GND: An Outdoor Global Navigation Dataset with Multi-Modal Perception and Traversability

Author Names Omitted for Anonymous Review. Paper-ID [add your ID here]

***Abstract*—Navigating large-scale outdoor environments requires complex reasoning in terms of geometric structures, environmental semantics, and terrain characteristics, which are typically captured by onboard sensors such as LiDAR and cameras. While current mobile robots can navigate such environments using pre-defined, high-precision maps based on hand-crafted rules catered for the specific environment, they lack commonsense reasoning capabilities, especially the traversability analysis, that most humans possess when navigating unknown outdoor spaces. To address this gap, we introduce the Global Navigation Dataset (GND), a large-scale dataset that integrates multi-modal sensory data, including 3D LiDAR point clouds and RGB and 360° images, as well as multi-category traversability maps (pedestrian walkways, vehicle roadways, stairs, off-road terrain, and obstacles) from ten university campuses. These environments encompass a variety of parks, urban settings, elevation changes, and campus layouts of different scales. The dataset covers approximately 2.7km^2 and includes at least 350 buildings in total. We also present a set of novel applications of GND to showcase its utility to enable global robot navigation, such as map-based global navigation, mapless navigation, and global place recognition. GND's website can be found at https://cs.gmu.edu/~xiao/Research/GND/.**

I. Introduction

Global navigation plays a critical role in enabling robots to traverse large-scale outdoor environments [9, 31, 37, 30, 29, 48]. It is widely used in real-world tasks like last-mile delivery [21, 8], remote exploration [5, 4], autonomous driving [37, 68], etc. Unlike navigation in structured indoor spaces [63, 38, 62, 51], global navigation needs to reason about a variety of environmental factors in complex outdoor scenarios [28, 56, 30, 42], including recognizing terrain characteristics for traversability analysis [13, 14, 15, 58, 24] and inferring navigational cues to determine the shortest path in large-scale open environments [51, 31].

One challenge of global navigation is the need of navigational reasoning at a very large scale, e.g., trajectories corresponding to hundreds or thousands of meters [9, 31, 37, 30, 29]. In practice, navigating from the south side of a university campus to a cafeteria on the north side without a prior map may require following major pedestrian walkways at the beginning, while taking shortcuts close to buildings later. Such a reasoning process requires not only geometric, but also semantic understanding of the large-scale outdoor scenes, which needs to be captured by various sensors like 3D LiDARs and RGB cameras.

Another challenge is making navigation decisions by reasoning beyond simple obstacles and free spaces, a common delineation of indoor workspaces, while considering different robot morphologies. For example, pedestrian walkways, vehicle roadways, and some off-road terrain may all appear as free spaces, but they may correspond to different categories of traversability in different scenarios [47], e.g., stairs should be treated as obstacles for wheeled robots, but they can be regarded as free spaces for legged robots [31].

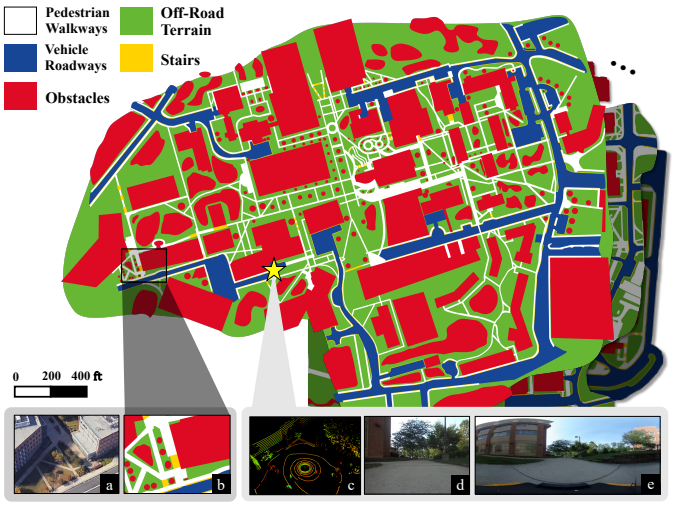


Fig. 1: **Traversability maps of the Global Navigation Dataset (GND)**: The legend shows five different categories of terrain traversability with different colors. Inset (a) shows the 3D satellite image of inset (b), an enlarged traversability map. GND contains multi-modal sensory data including 3D LiDAR point clouds (c), RGB images (d), and 360° images (e).

Although robotics practitioners have tackled these challenges with high-precision prior maps and hand-crafted navigation rules for specific environments, recent research has focused on leveraging machine learning to equip robots with generalizable, human-like reasoning capabilities during outdoor global navigation. Imitation and reinforcement learning techniques [28, 56] have been used for semantic understanding [22, 36, 11, 5, 40, 33], traversability analysis [22, 36, 11, 5, 40, 33], topological modeling [42, 45], trajectory generation [29, 30], heuristic estimation [51, 50], parameter tuning [61, 57, 53, 55, 65], and policy learning [39, 17]. One common requirement of all these data-driven approaches is high-quality ground truth or trial-and-error data for training.

Main Results: Motivated by such difficulties of global navigation and research needs of training data, we present a novel, large-scale Global Navigation Dataset (GND), which includes multi-modal perception and multi-category traversability in outdoor campus environments (Fig. 1). GND comprises almost 11 hours of navigation data captured using two Clearpath Jackal robots, accompanied by 3D LiDAR point clouds, RGB

TABLE I: **Comparison of State-of-the-Art Navigation Datasets**: Our GND dataset features the most comprehensive range of sensors and provides multi-category traversability maps for global navigation in various campuses.

Dataset	Traversability Labels	Distance (km)	Duration (min.)	Sensors	Platform	Purpose
SCAND[23]	×	40	522	3D LiDAR, RGB Camera, RGB-D Camera, Wheel Odometry, Visual Odometry	Robot	Social Navigation
MuSoHu[35]	×	100	1200	3D LiDAR, RGB-D Camera, 360° Camera, IMU, Microphone, Visual Odometry	Human	Social Navigation
SiT[2]	✓	N/A	24.3	3D LiDAR, RGB Camera (covering 360°), IMU, GPS, Wheel Odometry	Robot	Social Navigation, Human Detection & Tracking
JRDB[34]	×	N/A	64	3D LiDAR, 2D LiDAR, RGB Camera, RGB-D Camera, 360° Camera, IMU, GPS, Microphone, Wheel Odometry	Robot	Human Detection & Tracking
NCLT[7]	×	147.4	2094	3D LiDAR, RGB Camera, Fisheye Camera, IMU, GPS, Wheel Odometry	Robot	Global Navigation
GND (Ours)	✓	53	668	3D LiDAR, RGB Camera, 360° Camera, IMU, GPS, Wheel Odometry	Robot	Global Navigation

and 360° camera images, inertia Measurement Unit (IMU) information, GPS data, as well as robot odometry and actions, collected across **10** university campuses in city and village areas, including a variety of park areas, vegetation types, elevation changes, diverse campus layouts and objects in the campuses. In total, We covered around **2.7km**^{2} with at least **350** buildings in the datasets over 11 hours of recorded data (in rosbags). All raw perception data are post-processed into ten large-scale global campus maps labeled with five categories of traversability (pedestrian walkways, vehicle roadways, stairs, off-road terrain, and obstacles) and associated with multi-modal perception (e.g., first-person and 360° view) on the robot trajectories. Some of our main contributions include:

- The first large-scale, long-range, across-campus global navigation dataset with multi-modal perception data and multi-category traversability maps;
- A standardized and streamlined data collection and post-processing pipeline designed to encourage broader contributions from all users to the dataset; and
- Novel dataset applications showing GND's utility in enabling outdoor global navigation tasks with different types of robots (wheeled and legged robots), as shown in Section IV, including global map-based navigation (path planing), mapless navigation (trajectory and motion generation), and global place recognition.

II. Related Work

In this section, we review related literature on global robot navigation and state-of-the-art robot navigation datasets.

A. Global Robot Navigation

Navigating robots in large, outdoor environments presents multiple challenges, including the need to assess terrain traversability and perform large-scale navigational reasoning. Global robot navigation can be divided into two main approaches: map-based and mapless. Map-based approaches rely on a comprehensive cost map for path planning [18, 25] and precise robot localization [1, 32] to ensure accurate path execution. However, these methods can be computationally expensive and require significant overhead to maintain up-to-date maps. To address these limitations, ViNT [45] and NoMaD [49] proposed generating topological maps and using vision-based images as subgoals for navigation, though they still require initial runs to collect subgoal images. On the other hand, mapless navigation techniques eliminate the need for maps entirely. AdaptiveON [28] focused on generating actions in a mapless fashion but was limited to local planning without addressing long-distance navigation. More recent advancements, such as MTG [29], enable long-distance navigation, while DTG [30] further optimizes traversable trajectories in large-scale outdoor settings. Both map-based and mapless approaches require extensive datasets with highly accurate traversability maps and multi-modal sensory data for effective reasoning and training.

B. Datasets for Robot Navigation

Over the past decade, large-scale navigational datasets have proven invaluable across various research domains, including social robot navigation [20], human trajectory prediction [41], autonomous driving [37, 68], vision-based navigation [44], and global navigation [30].

KITTI [19] was one of the first large-scale datasets that emphasized the importance of well-organized, real-world data for advancing machine learning and computer vision research. It not only significantly impacted autonomous driving but also influenced the broader field of computer vision. Although many follow-up datasets have been introduced [3, 52, 10], early efforts were predominantly focused on autonomous vehicle applications, particularly in perception. Similarly, in robotics, multi-sensory datasets have been released. However, current datasets focused on perception or mapping tasks [66, 34, 6], considered specific local navigation tasks [23, 35, 2], or were limited in scale and sensor modalities [43]. As a result, there remains a gap in the datasets among adequately covering large-scale environments and traversability analysis across different sensor and robot modalities. GND addresses this gap by offering rich multi-modal robot sensory data collected in outdoor campus environments, complemented by human-

TABLE II: **Five Example Campuses in GND**: The table outlines details of five example campuses, including the University of Maryland (UMD), George Mason University (GMU), Catholic University of America (CUA), Georgetown University, and George Washington University (GWU). We list the covered areas, number of buildings, trajectory length, number of RGB and 360° images, number of LiDAR point clouds, and ratio of different traversability categories in the campus map. P, O, V, and S represent pedestrian walkways, off-road terrain, vehicle roadways, and stairs, respectively. For example, 0.84 km^2 is covered on the UMD campus with 60 buildings, and the ratio of pedestrian walkways in the campus is 10.66% of the UMD campus map. More campus datasets are on GND's website.

Campuses	Covered Areas (km^2)	# of Buildings	Trajectory Length (km)	# of RGB Images	# of 360° Images	# of LiDAR Clouds	Ratio of Traversability (%)			
							P	O	V	S
UMD	0.84	60	23.26	214768	N/A	146703	10.66	16.29	25.84	1.67
GMU	0.46	51	13.67	137948	137027	91500	17.31	25.04	17.11	0.41
CUA	0.40	32	2.87	29921	30266	20025	7.86	42.29	18.78	1.81
Georgetown	0.25	40	3.25	33244	33325	22050	7.16	21.42	13.96	1.51
GWU	0.15	39	3.00	33156	32714	22190	8.95	14.04	28.09	1.99

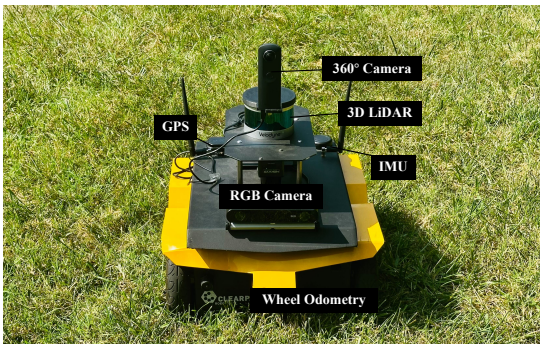


Fig. 2: **Robot Setup**: We use a Clearpath Jackal for dataset collection, which is equipped with various sensors, including 3D LiDAR, RGB camera, 360° camera, IMU, and GPS. It is capable of traversing diverse terrains, including pedestrian roads, roadways, off-road areas, ramps, and woods.

labeled multi-category traversability maps. Table I compares our work with relevant state-of-the-art datasets.

While previous datasets have proven to be valuable for studying perception and various navigation challenges, they fall short in providing demonstrations on global navigation, particularly regarding traversability information. While SiT [2] offered 12-layered semantic maps, its primary focus was on providing benchmarks and analysis for pedestrian detection and tracking, falling short in demonstrating practical applications that leverage the semantic maps. We aim to showcase applications of GND by utilizing various data and methods to demonstrate their practical applications.

III. Dataset

In this section, we first describe the data collection procedure. We then describe the details of our dataset, particularly on the traversability map.

A. Data Collection

We manually operate the robot to navigate various campus environments for data collection. We guide the robot considering the traversability of the road. The robot primarily navigates pedestrian walkways; However, when necessary, it also traverses vehicle roadways, such as when crossing streets or accessing specific areas. As shown in Fig. 2, the robot is equipped with the following sensors:

- **3D LiDAR**: Velodyne VLP-16 with 16 channels or Ouster OS1-32 with 32 channels, both covering a 360-degree field of view and operating at 10 Hz;
- **RGB Camera**: ZED2 with image resolution of $1080p$ facing front and operating at 15 Hz;
- **360° Camera**: RICOH Theta V operating at 15 Hz;
- **IMU**: 6D 3DM-GX5-10 operating at 355 Hz; and
- **GPS**: u-blox F9P operating at 20 Hz.

Our robot operates on Ubuntu 20.04 and Robot Operating System (ROS) Noetic. The data captured by the sensors are recorded in the rosbag file format. We also provide both intrinsic and extrinsic calibration parameters for the LiDARs and the cameras. We gathered datasets from 10 university campuses with approximately 2.7 km^2 of campus area, including at least 350 buildings, with over 11 hours of recorded rosbag data. These campus datasets encompass a variety of environments, such as parks, different types of vegetation, elevation changes, diverse campus layouts, and objects. As shown in Table II, we list five example campus datasets with their details. The campus datasets cover a various size of campus areas ranging from 0.15 km^2 to 0.84 km^2, with 32 to 60 buildings. The table also underscores the variety of traversability category ratios in different campuses. For examples, the pedestrian walkways range from 7.16% to 17.31% over the covered campus area. More campus datasets are in the website in the Abstract.

B. Standardized Data Processing

To encourage broader contributions from dataset users, we standardize our data processing workflow. First, the raw rosbag data is processed using LIO-SAM [46] or FAST-LIO [64] to generate both the trajectories and 3D local maps. Next, we process the point cloud maps by removing the ground, which enhances the visibility of significant features, such as buildings, in a top-down view. This ground removal also improves localization performance. Then, these local maps are registered using TEASER++ [67] to create a global map, where all trajectories and maps are transformed into the global map's coordinate system. For each campus, we generate a single global map within the dataset. This global map is

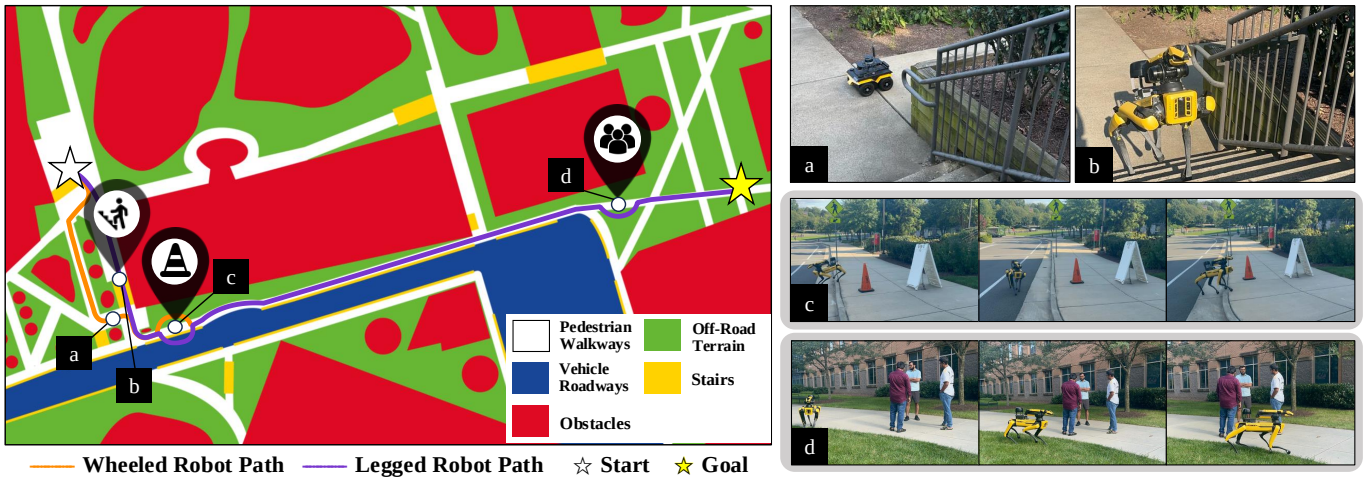
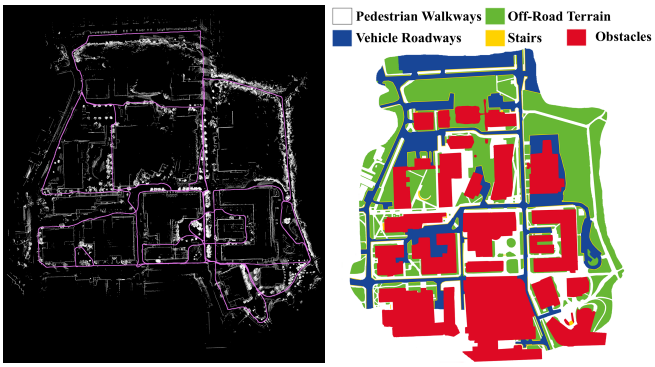


Fig. 3: **Map-based Global Navigation using Multi-Category Traversability Map in GND**: The white and yellow stars indicate the start and goal positions, respectively. The orange line shows the path of the wheeled robot, and the purple line shows the path of the legged robot. (a) and (b) shows Scenario 1, where there are stairs. (c) shows Scenario 2, where the robot is going to the right bottom of the figure. When the road is blocked and the legged robot can only go through the roadway due to narrow passages in off-road terrain. (d) shows Scenario 3, when the road is blocked and the robots have to go around the path using the off-road terrain.



(a) Point Cloud Map (b) Multi-category Traversability Map

Fig. 4: **Partial Map of the University of Maryland**: Inset (a) presents the processed point cloud map, featuring the robot's trajectory for data collection marked by a pink line; (b) illustrates the multi-category traversability map corresponding to the dataset shown in (a).

a comprehensive 3D representation of the entire campus, designed for use in map-based navigation approaches and as ground truth for mapless ones.

To create 2D traversability maps, we first utilize semantic segmentation approaches to segment different traversability areas in RGB images and correlate the traversabilities to Lidar point cloud through Lidar-camera calibration. Then according to the robot posese, we merge all the single frame colored point clouds together in 3D global coordinate and project the 3D point cloud global map onto the 2D plane along the Z-axis. Then we manually annotate the five distinct traversability types, each represented by different colors. The standardized data processing pipeline is also published with the dataset.

C. Multi-Category Traversability Map

The multi-category traversability map consists of five traversability types. White indicates areas traversable by most robots, such as sidewalks, concrete surfaces, and brick roads. Red represents areas that are non-traversable for all robots, including buildings, rivers, construction sites, and poles. Between these extremes, we define three additional categories: green marks off-road or vegetated areas, yellow represents stairs or curbs, and blue highlights roadways and parking lots. Different types of robots can navigate through different traversable areas. For example, legged robots can handle stairs and curbs, while wheeled robots cannot. However, fast-moving wheeled robots are capable of traveling on roadways, whereas slow legged robots are not safe in high-traffic environments.

Figure 4 shows an example of the multi-category traversability map, alongside the point cloud map. It demonstrates that GND provides not only geometric but also semantic information about the environment, closely aligned with real-world conditions.

IV. Applications

In this section, we present three applications for the GND dataset, emphasizing its unique characteristics: *globalness* and *traversability*, which do not present in existing navigation datasets. This dataset is collected mostly by Jackal robot, but it can be used for navigation tasks with different types of robots, such as legged robots and wheeled robots. We implement map-based global navigation, mapless navigation, and global place recognition.

A. Map-based Global Navigation

The primary objective of the GND dataset is to provide precise map data for global robot navigation. To demonstrate its utility, we conduct an experiment comparing the navigation

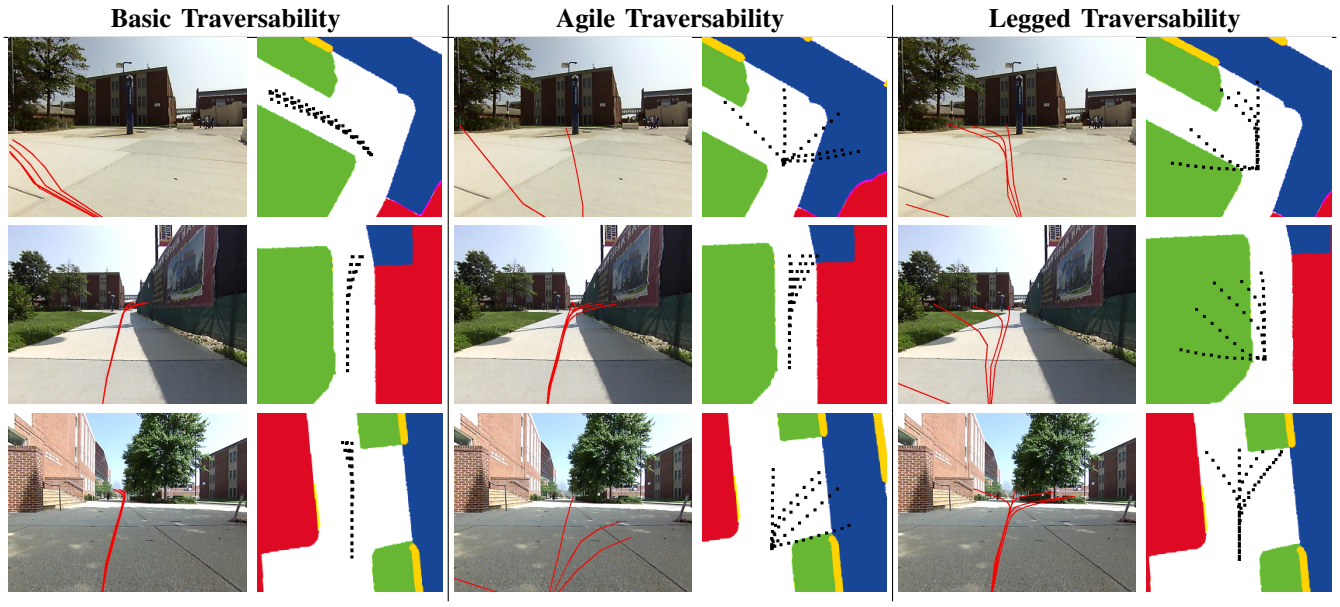


Fig. 5: **T-MTG Generated Trajectories**: T-MTG generates trajectories to cover traversable areas across three levels: basic, agile, and legged traversability levels. Each level imposes different constraints and utilizes different types of robots for navigation. Basic traversability level contains only pedestrian roads for all types of robots, agile traversability level includes both pedestrian roads and vehicle roadways for fast-moving wheeled robots, and legged traversability level has pedestrian walkways and off-road terrains for legged robots. For each cell, the left side displays the generated trajectories (red solid lines) overlaid on the robot's RGB view image, while the right side shows the generated trajectories (black dotted lines) overlaid on the cropped multi-category traversability map.

of two robots with different modalities and traversabilities, wheeled and legged. Using the map, path planning methods such as A* or RRT* can generate a path based on the GPS coordinates of the start and goal positions. As the robot moves, motion planning methods like the classical approaches [16, 27], like Dynamic Window Approach [16], or learning-based approaches [28, 26, 12, 60, 59, 54] can be employed to observe the real-time environment changes and guide the robot's actions. Both robots will initially navigate along the sidewalk, but if the path becomes non-traversable for a particular robot type, the motion planner will select alternative traversable areas, adjusting the robot's course to reach the next waypoint along the trajectory.

As shown in Figure 3, the traversability map illustrates various scenarios during the robots' navigation from the white star to the yellow star. The purple trajectory represents the path of the legged robot, while the orange trajectory shows the wheeled robot's path. When the path encounters stairs, the wheeled robot deviates to a nearby ramp before returning to the next waypoint, as depicted in the RGB image (a) on the right side of the figure. Meanwhile, the legged robot continues on its original path, walking directly up the stairs. For other obstacles, such as construction cones and groups of people blocking the sidewalk, as shown on the right side of Figure 3, the legged robot steps down the curb or navigates through off-road terrain to avoid the blockage.

B. Mapless Navigation with Traversability Analysis

To assess the efficacy of various traversability types in the dataset for learning-based mapless navigation algorithms, we extend the MTG algorithm [29] with multiple traversability levels, referred to as T-MTG (Fig. 6). The problem formulation in MTG is given by Equation 1:

$$p(\tau|\mathbf{c}) \approx \frac{1}{S}\sum_{s=1}^{S} p(\tau|\mathbf{z}^{(s)}, \mathbf{c}), \quad \mathbf{z}^{(s)} \sim p_\theta(\mathbf{z}|\mathbf{o}), \tag{1}$$

where τ represents the generated trajectories under the condition $\mathbf{c} = f_c(\mathbf{o})$. $f_c(\cdot)$ is a sequence of linear layers, and $\mathbf{o}$ denotes the observation information. Here, $\mathbf{z} = f_z(\mathbf{c})$ is the embedding vector of the encoded observation and $f_z(\cdot)$ represents a sequence of linear layers. During training, $\mathbf{z} = \{\mathbf{z}^{(s)}\}$ is sampled from the distribution $p_\theta(\mathbf{z}|\mathbf{o})$, where $p_\theta(\cdot)$ is the distribution of $\mathbf{z}$ and θ represents the parameters of the encoder model. S indicates the number of waypoints.

For different traversability levels, we aim for the observation embeddings to capture the current traversabilities. To achieve this, a model is employed to reprocess the traversability: $\mathbf{z}_t = f_t^k(\mathbf{z}_t|\mathbf{z}, \mathbf{o})$, where $\mathbf{o}$ is reused to provide residual information, enhancing the calculation of $\mathbf{z}_t$. k represents the traversability level. Thus, the MTG model in Equation 1 is extended to the T-MTG model in Equation 2:

$$\begin{aligned} p(\tau_k|\mathbf{c}) &\approx \frac{1}{S}\sum_{s=1}^{S} p(\tau_k|\mathbf{z}_t^{k,(s)}, \mathbf{c}), \\ \mathbf{z}_t^k &\sim p_\theta(\mathbf{z}_t^k|\mathbf{f}_z(\mathbf{z}|\mathbf{o}), \mathbf{o}), \end{aligned} \tag{2}$$

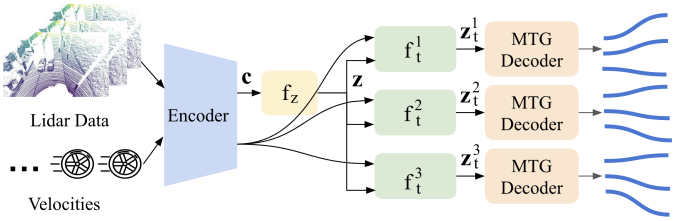


Fig. 6: **T-MTG**: As shown in Equation 2, for different traversabilities, we generate corresponding embedding vectors $\mathbf{z}_t^k$ from $\mathbf{z}$, under the condition of observation $\mathbf{o}$. The MTG Decoder is used to decode the embedding vectors into trajectories. Finally, the output of the model includes trajectories in all the traversability levels.

where $\mathbf{f}_z(\cdot)$ represents the encoder and $\mathbf{z}_t^{k,(s)}$ is the embedding of the waypoint s in $\mathbf{z}_t^k$.

As shown in Fig. 5, we implement three traversability levels: Basic traversability includes only pedestrian walkways, where robots can move in various speed on in the areas; agile traversability level is designed primarily for fast-moving wheeled robots and includes both pedestrian walkways and vehicle roadways, where the robot is required to move fast to keep up with traffic; and legged traversability level is suited for legged robots, allowing traversal on pedestrian walkways and off-road terrain, though it is not safe for use on vehicle roadways. Our approach, T-MTG, generates trajectories to cover the 200° field of view (FOV) in front of the robot. In Fig. 5, the RGB images display the generated trajectories from the front camera's perspective (70° FOV), and the traversability map highlights each waypoint of the trajectories in their respective traversability regions. For each traversability level, our approach successfully generates trajectories that lie within the appropriate regions, effectively covering the areas in front of the robot.

C. Vision-based Place Recognition

We collect both RGB and 360° camera images to offer vision-based global navigation in the GND dataset. To demonstrate the usability of 360° image data, we conduct an experiment using the NoMaD [49] algorithm for goal detection, which compares the current observation with topological image nodes to recognize the best target to follow the recorded topological nodes. For goal-directed navigation, NoMaD [49] encodes images of the robot's current RGB observations as vectors, and then uses the vectors to predict the temporal distance to the goal, by calculating the similarities with the vectors of the sub-goal images of the topological nodes. The sub-goal image with the highest similarity is then chosen as the closest node for goal-directed navigation.

We use images in four views (front, left, right, and back) from the 360° camera to implement the NoMaD [49] algorithm. Initially, we collect images with different time intervals to generate a sparse topological map. The topological map is constructed using only the front view of the 360° images. Then we use the four views from the 360° images as the robot's current observations to generate four embedded vectors. Among these vectors, we determine the direction of the context images that has the closest distance to the sub-goals and select the subgoal with the highest similarity as the closest node for further navigation, as shown in Fig. 7(a).

We notice that the average similarity score of all four views from the real-time image observations and the goal (blue line) is higher than the comparisons with only one of the image, as shown in Fig. 7(b). Additionally, the temporal distance between observed images and sub-goals also increases with a sparser topological map. These findings highlight the advantages of utilizing all available directional information from 360° images for more accurate and efficient goal-directed navigation.

V. Conclusion, Limitations, and Future Work

We introduce GND, a large-scale, long-range, global robot navigation dataset that includes multi-modal perception data and multi-category traversability maps. This dataset enables robots to undertake long-range navigation while accounting for geometric and semantic traversability. We are making all data available online, accompanied by a standardized post-processing pipeline, and we encourage contributions from the wider research community to enhance the dataset. The primary features of GND—*globalness* and *traversability*—are illustrated through three distinct experiments, highlighting its potential applications. Although GND offers valuable sensory data, we recognize that there is still work to be done in providing benchmark scenarios and evaluation metrics for global robot navigation, which remain underexplored in the existing literature.

Acknowledgment: This work was supported in part by ARO Grant W911NF2310046 and W911NF2310352 and Army Cooperative Agreement W911NF2120076.

References

[1] Naoki Akai, Luis Yoichi Morales, Eijiro Takeuchi, Yuki Yoshihara, and Yoshiki Ninomiya. Robust localization using 3d ndt scan matching with experimentally determined uncertainty and road marker matching. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1356–1363. IEEE, 2017.

[2] Jong Wook Bae, Jungho Kim, Junyong Yun, Changwon Kang, Jeongseon Choi, Chanhyeok Kim, Junho Lee, Jungwook Choi, and Jun Won Choi. Sit dataset: socially interactive pedestrian trajectory dataset for social navigation robots. *Advances in Neural Information Processing Systems*, 36, 2024.

[3] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[4] Chao Cao, Hongbiao Zhu, Howie Choset, and Ji Zhang. Tare: A hierarchical framework for efficiently exploring

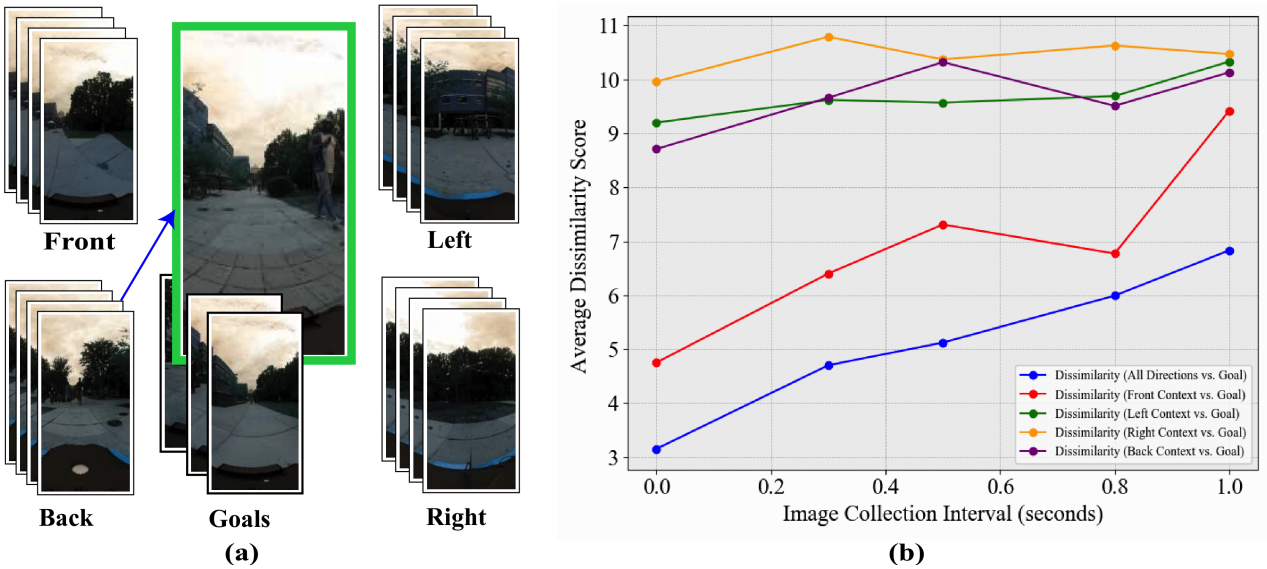


Fig. 7: **Vision-based Place Recognition using 360° images in GND**: Inset (a) shows the application of 360° context images for goal detection. In this particular example, when all four directional context images are available, the back context image is selected as the closest to the goal image (green-bordered), highlighting the advantage of 360° images in goal detection. (b) represents how the average dissimilarity between all directional context images and the goal (blue line) decreases compared to using only one directional context image.

complex 3d environments. In *Robotics: Science and Systems*, volume 5, page 2, 2021.

[5] Chao Cao, Hongbiao Zhu, Zhongqiang Ren, Howie Choset, and Ji Zhang. Representation granularity enables time-efficient autonomous exploration in large, complex worlds. *Science Robotics*, 8(80):eadf0970, 2023.

[6] Nicholas Carlevaris-Bianco, Arash K. Ushani, and Ryan M. Eustice. University of Michigan North Campus long-term vision and lidar dataset. *International Journal of Robotics Research*, 35(9):1023–1035, 2015.

[7] Nicholas Carlevaris-Bianco, Arash K Ushani, and Ryan M Eustice. University of michigan north campus long-term vision and lidar dataset. *The International Journal of Robotics Research*, 35(9):1023–1035, 2016.

[8] Cheng Chen, Emrah Demir, Yuan Huang, and Rongzu Qiu. The adoption of self-driving delivery robots in last mile logistics. *Transportation research part E: logistics and transportation review*, 146:102214, 2021.

[9] Benjamin Congram and Timothy D Barfoot. Relatively lazy: Indoor-outdoor navigation using vision and gnss. In *2021 18th Conference on Robots and Vision (CRV)*, pages 25–32. IEEE, 2021.

[10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[11] Lalitha Dabbiru, Suvash Sharma, Chris Goodin, Sam Ozier, Christopher Hudson, Daniel Carruth, Matthew Doude, George Mason, and John Ball. Traversability mapping in off-road environment using semantic segmentation. In *Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2021*, volume 11748, pages 78–83. SPIE, 2021.

[12] Dibyendu Das, Yuanjie Lu, Erion Plaku, and Xuesu Xiao. Motion memory: Leveraging past experiences to accelerate future motion planning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 16467–16474. IEEE, 2024.

[13] Aniket Datar, Chenhui Pan, and Xuesu Xiao. Learning to model and plan for wheeled mobility on vertically challenging terrain. *arXiv preprint arXiv:2306.11611*, 2023.

[14] Aniket Datar, Chenhui Pan, Mohammad Nazeri, Anuj Pokhrel, and Xuesu Xiao. Terrain-attentive learning for efficient 6-dof kinodynamic modeling on vertically challenging terrain. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024.

[15] Aniket Datar, Chenhui Pan, Mohammad Nazeri, and Xuesu Xiao. Toward wheeled mobility on vertically challenging terrain: Platforms, datasets, and algorithms. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 16322–16329. IEEE, 2024.

[16] Dieter Fox, Wolfram Burgard, and Sebastian Thrun. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1):23–33, 1997.

[17] Anthony Francis, Aleksandra Faust, Hao-Tien Lewis Chiang, Jasmine Hsu, J Chase Kew, Marek Fiser, and Tsang-Wei Edward Lee. Long-range indoor navigation with prm-rl. *IEEE Transactions on Robotics*, 36(4):1115–1134, 2020.

[18] S Julius Fusic, G Kanagaraj, K Hariharan, and

S Karthikeyan. Optimal path planning of autonomous navigation in outdoor environment via heuristic technique. *Transportation research interdisciplinary perspectives*, 12:100473, 2021.

[19] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11): 1231–1237, 2013.

[20] Noriaki Hirose, Dhruv Shah, Ajay Sridhar, and Sergey Levine. Sacson: Scalable autonomous control for social navigation. *IEEE Robotics and Automation Letters*, 2023.

[21] Thomas Hoffmann and Gunnar Prause. On the regulatory framework for last-mile delivery robots. *Machines*, 6(3): 33, 2018.

[22] Sadegh Hosseinpoor, Jim Torresen, Mathias Mantelli, Diego Pitto, Mariana Kolberg, Renan Maffei, and Edson Prestes. Traversability analysis by semantic terrain segmentation for mobile robots. In *2021 IEEE 17th international conference on automation science and engineering (CASE)*, pages 1407–1413. IEEE, 2021.

[23] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Sören Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas, and Peter Stone. Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation. *IEEE Robotics and Automation Letters*, 2022.

[24] Haresh Karnan, Kavan Singh Sikand, Pranav Atreya, Sadegh Rabiee, Xuesu Xiao, Garrett Warnell, Peter Stone, and Joydeep Biswas. Vi-ikd: High-speed accurate off-road navigation using learned visual-inertial inverse kinodynamics. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3294–3301. IEEE, 2022.

[25] Jing Li, Hui Qin, Junzheng Wang, and Jiehao Li. Openstreetmap-based autonomous navigation for the four wheel-legged robot via 3d-lidar and ccd camera. *IEEE Transactions on Industrial Electronics*, 69(3):2708–2717, 2021.

[26] Jing Liang, Utsav Patel, Adarsh Jagan Sathyamoorthy, and Dinesh Manocha. Crowd-steer: Realtime smooth and collision-free robot navigation in densely crowded scenarios trained using high-fidelity simulation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 4221–4228, 2021.

[27] Jing Liang, Yi-Ling Qiao, Tianrui Guan, and Dinesh Manocha. Of-vo: Efficient navigation among pedestrians using commodity sensors. *IEEE Robotics and Automation Letters*, 6(4):6148–6155, 2021.

[28] Jing Liang, Kasun Weerakoon, Tianrui Guan, Nare Karapetyan, and Dinesh Manocha. Adaptiveon: Adaptive outdoor local navigation method for stable and reliable actions. *IEEE Robotics and Automation Letters*, 8(2): 648–655, 2022.

[29] Jing Liang, Peng Gao, Xuesu Xiao, Adarsh Jagan Sathyamoorthy, Mohamed Elnoor, Ming C Lin, and Dinesh Manocha. Mtg: Mapless trajectory generator with traversability coverage for outdoor navigation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2396–2402. IEEE, 2024.

[30] Jing Liang, Amirreza Payandeh, Daeun Song, Xuesu Xiao, and Dinesh Manocha. Dtg: Diffusion-based trajectory generation for mapless global navigation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024.

[31] Bo Liu, Xuesu Xiao, and Peter Stone. Team orienteering coverage planning with uncertain reward. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9728–9733. IEEE, 2021.

[32] Stephanie Lowry, Niko Sünderhauf, Paul Newman, John J Leonard, David Cox, Peter Corke, and Michael J Milford. Visual place recognition: A survey. *ieee transactions on robotics*, 32(1):1–19, 2015.

[33] Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, Wanli Ouyang, and Xin Fan. Accurate monocular 3d object detection via color-embedded 3d reconstruction for autonomous driving. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6851–6860, 2019.

[34] Roberto Martin-Martin, Mihir Patel, Hamid Rezatofighi, Abhijeet Shenoi, JunYoung Gwak, Eric Frankel, Amir Sadeghian, and Silvio Savarese. Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments. *IEEE transactions on pattern analysis and machine intelligence*, 45(6):6748–6765, 2021.

[35] Duc M Nguyen, Mohammad Nazeri, Amirreza Payandeh, Aniket Datar, and Xuesu Xiao. Toward human-like social robot navigation: A large-scale, multi-modal, social human navigation dataset. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 7442–7447, 2023.

[36] Minho Oh, Euigon Jung, Hyungtae Lim, Wonho Song, Sumin Hu, Eungchang Mason Lee, Junghee Park, Jaekyung Kim, Jangwoo Lee, and Hyun Myung. Travel: Traversable ground and above-ground object segmentation using graph representation of 3d lidar scans. *IEEE Robotics and Automation Letters*, 7(3):7255–7262, 2022.

[37] Teddy Ort, Liam Paull, and Daniela Rus. Autonomous vehicle navigation in rural environments without detailed prior maps. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 2040–2047. IEEE, 2018.

[38] Jin-Soo Park, Xuesu Xiao, Garrett Warnell, Harel Yedidsion, and Peter Stone. Learning perceptual hallucination for multi-robot navigation in narrow hallways. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10033–10039. IEEE, 2023.

[39] Mark Pfeiffer, Michael Schaeuble, Juan Nieto, Roland Siegwart, and Cesar Cadena. From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots. In *2017 ieee*

international conference on robotics and automation (icra), pages 1527–1533. IEEE, 2017.

[40] Thomas Roddick and Roberto Cipolla. Predicting semantic map representations from images using pyramid occupancy networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11138–11147, 2020.

[41] Andrey Rudenko, Tomasz P Kucner, Chittaranjan S Swaminathan, Ravi T Chadalavada, Kai O Arras, and Achim J Lilienthal. Thör: Human-robot navigation data collection and accurate motion trajectories dataset. *IEEE Robotics and Automation Letters*, 5(2):676–682, 2020.

[42] Dhruv Shah and Sergey Levine. Viking: Vision-based kilometer-scale navigation with geographic hints. *arXiv preprint arXiv:2202.11271*, 2022.

[43] Dhruv Shah, Benjamin Eysenbach, Gregory Kahn, Nicholas Rhinehart, and Sergey Levine. Rapid exploration for open-world navigation with latent goal models. *arXiv preprint arXiv:2104.05859*, 2021.

[44] Dhruv Shah, Ajay Sridhar, Arjun Bhorkar, Noriaki Hirose, and Sergey Levine. Gnm: A general navigation model to drive any robot. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7226–7233. IEEE, 2023.

[45] Dhruv Shah, Ajay Sridhar, Nitish Dashora, Kyle Stachowicz, Kevin Black, Noriaki Hirose, and Sergey Levine. Vint: A foundation model for visual navigation. *arXiv preprint arXiv:2306.14846*, 2023.

[46] Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Daniela Rus. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 5135–5142. IEEE, 2020.

[47] Kavan Singh Sikand, Sadegh Rabiee, Adam Uccello, Xuesu Xiao, Garrett Warnell, and Joydeep Biswas. Visual representation learning for preference-aware path planning. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 11303–11309. IEEE, 2022.

[48] Daeun Song, Jing Liang, Xuesu Xiao, and Dinesh Manocha. Vl-tgs: Trajectory generation and selection using vision language models in mapless outdoor environments. *IEEE Robotics and Automation Letters*, 2025.

[49] Ajay Sridhar, Dhruv Shah, Catherine Glossop, and Sergey Levine. Nomad: Goal masked diffusion policies for navigation and exploration. *arXiv preprint arXiv:2310.07896*, 2023.

[50] Gregory Stein. Generating high-quality explanations for navigation in partially-revealed environments. *Advances in neural information processing systems*, 34:17493–17506, 2021.

[51] Gregory J Stein, Christopher Bradley, and Nicholas Roy. Learning over subgoals for efficient navigation of structured, unknown environments. In *Conference on robot learning*, pages 213–222. PMLR, 2018.

[52] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin 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*, pages 2446–2454, 2020.

[53] Zizhao Wang, Xuesu Xiao, Bo Liu, Garrett Warnell, and Peter Stone. Appli: Adaptive planner parameter learning from interventions. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 6079–6085. IEEE, 2021.

[54] Zizhao Wang, Xuesu Xiao, Alexander J Nettekoven, Kadhiravan Umasankar, Anika Singh, Sriram Bommakanti, Ufuk Topcu, and Peter Stone. From agile ground to aerial navigation: Learning from learned hallucination. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 148–153. IEEE, 2021.

[55] Zizhao Wang, Xuesu Xiao, Garrett Warnell, and Peter Stone. Apple: Adaptive planner parameter learning from evaluative feedback. *IEEE Robotics and Automation Letters*, 6(4):7744–7749, 2021.

[56] Kasun Weerakoon, Adarsh Jagan Sathyamoorthy, Jing Liang, Tianrui Guan, Utsav Patel, and Dinesh Manocha. Graspe: Graph based multimodal fusion for robot navigation in outdoor environments. *IEEE Robotics and Automation Letters*, 2023.

[57] Xuesu Xiao, Bo Liu, Garrett Warnell, Jonathan Fink, and Peter Stone. Appld: Adaptive planner parameter learning from demonstration. *IEEE Robotics and Automation Letters*, 5(3):4541–4547, 2020.

[58] Xuesu Xiao, Joydeep Biswas, and Peter Stone. Learning inverse kinodynamics for accurate high-speed off-road navigation on unstructured terrain. *IEEE Robotics and Automation Letters*, 6(3):6054–6060, 2021.

[59] Xuesu Xiao, Bo Liu, and Peter Stone. Agile robot navigation through hallucinated learning and sober deployment. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 7316–7322. IEEE, 2021.

[60] Xuesu Xiao, Bo Liu, Garrett Warnell, and Peter Stone. Toward agile maneuvers in highly constrained spaces: Learning from hallucination. *IEEE Robotics and Automation Letters*, 6(2):1503–1510, 2021.

[61] Xuesu Xiao, Zizhao Wang, Zifan Xu, Bo Liu, Garrett Warnell, Gauraang Dhamankar, Anirudh Nair, and Peter Stone. Appl: Adaptive planner parameter learning. *Robotics and Autonomous Systems*, 154:104132, 2022.

[62] Xuesu Xiao, Zifan Xu, Zizhao Wang, Yunlong Song, Garrett Warnell, Peter Stone, Tingnan Zhang, Shravan Ravi, Gary Wang, Haresh Karnan, et al. Autonomous ground navigation in highly constrained spaces: Lessons learned from the benchmark autonomous robot navigation challenge at icra 2022 [competitions]. *IEEE Robotics & Automation Magazine*, 29(4):148–156, 2022.

[63] Xuesu Xiao, Zifan Xu, Garrett Warnell, Peter Stone, Ferran Gebelli Guinjoan, Rômulo T Rodrigues, Herman

Bruyninckx, Hanjaya Mandala, Guilherme Christmann, Jose Luis Blanco-Claraco, et al. Autonomous ground navigation in highly constrained spaces: Lessons learned from the second barn challenge at icra 2023 [competitions]. *IEEE Robotics & Automation Magazine*, 30(4): 91–97, 2023.

[64] Wei Xu and Fu Zhang. Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter. *IEEE Robotics and Automation Letters*, 6(2):3317–3324, 2021.

[65] Zifan Xu, Gauraang Dhamankar, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Bo Liu, Zizhao Wang, and Peter Stone. Applr: Adaptive planner parameter learning from reinforcement. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 6086–6092. IEEE, 2021.

[66] Zhi Yan, Simon Schreiberhuber, Georg Halmetschlager, Tom Duckett, Markus Vincze, and Nicola Bellotto. Robot perception of static and dynamic objects with an autonomous floor scrubber. *Intelligent Service Robotics*, 13(3):403–417, 2020.

[67] Heng Yang, Jingnan Shi, and Luca Carlone. Teaser: Fast and certifiable point cloud registration. *IEEE Transactions on Robotics*, 37(2):314–333, 2020.

[68] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443–58469, 2020.