

# Ground-Challenge: A Multi-sensor SLAM Dataset Focusing on Corner Cases for Ground Robots

Jie Yin <sup>†</sup>, Hao Yin <sup>†</sup>, Conghui Liang\* <sup>‡</sup> and Zhengyou Zhang <sup>‡</sup> (IEEE Fellow & ACM Fellow)

**Abstract**—High-quality datasets can speed up breakthroughs and reveal potential developing directions in SLAM research. To support the research on corner cases of visual SLAM systems, this paper presents Ground-Challenge: a challenging dataset comprising 36 trajectories with diverse corner cases such as aggressive motion, severe occlusion, changing illumination, few textures, pure rotation, motion blur, wheel suspension, etc. The dataset was collected by a ground robot with multiple sensors including an RGB-D camera, an inertial measurement unit (IMU), a wheel odometer and a 3D LiDAR. All of these sensors were well-calibrated and synchronized, and their data were recorded simultaneously. To evaluate the performance of cutting-edge SLAM systems, we tested them on our dataset and demonstrated that these systems are prone to drift and fail on specific sequences. We will release the full dataset and relevant materials upon paper publication to benefit the research community. For more information, visit our project website at <https://github.com/sjtuyinjie/Ground-Challenge>.

**Index Terms**—Data Sets for SLAM, Data Sets for Robotic Vision

## I. INTRODUCTION

INTELLIGENT ground robots have been widely used in industrial production and daily life, such as logistics, cleaning, warehouses, security, and food delivery. And navigation is the fundamental capability for these robots to execute these diverse tasks. To achieve reliable navigation, visual SLAM (Simultaneous Localization and Mapping) problem

has been researched for decades, with quite a few classical methods proposed [1].

A recent developing trend in visual SLAM is low-cost multi-sensor fusion, which has been verified to be a practical approach [2] to enhance the robustness to diverse scenarios. Different sensors can complement each other, maximizing the perceptual awareness of environments. One of the best example is that visual-inertial odometry (VIO) algorithms can significantly improve the tracking stability and accuracy in aggressive motion and textureless scenarios. While VIO systems have performed well in most cases, [3] has proven that this does not apply to ground vehicles. For generic movement patterns, a VIO system has only four unobservable directions (three for global translation and one for global yaw). However, ground vehicles are restricted from moving in a 2D plane, mostly along a straight line or a circular arc, and thus the IMU is not sufficiently activated. Therefore, the VIO system on the ground robot will suffer from additional DoF unobservability, such as the scale. To address this issue, [4] extends VINS-Mono [5] to incorporate low-frequency wheel-encoder data and keep the scale observable. Similarly, [6] proposes a RGB-D Encoder SLAM system for differential-drive robots. Most recently, [7] proposes an optimization-based visual-inertial-wheel tightly coupled odometry, which claims to work robustly in dark or overexposed conditions. Nonetheless, its performance has not been tested on any public dataset with ground truth trajectories.

We believe that progress in SLAM, like in the AI field, is highly data-driven [8]. Although there have

Authors <sup>†</sup> are independent researchers. Authors <sup>‡</sup> are with Tencent Robotics X Lab, Shenzhen, China. \* Corresponding Author: Conghui Liang ([isaac.liang@tencent.com](mailto:isaac.liang@tencent.com))

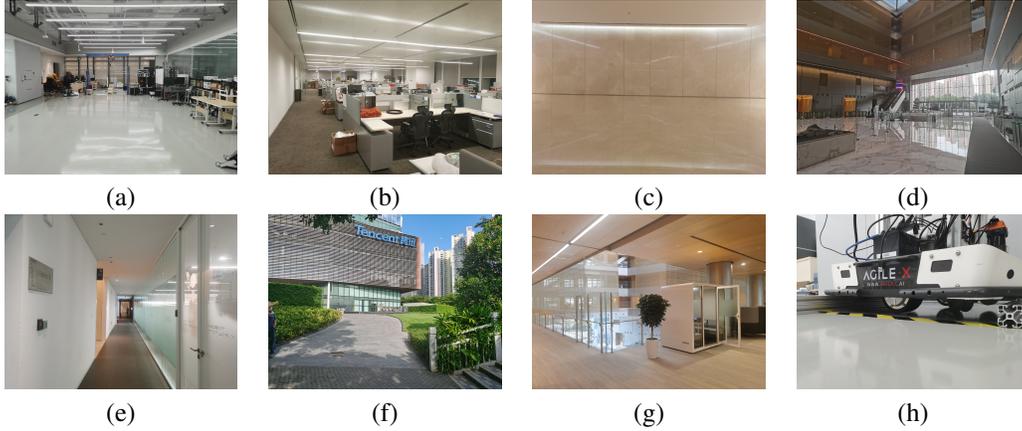


Fig. 1. Diverse scenarios included in our datasets: (a). A room under the motion capture system. (b). a richly-textured and well-lit office. (c). A wall that lacks texture. (d). A hall with smooth floors. (e). A narrow corridor. (f). An outdoor slope. (g). Aisles with carpet. (h). Hanging the robot on a bracket.

been extensive public datasets available to evaluate different SLAM algorithms, most of these datasets are outdated and do not challenge cutting-edge SLAM algorithms. In our opinion, those datasets focusing on challenging cases can more efficiently reveal the defects and limitations of existing algorithms. We notice that corner case detection in autonomous driving receive extensive concern from researchers [9] [10] because such cases could easily cause the navigation system to drift. Similarly, once the localization module of the robot fails, it might cause industrial accidents and even pose potential threats to human safety as well. Nonetheless, to our knowledge, there is currently not much literature discussing the corner cases of robot navigation, which is not conducive to the safety of real-world robot applications.

To fill this gap, we present a novel SLAM dataset for ground robots, which aims to challenge existing cutting-edge SLAM systems with corner cases and thus promotes the progress of the multi-sensor fusion SLAM algorithm. The challenges of our datasets lie in two areas: specific movement patterns and sensor failures, which will be elaborated in subsequent sections. Some scenarios covered in our datasets are visualized in Figure 1. Our major contributions are summarized as follows:

- We collect a novel visual SLAM dataset for ground robots with a rich pool of sensors in diverse environments both indoors and outdoors. Particularly, the dataset covers a series of challenging sequences including sensor failures and specific movement patterns.
- State-of-the-art SLAM algorithms of different settings are tested on our benchmark. And the results indicate these systems are not robust enough for situations such as sensor failures.
- To facilitate the research on corner cases of robot navigation, we will release the full dataset with ground truth trajectories and the configuration file of each tested algorithm upon paper publication.

## II. RELATED WORKS

### A. SLAM Datasets for Ground Robots

Most existing SLAM datasets are collected by UAVs [11] or cars [12], but only a few are targeted at ground robots. For instance, Rawseeds [13] and UTIAS [14] provide RGB images only, thus making them unsuitable for evaluating multi-sensor fusion systems. The Rosario dataset [15] is rich in sensor variety, yet is specifically designed for agricultural environments. M2DGR [2] captures diverse indoor and outdoor scenarios, including some challenging

TABLE I  
COMPARISON OF SLAM DATASETS ON GROUND ROBOTS

Dataset	Environment	RGB	Depth	IMU	Odom
Rawseeds [13]	In/Outdoors	✓			
UTIAS [14]	Indoors	✓			
TUM RGBD [17]	Indoors	✓	✓	✓	
NCLT [18]	In/Outdoors	✓		✓	
Rosario [15]	Outdoors	✓		✓	✓
OpenLORIS [16]	Indoors	✓	✓	✓	✓
SubT-Tunnel [19]	Outdoors	✓		✓	
M2DGR [2]	In/Outdoors	✓		✓	
Ours	In/Outdoors	✓	✓	✓	✓

scenes like elevators and darkrooms, but doesn't contain wheel odometer information which is essential for multi-sensor fusion SLAM algorithms due to its low cost and high precision. OpenLORIS [16] offers rich sensor types in visual challenging scenarios such as highly dynamic markets and poorly exposed corridors, but wheel challenges or motion challenges are not included.

### B. Corner Cases

Corner cases, i.e., extreme and non-predictable situations, are a popular research topic in autonomous driving [20]. Although infrequent, these cases can potentially threaten the security and reliability of autonomous navigation systems. Corner cases exist in robot navigation tasks as well. To address such challenging scenarios, researchers have proposed various methods, such as RGB-D SLAM [21] and DS-SLAM [22], to handle dynamic environments, and GVINS [23] to deal with degenerate cases including low-speed movement, less than four visible satellites, and GNSS-denial environments. Additionally, [24] proves that their method is robust in aggressive motions and a visual texture-less white wall. Nonetheless, we note that there are still plenty of corner cases that tend to be overlooked, such as wheel slippage, motion blur, and complete visual occlusion. There is a lack of SLAM datasets specifically designed for studying these corner cases, which is a gap yet to be filled. To sum up, it is urgent and critical to collect a novel SLAM dataset with rich sensor types, precise calibration, and sufficient challenge to

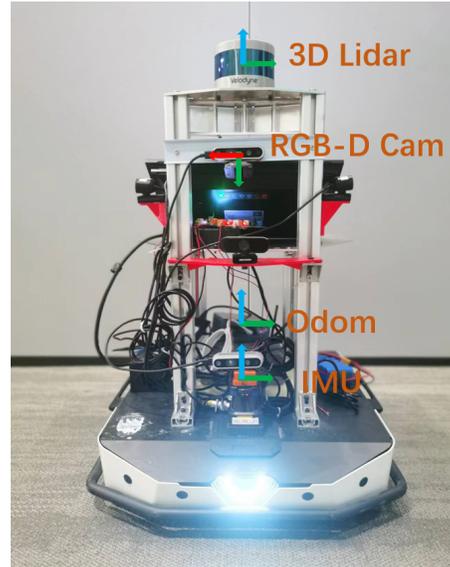


Fig. 2. Our ground robot for data collection. Red is x-axis, green is y-axis, and blue is z-axis.

support studies on corner cases, particularly sensor failures.

## III. THE GROUND-CHALLENGE DATASET

### A. Sensor setup

We construct a ground robot for data collection and the sensor locations on the robot are shown in Figure 2. The chassis is equipped with a front-view VI-Sensor (Visual-Inertial Sensor) that captures RGB and depth images along with 6-axis IMU's measurements. Driven by two driving wheels providing odometer information and four assisting wheels, the robot also has a high-precision 9-axis Xsens IMU and a 16-beam 3D LiDAR.

The ground truth trajectories and point clouds are generated by the Velodyne LiDAR and the Xsens IMU using Fast-LIO2 [25], a state-of-the-art LiDAR-based SLAM system. To evaluate its performance, we compared the high-precision trajectories generated by a motion capture system with 16 infrared cameras to those generated by Fast-Lio2. The experiment revealed that Fast-LIO2 can reach a positioning accuracy of 3cm in a small-scale (15m x 15m) indoor room. Additionally, as reported

TABLE II  
SPECIFICATIONS OF SENSORS

Device	Type	Spec.	Freq.(Hz)
VI-sensor	Realsense D435I	RGB: 640*480, 69 H-FOV, 42.5V-FOV	15
		Depth: 640*480, 0.1~10 meters	15
IMU	Xsens Mti-300	IMU: 6-axis	200
Wheel Odometer	AgileX	9-axis	400
		2D	25
LiDAR	Velodyne VLP-16	16 beam,360 H-FOV,30V-FOV	10

in [25], Fast-LIO2 can achieve less than 0.1m end-to-end error in an outdoor trajectory spanning 1000 meters. Thus, considering that it is difficult for visually-based SLAM algorithms to achieve similar accuracy in challenging scenarios, we use the result of Fast-LIO2 as the pseudo-ground-truth trajectory.

[28], which includes the white noise and random walk of both the gyroscopic and accelerometer measurements. We choose the IMU frame as the reference to calibrate the extrinsic parameters (relative poses) between sensors, and employ the toolbox from [29] for calibrating the extrinsic parameters between cameras and IMU.

TABLE III  
AN OVERVIEW OF SCENARIOS IN OUR DATASET.

Scenario	Seq. Num	Dist./m	Vel./(m/s)	Size/GB
Office	3	75.5	0.46	4.7
Room	3	102.1	0.66	4.6
Darkroom	3	92.0	0.45	6.1
Wall	3	86.7	0.46	5.6
Motionblur	3	166.6	1.15	4.3
Occlusion	4	273.8	0.82	9.9
Roughroad	3	68.1	0.37	5.4
Slope	2	128.5	0.66	5.7
Hall	3	263.3	0.87	8.7
Loop	2	371.8	1.12	9.9
Corridor	2	68.1	0.37	5.4
Rotation	3	12.4	0.06	5.4
Static	2	1.9	0.00	2.7
Total	36	1780.0	—	78.8

### B. Synchronization and Calibration

We capture all the data using the ROSbag tool in the Robot Operating System (ROS). The RGB camera and 6-axis IMU embedded in the Realsense D435I are hard-synchronized, while the depth images are pixel-by-pixel aligned to the RGB images. The 3D LiDAR and 9-axis IMU are software-synchronized by triggering data capture at the same instance. To calculate the camera intrinsics of pinhole cameras, we use the MATLAB Camera Calibration Toolbox. To calibrate the internal parameters of the IMU, we use the toolbox from

### C. Data collection

We provide an overview of our dataset in Table III. All data was captured using the Rosbag tool within the Robot Operating System (ROS). The recording process is as follows: First, we recorded Office and Room sequences, where the robot moves slowly in a well-lit and textured office or room respectively, to test the performance of different algorithms in normal situations. Subsequently, we designed a series of corner case experiments from three aspects: visual challenge, wheel odometer challenge, and particular movement pattern, which are presented as follows:

1) *Visual Challenge*: In our experiments, we manipulate the robot to move in a room with poor illumination (Darkroom sequences), back and forth in front of walls lacking texture (Wall sequences), and through scenarios of varying degrees of occlusion (Occlusion sequences). Figure 3 (a) shows sequences Occlusion1~2, which involves a person walking in front of the robot and causing intermittent partial occlusion. Figure 3 (b) displays sequence Occlusion3, in which the camera is covered with the palm repeatedly. In sequence Occlusion4 (Figure 3 (c)), a piece of black tape is attached to the camera’s lens to completely block its view, disabling feature extraction and matching for visual SLAM. Furthermore, Motionblur sequences are generated by rapidly translating and rotating the

TABLE IV  
ATE RMSE (M) OF SLAM SYSTEMS ON SAMPLE SEQUENCES

Type	Sequence / Method	VINS-Mono [5]	VINS-RGBD [26]	VIW-Fusion [7]	EKF [27]	Raw Odometer
Normal	Office3	0.35	0.31	0.18	0.16	<b>0.16</b>
	Darkroom2	1.66	0.82	0.53	<b>0.22</b>	0.30
Visual Challenge	Wall2	1.21	1.00	<b>0.15</b>	32.91	0.57
	Motionblur3	9.37	32.31	0.78	<b>0.42</b>	0.79
	Occlusion4	* <sup>a</sup>	*	*	<b>0.22</b>	0.23
Wheel Challenge	Roughroad3	0.17	25.52	0.14	<b>0.11</b>	<b>0.11</b>
	Slope1	9.41	2.84	<b>0.65</b>	0.89	0.89
	Hall1	7.06	94.27	<b>0.85</b>	2.16	2.79
	Loop2	6.09	<b>3.44</b>	9.23	16.70	17.61
Motion Challenge	Corridor1	4.48	<b>0.85</b>	1.12	1.78	2.17
	Rotation3	29.12	0.19	0.18	<b>0.14</b>	<b>0.14</b>
	Static1	*	*	*	5.61	<b>3.54</b>

<sup>a</sup>If a SLAM system fails to initialize or track frames less than a half of total frames, we mark it \*.

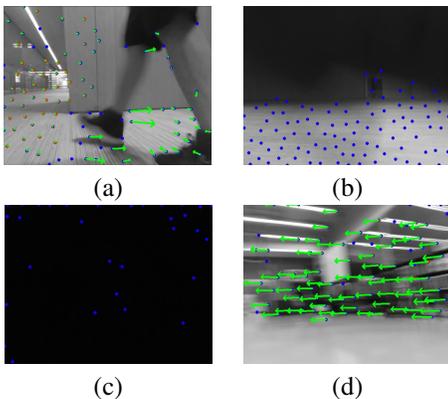


Fig. 3. (a) Moving feet. (b) occluding the camera with a palm. (c) complete occlusion. (d) motion blur

robot, creating motion blur for cameras (Figure 3 (d)).

2) *Wheel Odometer Challenge*: The Hall and Loop sequences are collected in a hall with smooth ground and a heavily carpeted aisle loop, respectively, where the wheels slip significantly. Moreover, we record Roughroad sequences to test the performance of the localization algorithm on rough roads.

3) *Particular Moving Patterns*: In the Sequences Corridor1 and Corridor2, the robot moves forward in a zigzag shape and straight forward, respectively. In the zigzag route, motion blur and less overlapping between adjacent image frames will lead to errors in feature matching. In the Rotation sequence,

the robot only rotates and hardly translates, which makes it difficult for vision-based algorithms to estimate the depth of feature points by triangulation. In the Static sequences, the robot stands still on a bracket, and we control its wheels to move in different directions through the handle. This experiment aims to test whether SLAM systems coupled with the wheel odometer can work well when the robot wheel is suspended. Finally, we operate the robot from a flat surface to another, passing through a slope. In this experiment, since the wheel odometer only provides two-dimensional speed observations, it could be misleading to estimate three-dimensional trajectories.

#### IV. EVALUATION

The features of all the sequences are described on our project website. We evaluated some SLAM systems with different sensor configurations on twelve representative sequences from our dataset. The tested algorithms are ORB-SLAM3 [30], an optimization-based SLAM system; VINS-Mono [5], one of the state-of-the-art monocular visual-inertial systems; VINS-RGBD [26], a fusion algorithm of RGB-D and IMU information based on the VINS-Mono [5] framework; and VIW-Fusion [7], a tightly-coupled visual-inertial-wheel system featuring online extrinsic calibration and wheel-aided initialization. Also, we use an EKF algorithm [27] for fusion of IMU and wheel odometer.

The EVO tool [31] was used to align all the estimated trajectories with ground truth trajectories

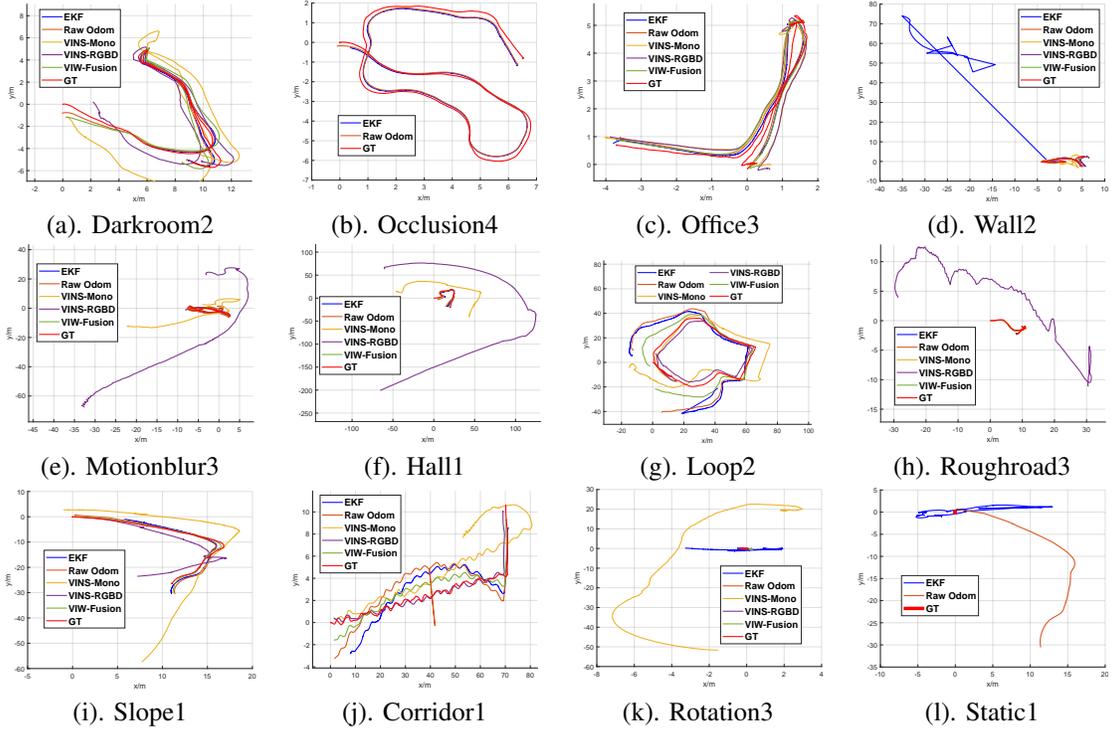


Fig. 4. Estimated and ground-truth (GT) trajectories of 12 sample sequences are visualized on the x-y plane.

to obtain the ATE RMSE [17]. The quantitative results are shown in Table IV, with the estimated trajectories in 2D plotted in Figure 4. Since most of the selected sequences are highly challenging (even with sharp turns), ORB-SLAM3 (both monocular-inertial and RGBD-inertial version) performed poorly on most of our test sequences, with frequent tracking failures (less than 50% of successfully tracked frames), initialization failure, or scale drift. In contrast, SLAM algorithms with multi-sensor fusion (like VIW-Fusion [7]) achieved better localization results but failed in some specific scenarios as well. We discuss the experiment results in detail as follows:

*a) Normal Situation:* The ATE RMSE results on Sequence Office3 indicate that existing localization methods can perform well when the motion mode matches the assumptions of these algorithms and all the sensors work well.

*b) Vision Challenge:* In Sequence Darkroom2 and Motionblur3, VINS-Mono [5] and VINS-RGBD [26] drift a lot due to visual failures, while Wheel odometer based algorithms work more robustly in this case. In Sequence Occlusion4, all the vision-based methods including VIW-Fusion [7] fail to initialize because of poor feature extraction. This finding indicates that VIW-Fusion [7] has not been adequately designed to handle adverse conditions. A more prudent strategy may be to combine the wheel odometer and IMU to output a trajectory when a visual sensor failure is detected.

*c) Wheel Odometer Challenge:* In the sequences Roughroad3 and Slope1, vision-based systems perform worse than wheel odometer-based algorithms due to inaccurate scale estimation in aggressive motion. In Sequence Hall1, VINS-Mono [5] and VINS-RGBD [26] drift significantly due to ground reflection and faraway feature points.

Here, VIW-Fusion [7] maintains satisfactory positioning performance even with slight wheel slippage, demonstrating the advantages and necessity of multi-sensor fusion in complex scenarios. However, when the wheels slip more severely in Sequence Loop2, the significant deviation caused by the wheel odometer increases the localization error of estimated trajectories. This can be attributed to two main reasons: current algorithms lack the ability to detect wheel slippage, and the angular velocity provided by the wheel speedometer is not accurate, leading to the long-term divergence of the estimated trajectory. To reduce the accumulation of errors, it is suggested that IMU's angular velocity measurement be used instead of the wheel odometer's.

*d) Particular Movement Patterns:* In Sequence Corridor1, the zigzag movement of the robot not only fails the feature extraction but also leads to severe wheel slippage. Therefore, all the tested algorithms cannot accurately estimate the trajectory. In Sequence Rotation1, pure rotation causes severe errors in depth estimation by VINS-Mono's triangulation, while the remaining tested systems perform well thanks to measurements from other sensors. Finally, in Sequence Static1, VIO systems cannot be initialized successfully due to the lack of IMU excitation. Since the wheels are still moving after suspension, the wheel odometer-based methods mistake the robot being in motion.

In summary, VINS-Mono [5] is most likely to generate catastrophic localization results in corner cases, and VINS-RGBD [26] can also inevitably fail when severe camera failures occur. We have noticed that the wheel odometer alone can achieve good results in most situations, except for severe wheel slippage. Integrating the IMU and the wheel odometer through the EKF [27] can achieve higher accuracy than the raw odometer. Nonetheless, the trajectory of the EKF can shake violently in the initialization phase due to the inaccuracy in the initial covariance estimation (this part was manually eliminated in our experiment). VIW-Fusion [7] can achieve satisfying accuracy and robustness in most sequences, but its initialization in visual failure needs improvement. Furthermore, it lacks consideration for wheel slippage, and its adopted dead

reckoning model will diverge in a long trajectory due to inaccurate angular velocity estimates.

The experiments conducted demonstrate the validity and value of our dataset as a benchmark for existing SLAM systems. The results further suggest that there is still much room for improvement in current cutting-edge multi-sensor fusion algorithms for real-world applications. Sensor failures, such as complete occlusion and wheel suspension, can be fatal for single-sensor-based methods; however, multi-sensor fusion systems should be designed to be more robust in these cases. For instance, we posit that a reliable visual-IMU-wheel system should be able to explicitly identify scenarios where visual observations are inaccurate and respond accordingly (e.g. disable visual information and rely only on wheel odometer and IMU). Nevertheless, to our knowledge, corner case identification and troubleshooting have been scarcely addressed in prior work. Therefore, we provide this dataset to support relevant researches.

## V. CONCLUSION

We present Ground-Challenge, a novel ground robot dataset to encourage breakthroughs in multi-sensor fusion SLAM algorithms. Specifically, we have crafted a series of corner case experiments, including sensor failures in diverse environments, to challenge current cutting-edge SLAM systems. We have tested these systems on our dataset and analyzed their limitations in various scenarios, thus providing potential developing directions for SLAM. We are committed to continually updating our benchmark dataset. Specifically, we will mount 2D and 3D LiDAR on the robot, design experiments to invoke corner cases, and utilize higher-precision equipment such as motion capture systems to ensure accurate ground truth for LiDAR SLAM in our future work.

**Acknowledgement** Thank Tencent Robotics X Lab for support to this work.

## REFERENCES

- [1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

- [2] J. Yin, A. Li, T. Li, W. Yu, and D. Zou, "M2dgr: A multi-sensor and multi-scenario slam dataset for ground robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2266–2273, 2021.
- [3] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Consistency analysis and improvement of vision-aided inertial navigation," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 158–176, 2013.
- [4] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, "Vins on wheels," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 5155–5162.
- [5] 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.
- [6] D. Yang, S. Bi, W. Wang, C. Yuan, W. Wang, X. Qi, and Y. Cai, "Dre-slam: Dynamic rgb-d encoder slam for a differential-drive robot," *Remote Sensing*, vol. 11, no. 4, p. 380, 2019.
- [7] T. Zhuang, "Viw-fusion," <https://github.com/TouchDeeper/VIW-Fusion>, 2022.
- [8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [9] J.-A. Bolte, A. Bar, D. Lipinski, and T. Fingscheidt, "Towards corner case detection for autonomous driving," in *2019 IEEE Intelligent vehicles symposium (IV)*. IEEE, 2019, pp. 438–445.
- [10] K. Muhammad, A. Ullah, J. Lloret, J. Del Ser, and V. H. C. de Albuquerque, "Deep learning for safe autonomous driving: Current challenges and future directions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4316–4336, 2020.
- [11] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.
- [12] A. Ligocki, A. Jelinek, and L. Zalud, "Brno urban dataset-the new data for self-driving agents and mapping tasks," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3284–3290.
- [13] A. Bonarini, W. Burgard, G. Fontana, M. Matteucci, D. G. Sorrenti, and J. D. Tardos, "Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets," in *In proceedings of IROS*, vol. 6, 2006, p. 93.
- [14] K. Y. Leung, Y. Halpern, T. D. Barfoot, and H. H. Liu, "The utias multi-robot cooperative localization and mapping dataset," *The International Journal of Robotics Research*, vol. 30, no. 8, pp. 969–974, 2011.
- [15] T. Pire, M. Mujica, J. Civera, and E. Kofman, "The rosario dataset: Multisensor data for localization and mapping in agricultural environments," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 633–641, 2019.
- [16] 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.
- [17] 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 (IROS)*. IEEE, 2012, pp. 573–580.
- [18] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and lidar dataset," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.
- [19] J. G. Rogers, J. M. Gregory, J. Fink, and E. Stump, "Test your slam! the sub-tunnel dataset and metric for mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 955–961.
- [20] D. Bogdoll, J. Breitenstein, F. Heidecker, M. Bieshaar, B. Sick, T. Fingscheidt, and M. Zöllner, "Description of corner cases in automated driving: Goals and challenges," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1023–1028.
- [21] J. Liu, X. Li, Y. Liu, and H. Chen, "Rgb-d inertial odometry for a resource-restricted robot in dynamic environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9573–9580, 2022.
- [22] C. Yu, Z. Liu, X.-J. Liu, F. Xie, Y. Yang, Q. Wei, and Q. Fei, "Ds-slam: A semantic visual slam towards dynamic environments," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1168–1174.
- [23] S. Cao, X. Lu, and S. Shen, "Gvins: Tightly coupled gnss-visual-inertial fusion for smooth and consistent state estimation," *IEEE Transactions on Robotics*, 2022.
- [24] C. Zheng, Q. Zhu, W. Xu, X. Liu, Q. Guo, and F. Zhang, "Fast-livo: Fast and tightly-coupled sparse-direct lidar-inertial-visual odometry," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 4003–4009.
- [25] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-liv2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, 2022.
- [26] Z. Shan, R. Li, and S. Schwertfeger, "Rgbd-inertial trajectory estimation and mapping for ground robots," *Sensors*, vol. 19, no. 10, p. 2251, 2019.
- [27] M. I. Ribeiro, "Kalman and extended kalman filters: Concept, derivation and properties," *Institute for Systems and Robotics*, vol. 43, p. 46, 2004.
- [28] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-696, Aug. 2007.
- [29] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2013, pp. 1280–1286.
- [30] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam," *IEEE Transactions on Robotics*, 2021.
- [31] MichaelGrupp, "evo," <https://github.com/MichaelGrupp/evo>, 2018.