2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

M3-GMN: A Multi-environment, Multi-LiDAR, Multi-task dataset for Grid Map based Navigation

Guanglei Xie, Hao Fu, Hanzhang Xue, Bokai Liu, Xin Xu, Xiaohui Li and Zhenping Sun*

***Abstract*— In this paper, we propose a multi-environment, multi-LiDAR, multi-task dataset to promote the grid map-based navigation capability for autonomous vehicles. The dataset comprises structured and unstructured environmental data captured by different types of LiDAR and contains various challenging scenarios, including moving objects, negative obstacles, steep slopes, cliffs, overhangs, etc. Further, we have devised an innovative method for generating ground truth, facilitating the creation of dense, accurate, and stable grid maps with a minimal requirement for human annotation efforts. A new baseline method and two existing approaches are evaluated on this dataset. Results indicate that existing approaches perform much worse than the proposed baseline. The dataset will be made publicly available at https://github.com/guanglei96/M3-GMN.**

I. INTRODUCTION

Traversability analysis is one of the most basic and critical tasks for autonomous driving [1]. In structured environments, this task is relatively easy as the terrain is mostly flat, and we only need to perform binary classification on the LiDAR points or 2D grid cells. Related tasks include ground segmentation [2], free-space detection, road detection [3], obstacle detection [4], etc. However, in unstructured environments where the terrain is irregular [5], a simple binary classification model may not work well, and we have to resort to a more complex model to fit the 3D terrain [6] [7]. Related tasks include terrain modeling [1] [8] [9] or traversability analysis [10] [11] [12].

In specific scenarios, autonomous vehicles are required to navigate through both structured and unstructured environments. In such cases, a unified environmental representation becomes necessary. In both structured and unstructured environments, the output of an autonomous vehicle's environment perception system is usually represented as a Bird's Eye View (BEV) grid map [13]. In the grid map, each grid cell can store different information, such as semantic attributes, elevation, reflectivity, speed, etc. Due to its low computational cost, the grid map is the de facto standard environmental representation method used by most autonomous vehicles, especially those operating in unstructured environments. It is the crucial link between the environment perception and path planning modules. Only if the environment perception system can output a perfect grid map, then we could expect the autonomous vehicle to exhibit outstanding performance.

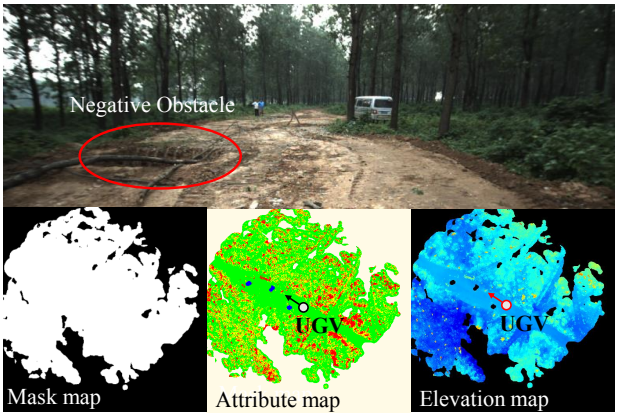


Fig. 1. The image shows a typical off-road scenario with negative obstacles in our dataset. The dataset provides three grid maps for this scenario: a mask map indicating the enlarged terrain region, an attribute map consisting of four semantic categories (negative obstacle, terrain, grey zone, and positive obstacle), and an elevation map that stores the maximum height of the grid cell.

So, how do we measure the quality of a grid map? We believe that an ideal grid map should possess three attributes: namely **accuracy, completeness**, and **stability**. The accuracy attribute requires the grid map to be semantically and geometrically accurate. The completeness attribute requires the grid map to be a dense representation. To achieve this, the algorithm must utilize historical observations and inferences about missing data. The stability attribute is to ensure consistent outputs from consecutive frames. A stable grid map will enable the path planning module to generate a more reliable path.

Although grid map plays an important role in autonomous driving, we were still looking for datasets that provide the ground truth of an accurate, dense, and stable grid map. To fill this gap, we construct a dataset that contains grid-level ground truth in this paper. The proposed dataset has the following characteristics:

- The dataset comprises sequences collected in both structured and unstructured environments to enable the smooth transition of autonomous navigation from structured to unstructured environment;
- The dataset is collected from multiple types of LiDAR to promote a generic algorithm that is not overfitted to a specific type of LiDAR;

The research was partly funded by the Natural Science Foundation of China under Grant 62003362.

Corresponding author: Zhenping Sun,sunzhenping@nudt.edu.cn.
The authors are with the College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China.

- The dataset contains several typical challenging scenarios, including moving objects, negative obstacles, steep slopes, cliffs, overhangs, etc.;
- The dataset provides frame-level dense ground truth in the form of an attribute grid map and an elevation grid map. These ground truth grid maps are designed to be accurate, complete, and stable, as shown in Fig. 1.

In summary, we make three contributions in this paper:

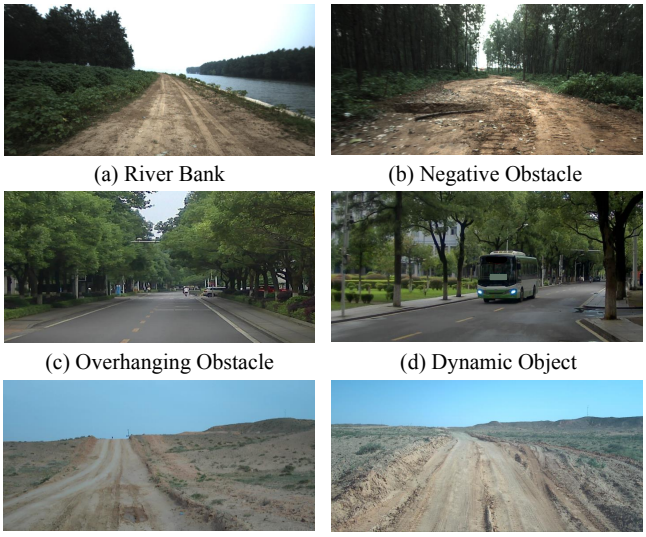
- For the first time, we propose **a dataset that provides grid-level(BEV) dense ground truth**. The dataset was collected in both structured and unstructured environments with different types of LiDAR.
- To generate grid-level ground truth, we propose **a semi-automatic method** that effectively fuses information from past, current, and future observations. The method requires minimal human labeling effort, thus enabling it to be scalable to larger datasets.
- We propose **new performance metrics to evaluate the accuracy, completeness, and stability attributes** of a grid map. Additionally, we implemented a strong baseline method that outperforms two existing methods on this dataset.

II. RELATED WORK

There has been a vast amount of datasets proposed for autonomous driving research. However, most of these datasets are focused on urban road environments, such as KITTI [14], nuScenes [15], and others. There are few datasets available that target non-structured environments in the wild. DeepScene [16] was the first publicly available off-road dataset for multi-spectral image segmentation. YCOR [17] collected more diverse and challenging environmental data, and ORFD [18] is the first off-road free space detection dataset. Table. I provides an overview of the wild datasets containing LiDAR data. While YCOR includes LiDAR data and 2D images, only the images have been annotated. Although RELLIS_3D [19] and ORFD have annotated LiDAR point clouds, their labels are obtained by directly projecting the LiDAR point cloud to the images. The image labels are then transferred to the LiDAR point cloud and thus have significant noise. Although the Dabbiru et al. [20] dataset provides point-wise semantic labels, they are generated from a simulator.

Annotating the LiDAR point cloud is laborious work. Annotating non-structured outdoor environment data is more challenging, owing to greater semantic ambiguity. Data annotation tools such as ImageTagger [21] and OneLabeler [22] are not effective in wild environments. Furthermore, directly transforming pixel-level or point-level semantic annotations to a grid map may produce artifacts, as to be shown in Fig. 7. As the grid map is a simplified environmental representation, it is not easy to annotate it directly. Therefore, we try to design a semi-automatic method to generate grid-level ground truth in this paper.

Although the grid cell could store multiple pieces of information, the most relevant ones to the vehicle's traversability



(a) River Bank (b) Negative Obstacle (c) Overhanging Obstacle (d) Dynamic Object (e) Steep Slope (f) Rough Road

Fig. 2. Our dataset contains a variety of typical scenes.

are the semantic attributes and elevation. The classification of drivable/non-drivable areas is the most basic semantic attribute for autonomous driving. On top of this binary classification, Gao et.al. [11] marked those areas that are difficult to classify as grey zone, thus producing a three-value semantic attribute map. This idea is further extended in [23], where the authors refined the drivable/non-drivable attribute into four categories: free, low cost, medium cost, and lethal. A semantic grid map is useful for autonomous vehicles to plan a safe path. However, the semantic attribute is a subjective concept and inevitably contains ambiguities. For example, a low-height curb may form a non-drivable area for a small-sized vehicle but may become drivable for a larger vehicle. In contrast to the semantic attribute grid map, the elevation grid map provides a more objective description of the local 3D environment, where each grid cell stores the maximum height value. Only storing the maximum height value in a grid cell has limitations: grid cells with hanging tree branches may be considered non-drivable. To overcome this, methods such as extended elevation map [24] or multi-level surface map [25] can be used. In this paper, we believe that the semantic attribute map and the elevation map contain complementary information, and are both functional for autonomous driving. So in our dataset, we generate an attribute grid map and an extended elevation grid map for each frame using our newly proposed ground truth generation method, which will be described in the next section.

III. DATASET

In this paper, we proposed a new dataset named M3-GMN (Multi-environment, Multi-LiDAR, Multi-task dataset for Grid Map based Navigation), that contains both structured and unstructured environmental data captured by different types of LiDAR.

TABLE I

OVERVIEW OF OFF-ROAD LIDAR DATASETS.

Dataset	Cities	Sensors	Annotation Type	LiDAR diversity	Scene diversity	Dense GT
YCOR [17]	2	camera	pixel-wise	✗	✗	✗
RELLIS_3D [19]	1	camera, LiDAR	pixel/point-wise	✗	✗	✗
ORFD [18]	1	camera, LiDAR	pixel-wise/point-wise	✗	✓	✗
Dabbiru et al. [20]	1	SimLiDAR	point-wise	✗	✗	✗
M3-GMN(Ours)	**3**	camera, LiDAR	**grid-wise(BEV)**	✓	✓	✓

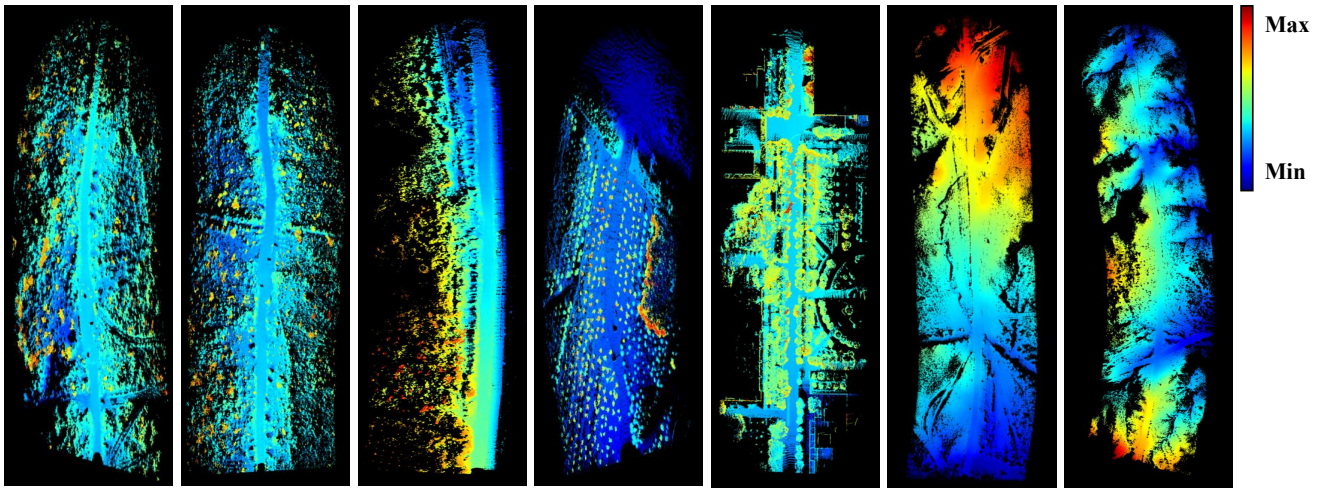


Fig. 3. The elevation maps generated from the seven sequences in our dataset.

TABLE II

DETAILS OF THE M3-GMN DATASET

Sequences	Features	Frames	LiDAR Type
1 (a)	3 Negative Obstacles	200	Velodyne HDL64
2 (b)	2 Negative Obstacles	200	Velodyne HDL64
3 (c)	River Bank	200	Velodyne HDL64
4 (d)	2 Negative Obstacles	200	Pandar40
5 (e)	Overhang/Moving Objects	200	Pandar64
6 (f)	Steep Slope	278	Pandar128
7 (g)	Rough Road	300	Pandar128

A. *Dataset Description*

The M3-GMN dataset features various challenging scenarios, including moving objects, negative obstacles, steep slopes, cliffs, overhangs, etc. Some typical scenes are shown in Fig. 2. Moreover, the dataset provides dense ground truth to promote grid map-based autonomous navigation capability. A comparison of our dataset with similar datasets is shown in Table. I.

The proposed dataset contains a total of seven sequences we collected. Each sequence contains LiDAR point clouds, RGB images, GPS/INS measurements, and corresponding configuration parameter files. To facilitate the validation of our semi-automatic ground truth generation method and to ease the comparison with state-of-the-art approaches, we incorporate the SemanticKITTI 07 sequence into our experiments. Each sequence in our dataset features specific challenges, as listed in Table. II. For each frame in the sequence, we also released the SLAM-optimized pose [26] using GPS/INS data, based on which we can easily assemble the frames to generate a global map, as shown in Fig. 3. We can also use other optimization methods to obtain the SLAM pose. Please note that the poses obtained after SLAM optimization differ from the original GPS/INS measurement results.

B. *Grid-level ground truth generation*

For each LiDAR frame, we try to generate its grid-level ground truth, which consists of an attribute grid map and an elevation grid map. In the attribute map, we define four semantic categories: terrain, positive obstacles, negative obstacles, and grey zone [11]. The elevation grid map stores the maximum height of each grid cell. The elevation values are invalid in negative obstacle areas since the bottom of the negative obstacles may not be scanned by LiDAR.

Additionally, similar to [8] and [27], our algorithm only focuses on areas with terrain support. Therefore, in addition to the attribute map and the elevation map, we also generate a terrain mask. Only regions within this mask are considered as regions of interest. A similar idea also appears in [28].

To ensure the completeness and stability of the grid map, we not only need to process the currently observed grid cells, but also those cells observed in the past. Additionally,

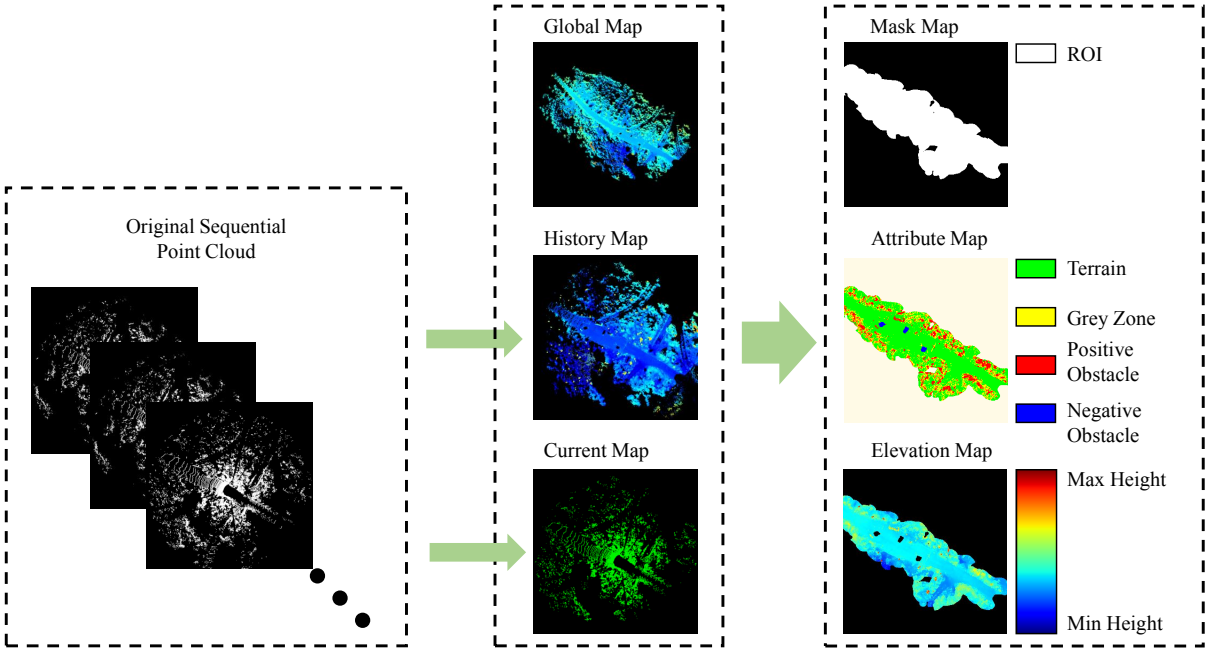


Fig. 4. An illustration of the proposed ground truth generation approach.

the algorithm should combine current and historical observations to predict the attributes of those unobserved grid cells.

From a grid cell's perspective, it contains past, current, and future observations. To calculate its ground truth state, three occasions might occur: if the grid cell is observed in the current frame, then we use both the current and historical observations to calculate its ground truth value; if a grid cell is not observed in the current frame but was observed in a previous frame, then we use historical observations to calculate its ground truth; otherwise, we can only use future observations to calculate its ground truth. This process is illustrated in Fig. 4.

We first discuss how to calculate the ground truth of a grid cell using all of its observations.

1) Using All Observations to Calculate the Global Map: Firstly, we use the pose of each frame to assemble all the data into a global map. To eliminate the long trail generated by moving objects, we adopt the approach proposed in [27] to distinguish between moving and static points. Those moving points are then removed from the global map.

We then adopt the terrain estimation method proposed in [27] to estimate a dense terrain map. The coverage of this dense terrain map defines our mask region, known as the global mask map. Only grid cells within the mask are considered as cells of interest and will be further processed.

Let z represent the global height of the point in the global map. By comparing it with the terrain height h_t obtained from the dense terrain map, we can easily judge its semantic attribute a_P use (1), where $a_P \in \{\text{ignored}, \text{noise}, \text{terrain}, \text{positive obstacles}, \text{overhang}\}$.

$$a_P = \begin{cases} \text{noise points}, & if\ z < h_t - th_{noise} \\ \text{terrain}, & if\ |z - h_t| \leq th_{terrain} \\ \text{positive obstacles}, & if\ h_t + th_{terrain} < z \\ & \quad < h_t + th_{overhangs} \\ \text{overhangs}, & if\ z > h_t + th_{overhangs} \end{cases} \tag{1}$$

where th_{noise}, $th_{terrain}$, $th_{overhangs}$ are threshold values. These threshold parameters should be closely related to the size and passability of the vehicle. Vehicles with stronger passability should have higher thresholds. Please refer to section IV for specific parameter values.

We then generate the global attribute map by only using points classified as terrain and positive obstacles, ignoring noise points and overhangs. The specific generation method is as follows: we first calculate the mean height μ, maximum height h_{max}, minimum height h_{min}, and standard deviation σ of each grid cell. If $\sigma < 0.1$ and the difference between μ and the terrain height h_t is less than 0.2m, then we consider the attribute of this grid cell to be terrain. If $|h_{max} - h_t| < th_{grey}$ and $|h_{min} - h_t| < th_{grey}$, the attribute of this grid cell is considered to be grey zone. Otherwise, it is considered as positive obstacle. Additionally, we manually label the negative obstacle area if there are any. In the end, we obtain the global attribute map a_g consists of four categories, i.e. negative obstacles, terrain, grey zone, and positive obstacles. The results obtained by applying the above method on SemanticKITTI sequence 07 are shown in Fig. 5.

Then we compute the global elevation map based on the global attribute map. The elevation value e_g is computed

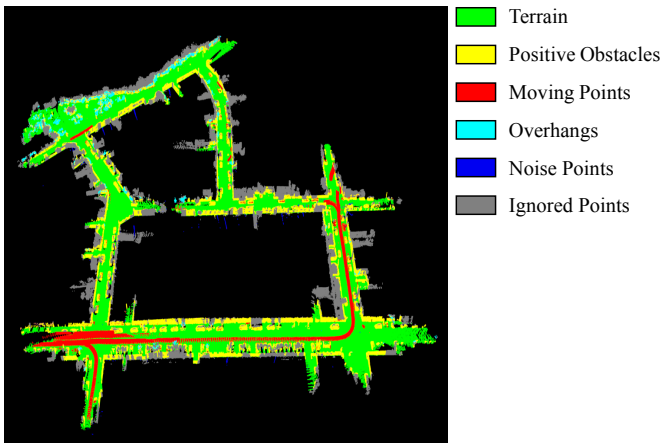


Fig. 5. By applying the terrain modeling and moving point identification method proposed in [27], we can classify each point in SemanticKITTI sequence 07 into ignored, noise, terrain, positive obstacle, and overhang.

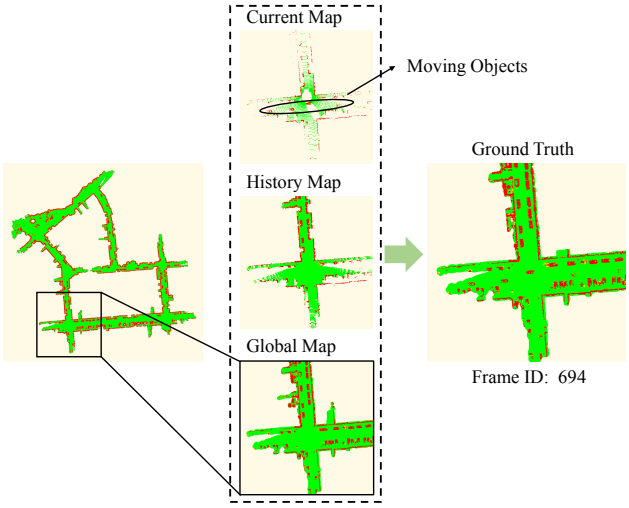


Fig. 6. The process of fusing current Map, history map, and global map to generate the ground truth. Moving points only appear in the current map.

using the following equation:

$$e_g = \begin{cases} \mu, & if\ a_g \in \text{terrain} \\ h_{max}, & if\ a_g \in \text{grey zone } or \text{ obstacle} \\ \text{INVALID}, & if\ a_g \in \text{ negative obstacle} \end{cases} \quad (2)$$

In the above equation, the reason we use mean height instead of the maximum height as the elevation value for the terrain cell is to ensure the grid map's stability.

Thus, by conducting a unified analysis of all observations, we have obtained the global mask map, global attribute map, and global elevation map.

2) Using Past Observations to Calculate the History Map: We utilize the rolling grid technique [8] to combine all previous observations to generate a history map. For each point in the history map, we search for its nearest neighbor in the global map and assign the attribute value of the nearest neighbor to it. In this way, we classify each point in the history map into moving points, ignored points, terrain points, noise points, positive obstacles, and overhangs. Next, we only utilize terrain points and positive obstacle points to calculate each grid cell's mean height, maximum height, minimum height, and standard deviation. The calculation of the history attribute map is similar to the global attribute map, and the calculation of the history elevation map also follows (2).

3) Using Current Observation to Calculate the Current Map: For the current frame, we also use the method of finding its nearest neighbor in the global map to classify the point's attribute. Unlike the history map, when calculating the current attribute map and current elevation map, we not only need to use points classified as terrain or positive obstacle class, but also those moving points. These moving points only appear in the current map.

4) Fusing Current Map, History Map, and Global Map to Calculate the Ground Truth: In the end, we combine information from the current grid map, history grid map, and global grid map to generate the ground truth. We prioritize the current map over the history map and global map. Let a_{gt}, a_c, a_h, a_g represent the ground truth attribute, the current attribute map, the history attribute map, and the global attribute map respectively, then a_{gt} can be calculated as:

$$a_{gt} = \begin{cases} a_c, & if\ a_c \neq \text{INVALID} \\ a_h, & if\ a_c = \text{INVALID}\ and\ a_h \neq \text{INVALID} \\ a_g, & if\ a_c = \text{INVALID}\ and\ a_h = \text{INVALID} \end{cases} \quad (3)$$

Similarly, the ground truth elevation map e_{gt} can be calculated as:

$$e_{gt} = \begin{cases} e_c, & if\ e_c \neq \text{INVALID}\ and\ e_h = \text{INVALID} \\ \max(e_c, e_h), & if\ e_c \neq \text{INVALID}\ and \\ & \quad e_h \neq \text{INVALID} \\ e_h, & if\ e_c = \text{INVALID}\ and\ e_h \neq \text{INVALID} \\ e_g, & if\ e_c = \text{INVALID}\ and\ e_h = \text{INVALID} \end{cases} \quad (4)$$

Fig. 6 illustrates this fusion process. It can be observed that moving points are only preserved in the current Map and are removed in the history map and global map.

IV. EXPERIMENTS

Our dataset comprises seven self-collected data sequences and sequence 07 from the SemanticKITTI dataset. A main contribution of this paper is the ground truth generation approach described in the previous section, which can automatically generate grid-level ground truth with minimal human involvement (only requiring manual labeling of negative obstacles). Therefore, in this section, we first conduct experiments on the SemanticKITTI sequence to verify the effectiveness of this ground truth generation approach. The parameters are set to $th_{noise} = 0.7$m, $th_{terrain} = 0.2$m, $th_{grey} = 0.5$m and $th_{overhangs} = 1.5$m. The resolution of the grid map we generated is 0.2m.

A. The Validity of The Ground Truth

Following [23] and [27], we first use the semantic labels of each point in the SemanticKITTI dataset to

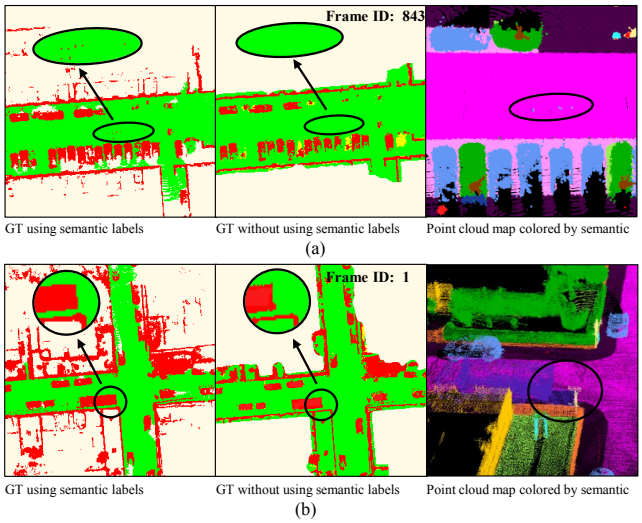


Fig. 7. Comparative results between using the semantic labels and without using the semantic labels on SemanticKITTI sequence 07 dataset.

generate the ground truth. Specifically, we ignore points labeled as 'noise' and 'unlabeled' and consider points labeled as 'lane marking,' 'road,' 'parking,' 'sidewalk,' 'other ground,' 'terrain,' and 'vegetation' (only those with a z-value below 1.3 m) as terrain. Points labeled as 'moving-car,' 'moving-bicyclist,' 'moving-person,' 'moving-motorcyclist,' 'moving-on-rails,' 'moving-bus,' 'moving-truck,' and 'moving-other-vehicle' are considered as moving points. Since SemanticKITTI 07 sequence does not contain negative obstacles, our ground truth generation method does not require any manual labeling.

Then, we use the same fusion strategy to generate ground truth for the grid map. The comparison of the ground truth generated entirely by our method, and the ground truth generated using semantic labels can be seen in Fig. 7.

In Fig. 7, we can observe that the ground truth generated entirely by our method closely matches the results generated using semantic labels. Upon closer inspection, we discovered that the ground truth generated using semantic labels contains some noise, as evidenced by Fig. 7(a), where some points on moving vehicles were not labeled as moving objects, resulting in wrong attributes in the grid map. In addition, Fig. 7(b) shows that labeling the truck tail plate as 'truck' mistakenly caused this region to be labeled as non-terrain. More qualitative results can be found in our supplementary material.

This experiment clearly demonstrates that using our method alone can produce ground truth similar to that generated using semantic labels. Moreover, our pure geometry-based ground truth generation approach avoids semantic ambiguities and possible errors from incomplete labeling of moving objects.

B. Evaluation Metrics

As previously mentioned, the generated grid map is a fusion of past, current, and future observations while maintaining accuracy, completeness, and stability. To quantitatively evaluate these attributes, we designed the following evaluation metrics:

1) ***Accuracy****:* To evaluate the accuracy of the attribute grid map, we use three commonly used metrics, namely the precision P, the recall R and $F1$-score,

$$\begin{cases} P = \dfrac{TP}{TP+FP} \\ R = \dfrac{TP}{TP+FN} \\ F_1 = \dfrac{2PR}{P+R} \end{cases} \tag{5}$$

for 'terrain' and 'positive obstacle' class. For the 'negative obstacle' class, we use the IOU(Intersection Over Union) metric. The 'grey zone' class is ignored for evaluation.

To evaluate the accuracy of the elevation grid map, we use the root mean square error

$$\text{RMSE} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{K}\sum_{j=1}^{K}\sqrt{\left(z_{ij}-z_{ij}^{GT}\right)^2}\right] \tag{6}$$

where N represents the total number of evaluated frames, K represents the number of terrain/positive obstacle grid cells in each frame, z_{ij} represents the estimated height of the j-th grid cell in the i-th frame, and z_{ij}^{GT} represents the ground truth height of the corresponding grid cell.

2) ***Completeness****:* The completeness of the grid map is evaluated by

$$Pc = \frac{N_P \cap N_{GT}}{N_{GT}} \tag{7}$$

where N_P represents the number of estimated grid cells and N_{GT} represents the total number of valid grid cells in the ground truth.

3) ***Stability****:* To assess the consistency of the predicted grid attributes across consecutive frames, we define

$$Ps = \frac{N_{stable}}{G_{last} \cap G_{current}} \tag{8}$$

In this equation, N_{stable} denotes the number of grid cells where the attributes remain unchanged between consecutive frames, while $G_{last} \cap G_{current}$ refers to the total number of cells that exist in both the last and current frames. Ideally, except for these sequences with moving objects, the attribute of a grid cell should remain unchanged between consecutive frames. Therefore, this stability attribute is only evaluated on sequences without moving objects.

C. Baseline approach

In constructing our dataset and designing the performance metrics, we have considered two key factors: Firstly, we require the environment perception algorithm to create a dense terrain while avoiding over-smoothing regions with negative obstacles. Therefore, our performance metrics include specific discrimination for negative obstacles. Secondly, we require the algorithm to produce a complete and stable attribute map. To achieve this, the algorithm has to

TABLE III

PERFORMANCE COMPARISON OF DIFFERENT APPROACHES IN OUR DATASETS

Approaches	Terrain				Positive Obstacle				Negative Obstacle	Completeness	Stability
	P(%)	*R(%)*	*F1(%)*	*RMSE(cm)*	*P(%)*	*R(%)*	*F1(%)*	*RMSE(cm)*	*IOU(%)*	P_c*(%)*	P_s*(%)*
GP [29]	72.235	13.877	23.281	31.137	44.049	22.500	29.786	/	/	18.924	73.317
GndNet [30]	66.408	24.064	35.326	50.874	55.176	41.483	47.359	/	/	19.493	79.853
Ours	**93.573**	**95.551**	**94.552**	**11.380**	**63.761**	**54.055**	**58.508**	**41.435**	**83.495**	**86.365**	**99.659**

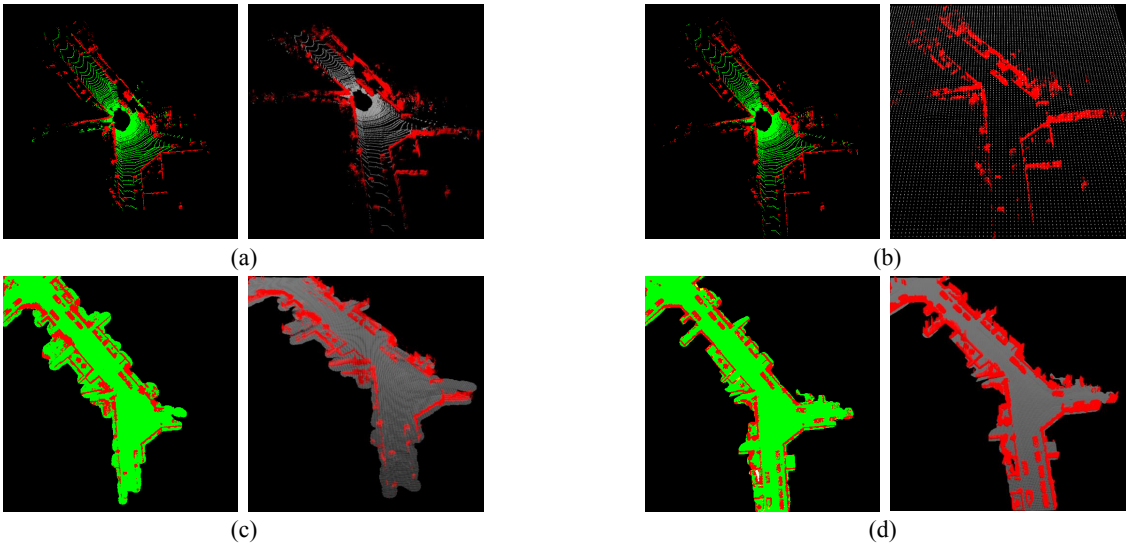


Fig. 8. Comparative results generated by different methods: (a) Gaussian Process, (b) GndNet, (c) Our method, and (d) Ground Truth. In each subfigure, the left image shows the attribute grid map, while the right image shows a combined view of the attribute map and elevation map in 3D.

fuse multiple frames while simultaneously having the ability to cope with moving objects.

However, existing methods often struggle to meet all of these requirements simultaneously. To establish a strong baseline method, we combined and improved upon the approaches of [31] and [8]. The method proposed in [31] can effectively handle moving objects while simultaneously enriching data completeness. The approach proposed in [8] could generate a dense terrain but could not detect negative obstacles. Therefore, we reduced the prediction radius used in this approach to prevent it from over-smoothing regions with negative obstacles. Those large empty regions on the estimated terrain map are marked as negative obstacles.

D. Experimental Results

We compare the aforementioned baseline method with a Gaussian Process (GP) based ground segmentation method [29] and a deep learning-based terrain modeling method called GndNet [30]. The results are presented in Table. III.

It can be observed that all three algorithms could obtain a good precision rate. But regarding the recall rate, the completeness measure, and stability measure, the proposed baseline method achieves a much higher accuracy than the two comparing approaches.

Fig. 8 shows some qualitative results. The proposed baseline method can produce a dense grid map. It performs well in both structured and unstructured environments. It can efficiently handle moving objects and has the ability to detect negative obstacles.

V. CONCLUSIONS

In this paper, we introduce a multi-environment, multi-LiDAR, multi-task dataset to promote the grid map-based navigation capability for autonomous vehicles. The dataset is composed of both structured and unstructured environmental data, which are captured by different types of LiDAR. It also features several typical challenging scenarios, including moving objects, negative obstacles, steep slopes, cliffs, overhangs, etc.

In addition, we have also developed a grid-level ground truth generation method that can effectively generate an accurate, dense, and stable grid map with minimum human labeling effort. We also build a strong baseline method that outperforms existing approaches. This dataset will be made publicly available at https://github.com/guanglei96/M3-GMN.

VI. ACKNOWLEDGMENT

The research was partly funded by the Natural Science Foundation of China under Grant 62003362.

REFERENCES

[1] H. Xue, H. Fu, L. Xiao, Y. Fan, D. Zhao, and B. Dai, "Traversability analysis for autonomous driving in complex environment: A lidar-based terrain modeling approach," *Journal of Field Robotics*, vol. 40, p. 17791803, June 2023.

[2] S. Lee, H. Lim, and H. Myung, "Patchwork++: Fast and robust ground segmentation solving partial under-segmentation using 3d point cloud," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13276–13283, IEEE, 2022.

[3] S. Gu, J. Yang, and H. Kong, "A cascaded lidar-camera fusion network for road detection," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 13308–13314, 2021.

[4] F. Gao, C. Li, and B. Zhang, "A dynamic clustering algorithm for lidar obstacle detection of autonomous driving system," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25922–25930, 2021.

[5] J. Frey, M. Mattamala, N. Chebrolu, C. Cadena, M. Fallon, and M. Hutter, "Fast traversability estimation for wild visual navigation," 2023.

[6] X. Meng, N. Hatch, A. Lambert, A. Li, N. Wagener, M. Schmittle, J. Lee, W. Yuan, Z. Chen, S. Deng, G. Okopal, D. Fox, B. Boots, and A. Shaban, "Terrainnet: Visual modeling of complex terrain for high-speed, off-road navigation," 2023.

[7] H. Karnan, E. Yang, D. Farkash, G. Warnell, J. Biswas, and P. Stone, "Sterling: Self-supervised terrain representation learning from unconstrained robot experience," 2023.

[8] H. Xue, H. Fu, R. Ren, J. Zhang, B. Liu, Y. Fan, and B. Dai, "Lidar-based drivable region detection for autonomous driving," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1110–1116, IEEE, 2021.

[9] L. Sharma, M. Everett, D. Lee, X. Cai, P. Osteen, and J. P. How, "Ramp: A risk-aware mapping and planning pipeline for fast off-road ground robot navigation," *arXiv preprint arXiv:2210.06605*, 2022.

[10] M. G. Castro, S. Triest, W. Wang, J. M. Gregory, F. Sanchez, J. G. Rogers, and S. Scherer, "How does it feel? self-supervised costmap learning for off-road vehicle traversability," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 931–938, 2023.

[11] B. Gao, A. Xu, Y. Pan, X. Zhao, W. Yao, and H. Zhao, "Off-road drivable area extraction using 3d lidar data," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1505–1511, IEEE, 2019.

[12] J. Seo, S. Sim, and I. Shim, "Learning off-road terrain traversability with self-supervisions only," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4617–4624, 2023.

[13] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, *et al.*, "Stanley: The robot that won the darpa grand challenge," *Journal of field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[14] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 3354–3361, IEEE, 2012.

[15] 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*, pp. 11621–11631, 2020.

[16] A. Valada, G. L. Oliveira, T. Brox, and W. Burgard, "Deep multispectral semantic scene understanding of forested environments using multimodal fusion," in *2016 International Symposium on Experimental Robotics*, pp. 465–477, Springer, 2017.

[17] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Field and Service Robotics: Results of the 11th International Conference*, pp. 335–350, Springer, 2018.

[18] C. Min, W. Jiang, D. Zhao, J. Xu, L. Xiao, Y. Nie, and B. Dai, "Orfd: A dataset and benchmark for off-road freespace detection," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 2532–2538, IEEE, 2022.

[19] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," in *2021 IEEE international conference on robotics and automation (ICRA)*, pp. 1110–1116, IEEE, 2021.

[20] L. Dabbiru, C. Goodin, N. Scherrer, and D. Carruth, "Lidar data segmentation in off-road environment using convolutional neural networks (cnn)," *SAE International Journal of Advances and Current Practices in Mobility*, vol. 2, no. 2020-01-0696, pp. 3288–3292, 2020.

[21] N. Fiedler, M. Bestmann, and N. Hendrich, "Imagetagger: An open source online platform for collaborative image labeling," in *RoboCup 2018: Robot World Cup XXII 22*, pp. 162–169, Springer, 2019.

[22] Y. Zhang, Y. Wang, H. Zhang, B. Zhu, S. Chen, and D. Zhang, "Onelabeler: A flexible system for building data labeling tools," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–22, 2022.

[23] A. Shaban, X. Meng, J. Lee, B. Boots, and D. Fox, "Semantic terrain classification for off-road autonomous driving," in *Conference on Robot Learning*, pp. 619–629, PMLR, 2022.

[24] P. Pfaff, R. Triebel, and W. Burgard, "An efficient extension to elevation maps for outdoor terrain mapping and loop closing," *The International Journal of Robotics Research*, vol. 26, no. 2, pp. 217–230, 2007.

[25] R. Triebel, P. Pfaff, and W. Burgard, "Multi-level surface maps for outdoor terrain mapping and loop closing," in *2006 IEEE/RSJ international conference on intelligent robots and systems*, pp. 2276–2282, IEEE, 2006.

[26] R. Ren, H. Fu, H. Xue, Z. Sun, K. Ding, and P. Wang, "Towards a fully automated 3d reconstruction system based on lidar and gnss in challenging scenarios," *Remote Sensing*, vol. 13, no. 10, p. 1981, 2021.

[27] H. Fu, H. Xue, and G. Xie, "Mapcleaner: Efficiently removing moving objects from point cloud maps in autonomous driving scenarios," *Remote Sensing*, vol. 14, no. 18, p. 4496, 2022.

[28] G. Volk, J. Gamerdinger, A. von Bernuth, and O. Bringmann, "A comprehensive safety metric to evaluate perception in autonomous systems," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8, IEEE, 2020.

[29] T. Chen, B. Dai, R. Wang, and D. Liu, "Gaussian-process-based real-time ground segmentation for autonomous land vehicles," *Journal of Intelligent & Robotic Systems*, vol. 76, pp. 563–582, 2014.

[30] A. Paigwar, Ö. Erkent, D. Sierra-Gonzalez, and C. Laugier, "Gndnet: Fast ground plane estimation and point cloud segmentation for autonomous vehicles," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2150–2156, IEEE, 2020.

[31] H. Fu, H. Xue, X. Hu, and B. Liu, "Lidar data enrichment by fusing spatial and temporal adjacent frames," *Remote Sensing*, vol. 13, no. 18, p. 3640, 2021.