What Breaks Monocular SLAM in Microgravity? An Initial Benchmark on Rotation-Dominant Astrobee ISS Sequences

Jungil Ham[1], Brian Coltin[2], Ryan Soussan[2], and Pyojin Kim[1,*]

Fig. 1. **Representative examples from our visual challenge taxonomy.** The rightmost column illustrates rapid attitude change, as evidenced by the motion blur. This paper focuses on the rotation-dominant subset.

***Abstract*— We present a challenge-oriented benchmark for Astrobee free-flyer monocular SLAM inside the International Space Station (ISS), guided by a visual challenge taxonomy. This initial study focuses on the rotation-dominant subset of this taxonomy. We evaluate monocular methods spanning the dense 3D Gaussian Splatting (3DGS) Simultaneous Localization and Mapping (SLAM) family and geometric foundation model based approaches on rotation-dominant monocular sequences, emphasizing dense monocular 3DGS pipelines for their explicit scene representation, which allows tracking, mapping, and rendering failures to be directly examined, and contrasting feed-forward geometric foundation models with tightly coupled SLAM systems built on top of them. Our evaluations reveal that robustness is highly local and motion-conditioned; optimized 3DGS pipelines improve stability, whereas geometric foundation models require tight coupling with temporal tracking to be effective. By exposing these algorithmic vulnerabilities, this work enables a more targeted approach to space robotics autonomy.**

I. Introduction

The Astrobee [1–3] is the free-flying robot to perform intra-vehicular activity tasks aboard the ISS U.S. Orbital Segment. To enable autonomous operation aboard the ISS, Astrobee's onboard localization and navigation software provides position estimates in the absolute ISS coordinate frame [4, 5]. This software relies heavily on robust visual perception, since localization must be performed from monocular imagery in a cluttered, fully three-dimensional interior under the representative ISS visual challenges summarized in Fig. 1.

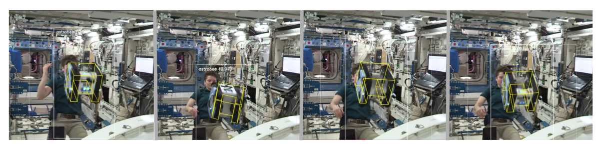

Fig. 2. **Unrestricted 360° rotational motion of Astrobee inside the ISS.** Astronaut Anne McClain is visible in the scene.

Visual localization for the Astrobee aboard the ISS has been actively studied to address the challenges of its long-term autonomy, including semantic localization, sparse-map localization, change-aware perception, and improved onboard localizers [4–9]. AstroLoc2 [5] is the current operational baseline, relying on prior-map matching with feature-based visual-inertial estimation for flight-proven onboard deployment. This motivates a challenge-aware benchmark: we need a systematic way to characterize which ISS visual conditions cause failures in practice and which visual mechanisms improve robustness in the spirit of challenge-aware SLAM evaluation [10–12].

We address this need by introducing a challenge-oriented benchmark for Astrobee ISS data, guided by our visual challenge taxonomy. Within this taxonomy, rotation-dominant motion is a particularly important stressor for space visual navigation, since microgravity free-flyers undergo un-

*Corresponding author.
[1]Jungil Ham and Pyojin Kim are with the Department of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea. {jungilham, pjinkim}@gist.ac.kr
[2]Brian Coltin and Ryan Soussan are with the Intelligent Robotics Group (KBR Inc.), NASA Ames Research Center, Moffett Field, CA 94035, USA. {brian.coltin, ryan.soussan}@nasa.gov

TABLE I
TRACKING ACCURACY ATE RMSE [M] ON
ASTROBEE MONOCULAR IMAGE SEQUENCES

Method	iva_kibo_rot	ff_return_rot
HI-SLAM2 [16]	**0.013**	**0.006**
WildGS-SLAM [17]	0.087	0.015
MUSt3R [18]	0.169	1.001
MASt3R-SLAM [19]	0.196	0.682
MonoGS [15]	0.462	1.223

restricted 3D attitude changes and the Astrobee ISS dataset contains abrupt in-place rotations and large viewpoint variation (see Fig. 2 and the rightmost column of Fig. 1). Such extreme rotations severely degrade operational baselines like AstroLoc2: a full 360-degree rotation removes prior-map landmarks from view, breaking global structural constraints [10, 11], while motion blur and severe viewpoint changes destroy the local correspondences needed to prevent localization drift [12, 13].

We therefore focus on rotation-dominant sequences and use local diagnostics, beyond sequence-level averages, to assess method-specific robustness to severe rotational motion. To isolate visual robustness, we restrict the study to monocular methods and further focus on dense monocular pipelines built on explicit scene representations, in which tracking, mapping, and rendering failures can be directly examined.

To examine which stabilizing mechanisms improve local robustness, we evaluate recent dense monocular SLAM methods based on 3DGS [14] (e.g., MonoGS [15], HI-SLAM2 [16], WildGS-SLAM [17]) alongside recent feed-forward geometric foundation model based methods (e.g., MUSt3R [18], MASt3R-SLAM [19]). Benchmarking these diverse approaches allows us to analyze the impact of global optimization, geometric foundation models, and uncertainty modeling on rotation-dominant sequences. Our main contributions are as follows:

- We introduce a visual challenge taxonomy for Astrobee ISS imagery as an organizing framework for challenge-aware benchmark design.
- We present an initial challenge-oriented benchmark study comparing dense monocular 3DGS SLAM methods and geometric foundation model based approaches on rotation-dominant ISS sequences.

II. MONOCULAR DENSE SLAM BENCHMARK

A. Metrics

We evaluate the compared methods on two representative Astrobee ISS monocular image sequences from the rotation-dominant subset of our challenge taxonomy, drawn from the publicly available Astrobee ISS Free-Flyer Dataset [20]. We report global absolute trajectory error (ATE) RMSE [21, 22] in Table I. However, because the global ATE can obscure short-duration failures, we additionally analyze local pose consistency using local windowed ATE, defined as the ATE RMSE after Sim(3) (3D similarity transform) alignment over a sliding window of $\Delta = 51$ frames (corresponding to a few seconds of motion at the dataset's NavCam frame rate, long enough to encompass abrupt rotational segments while remaining local), and report the result in Fig. 3.

B. Experimental Baselines

- **MonoGS** [15]: Vanilla monocular 3DGS SLAM baseline.
- **HI-SLAM2** [16]: 3DGS SLAM utilizing geometry-aware monocular priors and global consistency.
- **WildGS-SLAM** [17]: 3DGS SLAM featuring uncertainty-aware geometric mapping.
- **MUSt3R** [18]: Feed-forward geometric foundation model for multi-view 3D reconstruction.
- **MASt3R-SLAM** [19]: Real-time dense SLAM leveraging a geometric foundation model [23].

C. Results and Discussion

In Table I, HI-SLAM2 achieves the lowest overall trajectory error, followed by WildGS-SLAM, while MASt3R-SLAM and MUSt3R show larger global ATE and MonoGS performs worst on the selected ISS sequences.

As shown in Fig. 3, HI-SLAM2 and WildGS-SLAM maintain the lowest local error across most frames, including intervals of elevated rotation intensity, indicating stronger robustness to rotational stress. MonoGS is readily stressed by rotation-heavy and weak-overlap motion. MUSt3R shows the largest variance and the most frequent spikes, indicating that a geometric foundation model alone does not guarantee stable local estimation. Although MUSt3R preserves reasonable global ATE, its lack of temporal integration leads to large local ATE spikes. MASt3R-SLAM demonstrates that tightly coupling such geometric foundation models with sequential tracking significantly stabilizes local pose consistency.

Qualitatively (Fig. 4), the clearest differences are scene completeness and artifact quality. In the global and final views, the geometric foundation model based methods fail to recover much of the rear JEM section, likely because this region exceeds the reliable depth range of these models, which are trained predominantly on indoor datasets (e.g., ScanNet++, Habitat, ARKitScenes, BlendedMVS, Co3D) with scene depths concentrated within $\sim$5 m, whereas the JEM Pressurized Module spans 11.2 m end-to-end. In the close-up hatch view, MASt3R-SLAM and HI-SLAM2 recover finer local structure than MUSt3R and WildGS-SLAM, with MASt3R-SLAM benefiting from temporal coupling and HI-SLAM2 from geometry-aware monocular priors and global consistency. However, when MASt3R-SLAM revisits a similar scene after a large rotation, its loop closure fails and the same region is reconstructed as two disconnected submaps; tracking itself remains intact. WildGS-SLAM recovers a broader scene extent but exhibits substantial geometric distortion, floaters, and blur. Consistent with its lower local error in Fig. 3, HI-SLAM2 remains the most stable and visually coherent.

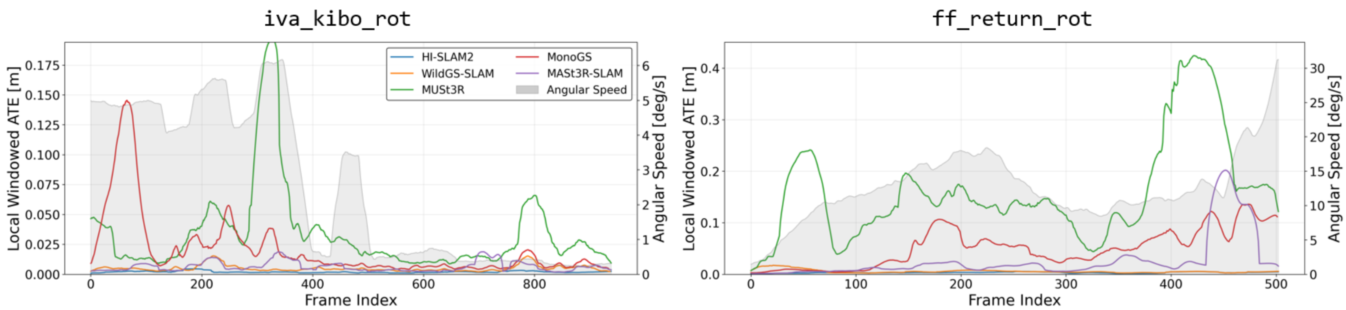

Fig. 3. **Local Windowed ATE [m] on the same two representative rotation-dominant Astrobee ISS sequences evaluated in Table I**. The overlaid gray curve shows the angular speed [deg/s] computed from ground-truth orientations. Elevated local pose errors concentrate around high-rotation intervals. HI-SLAM2 and WildGS-SLAM remain the most stable overall, while MUSt3R exhibits the largest variance and the most frequent spikes; MASt3R-SLAM partially mitigates these failures through temporal coupling.

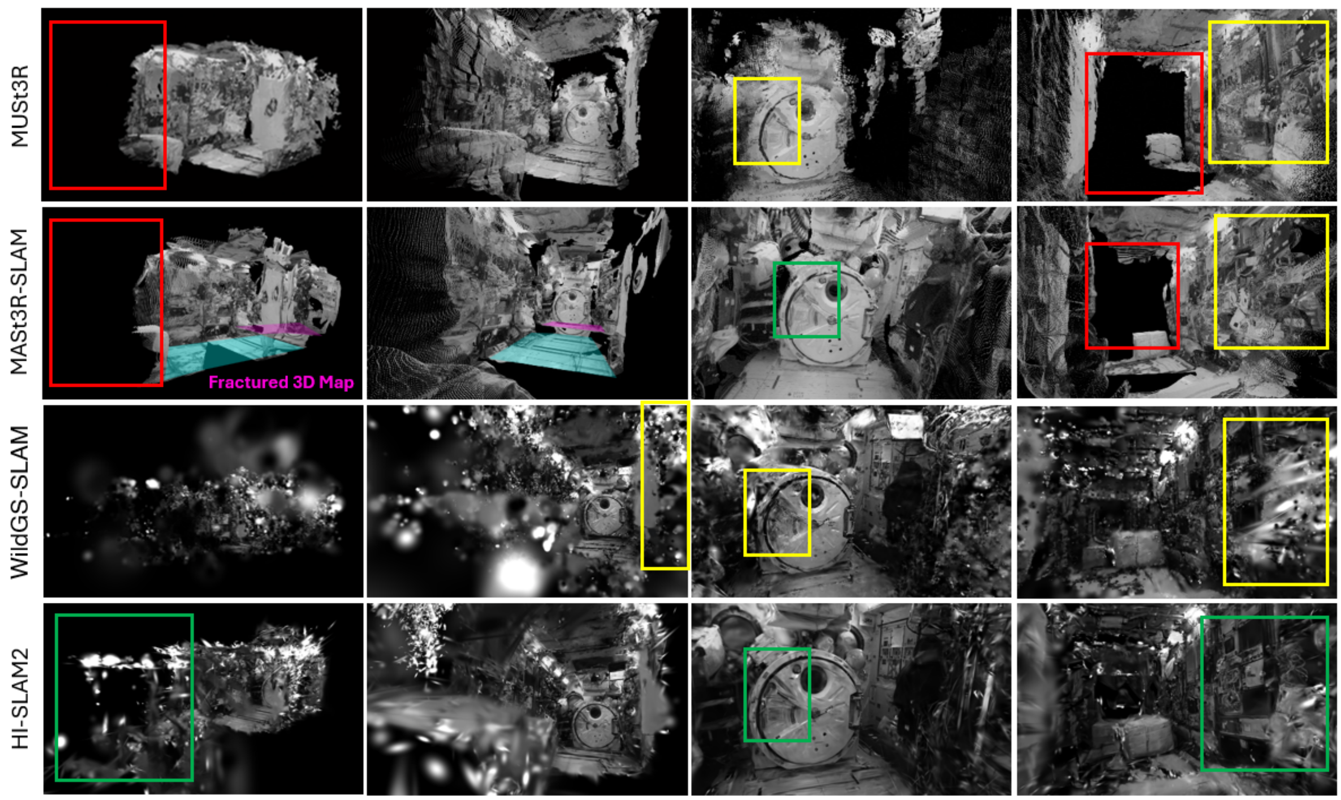

Fig. 4. **Reconstruction results on the iva_kibo_rot sequence.** Red boxes indicate unreconstructed regions, and yellow boxes highlight severe floaters or blurred artifacts. The geometric foundation model based methods fail to recover much of the rear Japanese Experiment Module (JEM) section, and MASt3R-SLAM reconstructs it as two disconnected sub-maps because loop closure fails on the post-rotation revisit. WildGS-SLAM exhibits substantial distortion and floaters. The green boxes show that MASt3R-SLAM and HI-SLAM2 recover the hatch details most faithfully, while HI-SLAM2 remains the most stable overall. MonoGS produced highly degraded reconstructions and is therefore omitted from this comparison.

Key Takeaways and Future Work. Our evaluations identify three failure modes that break monocular SLAM under rotation-dominant motion: (i) the plain MonoGS baseline loses tracking in rotation-heavy, weak-overlap segments; (ii) feed-forward foundation models without temporal integration (MUSt3R) exhibit large local-pose spikes despite reasonable global ATE; and (iii) tightly coupled hybrid systems (MASt3R-SLAM) stabilize tracking but expose loop closure as the dominant bottleneck under large-rotation revisits, fragmenting reconstructions into disconnected sub-maps—a critical dilemma in modern hybrid SLAM [24, 25]. These failures are primarily local and motion-conditioned rather than uniform over the trajectory. Augmenting the plain monocular baseline with structural constraints and global optimization [16, 17] substantially improves tracking stability and mitigates scale drift; HI-SLAM2 achieves the strongest local robustness and provides the most stable and visually coherent reconstruction overall.

Future work will expand our benchmark to the full spectrum of our ISS visual challenge taxonomy and develop adaptive SLAM mechanisms to support Astrobee's long-term autonomy.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2024-00358374) and the GIST Future-Leading Specialized Research Project grant funded by the GIST in 2026, and GIST-IREF from Gwangju Institute of Science and Technology (GIST). We also gratefully acknowledge the Astrobee team at NASA Ames Research Center for their continued support and contributions to space robotics research.

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] T. Smith, O. Alexandrov, J. Barlow, J. Benavides, M. Bualat, R. Carlino, B. Coltin, J. Cortez, E. Daley, J. Feller, *et al.*, "Astrobee: Free-flying robots for the international space station," *IEEE Transactions on Field Robotics*, 2026.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] X. Shi, D. Li, P. Zhao, Q. Tian, Y. Tian, Q. Long, C. Zhu, J. Song, F. Qiao, L. Song, *et al.*, "Are we ready for service robots? the openloris-scene datasets for lifelong slam," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 3139–3145.

[11] M. Bujanca, X. Shi, M. Spear, P. Zhao, B. Lennox, and M. Luján, "Robust slam systems: Are we there yet?" in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 5320–5327.

[12] W. Wang, D. Zhu, X. Wang, Y. Hu, Y. Qiu, C. Wang, Y. Hu, A. Kapoor, and S. Scherer, "Tartanair: A dataset to push the limits of visual slam," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4909–4916.

[13] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE transactions on robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
[14] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, *et al.*, "3d gaussian splatting for real-time radiance field rendering." *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.
[15] 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.
[16] W. Zhang, Q. Cheng, D. Skuddis, N. Zeller, D. Cremers, and N. Haala, "Hi-slam2: Geometry-aware gaussian slam for fast monocular scene reconstruction," *IEEE Transactions on Robotics*, vol. 41, pp. 6478–6493, 2025.
[17] J. Zheng, Z. Zhu, V. Bieri, M. Pollefeys, S. Peng, and I. Armeni, "Wildgs-slam: Monocular gaussian splatting slam in dynamic environments," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025, pp. 11 461–11 471.
[18] Y. Cabon, L. Stoffl, L. Antsfeld, G. Csurka, B. Chidlovskii, J. Revaud, and V. Leroy, "Must3r: Multi-view network for stereo 3d reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025, pp. 1050–1060.
[19] R. Murai, E. Dexheimer, and A. J. Davison, "Mast3r-slam: Real-time dense slam with 3d reconstruction priors," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 16 695–16 705.
[20] 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.
[21] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 573–580.
[22] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry," in *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2018, pp. 7244–7251.
[23] V. Leroy, Y. Cabon, and J. Revaud, "Grounding image matching in 3d with mast3r," in *European conference on computer vision*. Springer, 2024, pp. 71–91.
[24] Z. Teed and J. Deng, "DROID-SLAM: Deep visual slam for monocular, stereo, and rgb-d cameras," *NeurIPS*, 2021.
[25] L. Koestler, N. Yang, N. Zeller, and D. Cremers, "Tandem: Tracking and dense mapping in real-time using deep multi-view stereo," in *Conference on Robot Learning (CoRL)*, 2021.