2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

STONE Dataset: A Scalable Multi-Modal Surround-View 3D Traversability Dataset for Off-Road Robot Navigation

Konyul Park[1*], Daehun Kim[2*], Jiyong Oh[2], Seunghoon Yu[2], Junseo Park[1], Jaehyun Park[2], Hongjae Shin[1], Hyungchan Cho[1], Jungho Kim[1], and Jun Won Choi[2†]

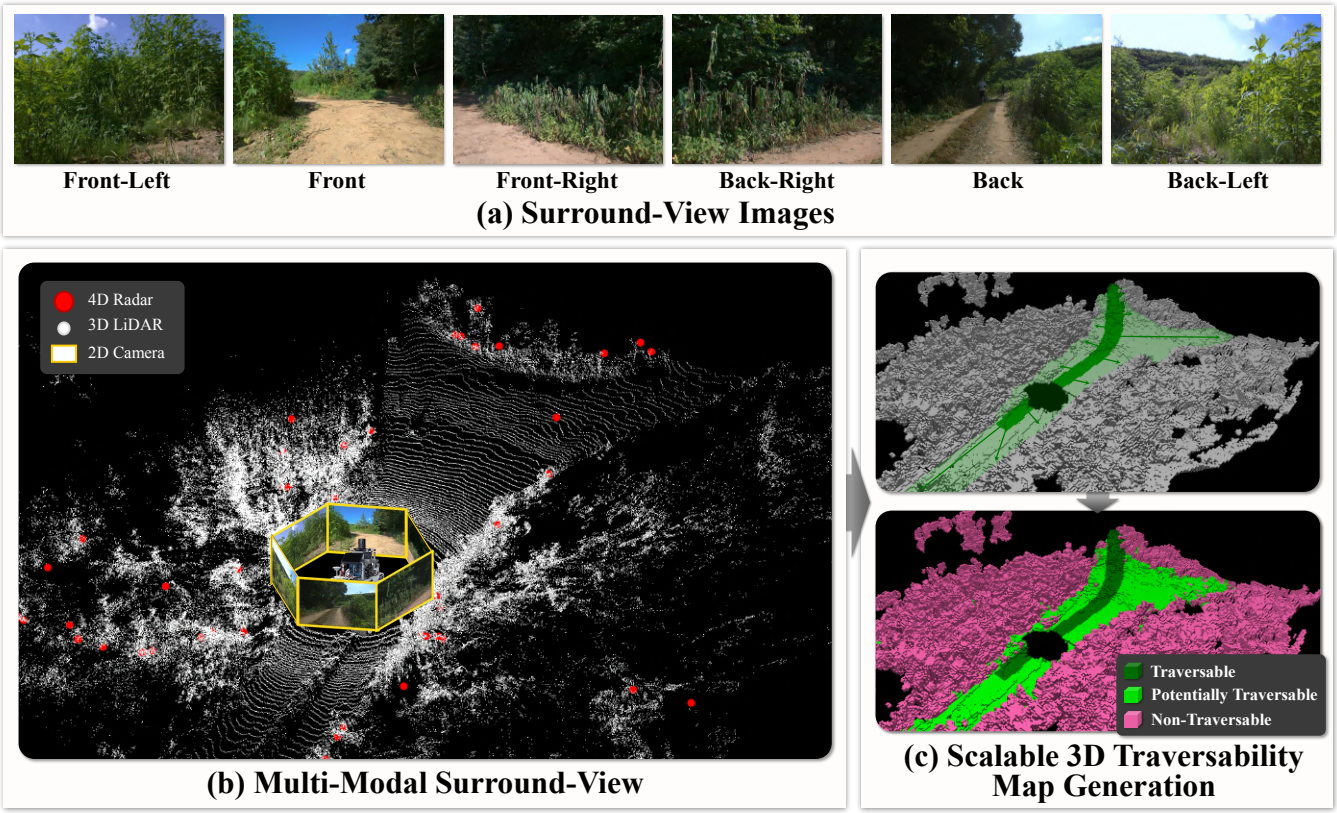


Fig. 1. **Overview of the STONE dataset.** The STONE dataset is a multi-modal 3D traversability dataset collected in off-road environments, which provides ground-truth annotations of 3D traversable areas automatically without human effort. (a) illustrates surround-view images captured around the robot. (b) shows the 3D scene captured by the multi-modal surround-view obtained using LiDAR, cameras, and 4D radar. (c) presents the process of automatically generating scalable 3D traversability maps based on robot trajectories.

***Abstract*— Reliable off-road navigation requires accurate estimation of traversable regions and robust perception under diverse terrain and sensing conditions. However, existing datasets lack both scalability and multi-modality, which limits progress in 3D traversability prediction. In this work, we introduce STONE, a large-scale multi-modal dataset for off-road navigation. STONE provides (1) trajectory-guided 3D traversability maps generated by a fully automated, annotation-free pipeline, and (2) comprehensive surround-view sensing with synchronized 128-channel LiDAR, six RGB cameras, and three 4D imaging radars. The dataset covers a wide range of environments and conditions, including day and night, grasslands, farmlands, construction sites, and lakes. Our auto-labeling pipeline reconstructs dense terrain surfaces from LiDAR scans, extracts geometric attributes such as slope, elevation, and roughness, and assigns traversability labels beyond the robot's trajectory using a Mahalanobis-distance-based criterion. This design enables scalable, geometry-aware ground-truth construction without manual annotation. Finally, we establish a benchmark for voxel-level 3D traversability prediction and provide strong baselines under both single-modal and multi-modal settings.**

*These authors contributed equally to this work.
†Corresponding author.
[1]Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, 08826, Korea.
[2]Department of Electrical and Computer Engineering, Seoul National University, Seoul, 08826, Korea.
{kypark, dhkim, jyoh, shyu, jspark, jhpark, hjshin, hccho, jhkim}@adr.snu.ac.kr
junwchoi@snu.ac.kr

I. INTRODUCTION

Robot navigation in off-road environments plays a vital role in applications ranging from military operations to agriculture, construction, and logistics, enabling safer and more efficient human activities. Reliable navigation in such settings hinges on accurate estimation of traversable regions. Unlike on-road environments, where structural cues such as

TABLE I
COMPARISON OF SEVERAL OFF-ROAD DATASETS

Dataset	Sensors	4D Radar	Camera FOV	# of Cameras	Image Resolution	LiDAR Resolution	Annotation
Freiburg Forest [13]	Camera, NIR	✗	Front-view	2	1024×768	-	2D semantic
YCOR [14]	Camera, LiDAR, INS	✗	Front-view	1	1024×544	64	2D semantic
RUGD [15]	Camera, LiDAR, IMU, GPS	✗	Front-view	1	688×550	32	2D semantic
RELLIS-3D [6]	Camera, LiDAR, INS	✗	Front-view	1	1920×1200	32, 64	2D/3D semantic
ORFD [7]	Camera, LiDAR	✗	Front-view	1	1280×720	40	2D traversability
TartanDrive 2.0 [16]	Camera, LiDAR, INS	✗	Front-view	1	1024×512	2×32, 70	-
GOOSE [17]	Camera, NIR, LiDAR, INS	✗	Front-view*	4	2048×1000	2×32, 128	2D/3D semantic
TOMD [10]	Camera, LiDAR, INS	✗	Front-view	1	1920×1080	128	2D traversability
STONE (Ours)	**4D Radar, Camera, LiDAR, INS**	✓	**Surround-view**	**6**	**1920×1200**	**128**	**3D traversability**

*indicates papers that employ a multi-camera setup but release only front-view images.

lane markings, curbs, and road boundaries clearly delineate drivable areas, off-road terrains are unstructured and often deformable (e.g., soil, vegetation, and gravel). These terrains lack well-defined boundaries and provide weaker geometric cues, making the reliable identification of traversable regions more challenging.

Existing off-road datasets remain insufficient for reliable traversability estimation. While RELLIS-3D [6] and RUGD [9] provide semantic annotations, such labels are costly to obtain, and semantics alone are insufficient to determine drivability, as regions sharing the same label may differ significantly due to geometric variations. TartanDrive [8] offers top-down height maps, but elevation alone provides limited cues for accurate traversability assessment. ORFD [7] and TOMD [10] rely on manually annotated 2D traversability maps, which are expensive to produce and difficult to scale. Trajectory-based approaches [18], [20], [12] generate self-supervised 2D pixel-level labels from robot paths, while Scate [34] derives ground truth solely from vehicle–terrain interaction signals projected onto point clouds, without explicitly exploiting geometric features. Overall, effective traversability assessment requires rich 3D geometric cues as well as compact representations that can be readily integrated into downstream tasks in the autonomous driving pipeline.

In addition to 3D traversability estimation, reliable off-road navigation requires full-scene 3D perception with comprehensive coverage and resilience under adverse sensing conditions. Maneuvers such as detours, turning, and reversing demand awareness of the entire 360° environment, making forward-view-only sensing insufficient. Radar, in particular, provides robustness under adverse weather where cameras and LiDAR often fail, offering a critical complement for reliable perception. Yet existing datasets [6], [9], [14], [7], [10] remain limited to front-view cameras or vision–LiDAR modalities, resulting in blind spots and degraded performance under challenging conditions. A detailed comparison is provided in Table I.

To address the limitations of existing datasets and advance research in off-road navigation, we present STONE, a large-scale dataset for off-road robot navigation collected across diverse environments. As shown in Fig. 1, STONE offers (1) 3D traversability maps automatically derived through a scalable, annotation-free pipeline, and (2) comprehensive multi-modal surround-view data that integrates cameras, LiDAR, and 4D radar.

In STONE, we collect data by manually driving an unmanned ground vehicle (UGV) equipped with a high-resolution 128-channel LiDAR, six RGB cameras, three 4D imaging radars, an RTK-capable GNSS, and a high-rate IMU. Each sample is provided with intrinsic and extrinsic calibration parameters. To achieve precise temporal alignment between the cameras and LiDAR, we employ a trigger-based synchronization scheme in which data acquisition is initiated by a LiDAR pulse signal. This design effectively reduces both temporal and spatial misalignment caused by timing offsets between sensors. As illustrated in Fig. 3, the STONE dataset spans a broad spectrum of challenging terrains and includes both daytime and nighttime conditions.

Manual annotation of fine-grained semantic classes is inherently unscalable, requiring substantial human effort and cost. To address this limitation, the STONE dataset provides an automated framework for generating 3D traversability maps. The core idea is to model the geometric attributes of traversable regions based on the areas actually traversed by the robot during data collection. To this end, we represent these geometric attributes using a multivariate Gaussian distribution and estimate the probability of traversability. The proposed auto-labeling pipeline consists of three stages. First, LiDAR scans are temporally aggregated to reconstruct a dense and wide-area terrain surface. Second, from this surface, we extract geometric cues such as slope, elevation, and roughness, which directly affect vehicle mobility. Third, voxels along the robot's trajectory are labeled as traversable. Using their feature distribution as a reference, we propagate labels to neighboring voxels by computing the Mahalanobis distance [33]. This automated pipeline enables scalable construction of traversability maps across diverse environments while maintaining consistent annotation quality.

We establish a benchmark for the 3D traversability prediction task, where the goal is to estimate voxel-level traversability maps from single- or multi-modal sensor inputs. We provide the performance of reference methods for comparative evaluation across three settings: camera-only, LiDAR-only, and multi-modal.

The main contributions of our dataset are summarized as follows:

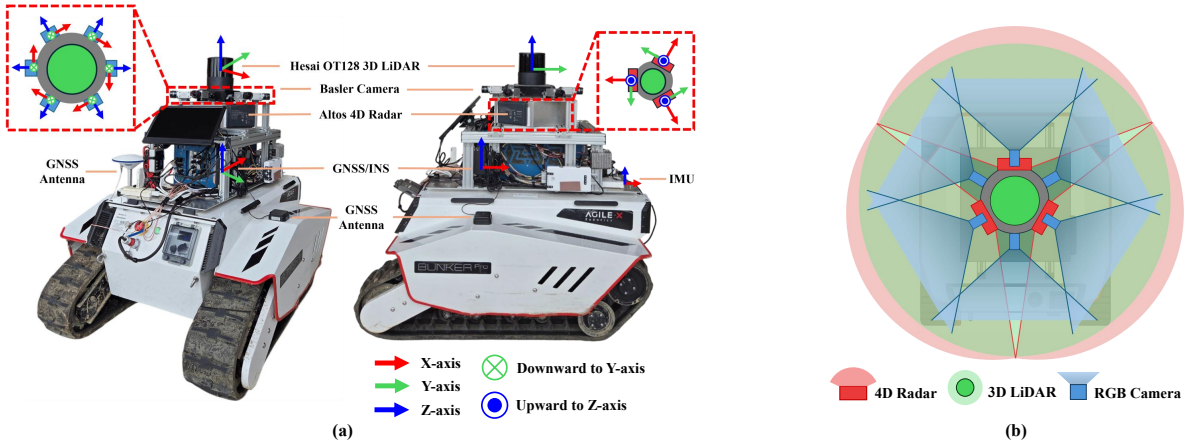


Fig. 2. **Sensor setup and coverage of the Bunker Pro UGV platform.** (a) shows the sensor placement including LiDAR, cameras, 4D radars, IMU and GPS. (b) illustrates the sensing range and coverage in off-road environments.

- **Trajectory-guided 3D traversability maps.** We introduce the first large-scale off-road dataset that provides 3D traversability maps as ground-truth labels. These labels are generated by a fully automated pipeline that reconstructs terrain geometry, extracts mobility-related features (e.g., slope, elevation, and roughness), and propagates traversability information beyond the robot's trajectory. This approach enables scalable construction of datasets for off-road traversability prediction tasks.
- **Comprehensive multi-modal surround-view sensing.** STONE is the first off-road dataset to integrate synchronized LiDAR, six RGB cameras, and three 4D imaging radars in a surround-view configuration. This setup supports full 360° scene perception and enhances robustness under adverse conditions.
- **Diverse environments and conditions.** The dataset spans diverse terrains (e.g., grasslands, farmlands, construction sites, and lakes) and covers both daytime and nighttime conditions, capturing the real-world variability of off-road environments.
- **Benchmark with strong baselines.** We establish a benchmark for 3D traversability prediction encompassing both single- and multi-modal settings, and provide strong baseline models. This enables standardized evaluation and fosters comparative studies within the community.
- We will release the dataset publicly.

II. RELATED WORK

A. Traversability in Off-Road Environments

RUGD [9], YCOR [14], RELLIS-3D [6], GOOSE [17], and WildOcc [19] have supported progress by providing pixel-, point-, or voxel-level semantic labels for off-road environments. Despite this progress, category-level semantics is insufficient for traversability estimation. For example, regions with the same label (e.g., grass, soil, rubble, or reeds) can vary in traversability depending on elevation, slope, and surface roughness.

ORFD [7] and TOMD [10] provide manually annotated pixel-level traversability labels. To reduce annotation effort, FtFoot [18] projects robot trajectories into images to supervise traversability, and V-STRONG [20] combines projected trajectories with SAM-based [21] instance masks for pixel-level contrastive costmap learning. Nonetheless, these methods remain limited to 2D supervision. TartanDrive [8], [16] provides height maps with images and vehicle dynamics (throttle, steering, IMU), but these signals are insufficient to serve as ground truth for traversability.

B. Sensor Coverage in Off-Road Datasets

Representative on-road datasets [2], [3], [4], [5], [22] offer multi-modal sensing through surround-view cameras, multiple radars, and 360° spinning LiDAR, providing comprehensive scene coverage. In contrast, as summarized in Table I, most off-road datasets [13], [14], [15], [6], [7], [16], [17], [10] are limited to a front-view camera and at most a single LiDAR, with no incorporation of radar sensing that is crucial under adverse weather.

III. STONE DATASET

A. Sensor Setup

We collected data across diverse off-road environments using the Bunker Pro platform [23], a tracked UGV designed for versatile industrial applications. As illustrated in Fig. 2, the platform was equipped with multiple sensors, providing 360° perception coverage without blind spots.

- **360° Rotating LiDAR**: 1 × Hesai OT128 with 128 channels, a maximum range of 200 m, a field of view of 360° (H) × 40° (V), an angular resolution of 0.1° (H) × 0.125° (V), and a scanning frequency of 10 Hz.
- **Multi-view RGB Cameras**: 6 × Basler ACE2 2A1920-51gcPRO with a resolution of 1920 × 1200 and a frame rate of 10 Hz.
- **4D Imaging Radars**: 3 × Continental ARS 548 RDI with a scanning frequency of 20 Hz.

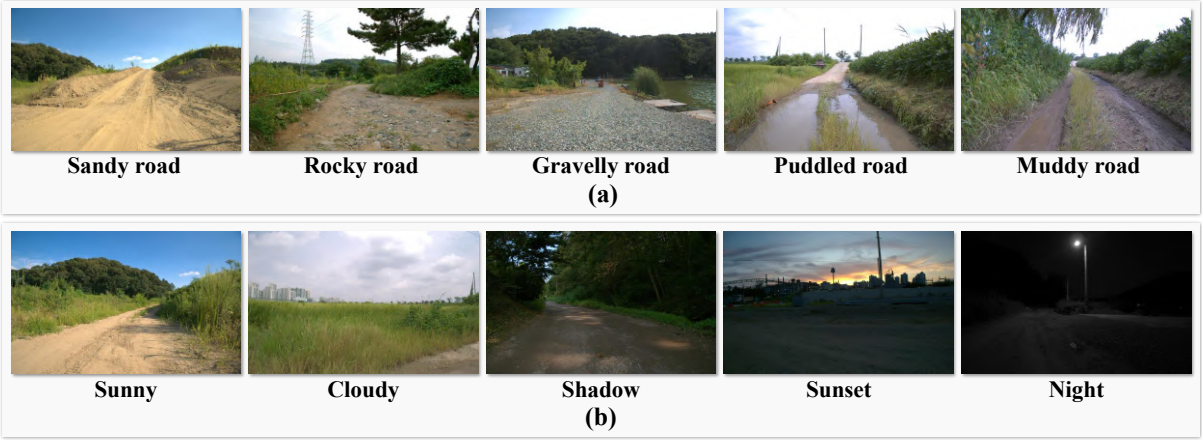


Fig. 3. **Various conditions in which the dataset was collected.** (a) illustrates off-road terrains such as sandy, rocky, gravelly, puddled, and muddy. (b) shows diverse illumination conditions in the dataset: sunny, cloudy, shadow, sunset, and night.

- **Global Navigation Satellite System (GNSS)**: NovAtel PIM222A dual-antenna GNSS/INS with RTK capability and an update rate of 20 Hz.
- **Inertial Measurement Unit (IMU)**: EPSON G366P IMU providing inertial measurements at 200 Hz.

The multi-modal system was integrated and operated on Ubuntu 22.04 using the ROS 2 Humble framework.

B. Sensor Calibration

The STONE dataset provides both extrinsic and intrinsic calibration parameters. To ensure consistency across coordinate frames, all extrinsic parameters are defined with respect to the LiDAR reference frame. Camera–LiDAR and radar–LiDAR calibrations were performed using open-source tools [31], [32], while the IMU–LiDAR extrinsic translation was obtained by measuring the relative sensor positions, and the relative rotation was set to the identity, assuming a co-aligned mounting. We calibrated the intrinsic parameters of the six cameras using a checkerboard-based method [30].

C. Time Synchronization

As the LiDAR spins, each camera is triggered when the LiDAR scan reaches the azimuth angle corresponding to the camera's viewing direction. To this end, we employed a custom trigger board that converted LiDAR pulse signals into angle-specific triggers to control camera shutters, aligning image captures with LiDAR scan angles and maintaining temporal synchronization. All other sensors were synchronized to LiDAR timestamps: RTK signals were linearly interpolated in (x, y, z) coordinates as well as quaternions [24], while radar and IMU measurements were temporally aligned by selecting the samples closest to each LiDAR frame stamp.

D. Data Collection Environments

The STONE dataset was collected in rural areas surrounding Seoul, South Korea. As illustrated in Fig. 3, it encompasses diverse off-road scenarios with varying levels of difficulty, including irregular unpaved roads along lakesides and narrow paddy embankments. The dataset also includes construction sites with heavy machinery and materials, as well as densely vegetated areas where drivable boundaries are poorly defined. In addition to daytime runs, we collected nighttime sequences to evaluate algorithm robustness under varying illumination conditions.

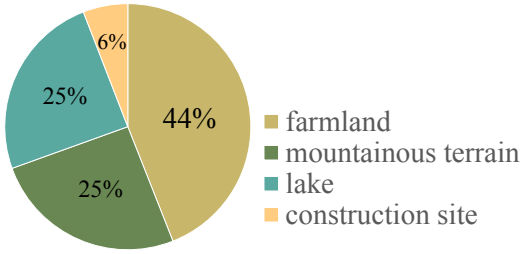


Fig. 4. Dataset composition across different environments.

E. Dataset Composition and Trainval Split

The dataset consists of 43 sequences with a total of 50,878 frames. Each sequence denotes a temporally continuous run of the robot recorded with the full multi-modal sensor suite at 10 Hz in a specific environment or route. Sequences are at least 20 seconds long to ensure sufficient temporal context for learning and evaluation. The data was collected across four environments: farmland, mountainous terrain, lakes, and construction sites. The collected sequences are distributed across these environments, as illustrated in Fig. 4. For benchmarking, the dataset is allocated into 31 sequences (≈36,000 frames) for training, 7 sequences (≈8,000 frames) for validation, and 5 sequences (≈7,000 frames) for testing. The split is structured to ensure representation of both daytime and nighttime conditions as well as diverse terrain types in each set. The test set was collected in regions different from those used for the training set.

IV. 3D Traversability Map Generation

In this section, we present a scalable framework that automatically generates 3D traversability map from LiDAR

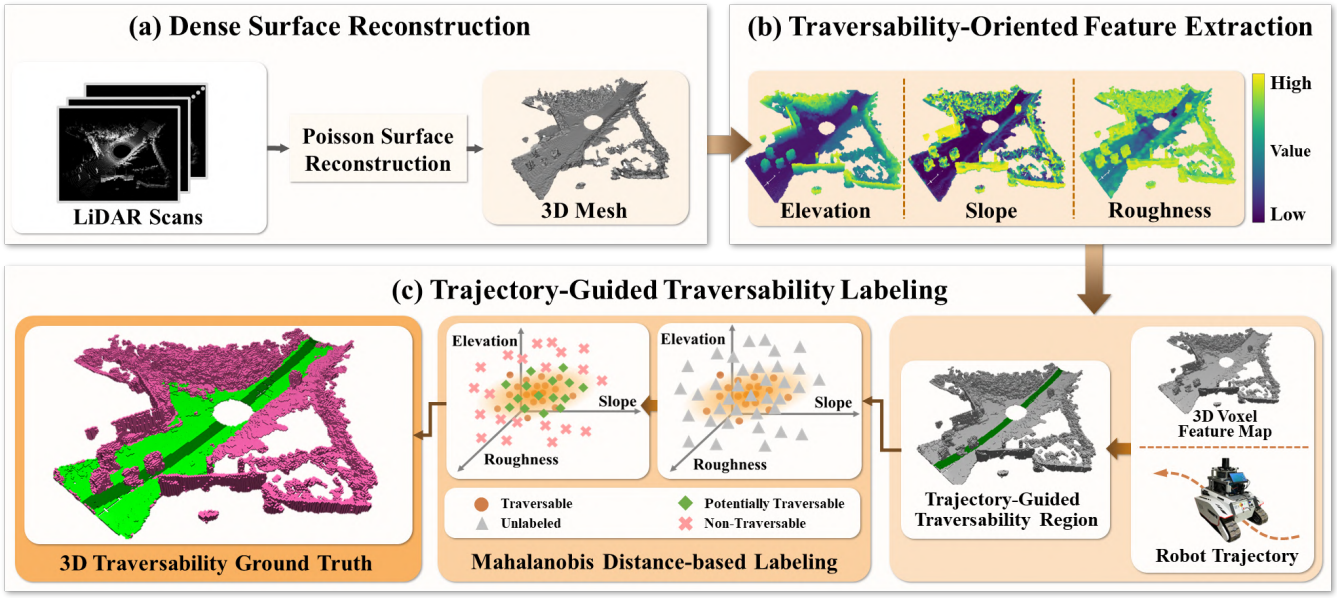


Fig. 5. **Overview of automated 3D traversability map generation process.** (a) LiDAR point clouds are accumulated over multiple time steps in the global coordinate system and reconstructed into a 3D mesh using Poisson surface reconstruction. (b) For each vertex of the 3D mesh, geometric features such as elevation, slope, and roughness are extracted to construct a traversability map. (c) A reference distribution is computed from the geometric features of voxels along the robot's driving trajectories using a multivariate Gaussian, and a 3D traversability ground-truth label is automatically generated for each voxel whose Mahalanobis distance falls below a predefined threshold.

data and robot trajectory records. As illustrated in Fig. 5, the proposed GT generation pipeline consists of three stages: (i) dense surface reconstruction, (ii) traversability-oriented feature extraction, and (iii) trajectory-guided traversability auto-labeling.

A. Dense Surface Reconstruction

Sparse and noisy LiDAR returns often hinder the extraction of reliable geometric features. To mitigate this issue, we aggregate multiple consecutive LiDAR scans during surface reconstruction. Specifically, individual LiDAR frames are aligned in a global coordinate frame using the robot's odometry and fused into a denser point cloud. This aggregation increases point density and yields a more complete representation of the 3D scene. We then apply the Poisson surface reconstruction method [35] to convert the aggregated point cloud into a watertight 3D mesh, which fills in missing regions and suppresses noise. The reconstructed 3D mesh is defined as $M = (\mathcal{V}, \mathcal{F})$, where $\mathcal{V}$ and $\mathcal{F}$ denote the sets of vertices and faces, respectively.

B. Traversability-Oriented Feature Extraction

For each vertex $v_i \in \mathcal{V}$ of the reconstructed mesh, we compute geometric parameters—elevation, slope, and roughness—to provide complementary cues for identifying drivable regions.

- **Elevation** (h_i): A UGV may fail to traverse regions where the elevation exceeds its climbing capability. We define h_i as the z-coordinate of vertex $v_i = (x_i, y_i, z_i)^\top$:

$$h_i = z_i. \tag{1}$$

- **Slope** (θ_i): A UGV struggles to traverse regions with excessively steep slopes due to vehicle dynamics, such as mass, speed, and wheel ground interaction. We extract the per-vertex normal vectors $\mathbf{n}_i$ from the mesh reconstructed via Poisson surface reconstruction, where each $\mathbf{n}_i$ is computed as the normalized average of the normals of faces incident to vertex v_i. The Slope θ_i is then defined as the angle between the normal vector $\mathbf{n}_i$ and the global vertical axis $\mathbf{z} = [0, 0, 1]^\top$:

$$\theta_i = \arccos(\mathbf{n}_i \cdot \mathbf{z}), \tag{2}$$

- **Roughness** (r_i): Roughness measures local deviations from planarity on the terrain surface. For each vertex v_i, it is defined as the logarithm of the mean squared error (MSE) between the neighboring vertices and their best-fit plane Π_i:

$$r_i = \log\left(\frac{1}{|N_i|}\sum_{v_j \in N_i} d(v_j, \Pi_i)^2\right), \tag{3}$$

where N_i denotes the k-nearest neighbor set of v_i, Π_i is the best-fit plane estimated from N_i, and $d(\cdot)$ represents the orthogonal distance from a vertex to the plane.

For each vertex v_i, we generate the geometric feature vector $\mathbf{f}_i$ as

$$\mathbf{f}_i = [h_i,\, \theta_i,\, r_i]^\top, \tag{4}$$

The entire 3D space is discretized into a set of voxels $\mathcal{X} = \{X_k\}_{k=1}^{K}$, where K denotes the total number of voxels. The geometric feature vector $\mathbf{F}_k$ for the kth voxel is defined as

$$\mathbf{F}_k = \frac{1}{|X_k|}\sum_{v_i \in X_k} \mathbf{f}_i, \tag{5}$$

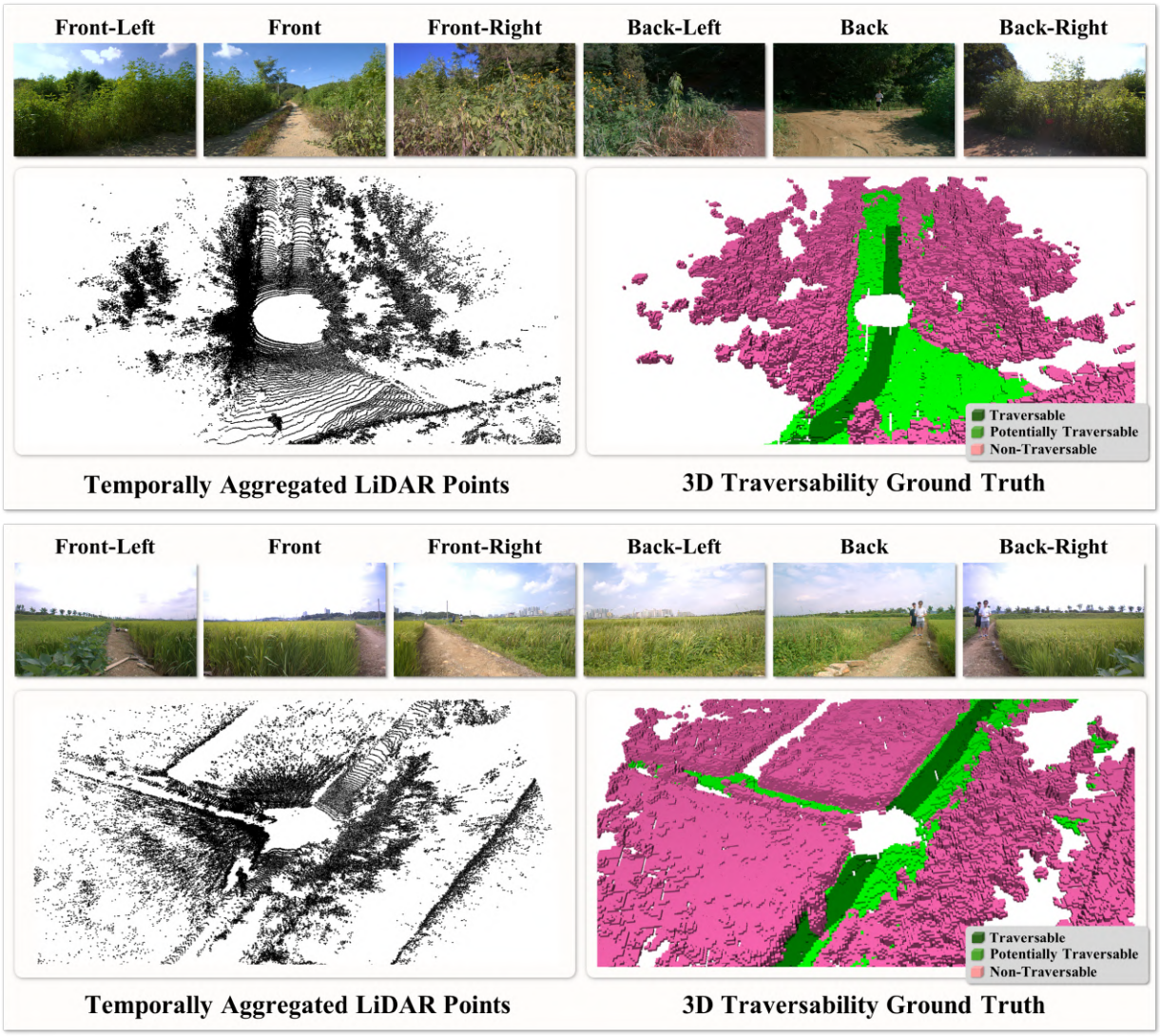


Fig. 6. **Visualization of the STONE dataset.** The top shows 360-degree images of the environment captured around the robot, the bottom left presents temporally aggregated LiDAR points, and the bottom right illustrates the automatically generated 3D traversability ground-truth map.

where $|X_k|$ is the number of vertices contained in the kth voxel X_k.

C. Trajectory-Guided Traversability Auto-Labeling

In practice, traversability is directly observed only along the robot's trajectories, as these regions are physically traversed. However, the goal of traversability prediction is to infer traversability beyond the observed trajectories and generalize to previously unseen areas. To this end, we model the distribution of geometric features extracted from trajectory voxels as a reference distribution. Formally, the set of features along the robot's trajectory, $\{\mathbf{F}_i\}_{i\in\mathcal{T}}$, is obtained from an underlying traversable distribution. We approximate this distribution with a multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the empirical mean and covariance of the geometric features along the trajectory.

We define the squared Mahalanobis distance [33] between a candidate voxel X_k and the reference distribution as

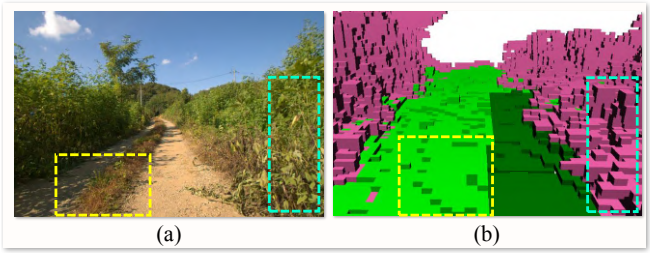
$$D^2(X_k) = (\mathbf{F}_k - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{F}_k - \boldsymbol{\mu}). \tag{6}$$

Under the Gaussian assumption, $D^2(X_k)$ follows a chi-squared distribution with d degrees of freedom, i.e., $D^2(X_k) \sim \chi^2_d$, where $d = 3$ in our setting. Using the trajectory distribution as reference, we set the threshold $\chi^2_{d,1-\alpha}$, which defines the $100(1-\alpha)\%$ confidence region of traversable geometry. Voxels inside this region are considered potentially traversable, while those outside are classified as non-traversable.

- **Traversable** (T): Voxels along the logged trajectory.
- **Potentially Traversable** (P): Off-trajectory voxels within the confidence region $\chi^2_{d,1-\alpha}$ of the trajectory distribution (e.g., $\alpha = 0.05$).
- **Non-Traversable** (N): Off-trajectory voxels outside this region.

D. Visualization of 3D Traversability

As illustrated in Fig. 6, the generated ground-truth labels are visually coherent, clearly delineating traversable and non-traversable regions. The consistency across different scenes



(a) (b)

Fig. 7. **Examples of regions with the same semantic class (vegetation) but different traversability labels.** (a) presents a front-view image, and (b) presents the corresponding ground-truth. Green and magenta regions denote traversable and non-traversable regions, respectively. Although the cyan and yellow boxes both correspond to vegetation, the cyan box is correctly marked as non-traversable, and the yellow box remains labeled as traversable.

TABLE II
EVALUATION OF 3D TRAVERSABILITY BASELINES BASED ON IOU METRICS (%)

Model	Sensor	IoU_{occ}	IoU_T	IoU_P	IoU_N	mIoU
C-OpenOcc [26]	C	35.0	14.5	23.8	33.7	24.0
C-CoNet [26]	C	37.0	14.8	24.3	34.2	24.4
OccFormer [27]	C	37.8	16.3	29.0	34.1	26.5
TPVFormer [28]	C	39.6	16.9	29.6	35.5	27.3
L-OpenOcc [26]	L	61.2	21.1	33.4	55.2	36.5
L-CoNet [26]	L	61.6	22.5	34.8	57.0	38.1
OccFusion [25]	C+R	45.2	21.5	36.4	41.3	33.1
M-OpenOcc [26]	C+L	62.3	18.1	35.0	61.0	38.0
M-CoNet [26]	C+L	62.6	18.3	36.0	62.0	38.8
OccFusion [25]	C+L	**66.1**	**19.0**	**36.9**	**62.8**	**39.6**

R, C, and L denote Radar, Camera, and LiDAR, respectively.

indicates that the automatic generation pipeline produces reliable annotations that align well with the underlying terrain geometry.

To illustrate the importance of incorporating geometric features, Fig. 7 presents a scenario in which semantic appearance alone does not reliably indicate traversability. Although both the yellow and cyan regions correspond to vegetation, their physical properties differ substantially: the low vegetation highlighted in yellow remains physically passable, whereas the dense bush in cyan obstructs any feasible path. Our traversability map (right) captures this distinction by labeling the ground vegetation as traversable (green) and the roadside bush as non-traversable (magenta).

V. BENCHMARKS AND EVALUATION

In this section, we introduce the benchmarks designed for evaluating 3D traversability prediction methods. We also present the performance results of several baseline models on these benchmarks.

A. Evaluation Metrics

The performance of 3D traversability prediction is evaluated using standard IoU-based metrics.

- **Occupancy IoU** measures the structural accuracy of free and occupied space at the voxel level:

$$\text{IoU}_{\text{occ}} = \frac{TP_{\text{occ}}}{TP_{\text{occ}} + FP_{\text{occ}} + FN_{\text{occ}}}, \tag{7}$$

where TP_{occ}, FP_{occ}, and FN_{occ} denote voxel-level true positive, false positive and false negative, respectively.

- **Per-class IoU** evaluates the ability to distinguish traversability classes within occupied voxels:

$$\text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}, \quad c \in \{T, P, N\}, \tag{8}$$

where TP_c, FP_c, and FN_c denote per-class statistics for Traversable (T), Potentially Traversable (P), and Non-Traversable (N).

- **Mean IoU (mIoU)** is the average of per-class IoUs:

$$\text{mIoU} = \frac{1}{|C|} \sum_{c \in C} \text{IoU}_c, \tag{9}$$

where C denotes the set of classes.

B. Experimental Settings

All reference methods were trained and evaluated on the splits of the STONE dataset. The traversability ground truth (GT) covers a range of [-25.6m, -25.6m, -2m, 25.6m, 25.6m, 4.4m] with a voxel size of [0.2m, 0.2m, 0.2m] in the ego coordinate system. All experiments were conducted using the PyTorch-based mmdetection3d framework on four NVIDIA RTX 3090 GPUs, and the hyperparameters of each model (e.g., optimizer, learning rate, batch size, etc.) followed the settings specified in their official implementations.

C. Experimental Results

Table II presents the performance of several 3D traversability prediction methods evaluated on our STONE dataset. To provide a comprehensive benchmark, we report results across different sensing modalities, including camera-only, LiDAR-only, and multi-modal fusion (camera+LiDAR, camera+radar). The results demonstrate that incorporating complementary sensors generally improves performance over single-modality baselines, establishing a solid reference point for future research on multi-modal traversability prediction in off-road environments.

VI. CONCLUSIONS

In this work, we introduce STONE, a scalable 3D traversability dataset designed to advance off-road autonomous navigation. While reliable off-road navigation requires a 360° field of view and robustness under adverse conditions, existing datasets fail to provide these essential capabilities. STONE integrates a 360° LiDAR, surround-view cameras, and surround-view 4D radars with precise time synchronization, and it covers diverse scenarios across varying terrains and illuminations. Manual annotation of fine-grained semantic classes is inherently unscalable and limits the potential of data-driven perception models. To address this, we propose a fully automated pipeline, which generates 3D traversability map GTs in volumetric space. The pipeline

leverages three geometric features extracted from LiDAR point clouds and uses the UGV's traversed trajectory as supervision, thereby producing consistent, geometry-aware labels without human annotation. We believe that STONE, together with the provided baselines, establishes a strong foundation for future research in 3D traversability prediction and will accelerate progress toward fully autonomous off-road robots.

VII. LIMITATIONS AND FUTURE WORK

Due to budget and time constraints, the current version of STONE does not cover a wide range of off-road regions or diverse weather conditions. In future work, we plan to expand both the size and geographic scope of the dataset, as well as include data collected under challenging conditions such as snow, rain, and fog. These extensions will further enable research on robust perception and planning in adverse environments.

VIII. ACKNOWLEDGEMENT

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] and the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2024-00421129).

REFERENCES

[1] Geiger, Andreas, et al. "Vision meets robotics: The kitti dataset." The international journal of robotics research 32.11 (2013): 1231-1237.

[2] Caesar, Holger, et al. "nuscenes: A multimodal dataset for autonomous driving." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020.

[3] Sun, Pei, et al. "Scalability in perception for autonomous driving: Waymo open dataset." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[4] Chang, Ming-Fang, et al. "Argoverse: 3d tracking and forecasting with rich maps." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[5] Bae, Jong Wook, et al. "Sit dataset: socially interactive pedestrian trajectory dataset for social navigation robots." Advances in neural information processing systems 36 (2023): 24552-24563.

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

[7] M. Chen, W. Jiang, D. Zhao, J. Xu, L. Xiao, Y. Nie, and B. Dai, "Orfd: A dataset and benchmark for off-road freespace detection." *2022 international conference on robotics and automation (ICRA). IEEE*, 2022.

[8] Triest, Samuel, et al. "Tartandrive: A large-scale dataset for learning off-road dynamics models." 2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022.

[9] Wigness, Maggie, et al. "A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments." 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019.

[10] Sun, Yixin, et al. "TOMD: A Trail-based Off-road Multimodal Dataset for Traversable Pathway Segmentation under Challenging Illumination Conditions." arXiv preprint arXiv:2506.21630 (2025).

[11] J. Behley, M. Garbade et al. "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences." 2019 IEEE/CVF International Conf. on Computer Vision (ICCV). IEEE, 2019.

[12] Seo, Junwon, Sungdae Sim, and Inwook Shim. "Learning off-road terrain traversability with self-supervisions only." IEEE Robotics and Automation Letters 8.8 (2023): 4617-4624.

[13] A. Valada, G. L. Oliveira, T. Brox, and W. Burgard, "Deep Multispectral Semantic Scene Understanding of Forested Environments Using Multimodal Fusion," in *Proc. Int. Symp. on Experimental Robotics (ISER)*, pp. 465–477, 2016.

[14] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-Time Semantic Mapping for Autonomous Off-Road Navigation," in *Field and Service Robotics: Proc. 11th Int. Conf. (FSR)*, pp. 335–350, 2018.

[15] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A RUGD Dataset for Autonomous Navigation and Visual Perception in Unstructured Outdoor Environments," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 5000–5007, 2019.

[16] Sivaprakasam, Matthew, et al. "Tartandrive 2.0: More modalities and better infrastructure to further self-supervised learning research in off-road driving tasks." 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024.

[17] P. Mortimer, R. Hagmanns, M. Granero, T. Lüttel, J. Petereit, and H.-J. Wuensche, "The GOOSE Dataset for Perception in Unstructured Environments," *2023 IEEE international conference on robotics and automation (ICRA). IEEE*, 2023.

[18] Jeon, Yurim, E. In Son, and Seung-Woo Seo. "Follow the Footprints: Self-supervised Traversability Estimation for Off-road Vehicle Navigation based on Geometric and Visual Cues." 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024.

[19] Zhai, Heng, et al. "WildOcc: A benchmark for off-road 3D semantic occupancy prediction." arXiv preprint arXiv:2410.15792 (2024).

[20] Jung, Sanghun, et al. "V-strong: Visual self-supervised traversability learning for off-road navigation." 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024.

[21] Kirillov, Alexander, et al. "Segment anything." Proceedings of the IEEE/CVF international conference on computer vision. 2023.

[22] Fent, Felix, et al. "Man truckscenes: A multimodal dataset for autonomous trucking in diverse conditions." Advances in Neural Information Processing Systems 37 (2024): 62062-62082.

[23] Bunker Pro, Agile X "https://global.agilex.ai/products/bunker-pro".

[24] Shoemake, Ken. "Animating rotation with quaternion curves." Proceedings of the 12th annual conference on Computer graphics and interactive techniques. 1985.

[25] Zhang, Ji, Yiran Ding, and Zixin Liu. "Occfusion: Depth estimation free multi-modal fusion for 3d occupancy prediction." Proceedings of the Asian Conference on Computer Vision. 2024.

[26] Wang, Xiaofeng, et al. "Openoccupancy: A large scale benchmark for surrounding semantic occupancy perception." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[27] Zhang, Yunpeng, Zheng Zhu, and Dalong Du. "Occformer: Dual-path transformer for vision-based 3d semantic occupancy prediction." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[28] Huang, Yuanhui, et al. "Tri-perspective view for vision-based 3d semantic occupancy prediction." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.

[29] Overbye, Timothy, and Srikanth Saripalli. "Radar-only off-road local navigation." arXiv preprint arXiv:2310.17620 (2023).

[30] Yan, Guohang and Liu, Zhuochun and Wang, Chengjie and Shi, Chunlei and Wei, Pengjin and Cai, Xinyu and Ma, Tao and Liu, Zhizheng and Zhong, Zebin and Liu, Yuqian and Zhao, Ming and Ma, Zheng and Li, Yikang, "OpenCalib: A Multi-modal Calibration Toolbox for Autonomous Driving", arXiv preprint arXiv:2205.14087, 2022

[31] Beltrán, Jorge and Guindel, Carlos and de la Escalera, Arturo and García, Fernando, "Automatic Extrinsic Calibration Method for LiDAR and Camera Sensor Setups", IEEE Transactions on Intelligent Transportation Systems, 2022

[32] Joris Domhof, Julian F. P. Kooij and Dariu M. Gavrila, "An Extrinsic Calibration Tool for Lidar, Camera and Radar", In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, Canada, 2019

[33] De Maesschalck, Roy, Delphine Jouan-Rimbaud, and Désiré L. Massart. "The mahalanobis distance." Chemometrics and intelligent laboratory systems 50.1 (2000): 1-18.

[34] Seo, Junwon, et al. "Scate: A scalable framework for self-supervised traversability estimation in unstructured environments." IEEE Robotics and Automation Letters 8.2 (2023): 888-895.

[35] KAZHDAN, Michael; BOLITHO, Matthew; HOPPE, Hugues. "Poisson surface reconstruction." In: Proceedings of the fourth Eurographics symposium on Geometry processing. 2006.