Drift-Free Visual Compass Leveraging Digital Twins for Cluttered Environments

Jungil Ham
Dept. of Mechanical and Robotics Eng.
GIST, Gwangju, Republic of Korea
jungilham@gm.gist.ac.kr

Ryan Soussan
Intelligent Robotics Group (KBR Inc.)
NASA Ames Research Center
ryan.soussan@nasa.gov

Brian Coltin
Intelligent Robotics Group (KBR Inc.)
NASA Ames Research Center
brian.coltin@nasa.gov

Hoyeong Chun
Dept. of Mechanical and Robotics Eng.
GIST, Gwangju, Republic of Korea
hychun.ug@gm.gist.ac.kr

Pyojin Kim[*]
Dept. of Mechanical and Robotics Eng.
GIST, Gwangju, Republic of Korea
pjinkim@gist.ac.kr

Abstract—**Drift-free and accurate rotational motion tracking is one of the most critical components for visual navigation of free-flying robots operating in microgravity environments, such as the International Space Station (ISS), where unrestricted 360-degree rotational motion is intrinsic. Traditional methods based on the Manhattan world (MW) assumption struggle in such environments due to occlusions and excessive outlier features. To address these issues, we present a novel Digital Twin-Based Outlier Rejection (DTOR) method that leverages the ISS 3D CAD model to improve the robustness of Manhattan world detection and drift-free 3-DoF rotational motion estimation. By matching observed line features against the digital twin, our approach effectively filters out clutter-induced outliers and extracts reliable structural features. The proposed method requires only a single line and plane to estimate absolute and drift-free orientation, enabling lightweight and efficient computation. Experimental evaluations on the Astrobee dataset demonstrate that our method achieves state-of-the-art performance with significantly lower rotation errors in highly cluttered environments.**

Index Terms—**Space Robotics, ISS Localization, Manhattan world, Digital Twin, Rotation Estimation.**
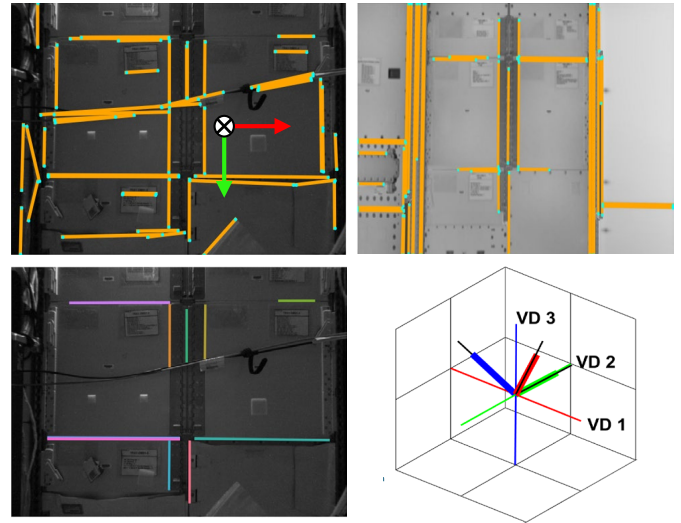
Fig. 1. All detected lines in the ISS image (top-left) and the corresponding rendered image from the 3D CAD digital twin (top-right) in an extremely cluttered scene. Unreliable features caused by clutter are distinguished from structural features that align with the Manhattan frame of the ISS (bottom-left), enabling drift-free camera orientation estimation (bottom-right).

I. INTRODUCTION

The Astrobee is the free-flying robot to perform intra-vehicular activity (IVA) tasks aboard the International Space Station (ISS) [1], [2]. It provides a microgravity robotic research facility in the ISS U.S. Orbital Segment (USOS). To enable autonomous operation aboard the ISS, motion plans for the Astrobee are typically generated within a CAD model of the ISS [3]. These plans are then executed onboard the robot using its localization and navigation software.

Visual localization methods have been actively studied to enhance the autonomy of Astrobee [3]–[8]. However, even recent state-of-the-art methods tend to accumulate drift error over time and require frequent re-initialization of localization.

It is well-known that the rotational drift error and non-linearity during the 3-DoF camera orientation estimation are the main sources of positioning inaccuracy [9] in visual odometry (VO) and visual simultaneous localization and mapping (VSLAM). This challenge is even more pronounced in

microgravity: unlike ground robots that primarily experience planar motion, Astrobee undergoes unrestricted 360-degree rotation in all axes. Therefore, achieving accurate and drift-free rotational motion estimation is critical for robust and long-term autonomy on the ISS.

[10]–[14] demonstrate that leveraging structural assumptions such as the Manhattan world [15], Atlanta world [16], and San Francisco world [14] enables achieving highly accurate and drift-free 3-DoF rotation estimation in diverse indoor and outdoor man-made environments. As shown in Fig 3, the International Space Station (ISS) is an example of the Manhattan world, which is characterized by three mutually orthogonal dominant planes.
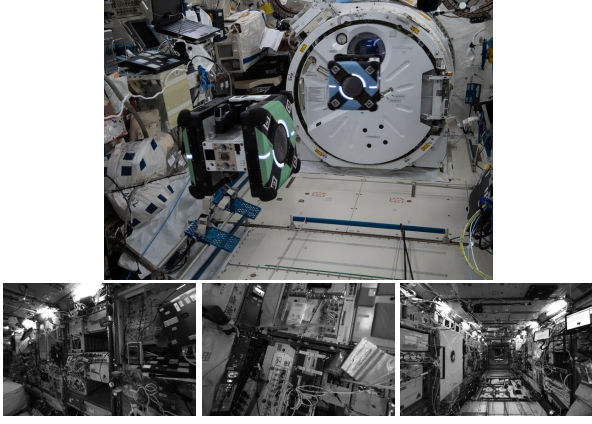
Fig. 2. Astrobee free-flying robots roaming on the ISS. The ISS is a representative example of a cluttered and dynamic environment, which causes localization errors for the Astrobee (top). Sample images from the Astrobee dataset are shown (bottom).
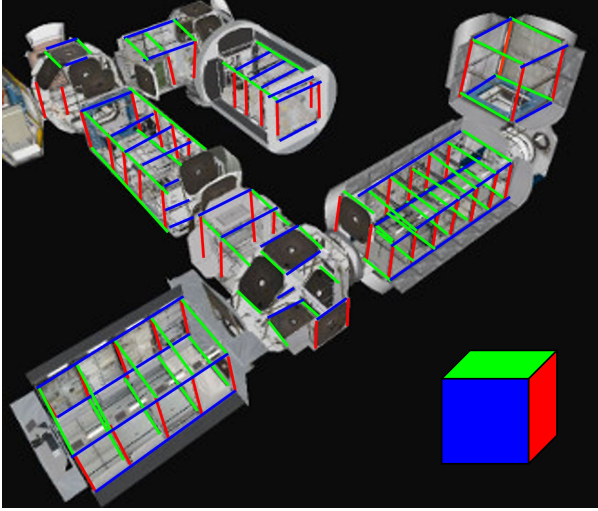

Fig. 3. The International Space Station (ISS) is one of the examples that satisfy the Manhattan world (MW) assumption, with three dominant orthogonal directions visualized in red, green, and blue.

However, it is challenging to detect geometric features (e.g., lines and planes) or maintain consistent data association since objects such as cargo bags, wires, laptops, and racks either occlude these features or introduce excessive outliers (see Fig. 2), which lead to the failure of existing MW-based algorithms. Identifying reliable landmarks in the presence of excessive outliers often requires computationally expensive iterative algorithms, such as deep learning-based methods or robust optimization techniques, which are not suitable for IVR systems with limited computational resources.

To address these issues, we propose a novel rotation estimation that fully leverages Digital Twin-Based Outlier Rejection (DTOR). As shown in Fig. 1, we run the line matching against the mesh representation of the digital twin (3D CAD model), eliminating outlier lines not associated with the global Manhattan world (MW) on the ISS. The proposed method can effectively extract reliable line landmarks and robustly

estimate 3-DoF camera orientation. Our main contributions are as follows:

- We propose a novel Digital Twin-Based Outlier Rejection (DTOR) method that utilizes a 3D CAD model as a clutter-free digital twin to eliminate unreliable landmarks in cluttered and dynamic environments.
- We develop a lightweight, drift-free 3-DoF rotational estimation framework based on robust Manhattan World (MW) tracking from a single line and plane, which is the minimal solution.
- We evaluate our method in a challenging ISS environment and demonstrate improved rotational accuracy and robustness compared to state-of-the-art SLAM and localization methods.

II. RELATED WORK

Digital Twin-based Data Association. In VSLAM, the localization process is based on visual data association between the current camera view and the map of the environment. Classical methods heavily rely on matching feature descriptors, e.g., the ORB descriptor [17]. However, inaccuracies in the 3D map and dynamic environments severely degrade the accuracy of this association.

To address this issue, several works leverage 3D CAD models of the environment, sometimes called digital twins, to provide high-fidelity geometric context for perception and localization. [18] has developed a method to match images captured by unmanned aerial vehicles (UAVs) with Google Street View images by back-projecting their textures onto CAD models of cities. Satellite images have also been used to localize the current camera view in a fixed global frame. [19] proposes a technique to match vehicle images with satellite images by aligning road features. Visual localization techniques provide accurate localization when visual data association can be accurately performed. However, they are highly sensitive to viewpoint variations and environmental conditions. [20] addresses this limitation by directly aligning the sparse 3D point cloud from a visual odometry or VSLAM system to a digital twin, eliminating the dependency on visual matching. Although their method demonstrates drift-free performance without relying on visual data, it fails to operate reliably when significant structural discrepancies exist between the real environment and the digital twin due to clutter.

Clutter-Resistance SLAM. [21] proposes a clutter-resistant SLAM algorithm for dynamic industrial environments, which utilizes point features generated from reflectors and line features to improve SLAM robustness. However, since it primarily relies on static features, it struggles to maintain reliability in the presence of dynamic objects and unexpected clutter. [22] employs a probabilistic approach using the probability hypothesis density filter to model clutter and missed detections. While this method effectively reduces the impact of clutter by down-weighting unreliable measurements, it suffers from high computational complexity and limited scalability for real-time industrial applications. [23] shows an episodic non-Markov
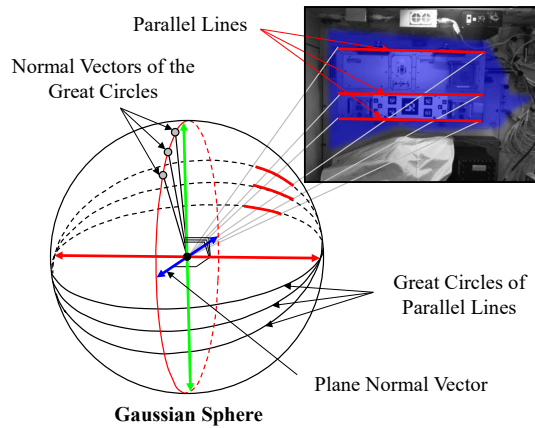
Fig. 4. Geometric relationships between the parallel lines and MW on the Gaussian sphere. We map the image lines onto the normal vectors of the great circles (gray dots). The normal vectors from parallel lines lie on a single dominant plane that is perpendicular to the vanishing direction (red axis). The plane normal vector (blue axis) also lies on that dominant plane.

localization framework that separates short-term and long-term features to enhance robustness in dynamic environments. However, since it does not incorporate geometric constraints or structural priors, its applicability remains limited in severely cluttered environments with high visual ambiguity.

MW Assumption-Based Rotation Estimation. Manhattan world (MW) is widely utilized for estimating drift-free camera rotation in visual odometry [10], [24] and SLAM [9], [25].

Traditional methods [26], [27] have utilized the distribution of sampled surface normals to estimate dominant orthogonal directions in an MW. Although these surface normal-based sampling approaches demonstrate a stable and accurate rotation estimation, they require a dense surface normal distribution, and at least two orthogonal planes must always be visible. [10] proposes a line and plane-based method that only requires a single dominant plane and a parallel line lying on that plane. This method is primarily specialized in texture-less environments and sensitive and unstable in the presence of spurious or noisy line segments.

Although parameter search-based methods [28], [29] guarantee global optimality using only line segments by maximizing the number of inliers, they require more than three seconds per image to compute the optimal solution. Recent studies [11], [12] have hybridized these two strategies, but they still depend on the computationally expensive Branch and Bound (BnB). [30] employs a hybrid strategy. Although they have reduced computational load significantly, their method still suffers when there is a significant amount of outliers.

In this work, we are inspired by the recent efforts to leverage digital twins for robust localization. We propose to utilize a static 3D CAD model of the ISS as a digital twin to reject outlier observations and enable accurate, drift-free rotation estimation under extreme visual clutter.
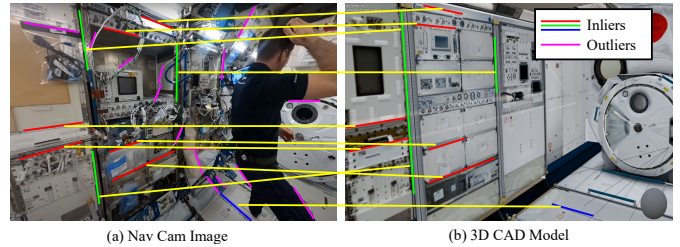

(a) Nav Cam Image   (b) 3D CAD Model

Fig. 5. Line matching and outlier rejection using the digital twin. Inliers (red, green, blue) represent structural features aligned with the Manhattan world (MW) and successfully matched with the digital twin. Outliers (magenta) indicate unreliable features caused by clutter or unstructured objects.

III. BACKGROUND

The Gaussian sphere is a foundation for the geometric interpretation of image lines and surface normals, representing a virtual unit sphere centered at the optical center of the camera. We project a line in an image onto the Gaussian sphere as a *great circle* (the intersection of the Gaussian sphere and the plane defined by the center of projection (COP) and the line, see Fig. 4).

The great circle of each line can be expressed as a unit normal vector (gray dots). We transform all image lines into the normal vectors of the great circles on a Gaussian sphere. The great circles from the parallel lines intersect at two antipodal points called a *vanishing direction (VD)* in the *Manhattan world (MW)*, which posits that every line and plane is perpendicular to one of the axes of a single coordinate system.

We call this fixed coordinate system a *Manhattan Frame (MF)*. We summarize the abbreviations in Table I. The orthogonal surface normals of the MW planes exactly match the three orthogonal VDs defined by the parallel lines in an ideal MW. All normal vectors from the parallel lines pointing in the same direction should lie on the same great circle.

IV. PROPOSED METHOD

We propose a new approach for estimating a camera's drift-free 3-DoF rotational motion from an RGB and depth image pair captured on the ISS. Our method filters out unreliable landmarks not aligning with the Manhattan world (magenta-colored lines in Fig. 5) on the ISS, leveraging the 3D CAD model as a digital twin. An overview of the proposed method is illustrated in Fig. 6.

TABLE I
ACRONYMS WITH COMPLETE WORDS

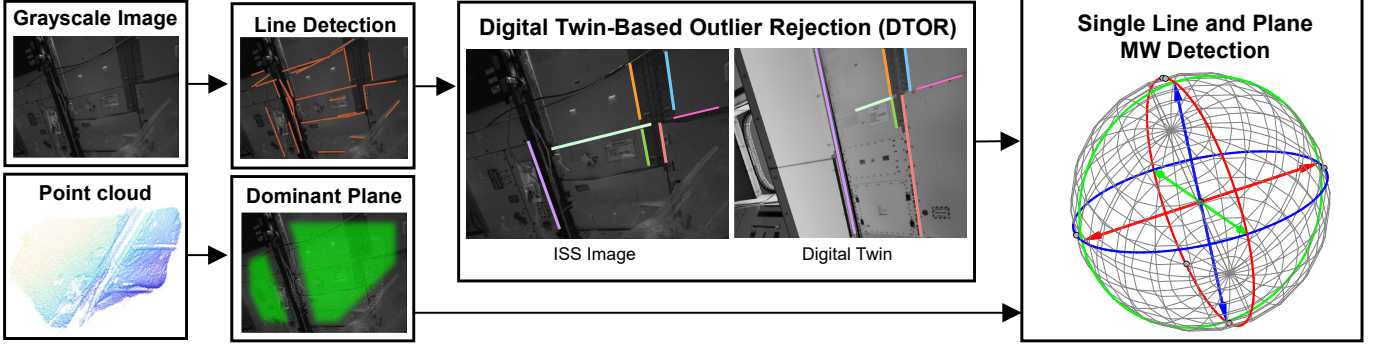| Acronym	Meaning
MW	Manhattan World
DD	Dominant Direction
PNV	Plane Normal Vector
MF	Manhattan Frame

Fig. 6. Overview of the proposed drift-free 3-DoF rotation estimation algorithm. The pipeline consists of four main steps: (1) detection of line features from the input RGB image, (2) rendering the simulated view from the digital twin and performing line matching to reject clutter-induced outliers, (3) detection and tracking of the dominant plane from the depth map, and (4) estimation of the Manhattan frame from a single line and plane pair using RANSAC.

A. Digital Twin-Based Outlier Rejection

We begin by detecting N line segments in the grayscale image using LSD [31]. We assume that the camera pose of the first frame is known. Using the first camera pose or the estimated camera pose from the previous image frame, we render the simulated camera view in the clutter-free digital twin. We then detect line segments in the rendered image captured from the digital twin. As shown in Fig. 5, we filter out clutter that does not appear in the digital twin using a deep learning-based line matching method [32]. Using this approach, we can effectively remove unreliable line segments generated in unstructured and cluttered environments.

Following the notation in [21], we refer to lines that do not comply with the Manhattan world as *unreliable landmarks* or *outliers*, while those matched with the landmarks detected from the digital twin are referred to as *structural landmarks* or *inliers*.

B. Dominant Plane Detection and Tracking

We subsequently detect a dominant plane from the depth image's 3D point cloud using the RANSAC plane detection method [33]. The algorithm finds the plane supported by the maximum number of 3D points inliers, i.e., how many 3D points fall within a given distance threshold from this plane. As shown in Fig 7, when the density distribution of the surface normal vectors around the currently tracked normal vector is too low, our method finds and re-initializes a new dominant plane. Please refer to [10], [27] for full details of the detection and tracking of the dominant plane.

C. Single Line and Plane Rotation Estimation

Our method leverages line and plane geometry to determine MF in the current view. As illustrated in Fig. 8, when a dominant plane and a parallel line lying on that plane are found, they uniquely determine the Manhattan world in the current frame. We utilize this geometric feature in a RANSAC framework to estimate the camera's drift-free 3-DoF rotation by tracking the MF of consecutive frames.

Each RANSAC iteration randomly selects one line among the reliable structural lines we filtered in Section IV-A. Using
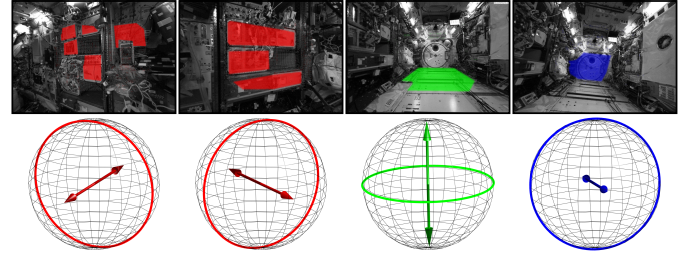

Fig. 7. Dominant plane detection and tracking. Our method detects a single dominant plane by analyzing the density distribution of the surface normal vectors. When the density distribution of the surface normal vectors around the currently tracked normal vector is too low, our method finds and re-initializes a new dominant plane.
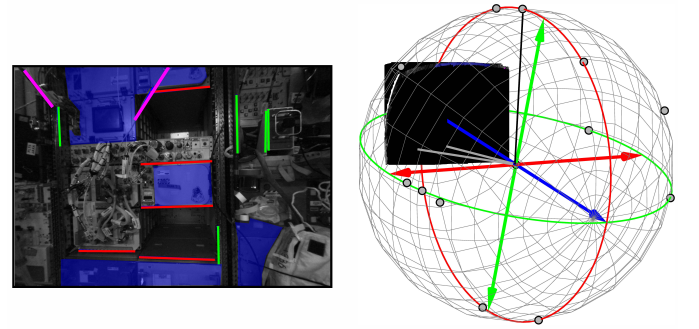

Fig. 8. Determination of the three vanishing points ($\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$) in the Manhattan world. $\mathbf{v}_1$ is set as the plane normal vector. If the sampled line corresponds to the thick red line, the cross product of its normal vector (black) and $\mathbf{v}_1$ determines $\mathbf{v}_2$. $\mathbf{v}_3$ is obtained by taking the cross product of $\mathbf{v}_1$ and $\mathbf{v}_2$.

the tracked plane normal vector $\mathbf{v}_1$ from Section IV-B, if the selected line normal corresponds to the black axis in Fig. 8, the second vanishing point $\mathbf{v}_2$ (the red axis) can be defined by computing the cross product with the selected line's normal. The third vanishing point $\mathbf{v}_3$ is then automatically defined as $\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2$.

V. EXPERIMENTS

We evaluate the proposed method on the Astrobee dataset [34], collected inside the ISS. The dataset includes

monocular grayscale image sequences during various intra-vehicular activities across multiple ISS modules, along with pseudo ground-truth 6-DoF camera poses and detailed 3D CAD models of the ISS interior. Only a limited number of sequences provide point cloud data acquired from the Pico Flexx depth sensor. For these sequences, we first align the grayscale images with the corresponding depth images to ensure proper cross-modal consistency before evaluation.

A. Manhattan World Detection

We demonstrate the effectiveness of our method in accurately detecting the Manhattan world under highly cluttered ISS environments. Fig. 9 illustrates the qualitative results of our method. The first column shows all detected line segments using the LSD without any filtering. It contains a large number of outliers caused by dynamic objects such as wires, bags, and laptops. In the second column, structural landmarks extracted from the digital twin CAD model are highlighted, The third column presents the matched structural landmarks in the grayscale images after outlier rejection. In the fourth column, lines are clustered according to their consistency with the estimated vanishing directions. Finally, the inferred Manhattan world orientation is visualized in the last column. Our approach can robustly identify the MW structure, even in complex and dynamic environments like the ISS.

B. Rotational Motion Tracking

1) Evaluation Criteria: We measure the mean value of the absolute rotation error (ARE) [27] 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 method against state-of-the-art approaches, including both structural model-based and general-purpose methods. We use 6-DoF ground-truth camera poses generated through an offline Structure-from-Motion (SFM) process, as described in [34].

2) Methods for Comparison:

- LPIC [10]: A lightweight visual compass that estimates camera orientation using a single dominant plane and parallel lines, designed specifically for MW environments. We include it as a representative of low-overhead rotation estimation methods.
- ManhattanSLAM [25]: A SLAM system that combines points, lines, and planes while modeling a mixture of Manhattan frames. It serves as a strong MW-based full SLAM baseline.
- U-ARE-ME [35]: A monocular rotation estimation method that leverages Manhattan structures without requiring a full SLAM pipeline. We compare it against a recent lightweight method specialized for MW.
- GS-SLAM [36]: A dense mapping approach using Gaussian Splatting and RGB-D input. It represents state-of-the-art dense visual SLAM under structured environments.
- ORB-SLAM3 [37]: A widely used point feature-based SLAM system supporting monocular, stereo, and RGB-

D sensors. We include it as a general-purpose SLAM baseline.

3) Experimental Results: Table II compares the average ARE results of the proposed and state-of-the-art methods, and Fig. 10 presents the absolute rotation error evaluated at each frame. The *iva_kibo_rot* sequence involves complex rotational camera motion, while *iva_kibo_trans* exhibits less rotational motion, with the total traveling rotation below 10 degrees.

While most methods successfully track the nearly static camera motion in the *iva_kibo_trans* sequence, structural model-based RGB-D approaches (LPIC, ManhattanSLAM, U-ARE-ME) fail to initialize and track structural models due to cluttered environments in the *iva_kibo_rot* sequence, resulting in inaccurate pose estimation. U-ARE-ME, relying solely on RGB images for Manhattan world detection and tracking, exhibits the largest ARE. General-purpose methods (ORB-SLAM3, GS-SLAM) suffer from accumulated errors in the *iva_kibo_rot* sequence.

The proposed method achieves accurate and robust rotation estimation in highly cluttered environments with the help of our digital twin-based outlier rejection.

C. Runtime Analysis

We analyze the runtime of each module that constitutes the proposed method, as shown in Table III. First, the image processing step, including line detection using LSD [31], takes the longest runtime in the proposed method, approximately ∼50 ms for processing 15 image lines. Note that the computational load of image processing, such as line detection (LSD [31]), can vary significantly depending on various conditions, including input image size and the length of detected lines. Subsequently, we compute surface normals from the depth images and track the dominant plane, which takes approximately ∼10 ms. The initial Manhattan world detection, finding the San Francisco world (SFW) model, and tracking the corresponding SFW frames take about ∼7 ms, ∼2 ms, and ∼2 ms, respectively. The total computation time of the proposed digital twin-based method, excluding image processing such as line detection, is approximately ∼20 ms per image frame.

D. Implementation Details

We have implemented and tested the proposed method in MATLAB R2023b on a desktop computer with an Intel Core i5-12400F (2.50 GHz) CPU and 32 GB memory. The system is also equipped with an NVIDIA GeForce RTX 3060 GPU, which was used for real-time differentiable rendering in GS-SLAM [35] and for running deep learning-based methods such as U-ARE-ME [36] and line matching [32] in our method.

VI. DISCUSSION

A. Limitations

While the proposed method effectively improves rotational motion estimation in cluttered environments using a digital twin, several practical limitations remain. The primary limitation is the reliance on the completeness and accuracy of

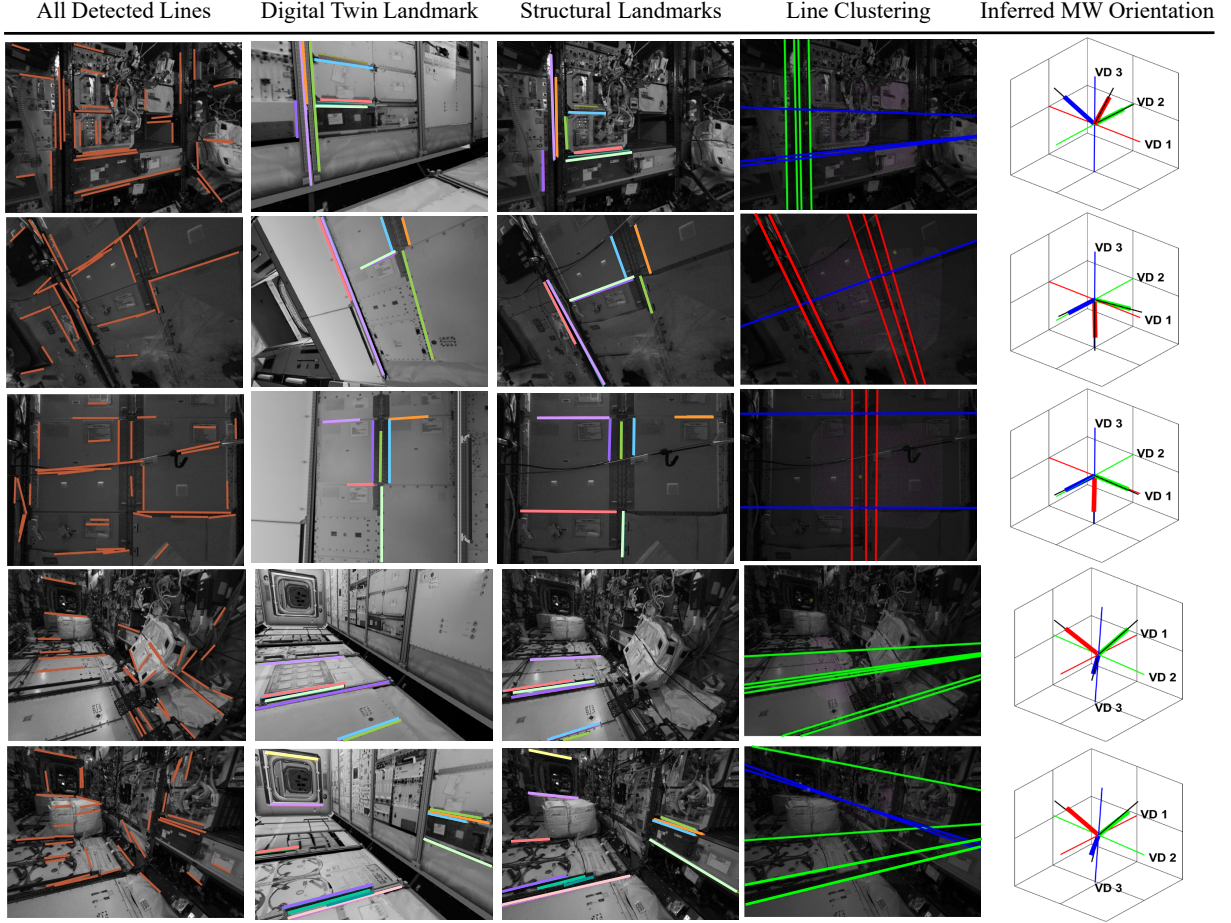| All Detected Lines | Digital Twin Landmark | Structural Landmarks | Line Clustering | Inferred MW Orientation |

Fig. 9. Qualitative results of Manhattan world detection using the proposed method. The line colors in the second and third columns indicate that lines with the same color are matched to each other. In the rightmost column, colored thick and thin lines denote the estimated 3-DoF camera orientation and the MW, and the black lines represent the ground-truth camera orientation.

TABLE II
COMPARISON OF METHODS ON ASTROBEE DATASETS.

Dataset		Structural Model-based Methods				General-purpose Methods		Length (deg)	# of frame
		Proposed	LPIC	ManhattanSLAM	U-ARE-ME	GS-SLAM	ORB-SLAM3		
Astrobee Dataset	kibo_rot 1	**1.75**	<u>2.32</u>	4.26	63.67	7.17	4.09	172	164
	kibo_rot 2	**1.09**	<u>2.74</u>	2.94	49.56	14.53	3.42	93	136
	kibo_trans	<u>1.44</u>	3.56	1.66	32.03	**1.17**	1.47	16	231
	Average	**1.43**	<u>2.87</u>	2.95	48.42	7.62	2.99	–	–

TABLE III
RUNTIME ANALYSIS OF PROPOSED METHOD

Module	Runtime
Preprocessing (Line Detection)	32.62 ms
Surface Normal & Mean Shift	10.02 ms
Digital Twin-Aided Outlier Rejection	3.07 ms
Single Line and Plane RANSAC	6.67 ms

the digital twin model. As illustrated in Fig. 12, mismatches between the digital twin 3D CAD model and the real ISS environment limit the effectiveness of our approach. These discrepancies arise because the CAD model only approximates the overall appearance of the ISS using simple textures, resulting in a lack of realistic details and missing temporary modifications. In some cases, the textures are also incorrectly oriented or misaligned, further reducing the accuracy of landmark matching and leading to false rejections or missed detections of valid line features. By refining the 3D CAD model with more detailed and realistic 3D textures, our method's efficiency and accuracy can be further maximized, showing high potential for enhanced performance in real-world applications.

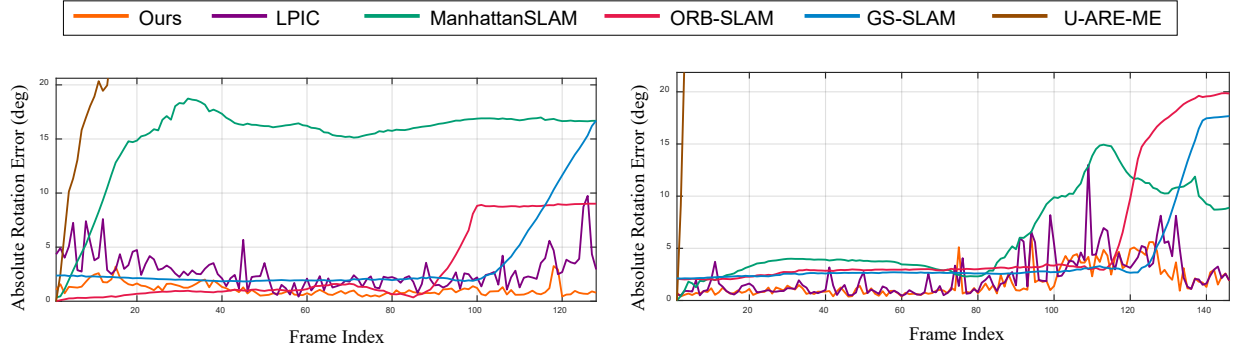In addition, the proposed method focuses solely on ro-

Fig. 10. Comparison of Absolute Rotation Error (ARE) in degrees against state-of-the-art methods on the Astrobee dataset.
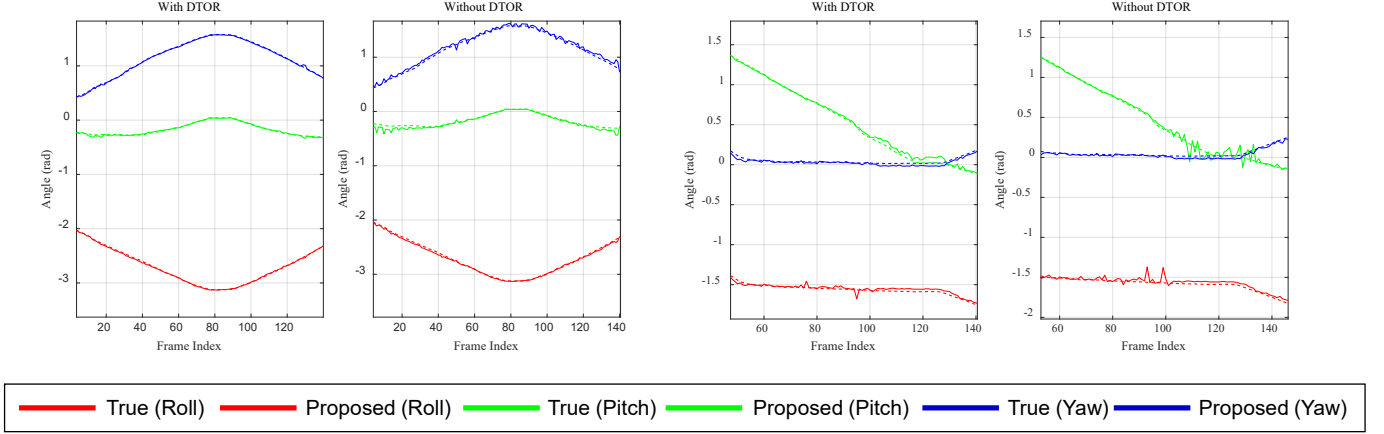

Fig. 11. Ablation study on the effectiveness of the proposed Digital Twin-Based Outlier Rejection (DTOR).
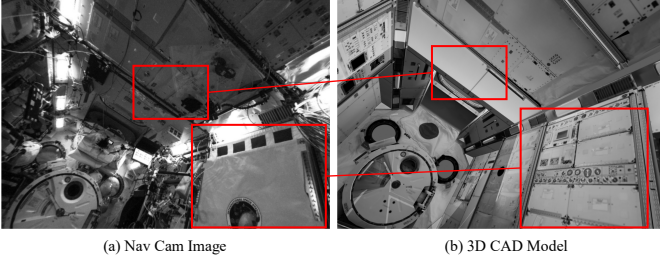

(a) Nav Cam Image   (b) 3D CAD Model

Fig. 12. The mismatched regions (marked by boxes) between the ISS digital twin 3D CAD model and the real environment prevent correct line feature matching, which ultimately limits the effectiveness of our proposed method.

tational motion estimation and does not address translation estimation, which is essential for achieving full 6-DoF pose estimation. Incorporating translation estimation remains challenging, particularly in environments where reliable point correspondences are difficult to obtain due to repetitive patterns and occlusions. Our digital twin-based outlier rejection (DTOR) approach can also be effectively extended to address translation estimation. For example, structural landmarks verified through the digital twin can provide reliable anchor points for constraining translation, especially in cluttered environments where point-based methods struggle.

Furthermore, the current evaluation is limited by the lack of sequences containing depth data, which prevented us from performing extensive validation across sufficiently diverse scenes. Recent datasets collected by the Intelligent Robotics Group at NASA Ames Research Center, including those from 2024, contain significantly more depth data. Our research group, the *Machine Perception and Intelligence Lab*, is currently working on constructing enhanced datasets based on these collections, including pseudo ground-truth annotations, to enable more comprehensive evaluations in future studies.

To address these limitations, future work will explore data-driven refinement of the digital twin using real-world sensor data collected from the ISS. By adapting the CAD model to more accurately reflect the current environment, we aim to improve feature correspondence reliability and enhance the effectiveness of digital twin-based outlier rejection.

B. Future Work

Future work will focus on enhancing the practical deployment of the proposed method and extending its capabilities toward full 6-DoF SLAM. Building on our successful validation in the ideal Manhattan World environment using the ICL-NUIM dataset [36], we plan to integrate the accurately estimated Manhattan frames from the ISS environment into the 3D Gaussian Splatting SLAM [36] framework. By aligning the

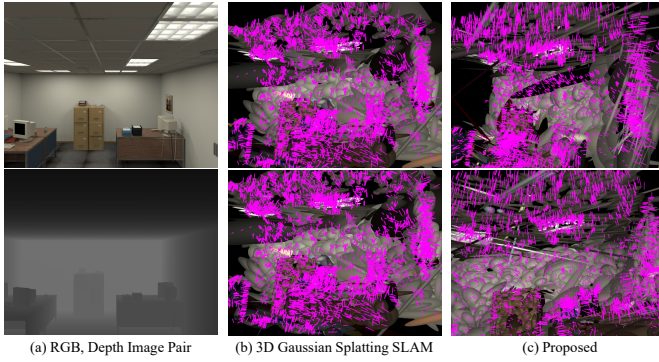(a) RGB, Depth Image Pair   (b) 3D Gaussian Splatting SLAM   (c) Proposed

Fig. 13. Visualization of 3D Gaussians aligned with the detected Manhattan frame. The normal vectors (magenta) of Gaussians, the shortest axis of 3D Gaussians, are regularized along the dominant structural directions, resulting in improved geometric consistency and more stable scene reconstruction.

normal vectors of 3D Gaussians with the dominant axes of the Manhattan World, we aim to regularize Gaussians' orientation, leading to improved scene structure representation and more stable mapping results. This effect is visually illustrated in Fig. 13, where the 3D Gaussians are successfully aligned according to the detected Manhattan frame, demonstrating enhanced structural consistency.

This normal alignment strategy addresses one of the key weaknesses of conventional Gaussian Splatting SLAM, which often suffers from poor normal estimation and fragmented reconstructions in complex environments.

In future experiments, we will apply this approach to real-world ISS datasets, leveraging the precise Manhattan World structures extracted from our system to guide normal alignment in highly cluttered and dynamic scenes. This extension will contribute to achieving real-time and drift-free navigation capabilities, ultimately advancing autonomous operations in space robotics platforms such as Astrobee.

CONCLUSION

We presented a digital twin-based outlier rejection method for robust Manhattan World (MW) detection and drift-free 3-DoF rotational motion estimation in cluttered and dynamic environments such as the International Space Station (ISS). By leveraging the ISS 3D CAD model as a digital twin, the proposed method effectively filters out clutter-induced outliers and extracts reliable structural features with minimal computational overhead. Experimental evaluations on the Astrobee dataset demonstrated that our method achieves state-of-the-art performance, significantly reducing rotation errors in highly cluttered environments.

ACKNOWLEDGMENT

REFERENCES

[1] L. Fluckiger and B. Coltin, "Astrobee robot software: Enabling mobile autonomy on the iss," Tech. Rep., 2019.

[2] M. G. Bualat, T. Smith, E. E. Smith, T. Fong, and D. Wheeler, "Astrobee: A new tool for iss operations," in *2018 SpaceOps Conference*, 2018, p. 2517.

[3] I. D. Miller, R. Soussan, B. Coltin, T. Smith, and V. Kumar, "Robust semantic mapping and localization on a free-flying robot in microgravity," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4121–4127.

[4] L. Mao, R. Soussan, B. Coltin, T. Smith, and J. Biswas, "Semantic masking and visual feature matching for robust localization," in *2024 International Conference on Space Robotics (iSpaRo)*. IEEE, 2024, pp. 1–7.

[5] H. Dinkel, J. Di, J. Santos, K. Albee, P. V. Borges, M. Moreira, R. Soussan, O. Alexandrov, B. Coltin, and T. Smith, "Astrobeecd: Change detection in microgravity with free-flying robots," *Acta Astronautica*, vol. 223, pp. 98–107, 2024.

[6] B. Coltin, J. Fusco, Z. Moratto, O. Alexandrov, and R. Nakamura, "Localization from visual landmarks on a free-flying robot," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4377–4382.

[7] R. Soussan, V. Kumar, B. Coltin, and T. Smith, "Astroloc: An efficient and robust localizer for a free-flying robot," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4106–4112.

[8] R. Soussan, M. Moreira, B. Coltin, and T. Smith, "Astroloc2: Fast sequential depth-enhanced localization for free-flying robots," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, p. To appear.

[9] P. Kim, B. Coltin, and H. J. Kim, "Linear rgb-d slam for planar environments," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 333–348.

[10] ——, "Indoor rgb-d compass from a single line and plane," in *CVPR*, 2018.

[11] H. Li, J. Zhao, J.-C. Bazin, and Y.-H. Liu, "Quasi-globally optimal and near/true real-time vanishing point estimation in manhattan world," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 3, pp. 1503–1518, 2020.

[12] H. Li, J. Zhao, J.-C. Bazin, W. Chen, Z. Liu, and Y.-H. Liu, "Quasi-globally optimal and efficient vanishing point estimation in manhattan world," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1646–1654.

[13] H. Li, K. Joo, and Y.-H. Liu, "Globally optimal and efficient vanishing point estimation in atlanta world," in *ECCV*, 2020.

[14] J. Ham, M. Kim, S. Kang, K. Joo, H. Li, and P. Kim, "San francisco world: Leveraging structural regularities of slope for 3-dof visual compass," *IEEE Robotics and Automation Letters*, 2024.

[15] J. M. Coughlan and A. L. Yuille, "Manhattan world: Compass direction from a single image by bayesian inference," in *ICCV*, 1999.

[16] 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.

[17] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: A versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[18] A. L. Majdik, Y. Albers-Schoenberg, and D. Scaramuzza, "Mav urban localization from google street view data," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 3979–3986.

[19] R. Sadli, M. Afkir, A. Hadid, A. Rivenq, and A. Taleb-Ahmed, "Map-matching-based localization using camera and low-cost gps for lane-level accuracy," *Sensors*, vol. 22, no. 7, p. 2434, 2022.

[20] R. Merat, G. Cioffi, L. Bauersfeld, and D. Scaramuzza, "Drift-free visual slam using digital twins," *IEEE Robotics and Automation Letters*, 2024.

[21] W. Wang, Y. Wu, Z. Jiang, and J. Qi, "A clutter-resistant slam algorithm for autonomous guided vehicles in dynamic industrial environment," *IEEE Access*, vol. 8, pp. 109 770–109 782, 2020.

[22] L. Gao, G. Battistelli, and L. Chisci, "Phd-slam 2.0: Efficient slam in the presence of missdetections and clutter," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1834–1843, 2021.

[23] J. Biswas and M. Veloso, "Episodic non-markov localization: Reasoning about short-term and long-term features," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 3969–3974.

[24] J. Straub, N. Bhandari, J. J. Leonard, and J. W. Fisher, "Real-time manhattan world rotation estimation in 3D," in *IROS*, 2015.

[25] R. Yunus, Y. Li, and F. Tombari, "Manhattanslam: Robust planar tracking and mapping leveraging mixture of manhattan frames," in *ICRA*, 2021.

[26] J. Straub, O. Freifeld, G. Rosman, J. J. Leonard, and J. W. Fisher, "The manhattan frame model—manhattan world inference in the space of surface normals," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 1, pp. 235–249, 2017.

[27] 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.

[28] J.-C. Bazin, Y. Seo, and M. Pollefeys, "Globally optimal consensus set maximization through rotation search," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 539–551.

[29] J.-C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys, "Globally optimal line clustering and vanishing point estimation in manhattan world," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 638–645.

[30] P. Kim and K. Joo, "Quasi-globally optimal and real-time visual compass in manhattan structured environments," *RA-L*, 2022.

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

[32] R. Pautrat, I. Suárez, Y. Yu, M. Pollefeys, and V. Larsson, "Gluestick: Robust image matching by sticking points and lines together," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9706–9716.

[33] 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

[34] S. Kang, R. Soussan, D. Lee, B. Coltin, A. M. Vargas, M. Moreira, K. Browne, R. Garcia, M. Bualat, T. Smith *et al.*, "Astrobee iss free-flyer datasets for space intra-vehicular robot navigation research," *IEEE Robotics and Automation Letters*, vol. 9, no. 4, pp. 3307–3314, 2024.

[35] A. Patwardhan, C. Rhodes, G. Bae, and A. J. Davison, "U-are-me: Uncertainty-aware rotation estimation in manhattan environments," *arXiv preprint arXiv:2403.15583*, 2024.

[36] H. Matsuki, R. Murai, P. H. Kelly, and A. J. Davison, "Gaussian splatting slam," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 18 039–18 048.

[37] C. Campos *et al.*, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *T-RO*, 2021.