M2DGR: A Multi-sensor and Multi-scenario SLAM Dataset for Ground Robots

Jie Yin, Ang Li, Tao Li, Wenxian Yu, and Danping Zou*,

***Abstract*— We introduce M2DGR: a novel large-scale dataset collected by a ground robot with a full sensor-suite including six fish-eye and one sky-pointing RGB cameras, an infrared camera, an event camera, a Visual-Inertial Sensor (VI-sensor), an inertial measurement unit (IMU), a LiDAR, a consumer-grade Global Navigation Satellite System (GNSS) receiver and a GNSS-IMU navigation system with real-time kinematic (RTK) signals. All those sensors were well-calibrated and synchronized, and their data were recorded simultaneously. The ground truth trajectories were obtained by the motion capture device, a laser 3D tracker, and an RTK receiver. The dataset comprises 36 sequences (about 1TB) captured in diverse scenarios including both indoor and outdoor environments. We evaluate state-of-the-art SLAM algorithms on M2DGR. Results show that existing solutions perform poorly in some scenarios. For the benefit of the research community, we make the dataset and tools public. The webpage of our project is https://github.com/SJTU-ViSYS/M2DGR.**

Keywords: Data Sets for SLAM, Data Sets for Robotic Vision

I. INTRODUCTION

Intelligent ground robots have been emerging in a wide range of applications such as logistics, security, warehouses, cleaning, and food delivery [1]. In those applications, the robots need to work reliably in indoor or a mixture of indoor and outdoor scenes. SLAM (Simultaneous Localization and Mapping) is the critical enabling technology that allows those robots to navigate in those complex scenes, which can construct a map of the environment while simultaneously tracking the location of the robot within the map. Though SLAM research has made a substantial progress in the past decades [2], [3], existing solutions frequently perform poorly in practice. For example, visual SLAM may fail at texture-less or dark backgrounds, while LiDAR SLAM could have trouble with long corridors or open areas. SLAM may also become invalid when the robot takes some unusual actions, for instance, a robot moves into a lift and goes out to the new floor. Those failure cases motivate us to construct a dataset that includes more practical scenarios to facilitate SLAM research.

High-quality datasets can speed up breakthroughs and enable a fair comparison between different algorithms. However, most existing SLAM datasets are designed for autonomous driving or aerial robots as pointed in [4]. Those datasets are not the best fit for developing and evaluating algorithms of ground robots. Autonomous cars travel fast on streets and roads, while the ground robots move at a much lower speed in remarkably different surroundings, including both indoor and outdoor scenes. The aerial robots fly freely in 3D space and are also quite different from ground robots. Although there are a few datasets targeting ground robots [5], [6], they include only a few specific sensors or particular scenes. For logistics robots, catering robots, and service robots, challenging scenarios are frequently faced, like going into a lift or complete darkness, or going from outdoors to indoors. These situations may easily make existing SLAM methods fail, while they are seldom included in existing SLAM benchmark tests.

In this paper, we introduce a new dataset for SLAM research of ground robots, which includes both indoor and outdoor environments and contains a rich suite of sensors. The dataset contains trajectories in highly diverse scenes such as halls, lifts, corridors, and roads. Based on this dataset, we evaluate the state-of-the-art SLAM algorithms, including both LiDAR SLAM and visual SLAM. The results show that existing SLAM systems perform poorly in at least one situation, indicating further efforts are required to improve the SLAM performance. We summarize major contributions as follows:

- We collected large-scale sequences for ground robots with a rich sensor suite, which includes six surround-view fish-eye cameras, a sky-pointing fish-eye camera, a VI-sensor, an event camera, an infrared camera, a 32-beam LiDAR, an IMU, and two GNSS receivers. To our knowledge, this is the first SLAM dataset focusing on ground robot navigation with such rich sensory information.
- We recorded trajectories in challenging situations like entering lifts and complete darkness which are commonly faced in practical applications, whereas they are not present in previous datasets.
- We launched a comprehensive benchmark where we evaluated existing state-of-the-art SLAM algorithms of various designs and analyzed their characteristics and defects.

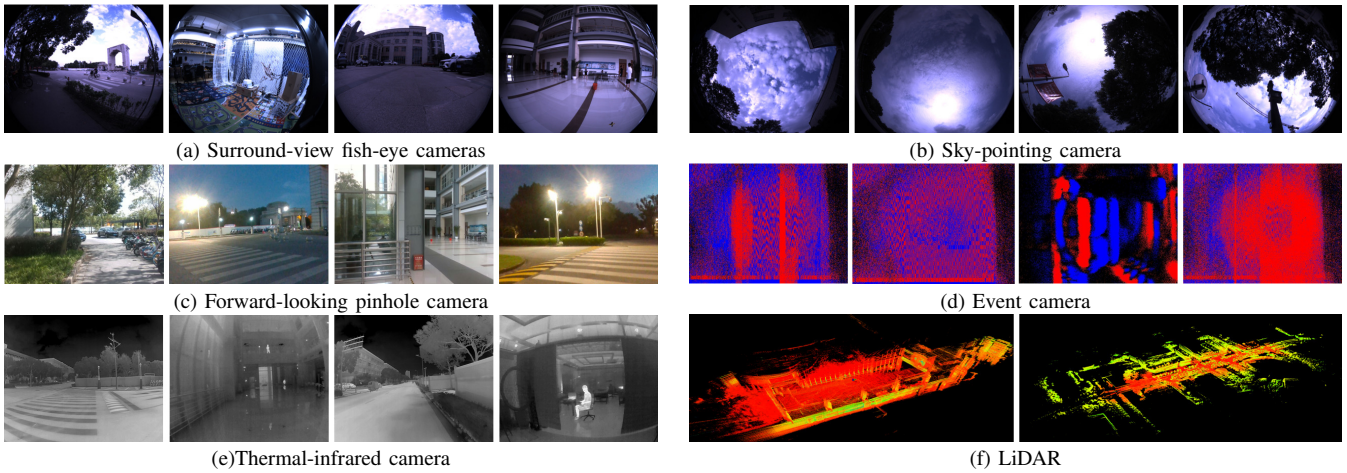
II. RELATED WORK

A. SLAM with different sensors

Generally speaking, SLAM can be categorized into vision-based and laser-based ones. Vision-based SLAM, or visual SLAM, can be divided into monocular, binocular, and multi-camera settings according to the number of cameras [3]. Though monocular visual SLAM [7] [8] is the most studied topic, it suffers from scale uncertainty and scale drift. Binocular visual SLAM [9] [10] takes advantage of known baseline distance to calculate metric depth by triangulation, and multi-camera SLAM [11] [12] yields a broader field of view by using more cameras to enhance the robustness of SLAM in dynamic environments such as streets. However,

All authors are with Shanghai Key Laboratory of Navigation and Location Based Services, Shanghai Jiao Tong University. This work was supported by NSFC(62073214).

* Corresponding Author: Danping Zou (dpzou@sjtu.edu.cn)



(a) Surround-view fish-eye cameras (b) Sky-pointing camera

(c) Forward-looking pinhole camera (d) Event camera

(e)Thermal-infrared camera (f) LiDAR

Fig. 1. Our dataset for ground robots was captured by a rich suite of sensors within various scenarios. Some sensory data are visualized.

it is difficult for pure vision-based SLAM to handle scenes with few textures or low illumination conditions. By contrast, LiDAR SLAM is often regarded as a more reliable choice in such challenges. Nonetheless, LiDAR SLAM also has trouble with long corridors, highly dynamic movements, and foggy scenes.

Recently, multi-sensor fusion has been successfully applied to existing SLAM systems to improve both the accuracy and robustness in practice. For example, ORB-SLAM3 [9] and VINS-Mono [7] integrate vision and IMU. LVI-SAM [13], and R2live [14] tightly integrate vision, LiDAR, and IMU. LIO-SAM [15] loosely integrates LiDAR, IMU, and GNSS. GVINS [16] tightly couples image information, IMU, and GNSS.

To further improve the SLAM performance in practice, a recent trend is to explore different sensor configurations. We list some of them as follows.

a) Multiple cameras: Multiple cameras have a broader field of view than monocular or stereo cameras, which can improve the robustness in dynamic scenes [17]. CoSLAM [18] uses independently moving multiple cameras and gains the capability of co-localization and robustness in dynamic movements. If multiple cameras are mounted on the same platform, not only a wider field of view can be achieved, but also the scale ambiguity can be resolved by the baselines between cameras [12]. For example, Multicol-SLAM [11] applies three fixed fish-eye cameras on a helmet, ROVO [19] uses four fish-eye cameras to achieve full coverage of 360° field of views. Panoramic SLAM [20] is another open-source omnidirectional SLAM system, which claims to achieve centimeter-level accuracy even in highly dynamic movements.

b) Thermal-infrared cameras: Thermal-infrared cameras have gained increasing attention for their perceptual capability beyond the visible spectrum and their robustness regarding environmental changes. In environments with low visibility such as fog, smoke, and darkness, visual SLAM with ordinary RGB image information becomes ineffective. By contrast, thermal-infrared images can effectively improve visibility in these scenarios. For example, J. Delaune [21] proposed a SLAM algorithm using thermal-infrared images, enabling autonomous flight of UAVs (Unmanned Aerial Vehicle) at night.

c) Event cameras: Event cameras measure changes in the brightness of pixels which are with low delay, low power consumption, and high dynamic measurement range. Therefore, they have unique advantages in quick motions. Henri et al. proposed an event-based Visual Odometry algorithm [22]. A. R. Vidal et al. proposed to tightly integrate events, images, and inertial information [23] to achieve better performance.

d) GNSS: GNSS is a valuable localization source that can achieve high-precision positioning outdoors. Coupling GNSS raw measurements into SLAM systems has been proven effective in advancing the localization performance in urban canyons, as shown in recent work [24]. Besides, a sky-pointing camera can help monitor satellite availability and further improve localization accuracy [25].

B. Existing benchmark datasets

a) Datasets for ground robots: Most existing SLAM datasets focus on autonomous driving [27], [26], [28] or UAVs [32]. A few datasets are targeted at ground robots. OpenLORIS [6] was collected by a wheeled robot in indoor environments, which was designed for visual SLAM, where LiDAR SLAM was used to generate the ground truth. As we will see later, LiDAR SLAM may have even more significant errors than visual SLAM in some situations, making the ground truth unreliable. TUM RGBD [5] partly used a robot as the acquisition platform, but it only contains RGB and depth cameras. Similar datasets include UTIAS MultiRobot [37] and PanoraMIS [38], which are not applicable to LiDAR SLAM as well. The aforementioned datasets have limitations such as lack of rich sensory sources, outdated data, and insufficient challenges.

b) Datasets with multiple cameras: Lafida [35] dataset contains three fish-eye cameras on a helmet, but the maximum recording time of its sequences is too short for long-term evaluation (usually longer than 20 minutes). The NCLT [36] dataset used an omnidirectional camera on a ground robot to capture images on campus at a frame rate of 5Hz. Such a low frame rate could cause a small overlap between adjacent

TABLE I

COMPARISON OF SLAM DATASETS

Dataset	Environment	Platform	Duration	RGB Cam.	LiDAR	IMU	Infrared	GNSS	Event
KITTI [26]	Urban	Car	Short-term[a]	2	✓	✓			
UrbanLoco [27]	Urban	Car	Long-term	6	✓	✓		✓	
Brno Urban [28]	Urban	Car	Long-term	4	✓	✓	✓	✓	
Kaist D/N [29]	Urban	Car	Long-term	2	✓	✓	✓		
Pit30M [30]	Urban	Car	Long-term	1	✓	✓			
USVinland [31]	Inland	USV	Short-term	2	✓				
EUROC [32]	Indoors	UAV	Short-term	2		✓			
UZH-FPV [33]	In/Outdoors	UAV	Short-term	2		✓			✓
TUM VI [34]	In/Outdoors	Hand-held	Short-term	2		✓			
LaFiDa [35]	In/Outdoors	Helmet	Short-term	3		✓			
NCLT [36]	In/Outdoors	Ground Robot	Long-term	6	✓	✓			
OpenLORIS [6]	Indoors	Ground Robot	Short-term	2	✓	✓			
Our dataset	In/Outdoors/ Transition/Lift	Ground Robot	Long-term	8	✓	✓	✓	✓	✓

[a]We identify a dataset as long-term if it has sequences longer than 20 minutes.

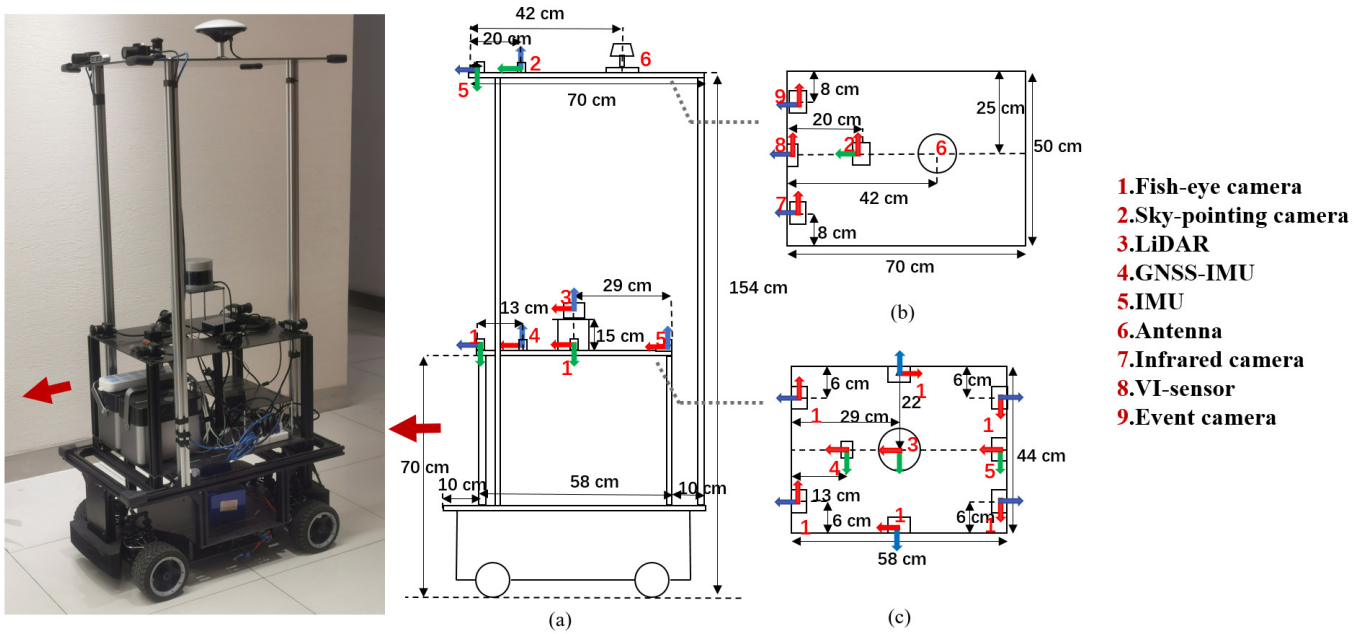


Fig. 2. Our ground robot for data collection. (a) Side view of the robot. (b) Sensors on the top layer. (c) Sensors on the middle layer.

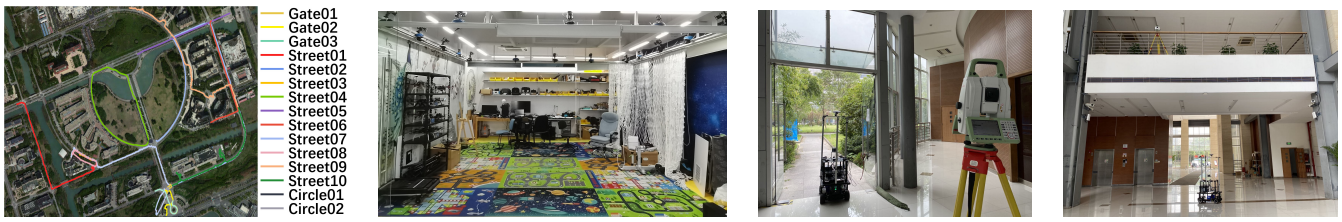
frames, which will be problematic to run visual SLAM algorithms. Nuscenes [39], Waymo [40], and A2D2 [41] datasets collected image data from multiple cameras with a 360° field of view in urban areas, but they did not provide ground truth of trajectories for evaluation of SLAM performance.

c) Datasets with thermal-infrared camera: Brno Urban dataset recorded the infrared camera information on a car [28]. The KAIST D/N [29] dataset collected data from a stereo camera and a thermal infrared camera on a car, as well as a 32-beam LiDAR. However, as far as we know, no public SLAM dataset contains the data of infrared cameras in indoor scenes, which could be useful for the research and development of indoor navigation algorithms in the night or the smoke.

d) Datasets with event Camera: E. Mueggler et al. [42] used the event camera to collect events of dynamic and static scenes inside and outside the room. The recording time of this dataset is too short to evaluate the performance of SLAM algorithms reliably. Interiornet dataset [43] simulates sensors including RGBD, IMU, stereo cameras, and event cameras, but there is a gap between the simulated and the real-world scenes.

e) SLAM datasets with GNSS: Currently, there are very few SLAM datasets containing raw GNSS information. UrbanLoco dataset [27] collected the raw GNSS data with Ublox M8T and used a fish-eye camera to capture the sky, but it did not publish the data of the camera. As far as we know, there is no public SLAM dataset containing both GNSS raw measurements and images of a sky-pointing camera, while such kind of data could be valuable to some research.

In summary, we review the current best-known SLAM datasets and analyze their limitations. Table I summarizes the defects and limitations of some mentioned datasets. From the table, we conclude that currently a dataset with rich scenarios and complete sensor information is urgently needed



(a) (b) (c) (d)

Fig. 3. (a) We visualize the trajectories of outdoor sequences in the map with different colors. (b) The ground truth of indoor sequences are acquired from a motion-capture system with twelve cameras. (c) A 3D laser tracker is used to track the robot through a door from indoors to outdoors for door sequences. (d) In lift sequences, the robot moved in a hall on the first floor and then went to the second floor by lift. A 3D laser tracker is used to track its ground-truth trajectory outside the lift.

TABLE II

SPECIFICATIONS OF SENSORS AND TRACKING DEVICES

Device	Type	Spec.	Freq.(Hz)
LiDAR	Velodyne VLP-32C	32 beam,360 H-FOV,40V-FOV	10
IMU	Handsfree A9	9-axis	150
GNSS Receiver	Ublox M8T	GPS/BeiDou	1
RGB Camera	FLIR Pointgrey CM3-U3-13Y3C	1280*1024, 190 H-FOV, 190 V-FOV	15
Infrared Camera	Gaode PLUG 617	640*512, 90.2 H-FOV,70.6 V-FOV	25
Event Camera	Inivation DVXplorer	640*480, 65.2 H-FOV,51.3 V-FOV	15
VI-sensor	Realsense d435i	RGB: 640*480, 69 H-FOV, 42.5V-FOV	15
		IMU: 6-axis	200
Mocap System	Vicon Vero 2.2	localization accuracy 1mm	50
Laser Tracker	Leica Nova MS60	localization accuracy 1mm + 1.5ppm	10
RTK/INS	Xsens Mti 680G	RTK: localization accuracy 2cm	100
		INS: 9-axis	100

for furthering the study of SLAM algorithms for ground robots. To fill this gap, we present a new dataset M2DGR. Compared with previous SLAM datasets, our dataset contains long-term trajectories in diverse real-world scenarios with a rich pool of sensory information, which facilitates tests and comparisons of various algorithm designs.

III. THE M2DGR DATASET

A. Acquisition platform

We construct a ground robot for data collection as shown in Figure 2. This robot has three layers. The bottom layer contains the power supply, the computer, and the display. The middle layer and the top layer include different sensors. The dimension figures of our robot are shown in Figure 2 (a) $\sim$ (c). To ensure the high-speed transmission of data, we connect the LiDAR to the Ethernet port of the host and other sensory devices to the USB3.0 port of the host. We record the data on a high-end laptop with a high-speed NMVe SSD.

B. Sensor setup

The location of all the devices mounted is shown in Figure 2. Six fish-eye cameras were used to capture the images of the surroundings in 360° field of view, including forward-looking stereo cameras, rear-looking stereo cameras, and two side-looking cameras. Additionally, a 32-beam Velodyne LiDAR was used to scan the surrounding environment and obtain the 3D point cloud. We also used an infrared camera to capture infrared thermal images, a VI-sensor to obtain forward-looking color images as well as inertial data, and an event camera to capture dynamic information. We also mounted a consumer-level IMU, a GNSS receiver to collect GNSS raw signals, and a sky-pointing fish-eye camera to monitor the sky. For ground truth of trajectories, we applied Xsens Mti 680G, a GNSS-IMU integrated navigation system, to track the robots in outdoor sequences, while used a motion-capture system and a laser scanner to track the robot in indoor sequences. All the sensors and tracking devices as well as their key parameters are listed in Table II.

C. Calibration and synchronization

We use the MATLAB camera calibration toolbox to obtain the camera intrinsics of pinhole cameras. For fish-eye cameras, we use Kannala Brandt model [48], Omnidirectional model [49], and MEI model [50] for calibration. To calibrate thermal-infrared cameras, we heat a checkerboard where black blocks and white blocks were made of materials with different heat capacities. Thus, those blocks can be identified by the infrared camera easily. We use toolbox [51] to calibrate internal parameters of IMU, including the white noise and random walk of both the gyroscopic and the accelerometer measurements.

We choose the LiDAR frame as the reference to calibrate the extrinsic parameters (relative poses) between sensors. We use the toolbox [52] to calibrate the extrinsic parameters between IMU and LiDAR, and Kalibr toolbox [53] to calibrate the extrinsic parameters between cameras and IMU, as well as Autoware software [54] to calibrate extrinsic parameters between LiDAR and cameras.

We do not use hardware signals to trigger all the sensors to capture data at the same time but record the data from different sensors using the same system time stamps. The cameras, including the six fisheye cameras and the sky-pointing camera, are synchronized by software - triggering data capturing at

TABLE III

AN OVERVIEW OF SCENARIOS IN OUR DATASET

Scenario	Street	Circle	Gate	Walk	Hall	Door	Lift	Room	Roomdark	TOTAL
Number	10	2	3	1	5	2	4	3	6	36
Size/GB	590.7	50.6	65.9	21.5	117.4	46.0	112.1	45.3	171.1	1220.6
Duration/s	7958	478	782	291	1226	588	1224	275	866	13688
Dist/m	7727.72	618.03	248.40	263.17	845.15	200.14	266.27	144.13	395.66	10708.67
GT	RTk/INS	RTk/INS	RTk/INS	RTk/INS	Leica	Leica	Leica	Mocap	Mocap	—

TABLE IV

SAMPLE SEQUENCES FOR EVALUATION

Sequence	Street02	Street06	Street07	Roomdark06	Hall05	Door01	Lift04
Duration/s	1227	494	929	172	402	461	299
Distance/m	1484.62	479.63	1104.07	72.53	79.28	285.51	142.78
Speed/(m/s)	1.21	0.97	1.19	0.42	0.71	0.31	0.27
Description of features	day, long-term,	night, straight line	night, zigzag route	room, complete darkness	long-term large overlap	outdoors to outdoors	first floor to second floor by lift

TABLE V

ATE(M) OF SLAM SYSTEMS ON SAMPLE SEQUENCES

Method / Sequence	Street02	Street06	Street07	Roomdark06	Hall05	Door01	Lift04
A-LOAM [44]	5.299	0.628	28.940	0.314	1.065	0.274	1.323
LeGO-LOAM [45]	20.021	1.246	35.437	0.373	1.030	0.253	1.370
LINS [46]	5.636	1.742	12.009	2.205	1.010	0.258	1.318
LIO-SAM [15]	4.063	0.417	28.642	0.324	1.047	0.268	1247.153
ORB3-Pinhole [9]	152.462	5.845	X[a]	X	3.291	7.662	X
ORB3-Fisheye [9]	X	95.056	X	X	X	2.295	8.131
ORB3-Thermal [9]	154.778	30.450	8.863	0.404	5.927	1.241	2.873
CubemapSLAM [8]	X	98.391	X	X	5.171	9.328	X
VINS-Mono [7]	24.157	124.357	143.725	1.001	0.646	0.694	5.582
RTKLIB [47]	7.072	6.749	13.096	X	X	5.344	X

[a]If a visual SLAM fails to initialize or track frames less than a half of total frames or a GNSS-based method fails to initialize, we mark it X

the same instance by calling the API. Our tests show that such a soft synchronization approach achieves accurate time synchronization of less than 10 ms between different cameras.

D. Data collection

We operated our robot traveling around various scenes. An overview of our dataset is given in Table III. The recording processes are described as follows:

For outdoor environments, we collected sequences on the campus of Shanghai Jiao Tong University. The satellite visibility was good so that the GNSS-RTK suite outputs high-accuracy ground truth of trajectories. To better test loop closing of visual SLAM, we particularly recorded sequences *Circle01* and *Circle02* while the robot was circling in repeated routes. All the outdoor trajectories are visualized in Figure 3 (a).

For indoor environments, we obtained the ground truth trajectories with a motion-capture system with twelve high-speed tracking cameras (50 Hz) in a room, as illustrated in Figure 3 (b). Particularly, we recorded a few sequences in complete darkness to test SLAM systems' robustness, as shown in Figure 5. Outside the room with the motion-capture system, we used a laser tracker to generate the ground truth of trajectories indoors. A prism reflector was mounted on our robot so that the laser tracker could well track it. Also, to better evaluate loop closing of visual SLAM, we recorded the sequence *Hall05* where the robot circled around a hall in repeated routes.

We also recorded sequences in a mixture of indoor and outdoor environments. We operated the robot traveling outdoors for some time until GNSS signals were well received. Then we got the robot entering a hall through a door. After the robot moved around the hall for a period, we got it going outdoors again. Those sequences are used to evaluate the performance of SLAM and GNSS positioning methods on the critical point between indoors and outdoors.

Lastly, we collected sequences to test the capability of entering and leaving the lift. More specifically, we manipulated the robot to travel around within a hall on the first floor and then enter a lift that carried the robot to the second floor, as Fig 3 (d) shows.

E. Data usage and tools

All the data were captured by rosbag in Robot Operation System (ROS), and the recorded topics are listed as follows.

- RGB camera:
  /camera/left/image_raw/compressed
  /camera/right/image_raw/compressed
  /camera/third/image_raw/compressed
  /camera/fourth/image_raw/compressed
  /camera/fifth/image_raw/compressed
  /camera/sixth/image_raw/compressed
  /camera/head/image_raw/compressed
- VI-sensor:
  /camera/color/image_raw/compressed
  /camera/imu

- Raw GNSS:
  /ublox/aidalm
  /ublox/rxmraw
  /ublox/fix
  /ublox/navstatus
- Event camera:
  /dvs_rendering/compressed
  /dvs/events0
- LiDAR:
  /velodyne_points
- Infrared Camera:
  /thermal_image_raw
- IMU:
  /handsfree/imu

For convenience, we provide scripts to export the data to other formats such as [32]. Ground-truth trajectories and calibration results are provided for each sequence. Furthermore, we give detailed instructions to evaluate the performance of different SLAM algorithms on our project page.

IV. Evaluation

We evaluate the state-of-the-art SLAM systems on seven representative sequences from our dataset. Those sequences are described in Table IV. The Absolute Trajectory Error (ATE) [5] is used for the evaluation metric. All the estimated trajectories are aligned with the ground truth by the EVO tool [55] to obtain the ATE errors.

For visual SLAM, we test ORB-SLAM3 [9], Cubemap-SLAM [8], Multicol-SLAM [11], and VINS-Mono [7]. The default setting of each method is used for evaluation. For ORB-SLAM3, we select the monocular camera mode (without IMU) with different types of cameras: a pinhole camera, a fish-eye camera, and a thermal camera for evaluation (denoted by ORB3-Pinhole, ORB3-Fisheye, ORB3-Thermal respectively). For LiDAR SLAM, we evaluate A-LOAM [44], LeGO-LOAM [45], LINS [46] , and LIO-SAM [15].

The quantitative results are shown in Table V. The estimated trajectories are visualized within the ENU frame as shown in Figure 4. As the robot travels on the ground, the visualization of most sequences is in 2D. The results show that LiDAR-based methods outperform vision-based methods generally, especially in large-scale outdoor scenarios, but both kinds of methods do not perform well in certain cases. We discuss the results in detail as follows.

a) Low illumination: Sequence *Roomdark06* and *Street07* are in environments with low illuminations. ORB-SLAM3 using both pinhole (ORB3-Pinhole) and fisheye (ORB3-Fisheye) cameras fail in those sequences. Though ORB-SLAM3 adopts the strategy of adaptive histogram equalization to address bad illuminations, it failed to extract enough feature points in those dark scenes. Moreover, most extracted feature points were from far-away bright objects like street lamps or light screens, leading to significant estimation errors. By contrast, using a thermal-infrared camera, ORB-SLAM3 achieved significantly better robustness in the same scene because thermal cameras can distinguish objects under low visibility. However, we observe that some objects that can be well recognized by an RGB camera may appear texture-less in a thermal-infrared camera, for example, a flat colorful curtain. This phenomenon indicates that thermal-infrared cameras do not necessarily perform better than ordinary RGB cameras in some scenes.

b) Entering and leaving the lift: Sequence *lift04* is a test sequence where the robot takes a lift to move across different floors as shown in Figure 6. Within a lift, SLAM systems with pure visual or laser information will consider the robot as being static, and only those with inertial information may recognize the robot moving upwards or downwards. Unfortunately, as shown in Figure 4 (g) and (h), none of the tested SLAM systems succeeded in tracking the whole trajectory or reconstructing a complete map. Particularly, LIO-SAM drifts severely after the robot enters the lift due to a mismatch between IMU pre-integration and LiDAR odometry. As working on different floors is quite common for robots in daily life, it is meaningful and urgent to address the localization problem when robots take a lift to switch floors.

c) Outdoor-Indoor switching: In Sequence *Door01*, GNSS signals degraded drastically as the robot got closer to the door. All the SLAM systems can complete this sequence, while those with pure visual inputs yield large errors. We evaluate the GNSS SPP (single point positioning) performance by using a GNNS localization software RTKLIB [47]. Though the APE [5] of RTKLIB still seems normal, actually indoor localization has failed, as Figure 4 (f) shows. GNSS positioning has a stable accuracy in environments with good satellite visibility, but they only work outdoors and usually suffer from signal loss, satellite ephemeris error, and multi-path effect. To better make use of GNSS signals, we provide a sky-pointing camera to apply the possible solutions such as [25].

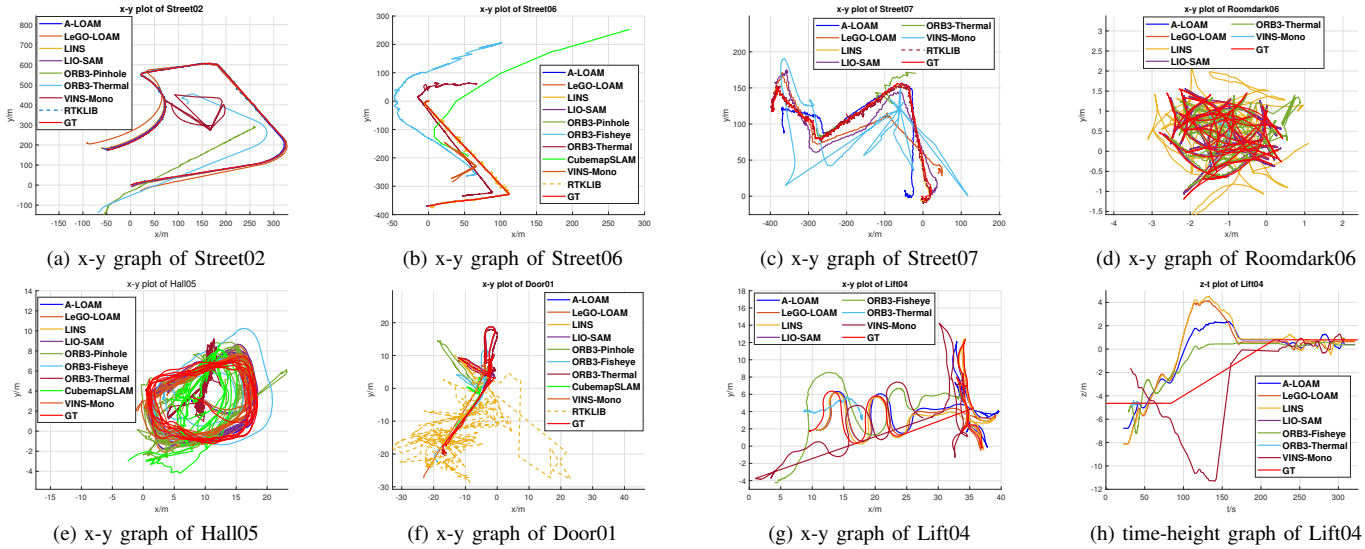
d) Dynamic motion: Sequence *Street07* was collected in a zigzag route with abrupt motions such as quick turning, braking, as well as speeding up and down. Neither visual SLAM nor LiDAR SLAM worked well on this test, as the results in Table V show. Most Visual SLAM methods (ORB3-Pinhole, ORB3-Fisheye, CubemapSLAM) failed. LiDAR SLAM also produced large ATE errors in this case.

e) Multiple cameras: Multiple-camera visual SLAM can take advantage of images in a wider field of view. We tested Multicol-SLAM [11] with three fisheye cameras (two in the front, one on the left side), but it lost its track in almost every sequence. The reason might be that it tries to extract and match features directly from highly distorted images, which may easily cause false matches [8].

The results indicate that the state-of-the-art SLAM systems, both visual and LiDAR ones, may perform well on existing benchmark tests, they still require significant improvement to be applied to ground robots in daily life. The results also indicate our dataset is a valid and valuable test field for existing SLAM systems. With rich sensory information and various scenarios, we believe our benchmark will promote the progress of robot navigation solutions.

V. Conclusion

We release M2DGR, a large-scale multi-sensor dataset focusing on ground robots' localization and mapping tasks.



(a) x-y graph of Street02 (b) x-y graph of Street06 (c) x-y graph of Street07 (d) x-y graph of Roomdark06

(e) x-y graph of Hall05 (f) x-y graph of Door01 (g) x-y graph of Lift04 (h) time-height graph of Lift04

Fig. 4. Estimated and ground-truth (GT) trajectories of 7 sample sequences are visualized in ENU (East-North-Up) coordinate system.

Fig. 5. Color image (**Left**) and Infra-red image (**Right**) in a complete dark scene.

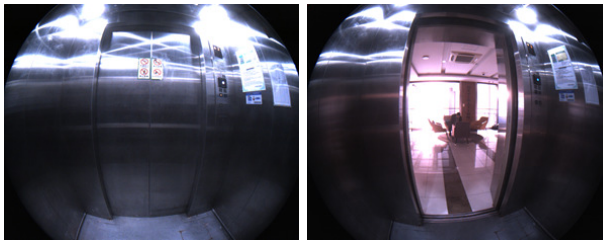


Fig. 6. The images captured by a fish-eye camera when the robot is leaving a lift.

Our dataset contains a large pool of sensory information to encourage breakthroughs in multi-sensor fusion on SLAM. Furthermore, we tested and evaluated a few state-of-the-art SLAM systems based on our dataset and analyzed the defects and limitations of existing systems in different scenarios, which may point out potential developing directions for SLAM. In the future, we plan to update and extend our project from time to time, striving to build a comprehensive SLAM benchmark similar to the KITTI dataset [26] for ground robots.

REFERENCES

[1] E. Appleton and D. J. Williams, *Industrial robot applications*. Springer Science & Business Media, 2012.

[2] J. Aulinas, Y. Petillot, J. Salvi, and X. Lladó, "The slam problem: a survey," *Artificial Intelligence Research and Development*, pp. 363–371, 2008.

[3] 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.

[4] Y. Liu, Y. Fu, F. Chen, B. Goossens, W. Tao, and H. Zhao, "Datasets and evaluation for simultaneous localization and mapping related problems: A comprehensive survey," *arXiv e-prints*, pp. arXiv–2102, 2021.

[5] 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.

[6] 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.

[7] 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.

[8] Y. Wang, S. Cai, S.-J. Li, Y. Liu, Y. Guo, T. Li, and M.-M. Cheng, "Cubemapslam: A piecewise-pinhole monocular fisheye slam system," in *Asian Conference on Computer Vision*. Springer, 2018, pp. 34–49.

[9] 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.

[10] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," *arXiv preprint arXiv:1901.03642*, 2019.

[11] S. Urban and S. Hinz, "Multicol-slam-a modular real-time multi-camera slam system," *arXiv preprint arXiv:1610.07336*, 2016.

[12] D. Z. Ang Li and W. Yu, "Robust initialization of multi-camera slam with limited view overlaps and inaccurate extrinsic calibration," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021.

[13] T. Shan, B. Englot, C. Ratti, and D. Rus, "Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping," *arXiv preprint arXiv:2104.10831*, 2021.

[14] J. Lin, C. Zheng, W. Xu, and F. Zhang, "R2live: A robust, real-time, lidar-inertial-visual tightly-coupled state estimator and mapping," *arXiv preprint arXiv:2102.12400*, 2021.

[15] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5135–5142.

[16] S. Cao, X. Lu, and S. Shen, "Gvins: Tightly coupled gnss-visual-inertial for smooth and consistent state estimation," *arXiv e-prints*, pp. arXiv–2103, 2021.

[17] Z. Zhang, H. Rebecq, C. Forster, and D. Scaramuzza, "Benefit of large field-of-view cameras for visual odometry," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 801–808.

[18] D. Zou and P. Tan, "Coslam: Collaborative visual slam in dynamic environments," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 2, pp. 354–366, 2012.

[19] H. Seok and J. Lim, "Rovo: Robust omnidirectional visual odometry for wide-baseline wide-fov camera systems," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6344–6350.

[20] S. Ji, Z. Qin, J. Shan, and M. Lu, "Panoramic slam from a multiple fisheye camera rig," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 169–183, 2020.

[21] J. Delaune, R. Hewitt, L. Lytle, C. Sorice, R. Thakker, and L. Matthies, "Thermal-inertial odometry for autonomous flight throughout the night," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1122–1128.

[22] H. Rebecq, T. Horstschäfer, G. Gallego, and D. Scaramuzza, "Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 593–600, 2016.

[23] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 994–1001, 2018.

[24] T. Li, L. Pei, Y. Xiang, Q. Wu, S. Xia, L. Tao, X. Guan, and W. Yu, "P 3-loam: Ppp/lidar loosely coupled slam with accurate covariance estimation and robust raim in urban canyon environment," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6660–6671, 2020.

[25] J. Marais, C. Meurie, D. Attia, Y. Ruichek, and A. Flancquart, "Toward accurate localization in guided transport: Combining gnss data and imaging information," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 188–197, 2014.

[26] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[27] W. Wen, Y. Zhou, G. Zhang, S. Fahandezh-Saadi, X. Bai, W. Zhan, M. Tomizuka, and L.-T. Hsu, "Urbanloco: a full sensor suite dataset for mapping and localization in urban scenes," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2310–2316.

[28] 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.

[29] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "Kaist multi-spectral day/night data set for autonomous and assisted driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 934–948, 2018.

[30] J. Martinez, S. Doubov, J. Fan, S. Wang, G. Máttyus, R. Urtasun, *et al.*, "Pit30m: A benchmark for global localization in the age of self-driving cars," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4477–4484.

[31] Y. Cheng, M. Jiang, J. Zhu, and Y. Liu, "Are we ready for unmanned surface vehicles in inland waterways? the usvinland multisensor dataset and benchmark," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3964–3970, 2021.

[32] 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.

[33] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are we ready for autonomous drone racing? the uzh-fpv drone racing dataset," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6713–6719.

[34] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, "The tum vi benchmark for evaluating visual-inertial odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1680–1687.

[35] S. Urban and B. Jutzi, "Lafida—a laserscanner multi-fisheye camera dataset," *Journal of Imaging*, vol. 3, no. 1, p. 5, 2017.

[36] 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.

[37] 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.

[38] H.-E. Benseddik, F. Morbidi, and G. Caron, "Panoramis: An ultra-wide field of view image dataset for vision-based robot-motion estimation," *The International Journal of Robotics Research*, vol. 39, no. 9, pp. 1037–1051, 2020.

[39] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2020, pp. 11 621–11 631.

[40] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2446–2454.

[41] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, *et al.*, "A2d2: Audi autonomous driving dataset," *arXiv preprint arXiv:2004.06320*, 2020.

[42] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.

[43] W. Li, S. Saeedi, J. McCormac, R. Clark, D. Tzoumanikas, Q. Ye, Y. Huang, R. Tang, and S. Leutenegger, "Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset," *arXiv preprint arXiv:1809.00716*, 2018.

[44] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and Systems*, vol. 2, no. 9, 2014.

[45] T. Shan and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4758–4765.

[46] C. Qin, H. Ye, C. E. Pranata, J. Han, S. Zhang, and M. Liu, "Lins: A lidar-inertial state estimator for robust and efficient navigation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 8899–8906.

[47] T. Takasu and A. Yasuda, "Development of the low-cost rtk-gps receiver with an open source program package rtklib," in *International symposium on GPS/GNSS*, vol. 1. International Convention Center Jeju Korea, 2009.

[48] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 8, pp. 1335–1340, 2006.

[49] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2006, pp. 5695–5701.

[50] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2007, pp. 3945–3950.

[51] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-696, Aug. 2007.

[52] J. Lv, J. Xu, K. Hu, Y. Liu, and X. Zuo, "Targetless calibration of lidar-imu system based on continuous-time batch estimation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9968–9975.

[53] 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.

[54] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada, "An open approach to autonomous vehicles," *IEEE Micro*, vol. 35, no. 6, pp. 60–68, 2015.

[55] MichaelGrupp, "evo," https://github.com/MichaelGrupp/evo, 2018.