Overcoming Nature: Perception for Autonomous Navigation in Dense Vegetation

Lukas Wimmer[*], Andre Koczka, Uros Petrovic, and Gerald Steinbauer-Wagner[†]

Abstract

Autonomous navigation in densely vegetated off-road environments remains challenging because conventional geometric perception often treats traversable vegetation as non-traversable obstacles. In this work, we present a modular semantic–geometric perception pipeline for vegetation-aware navigation. The approach combines camera-based semantic data with LiDAR to generate a local grid map containing geometric and semantic information. A subsequent filtering stage uses this representation to correct vegetation-induced artifacts in standard elevation maps while preserving rigid obstacles for navigation. The system is designed to be portable across multiple robot platforms and sensor configurations. The pipeline was evaluated in challenging alpine off-road environments on three robot platforms, indicating improved distinction between traversable vegetation and solid obstacles and supporting more reliable navigation in dense natural environments.

1 Introduction and Motivation

Perception and terrain modeling for navigation in challenging, unstructured environments have seen significant progress in recent years (1; 2; 3). However, navigation in densely vegetated off-road environments remains particularly difficult. Forest trails, alpine terrain, and overgrown vegetation contain a mixture of objects such as trees, rocks, bushes, and tall grass. While some of these objects must be avoided, others are physically traversable, creating a fundamental ambiguity for perception systems. Conventional geometric terrain representations, such as elevation maps, often fail in these settings because vegetation can appear as non-traversable structure, leading to overly conservative planning behavior. At the same time, semantic segmentation has become an important tool for scene understanding in robotics. Convolutional architectures such as U-Net (4) and DeepLabV3 (5), adaptive multimodal models such as AdapNet (6), and transformer-based architectures such as Mask2Former (7) have significantly improved semantic perception in outdoor environments. Their practical usefulness depends strongly on suitable training data, with datasets such as Freiburg Forest (8), RUGD (9), and WildScenes (10) providing increasingly relevant benchmarks for natural off-road scenes. In parallel, learning-based traversability approaches such as TerraPN (11) and the uncertainty-aware method of Lee et al. (3) have shown that combining semantics and geometry can improve off-road navigation. However, these methods often require large training datasets, complex learning pipelines, or terrain-specific motion models.

In this work, we pursue a hybrid alternative that does not learn traversability directly. Instead, semantic information obtained from image segmentation is fused with LiDAR-based terrain geometry to explicitly refine a local map representation. The key idea is a sector-based semantic grid representation with dedicated Ground, Obstacle, and Sky layers, which is then merged with a conventional elevation map. This allows traversable vegetation to be treated as passable terrain while preserving rigid objects such as rocks or tree trunks as obstacles. An overview of the pipeline is shown in Figure 1. The contributions of this work are the following: First, we present a sector-based semantic

[*] wimmer1luk@gmail.com

[†] Institute of Software Engineering and Artificial Intelligence, Graz University of Technology, Graz, Austria. {akoczka,uros.petrovic,gerald.steinbauer-wagner}@tugraz.at

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

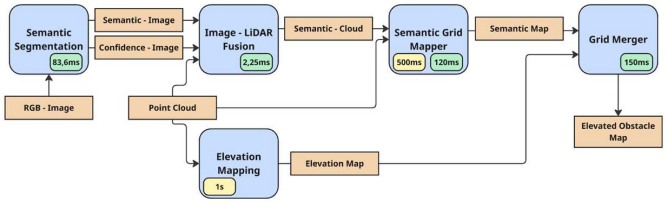

Figure 1: Performance Overview of the System. The green boxes in the modules are the execution times. The yellow boxes represent the execution period.

grid mapping method for vegetation-aware perception in dense natural environments. Second, we introduce a probabilistic fusion scheme for semantic classes and obstacle occupancy. Third, we propose a map-merging strategy that corrects vegetation-induced roughness and overhanging-branch artifacts in standard elevation maps. The resulting system is evaluated on three robot platforms in challenging alpine off-road environments, demonstrating practical feasibility across different sensor configurations.

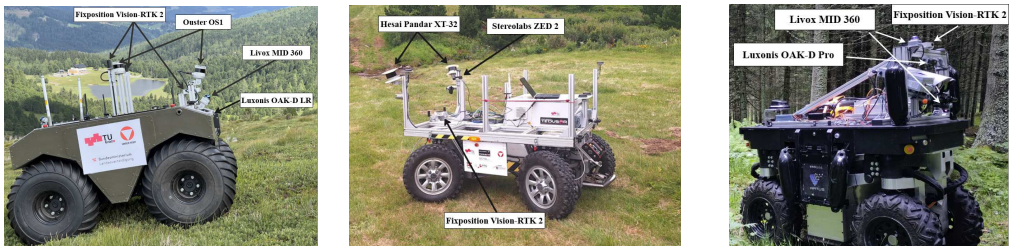

Figure 2: Warthog (left), Mercator (middle) and Artus[4](right) and their sensor setup shown on the mountain in the Seetaler Alps.

2 Methodology

The proposed system combines camera-based semantic perception with LiDAR-based terrain mapping to construct a semantically informed local map for vegetation-aware off-road navigation. The pipeline consists of four main steps: semantic image processing and image–LiDAR fusion, ground elevation estimation, semantic grid mapping, and final elevation-map correction.

First, RGB images are processed by a semantic segmentation model to obtain pixel-wise class predictions and confidence estimates. These semantic predictions are associated with LiDAR measurements by transforming each LiDAR point from the LiDAR frame into the camera frame. Points outside the camera field of view are discarded using angular constraints in the camera frame, and the remaining points are projected into the image plane using the intrinsics of the camera. Each valid 3D point is then assigned the semantic class and confidence value of the corresponding image pixel, producing a semantic point cloud. To reason about objects relative to the terrain, a robot-centered grid map is maintained. The raw LiDAR point cloud is first used to estimate a ground elevation layer. For each grid cell, the ground height is updated using the minimum-height rule $\hat{h}^+ = \min(\hat{h}^-, h(q))$, where $\hat{h}^-$ is the previous height estimate, q is a LiDAR point falling into the cell, and $h(q)$ is the height of that point in the map frame. The resulting elevation layer is smoothed and used as the reference surface for a height-based sector decomposition. Using the smoothed ground estimate, the space above the terrain is divided into three sectors: *Ground*, *Obstacle*, and *Sky*. Points between the ground elevation and the obstacle threshold are assigned to the Ground sector, points between the obstacle threshold and the robot height are assigned to the Obstacle sector, and points above the robot height are assigned to the Sky sector. Semantic observations from the fused point cloud are accumulated in the Ground and Obstacle sectors to estimate class labels for each grid cell. Let $n_c(i, j)$ denote the number of hits of class c in cell (i, j), and let C be the set of semantic classes.

[4]Artus robotic platform – CharismaTec og. https://charismatec.at/en/projects/

	Scenario 1	Scenario 2
Cell Resolution [m]	0.1	0.1
Total Area [m^2]	187	225
Ground Truth Traversable [m^2]	115.55	137.26
Total Detected Obstacles [m^2]	25.83	8.55
Intersecting Obstacles ↓ [m^2]	5.92	1.83
Intersecting ↓ [%]	22.92	21.4

Table 1: Results of the obstacle precision evaluation.

Route	Platform	Safety Interventions ↓	Functional Interventions ↓
Route 1	Artus	2	0
	Warthog	0	0
Route 2	Mercator	0	6
	Warthog	0	0
Route 3	Mercator	0	2
	Warthog	0	0

Table 2: Manual interventions during the autonomous navigation evaluation across all routes.

The total number of hits in the cell is defined as $N(i,j) = \sum_{c \in C} n_c(i,j)$. The probability that cell (i,j) belongs to class c is computed as $p_{i,j}(c) = \frac{n_c(i,j)}{N(i,j)}$.

To improve robustness against noisy single-frame predictions, these class estimates are fused over time using a log-odds representation, $\mathcal{L}_t(i,j,c) = \mathcal{L}_{t-1}(i,j,c) + L(p_{i,j}(c))$, where $L(p) = \log\left(\frac{p}{1-p}\right)$. In parallel, raw LiDAR measurements are used to update probabilistic obstacle and sky layers. The obstacle layer combines geometric measurements with the semantic obstacle class layer to determine whether a cell contains a rigid obstacle, while the sky layer captures structures above the robot height such as overhanging vegetation. The obstacle occupancy update is formulated as $\mathcal{L}_t(i,j) = \mathcal{L}_{t-1}(i,j) + L(p_{hit}(i,j)) + L(p_{miss}(i,j))$, where p_{hit} and p_{miss} are the probabilities associated with obstacle hits and misses, respectively, similarly to the approach in the book in the book Probabilistic Robotics (12). Finally, the semantic grid representation is merged with a conventional elevation map. Traversable ground classes are used to replace vegetation-induced disturbances with the smoothed ground estimate, while the sky layer allows correction of elevated cells caused by overhanging branches or foliage. The resulting local map is therefore both geometrically consistent and semantically aware, allowing the navigation stack to treat traversable vegetation as passable terrain while preserving rigid obstacles as non-traversable.

3 Implementation

The system is implemented as a modular ROS 2 perception pipeline running on Ubuntu 22.04. Semantic segmentation is performed using pretrained models, while the geometric mapping stages operate on a robot-centered rolling grid map that stores terrain height, semantic ground and obstacle classes, and probabilistic obstacle and sky layers. In the final experiments, Mask2Former with a Swin-L backbone was used due to its strong segmentation performance, while the full system was profiled offline using ROS 2 bagfiles before field deployment. The average CPU utilization was measured using the pidstat command from the Linux sysstat package. Figure 1 shows the timing of each component. The merged elevation map can be published at 1 Hz in the worst case, which is sufficient for local off-road navigation at around 1-1.5 m/s.

4 Experimental Setup and Results

The proposed perception pipeline is evaluated at both the component level and the system level. The evaluation is divided into three parts: obstacle precision assessment, merged map quality assessment, and fully autonomous navigation. The experiments were conducted in representative alpine and forest environments containing dense vegetation, uneven terrain, steep slopes, and overhanging branches. The system was deployed on three robotic platforms with different sensor setup. Warthog shown in Figure 2 left, was equipped with multiple LiDARs, including Ouster OS1 units and a front-mounted Livox MID-360, together with a Luxonis OAK-D LR camera and an Intel Core i7-13700HX PC with 64 GB RAM and an NVIDIA GeForce RTX 4070 Mobile. Mercator, shown in Figure 2 middle, used a Stereolabs ZED camera, two Hesai Pandar XT-32 LiDARs, and an AMD Ryzen 9 3900X PC with 64 GB RAM and an NVIDIA GeForce RTX 2070 SUPER Evo. Artus, shown in Figure 2 right, used a Luxonis Oak-D Pro camera, two Livox MID-360 LiDARs, and an Intel Core i7-13700HX PC with 64 GB RAM and an NVIDIA GeForce RTX 4070 Mobile. All platforms used Fixposition GNSS with RTK correction.

For the component-level evaluation, the obstacle precision and merged-map assessments were performed using teleoperated runs and recorded sensor data. In the obstacle precision evaluation, the robot footprint projected onto the grid map was used as a proxy for traversable ground truth. The

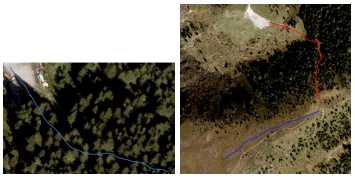

Figure 3: Pre-recorded Waypoints: (left) shows Route 1 through a forest road; (right) shows the two alpine paths. The first one in red (Route 2) is a steep hiking path. The second one in blue (Route 3) leads uphill through a less overgrown area.

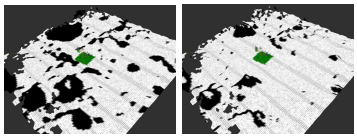

Figure 4: Merged Map Evaluation: (left) shows the Masked Slope of the standard Elevation Map; (right) shows the Masked Slope of the Merged Map; Black regions in the Slope Masks represent a slope greater than 45°. The robot's footprint is marked in green.

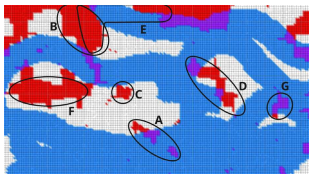

Figure 5: Evaluation Visualization Map. Blue: traversable area covered by the robot. Red: obstacles. Purple: Intersection between the obstacles and the robot. B, C, and F show true obstacles. A and D show intersections due to overhanging branches. G is a false positive.

results are shown in Table 1. In Scenario 1, the system detected $25.83\,\mathrm{m}^2$ of obstacles, with $5.92\,\mathrm{m}^2$ intersecting traversable terrain, corresponding to 22.92%. In Scenario 2, the detected obstacle area was $8.55\,\mathrm{m}^2$, with $1.83\,\mathrm{m}^2$ intersecting the traversable region, corresponding to 21.4%. These results indicate that the system detects relevant obstacles while keeping the overlap with traversable regions moderate in both scenarios. Scenario 1 is shown in Figure 5 as an example. The merged-map evaluation compares the standard elevation map with the semantically corrected map by counting cells with slope greater than $45°$ inside the traversed region. The number of such cells in the merged map is reduced to approximately one tenth of the corresponding number in the standard elevation map. This shows that the proposed semantic correction strongly reduces vegetation-induced artifacts while preserving the underlying terrain structure for planning. This is shown in Figure 4. For the autonomous evaluation, three predefined off-road routes were represented by GPS waypoints. Figure 3 shows the recorded routes. Route 1 was chosen to test transitions from normal roads into dense and narrow forest, Route 2 to test steep terrain with overhanging branches, and Route 3 to test disambiguation between grass, bushes, and small trees in more open terrain. Route 1 was evaluated with Artus and Warthog, while Routes 2 and 3 were evaluated with Mercator and Warthog. The main metric was the number of manual interventions, separated into safety and functional interventions. Table 2 shows the results of all routes. On Route 1, both Artus and Warthog was able to pass the narrow and densely vegetated forest, which was considered the hardest part of the route. Warthog completed the route without interventions, while Artus required two safety-related slowdowns due to its narrow footprint and high center of gravity. Using well-tuned classical elevation mapping, passing the narrow parts of this route was not possible at all. On Route 2, Warthog again completed the route without interventions, while Mercator required six functional interventions, two of them caused by branches that were still within robot height and therefore could not be cleared by the grid merger. On Route 3, Warthog completed the route without interventions, while Mercator required two functional interventions, mainly due to the combination of small and hard-to-segment vegetation and limited turning capability. Overall, the results show that the proposed semantic–geometric representation improves navigation in dense vegetation compared with conventional elevation-only mapping, while the main remaining limitations occur in edge cases.

5 Conclusion and Future Work

This work presented a sector-based semantic grid representation and map-merging strategy for autonomous navigation in densely vegetated off-road environments. By combining LiDAR-based elevation mapping with semantic information from image segmentation, the proposed system improves the distinction between traversable vegetation and rigid obstacles while reducing vegetation-induced artifacts in standard elevation maps. Field experiments on three robot platforms indicate that this semantic-geometric representation supports more reliable navigation in dense natural terrain compared to conventional elevation mapping. Future work will focus on reducing system overhead, improving obstacle modeling for more nuanced traversability cases, making ground-elevation estimation more robust under ambiguous visibility conditions, and integrating synchronized RGB-D sensing for tighter alignment between geometry and semantics.

Acknowledgments and Disclosure of Funding

This work was funded by the Austrian defense research program FORTE of the Federal Ministry of Finance (BMF) under the project PATH.

References

[1] P. Fankhauser, M. Bloesch, and M. Hutter, "Probabilistic terrain mapping for mobile robots with uncertain localization," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 4, pp. 3019–3026, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8392399

[2] P. Fankhauser and M. Hutter, "A Universal Grid Map Library: Implementation and Use Case for Rough Terrain Navigation," in *Robot Operating System (ROS) – The Complete Reference (Volume 1)*, A. Koubaa, Ed. Springer, 2016, ch. 5. [Online]. Available: http://www.springer.com/de/book/9783319260525

[3] H. Lee, J. Kwon, and C. Kwon, "Learning-based uncertainty-aware navigation in 3d off-road terrains," 2022. [Online]. Available: https://arxiv.org/abs/2209.09177

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597

[5] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017. [Online]. Available: https://arxiv.org/abs/1706.05587

[6] A. Valada, J. Vertens, A. Dhall, and W. Burgard, "Adapnet: Adaptive semantic segmentation in adverse environmental conditions," in *IEEE International Conference on Robotics and Automation (ICRA)*, 05 2017, pp. 4644–4651. [Online]. Available: https://doi.org/10.1109/ICRA.2017.7989540

[7] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," 2022. [Online]. Available: https://arxiv.org/abs/2112.01527

[8] A. Valada, G. Oliveira, T. Brox, and W. Burgard, "Deep multispectral semantic scene understanding of forested environments using multimodal fusion," in *International Symposium on Experimental Robotics (ISER)*, 03 2017, pp. 465–477. [Online]. Available: https://doi.org/10.1007/978-3-319-50115-4_41

[9] 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 *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 5000–5007. [Online]. Available: https://ieeexplore.ieee.org/document/8968283

[10] K. Vidanapathirana, J. Knights, S. Hausler, M. Cox, M. Ramezani, J. Jooste, E. Griffiths, S. Mohamed, S. Sridharan, C. Fookes, and P. Moghadam, "Wildscenes: A benchmark for 2d and 3d semantic segmentation in large-scale natural environments," *The International Journal of Robotics Research*, vol. 44, no. 4, p. 532–549, Sep. 2024. [Online]. Available: http://dx.doi.org/10.1177/02783649241278369

[11] A. J. Sathyamoorthy, K. Weerakoon, T. Guan, J. Liang, and D. Manocha, "Terrapn: Unstructured terrain navigation using online self-supervised learning," 2022. [Online]. Available: https://arxiv.org/abs/2202.12873

[12] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, ser. Intelligent Robotics and Autonomous Agents series. MIT Press, 2005. [Online]. Available: https://books.google.at/books?id=2Zn6AQAAQBAJ