An Open-Source LiDAR and Monocular Off-Road Autonomous Navigation Stack

Rémi Marsal*[1], Quentin Picard*[2], Adrien Poiré[1], Sébastien Kerbourc'h[1], Thibault Toralba[1], Clément Yver[1], Alexandre Chapoutot[1], David Filliat[2]

I. Introduction

Autonomous off-road navigation (Fig. 1) requires reliable 3D perception to detect and avoid obstacles in unstructured terrain. Most approaches rely on LiDAR sensors, which provide accurate geometric measurements but are expensive, power-intensive, and easily detectable when discretion is required. Stereo cameras offer an alternative but suffer from limited range, sensitivity to lighting, and poor performance on low-textured surfaces.

Recent advances in zero-shot monocular depth estimation using foundation models offer a promising lightweight alternative. Several approaches recover metric depth from monocular images: zero-shot methods infer it directly from a single image, depth completion fuses images with sparse depth, and depth rescaling estimates scale and shift to convert normalized predictions into metric depth. However, integrating these methods into full navigation pipelines for off-road environments remains largely unexplored. Compared to autonomous driving, few open-source stacks exist for off-road navigation. This paper addresses that gap by presenting an open-source navigation stack supporting both LiDAR and monocular 3D perception, without task-specific training or fine-tuning for new environments. By leveraging foundation models and geometric methods, the approach generalizes across diverse robotic platforms and sensor configurations.

The key contributions are: 1/ a ready-to-use open-source stack for off-road reactive navigation adaptable to any robotic platform without retraining, 2/ integration of monocular depth estimation alongside LiDAR for 3D perception, and 3/ extensive evaluation in photorealistic simulations and real-world unstructured environments.

II. Method

The navigation pipeline consists of three modules: 3D perception, ground segmentation and obstacle detection, and path planning as shown in Fig. 2.

A. 3D Perception

The pipeline takes either a LiDAR point cloud or monocular camera images. For the monocular configuration, normalized relative depth maps are estimated using the small Depth Anything V2 model, accelerated through FP16 quantization and TensorRT optimization. These are rescaled to metric depth using sparse 3D landmarks from VINS-Mono, a visual-inertial SLAM system, following the depth rescaling approach in [1]. This design avoids any task-specific training: the depth model operates in a zero-shot manner and the rescaling relies solely on geometric constraints from the SLAM. Two enhancements improve robustness: 1/ edge masking: a Sobel filter on the disparity map detects depth discontinuities at object borders, and pixels within a five-pixel distance around edges are masked. This removes phantom obstacles caused by blurry depth transitions that would otherwise block navigable space. 2/ Temporal smoothing: scale and shift parameters $\hat{\mathbf{s}}_t$ are smoothed via an exponential moving average ($\alpha = 0.8$) to mitigate instabilities from the SLAM system:

$$\tilde{\mathbf{s}}_t = \alpha \tilde{\mathbf{s}}_{t-1} + (1-\alpha)\hat{\mathbf{s}}_t. \tag{1}$$

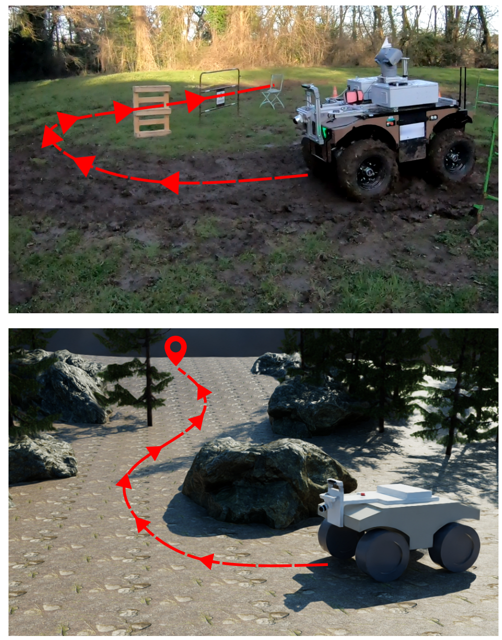

Fig. 1. Autonomous navigation with the wheeled ground robot in the real (top) and simulated (bottom) off-road environment.

B. Ground Segmentation and Obstacle Detection

The 3D point cloud is processed using a Cloth Simulation Filter (CSF) [2] that inverts the point cloud and simulates a rigid cloth falling onto the inverted surface to approximate the ground. The distance between points and the cloth

*Equal contribution

[1]U2IS, ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France. remi.marsal@ensta-paris.fr

[2]AMIAD, Pôle Recherche, France. quentin.picard@polytechnique.edu

All authors are members of LARIAD, a joint AMIAD-ENSTA laboratory.

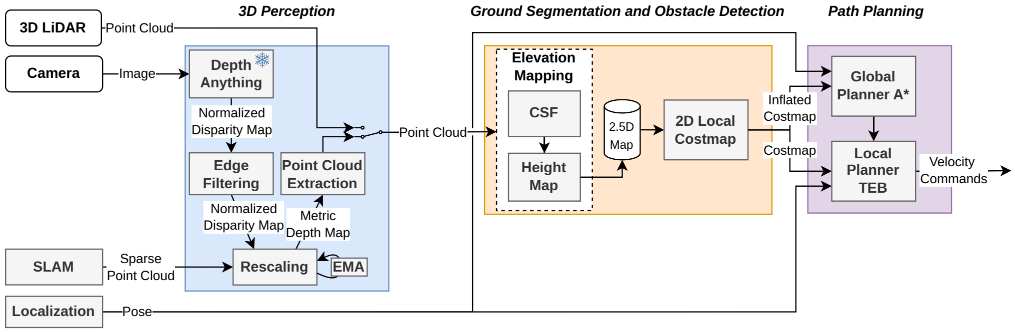

Fig. 2. Diagram of our navigation pipeline.

surface provides obstacle heights, stored in a robot-centric 2.5D elevation map [3]. A memory buffer persists obstacle positions as the robot moves, preventing obstacles from being forgotten when they leave the sensor's field of view. A binary costmap is generated using a 30 cm height threshold.

C. Path Planning

An A^* global planner operates on the inflated costmap, while a Timed-Elastic-Bands (TEB) [4] local planner computes flexible trajectories robust to perception errors. Localization is provided by fusing GNSS, IMU, and SLAM pose estimates through an extended Kalman filter.

III. Experiments

A. Evaluation Setup

Simulation: Three environments were created in Isaac Sim at increasing difficulty levels: easy (flat terrain with cube obstacles), medium (photorealistic trees and rocks), and hard (adding ground elevation and high grass). Each environment includes goals at 10, 20, and 30 meters. Each scenario was run 10 times per configuration (90 runs total per method).

Real-world: Three scenarios of increasing difficulty were tested using a Barakuda robot (see Fig. 1) equipped with an Ouster Dome 128-channel LiDAR, ZED2i camera, SBG IMU with dual GNSS, and a Jetson AGX Orin 64GB. Each scenario was run 3 times per configuration.

Three metrics were used: Success Rate (SR), Success weighted by Path Length (SPL), and Distance Ratio (DR).

B. Results

Simulation: The monocular configuration performs comparably to LiDAR across most scenarios. At 20 m goals, monocular even outperforms LiDAR with an average SR of 97% vs. 63% (sim-tuned). Performance degrades primarily in the hard environment at 30 m (10% SR), where dense high grass causes large blurry regions in depth maps. Unlike trees or rocks with sharp edges, high grass lacks clear depth discontinuities, making the edge filter ineffective and producing phantom 3D points behind obstacles that block the planner.

Real-world: Both LiDAR and monocular configurations achieve 100% success rate across all scenarios. The monocular configuration shows a 22% decrease in SPL (0.59 vs. 0.69) due to longer trajectories caused by later obstacle avoidance reactions, stemming from the longer processing chain. The monocular stack also exhibits occasional back-and-forth movements near circular obstacles, caused by the limited field of view and reduced depth accuracy compared to the LiDAR. In contrast, the LiDAR-based configuration produces smoother trajectories thanks to lower computational load, allowing the planner to anticipate obstacles earlier.

C. Computational Performance

On a Jetson AGX ORIN (MAXN profile), the LiDAR pipeline achieves higher frequencies across all modules: 3D perception runs at 20 Hz vs. 10 Hz, and elevation mapping at 16 Hz vs. 6 Hz. The path planning module operates at 12 Hz for both configurations. The frequency drop in the monocular setup is primarily caused by the GPU overhead from Depth Anything V2 inference, which leaves less compute available for the elevation mapping process. A potential improvement would be to distribute the depth estimation and elevation mapping across multiple GPUs.

IV. Conclusion

This work demonstrates that foundation-model-based monocular depth estimation can be a viable, low-cost alternative to LiDAR for off-road autonomous navigation. The proposed open-source stack achieves competitive navigation performance without requiring task-specific training: in real-world experiments, the monocular configuration matches LiDAR in success rate with only a moderate decrease in path efficiency. Future work targets the integration of traversability modules for classifying high grass as traversable, handling negative obstacles, and porting the stack from ROS1 to ROS2 for long-term maintainability and broader adoption.

References

[1] R. Marsal, A. Chapoutot, P. Xu, and D. Filliat, "A simple yet effective test-time adaptation for zero-shot monocular metric depth estimation," in *IROS*, 2025.

[2] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan, "An easy-to-use airborne lidar data filtering method based on cloth simulation," *Remote Sensing*, no. 6, 2016.

[3] T. Miki, L. Wellhausen, R. Grandia, F. Jenelten, T. Homberger, and M. Hutter, "Elevation mapping for locomotion and navigation using gpu," in *IROS*. IEEE, 2022.

[4] C. Rsmann, F. Hoffmann, and T. Bertram, "Integrated online trajectory planning and optimization in distinctive topologies," *Robot. Auton. Syst.*, no. C, 2017.