Apple ARKit.

2) TAMU RGB-D dataset [25]: The TAMU RGB-D dataset is collected in real-world structured environments. We test our DD estimation method on "Staircase-C-const" and "Staircase-B-vary" sequences.

B. DD/Vanishing Points Estimation

1) Evaluation Criteria: We assess the accuracy of our DD estimation method using precision and recall error metrics [26]. We compute the precision C/(C+W) and recall C/(C+M), where C, W, and M denote the number of correctly identified, wrongly identified, and missing inliers, respectively. We also compute the F_1-score that simultaneously encodes the precision and recall by $F_1 = \frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$. We compare the F_1-score on randomly selected scenes containing various staircases from the TAMU [25] and our GIST-SFW datasets.

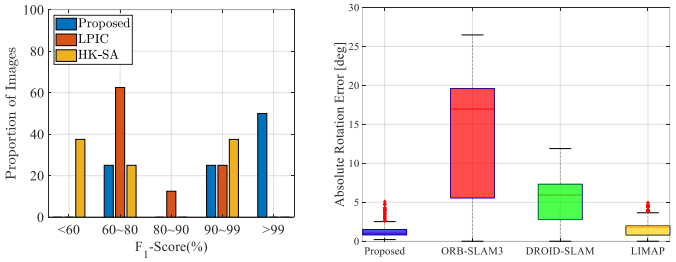
2) Methods for Comparison:

- LPIC [5]: Line and plane-based DD estimator designed for MW,
- HK-SA [8]: Sampling-based DD estimator designed for HKW.

3) Experimental Results: We present the accuracy comparisons in Figs. 12 and 13 (a). LPIC can only estimate

TABLE II
COMPARISON OF THE ABSOLUTE ROTATION ERROR[DEG] EVALUATION ON GIST-SFW DATASET.

Experiment	Proposed	ManhattanSLAM [14]	ORB-SLAM3 [27]	DROID-SLAM [28]	LIMAP [29]	Traveling Rotation
Half-Turn Stair 1	**0.68**	6.91	6.56	2.81	1.12	180°
Half-Turn Stair 2	**1.19**	18.23	12.40	3.33	1.38	360°
Quarter-Turn Stair 1	**0.96**	10.46	6.56	2.81	1.12	180°
Quarter-Turn Stair 2	**1.21**	12.21	15.98	5.76	1.41	360°
In-Place Rotation	**1.18**	20.11	20.26	10.23	×	1800°



(a) (b)

Fig. 13. (a) Accuracy comparison between various DDs estimation methods in terms of F_1-score of image line clustering on all the testing images of TAMU [25] and our GIST-SFW dataset. (b) The statistical distribution of the absolute rotation error from the 'Quarter-Turn Stair 360°'.

three orthogonal DDs. It neglects or misconceives sloping lines. HK-SA shows unsatisfactory accuracy. Over-clustering occurs owing to its high DoF. HK-SA is also a sampling-based method that necessitates a substantial number of vertical and horizontal line samples, which results in poorer outcomes in indoor environments with few HDD inliers. Thanks to the generality and compactness of our SFW model, our proposed method reliably identifies all the SDDs without over-/under-clustering. The F_1-Scores (Fig. 13 (a)) also demonstrate that other methods fail to achieve accurate DD estimation in randomly selected scenes with stairs, whereas our method achieves precision and recall of perfect score in the majority of SFW scenes.

C. Rotational Motion Tracking

1) Evaluation Criteria: We measure the mean value of the absolute rotation error (ARE) [12] in degrees and present the evaluation results in Table II. The smallest rotation error for each dataset is bolded. We compare our rotational motion tracking on our GIST-SFW dataset with state-of-the-art methods. We treat 6-DoF camera poses from Apple ARKit as the ground-truth [23].

2) Methods for Comparison:

- HK-SLAM [8]: Line-based SLAM designed for HKW.
- ManhattanSLAM [14]: Point, line, and plane-based SLAM leveraging a mixture of Manhattan frames.
- ORB-SLAM3 [27]: Point feature-based SLAM.
- DROID-SLAM [28]: Deep learning-based RGB-D SLAM.
- LIMAP [29]: Line and point association-based SLAM exploiting structural priors.

3) Experimental Results: Table II compares the average ARE results of the proposed and other methods. Our method can track accurate and drift-free camera rotational motion even in insufficient structural environments with a single line feature, while other line-based or line and plane-based approaches based on existing structural models (HK-SLAM, LPIC) fail rotational tracking. For LIMAP, we employ the fit-and-merge module that utilizes RGB-D images to ensure a fair comparison. Although it achieves the second-best results among the methods, the lack of a structural model leads to error drift. Other methods (ManhattanSLAM, ORB-SLAM3, DROID-SLAM) are significantly affected by the absence of the structure and texture components in the scenes.

The proposed method shows accurate and robust rotation estimation not only in abundant but also in very low structure and texture-less environments with the help of the minimal solution (one line and one plane). Our method also shows the lowest statistical distribution of the ARE, as illustrated by the boxplot from the 'Quarter-Turn 360°' dataset in Fig. 13 (b). Outliers marked as red cross are removed.

TABLE III
RUNTIME COMPARISON ON GIST-SFW DATASET

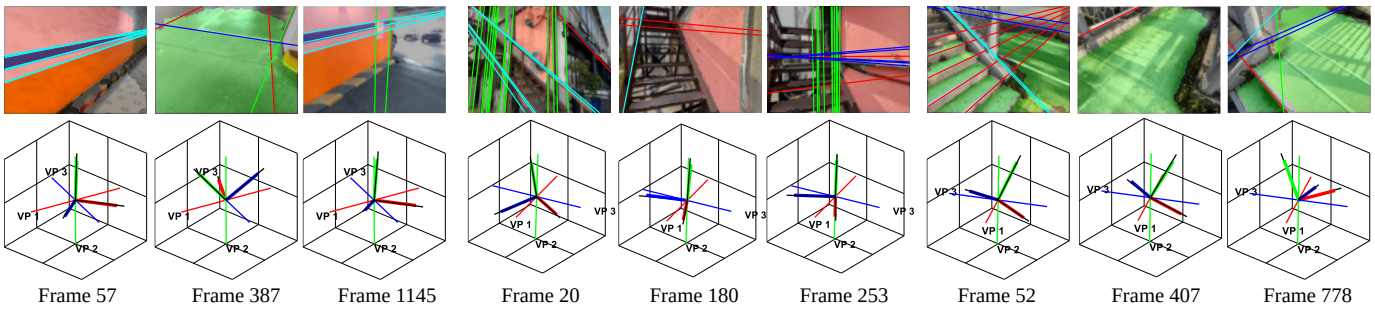
	Proposed	ManhattanSLAM	ORB-SLAM3	DROID-SLAM	LIMAP
Time (s)	0.061	0.128	0.070	0.119	0.208

We report the comparison of the average runtime in Table III. Our method is significantly faster than multi-feature-based or deep learning-based methods. We have also analyzed the runtime of each module that constitutes the proposed method. For more details, please refer to our supplementary materials.

Fig. 14 further demonstrates the generality and usefulness of our methods through experiments in additional test scenes, including various indoor/outdoor environments with slopes that we can easily find around us, not limited to stairs in Fig. 11, involving complex camera translations and rotations, as well as numerous outlier line features. Please refer to the video clips and supplementary materials submitted with this paper for more details about the experiments.

D. Limitations

Proposed *SLOPe* may result in significant errors or even fail if the sloping angle changes significantly between two consecutive staircases or if the symmetric and quarter-turn relationship is not maintained, as it relies on the geometric regularities of slopes and stairs in urban areas. However, in most typical man-made environments, stairs or slopes between floors are usually constructed with specific regulations to maintain consistent angles and structural patterns.



(a) Parking Building (b) Outdoor Fire Escape Staircase (c) Outdoor Pedestrian Bridge

Fig. 14. 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 [23] (bottom).

We further validate the robustness and sensitivity of our *SLOPe* against the sloping angle noise of synthetic lines. In these experiments, the error remains below one and two degrees in SFW detection and rotational estimation, respectively, when the standard deviation of random noise is within five degrees. For more details on these experiments with synthetic lines, please refer to our supplementary material on the project webpage.

VI. Conclusion

We propose a novel structural model, San Francisco world (SFW), designed for accurate, drift-free rotation estimation in structured indoor environments with slopes. SFW enables 3D inter-floor navigation in urban areas rather than limited to 2D intra-floor navigation of various robotics platforms. SFW is a more general model than MW and a more compact model than HKW. Leveraging the structural regularity of SFW, we design an efficient DDs estimator, San Francisco world detection, that overcomes the sparsity of environmental features. We further leverage the structural patterns of SFW for the 3-DoF visual compass from a single line and plane, which are the minimal sampling for 3-DoF rotation estimation. We establish and release the first dataset of sequential RGB-D images captured in SFW. Experiments show that our approach outperforms state-of-the-art methods in adaptability and accuracy. Future works involve extending the proposed *SLOPe* algorithm into VO/SLAM frameworks for robust inter-floor localization and mapping in urban environments, enabling advanced robotic applications such as seamless inter-floor navigation and efficient exploration of complex urban architectures.

References

[1] J. M. Coughlan and A. L. Yuille, "Manhattan world: Compass direction from a single image by bayesian inference," in *ICCV*, 1999.
[2] G. Schindler and F. Dellaert, "Atlanta world: An expectation maximization for simultaneous low-level edge grouping and camera calibration in complex man-made environments," in *CVPR*, 2004.
[3] J. Straub *et al.*, "A mixture of manhattan frames: Beyond the manhattan world," in *CVPR*, 2014.
[4] S. Gupta and J. Malik, "Perceptual organization and recognition of indoor scenes from RGB-D images," in *CVPR*, 2013.
[5] P. Kim, B. Coltin, and H. J. Kim, "Indoor rgb-d compass from a single line and plane," in *CVPR*, 2018.
[6] K. Joo *et al.*, "Linear rgb-d slam for atlanta world," in *ICRA*, 2020.
[7] H. Li, J. Yao, J.-C. Bazin, X. Lu, Y. Xing, and K. Liu, "A monocular slam system leveraging structural regularity in manhattan world," in *ICRA*, 2018.
[8] 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 T-PAMI*, 2023.
[9] J.-C. Bazin and M. Pollefeys, "3-line ransac for orthogonal vanishing point detection," in *IROS*, 2012.
[10] J.-P. Tardif, "Non-iterative approach for fast and accurate vanishing point detection," in *ICCV*, 2009.
[11] J.-C. Bazin, Y. Seo, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys, "Globally optimal line clustering and vanishing point estimation in manhattan world," in *CVPR*, 2012.
[12] Y. Zhou, L. Kneip, C. Rodriguez, and H. Li, "Divide and conquer: Efficient density-based tracking of 3D sensors in manhattan worlds," in *ACCV*, 2016.
[13] H. Li, Y. Xing, J. Zhao, J.-C. Bazin, Z. Liu, and Y.-H. Liu, "Leveraging structural regularity of atlanta world for monocular slam," in *ICRA*, 2019.
[14] R. Yunus, Y. Li, and F. Tombari, "Manhattanslam: Robust planar tracking and mapping leveraging mixture of manhattan frames," in *ICRA*, 2021.
[15] D. Zou, Y. Wu, L. Pei, H. Ling, and W. Yu, "Structvio: Visual-inertial odometry with structural regularity of man-made environments," *T-RO*, 2019.
[16] M. Antunes and J. P. Barreto, "A global approach for the detection of vanishing points and mutually orthogonal vanishing directions," in *CVPR*, 2013.
[17] H. Li, K. Joo, and Y.-H. Liu, "Globally optimal and efficient vanishing point estimation in atlanta world," in *ECCV*, 2020.
[18] R. G. Von Gioi and G. Randall, "Lsd: A fast line segment detector with a false detection control," *T-PAMI*, 2008.
[19] M. Y. Yang and W. Förstner, "Plane detection in point cloud data," Institute of Geodesy and Geoinformation (IGG), University of Bonn, Tech. Rep., 2010. [Online]. Available: https://ris.utwente.nl/ws/portalfiles/portal/103953896/Yang2010Plane.pdf
[20] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
[21] A. Handa and A. J. Davison, "A benchmark for rgb-d visual odometry, 3d reconstruction and slam," in *ICRA*, 2014.
[22] J. Sturm, N. Engelhard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *IROS*, 2012.
[23] P. Kim *et al.*, "A benchmark comparison of four off-the-shelf proprietary visual–inertial odometry systems," *Sensors*, 2022.
[24] E. Jeong *et al.*, "Linear four-point lidar slam for manhattan world environments," *RA-L*, 2023.
[25] "TAMU Dataset," http://telerobot.cs.tamu.edu/MFG/rgbd/livo/data.html, accessed: March 21, 2024.
[26] J. Straub, N. Bhandari, J. J. Leonard, and J. W. Fisher, "Real-time manhattan world rotation estimation in 3D," in *IROS*, 2015.
[27] C. Campos *et al.*, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *T-RO*, 2021.
[28] Z. Teed and J. Deng, "DROID-SLAM: Deep visual slam for monocular, stereo, and rgb-d cameras," *NeurIPS*, 2021.
[29] S. Liu *et al.*, "3d line mapping revisited," in *CVPR*, 2023.