Traversable Area Estimation for a Two-Wheeled Robot Based on Attention Mechanism

Soichiro Kamiyama[1]**, Takenori Obo**[1]**, Naoyuki Kubota**[1]

[1] Department of Mechanical Systems Engineering, Faculty of Systems Design, Tokyo Metropolitan University, Tokyo, Japan
E-MAIL: kamiyama-soichiro@ed.tmu.ac.jp

Abstract:

In unstructured environments such as construction sites, estimating navigable areas using 3D distance sensors is essential for the safe movement of autonomous mobile robots. However, sensor data downsampling for computational efficiency causes the loss of critical local terrain features vital for navigation decisions. Particularly for embodied robots, the impact of small local obstacles on actual traversability strongly depends on their physical characteristics, such as wheel size, kinematic constraints, and sensor placement. This research proposes an active navigable area estimation framework that utilizes the robot's embodiment to preserve locally important geometric information while considering both the distribution of observed data and the robot's behavior. It constructs an environmental map using a topological map (GNG-DT) and evaluates deviations from a locally fitted planar model to identify information-deficient regions. Furthermore, it actively directs the robot's viewpoint toward these regions to intensively improve local data density. Experiments in real environments demonstrated that this perception－ action framework enhances recognition accuracy for easily overlooked minute obstacles, establishing a foundation for robust autonomous navigation systems.

Keywords:

Traversability Estimation; Active Gaze-Driven Environment Representation; Topological Map

1 Introduction

In recent years, autonomous mobile robots have increasingly been deployed in unstructured environments such as construction sites, where labor shortages are becoming critical. Accurate traversable area estimation based on sensor data is therefore essential for safe autonomous navigation.

Three-dimensional LiDAR is widely used for outdoor perception because of its robustness to illumination and weather variations. However, real-time processing of large-scale point clouds on computationally constrained robotic platforms requires downsampling. While downsampling reduces computational cost, it often removes small yet safety-critical terrain features, such as minor steps and slender obstacles. This loss of local geometric detail can lead to terrain misclassification and compromise navigation safety.

Particularly for embodied robots, the influence of small obstacles on actual traversability strongly depends on physical characteristics such as wheel size, kinematic constraints, and sensor placement. Therefore, traversable area estimation should be formulated in an embodiment-aware manner that tightly couples perception with the robot' s physical capabilities. Conventional approaches— including camera-based semantic segmentation and LiDAR-based grid or voxel representations— primarily process static observations and cannot compensate for intrinsic information sparsity.

To address this limitation, this paper proposes an active traversable area estimation framework that utilizes the robot' s embodiment through viewpoint control. The environment is modeled using a topological map (GNG-DT), and information-deficient regions, referred to as attention nodes, are detected based on deviations from locally fitted planar models. The robot then actively adjusts its viewpoint toward these regions to reacquire dense observations, thereby preserving geometrically and physically relevant features.

The main contributions of this research are summarized as follows:

1. Reformulating traversable area estimation as an embodied perception－ action problem in which the robot actively reduces environmental uncertainty.

2. Integrating GNG-DT with planar deviation analysis to identify information-deficient regions in real time.

3. Demonstrating through real-world experiments that embodiment-aware active perception improves recognition of minute obstacles in unstructured environments.

The remainder of this paper is organized as follows. Related work is first reviewed, followed by a description of the proposed method, experimental validation, and conclusions.

2. Related Work

2.1 Traversable Area Estimation

Traversable area estimation is a fundamental component of autonomous mobile robot navigation. A comprehensive survey of terrain traversability analysis methods is provided by Papadakis [1]. Camera-based approaches, particularly semantic segmentation using deep learning, have been widely studied for terrain understanding [2, 3]. These methods exploit rich semantic and appearance information; however, they are inherently sensitive to illumination changes, shadows, and viewpoint variations. Such factors frequently occur in outdoor unstructured environments, leading to unstable predictions under real-world operating conditions.

In contrast, 3D LiDAR provides direct geometric measurements and is largely invariant to lighting and weather conditions. For this reason, LiDAR-based terrain representation methods are commonly adopted for outdoor navigation [4, 5]. Typical approaches convert raw point clouds into grid maps, voxel representations, or elevation maps to enable efficient computation. Nevertheless, due to limited onboard computational resources, downsampling is unavoidable when processing large-scale point cloud data in real time. During this downsampling process, small but safety-critical terrain features—such as minor steps or slender obstacles—may be removed [1]. This loss of local geometric detail can lead to terrain misclassification and directly compromise navigation safety.

2.2 Adaptive Representation and Topological Mapping

To mitigate the information loss caused by uniform-resolution representations, adaptive modeling approaches that reflect non-uniform data distributions have been proposed. Topological learning methods such as Growing Neural Gas (GNG) dynamically allocate nodes according to data density and can efficiently extract structural features from unknown environments [6]. Extensions of GNG have been applied to three-dimensional perception tasks, including terrain representation and affordance detection [7, 8].

Although such adaptive representations improve efficiency and structural expressiveness, they fundamentally depend on the density of the input data. When the observed point cloud of an object is itself extremely sparse—such as distant obstacles or thin structures—these methods cannot reconstruct sufficient geometric detail. In other words, adaptive representation alone cannot overcome intrinsic information deficiency at the sensing stage.

2.3 Positioning of This Work

Existing approaches predominantly focus on how to process given static observations more effectively. They rarely address the fundamental issue of insufficient data density at the sensing stage.

The proposed framework differs from prior work in that it explicitly integrates perception and action. Assuming downsampled point cloud input, the environment is modeled using a growing neural gas with different topologies (GNG-DT) [7]. Information-deficient regions are autonomously detected based on prediction errors derived from local planar models. The robot then actively adjusts its viewpoint to reacquire point cloud data in these regions. By closing the perception-action loop, the proposed method supplements missing geometric information and enables more robust traversable area estimation in unstructured environments.

3. Proposed Method

3.1 System Overview

The proposed framework formulates traversable area estimation as an embodied perception – action process rather than a static point cloud classification task. Perception, environment modeling, and robot motion are integrated in a closed-loop architecture, as illustrated in Fig. 1.

The system consists of three main components: (1) environment modeling using a topological map (GNG-DT), (2) attention node detection based on planar deviation analysis, and (3) active viewpoint control with adaptive relearning.

First, a topological environmental model is constructed from the input point cloud using GNG-DT. Local planar models are then fitted around each node, and the deviation between model prediction and actual observations is interpreted as geometric uncertainty. Nodes with large deviations are identified as attention nodes.

When the robot approaches these regions, it actively adjusts its sensor orientation and reacquires point cloud data,

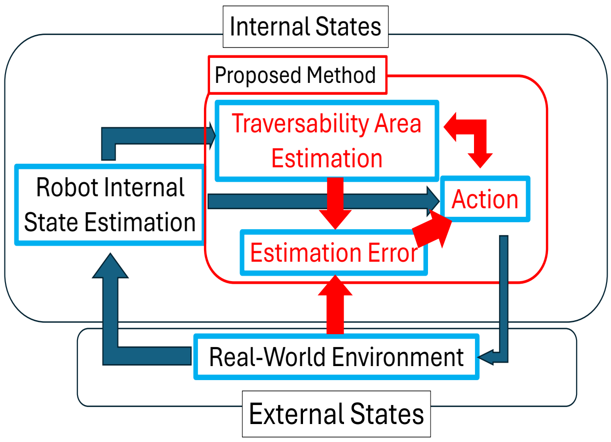

FIGURE 1. System architecture of the proposed active traversable area estimation framework.

increasing local observation density. The newly acquired data are incorporated into the GNG-DT model through adaptive relearning, thereby refining local geometric representation.

By iteratively reducing prediction error through physical action, the framework compensates for information loss caused by downsampling or unfavorable viewpoints. This method enables robust and accurate traversable area estimation in unstructured environments.

The detailed algorithms of each component are described in the following subsections. This design explicitly reflects the robot’s physical embodiment, allowing perception results to be interpreted in terms of actual traversability.

3.2 Environment Modeling with Topological Map

To construct an environment map that can adapt to unknown data structures from input 3D point clouds, we employ the unsupervised learning algorithm Growing Neural Gas with Different Topologies (GNG-DT) [7]. GNG-DT dynamically generates and updates nodes and edges based on the spatial distribution of input data, representing the data distribution as an efficient topological map.

In particular, GNG-DT is well suited for embodiment-aware traversability estimation. Its topological representation allows geometric properties, such as local inclination and surface continuity, to be evaluated directly at node locations. Traversability criteria in this framework, including slope estimation and obstacle detection based on PCA, are inherently defined with respect to the robot’s physical characteristics, such as wheel size, kinematic constraints, and contact stability. Because these criteria explicitly depend on the robot’s embodiment, GNG-DT provides a natural foundation for integrating perception with physically grounded traversability judgment. For this reason, the proposed method adopts GNG-DT as the core environmental representation.

To capture the three-dimensional geometric characteristics of the environment, we exploit the adjacency relationships learned by GNG-DT. Specifically, for each target node, Principal Component Analysis (PCA) [9] is applied to the coordinates of its directly connected neighboring nodes. This allows us to extract local geometric features from the learned topological structure.

Through eigenvalue decomposition of the covariance matrix, the eigenvector corresponding to the smallest eigenvalue indicates the direction of minimum variance. We interpret this direction as the normal vector of the local planar structure and assign it to each node.

The effectiveness of combining GNG-based structures with PCA for local geometric estimation has also been demonstrated in prior work [10]. By iteratively reducing prediction error through physical action, the framework compensates for information loss caused by downsampling or unfavorable viewpoints. This design explicitly reflects the robot’s physical embodiment, allowing perception results to be interpreted in terms of actual traversability. As a result, the proposed method enables robust and accurate traversable area estimation in unstructured environments.

The detailed algorithms of each component are described in the following subsections.

3.3 Attention Node Detection

To compensate for the loss of local terrain features caused by downsampling, the proposed framework autonomously detects information-deficient regions within the environment model constructed by GNG-DT. Specifically, a local planar model is estimated for each node classified as traversable, and deviation between the model and the input point cloud is evaluated.

Let the reference vector of node i be

$$\mathbf{h}_{tra,i} = (h_{tra,i,x},\, h_{tra,i,y},\, h_{tra,i,z}),$$

and let its normalized normal vector be

$$\mathbf{n}_i = (n_{i,x},\, n_{i,y},\, n_{i,z}).$$

The local plane equation at node i is defined as

$$n_{i,x}(x - h_{tra,i,x}) + n_{i,y}(y - h_{tra,i,y}) + n_{i,z}(z - h_{tra,i,z}) = 0 \tag{1}$$

$$n_{i,x}^2 + n_{i,y}^2 + n_{i,z}^2 = 1 \tag{2}$$

Let P denote the set of all input points. A local subset S_i is extracted within a radius r_{pos} on the xy-plane centered at node i:

$$S_i = \left\{(x_v, y_v, z_v) \in P \mid (x_v - h_{tra,i,x})^2 + (y_v - h_{tra,i,y})^2 < r_{pos}^2\right\} \tag{3}$$

For each point $(x_v, y_v, z_v) \in S_i$, the signed distance from the local plane is computed as

$$dist_{i,v} = n_{i,x}(x_v - h_{tra,i,x}) + n_{i,y}(y_v - h_{tra,i,y}) + n_{i,z}(z_v - h_{tra,i,z}) \tag{4}$$

If $|dist_{i,v}|$ exceeds a predefined threshold th_{dev}, the point is regarded as deviating from the planar model. The deviation set D_i is defined as

$$D_i = \{(x_v, y_v, z_v) \in S_i \mid |dist_{i,v}| > th_{dev}\} \tag{5}$$

Finally, if the number of deviation points $|D_i|$ exceeds a threshold th_{num}, the node is classified as an attention node:

$$h_{attn,i} = \begin{cases} 1 & (|D_i| \geq th_{num}) \\ 0 & \text{(otherwise)} \end{cases} \tag{6}$$

The extracted attention nodes are subsequently utilized as targets for active viewpoint control to reacquire dense observations.

3.4 Active Viewpoint Control and Adaptive Relearning

Attention nodes detected by planar deviation analysis ($h_{attn,i} = 1$) indicate regions where geometric information is insufficient due to measurement sparsity or absorption into dominant planar structures.

To compensate for this deficiency, the proposed framework exploits the robot' s embodiment through active viewpoint control and adaptive relearning.

During navigation, viewpoint adjustment is triggered when the robot enters within 500 mm of an attention node. When this distance threshold is satisfied, the robot modifies its pitch angle to actively observe the target region.

The nominal pitch angle during normal operation is $0°$, whereas attentive observation is performed at $15°$. This adjustment increases point cloud density around small-scale obstacles.

The reacquired data are then used to update the GNG-DT model with emphasis on the attention region. Specifically, learning prioritizes points near attention nodes, and the node insertion threshold is reduced from r_{pos} to r_{pos_att} ($r_{pos\_att} < r_{pos}$) to achieve higher local resolution.

Through this closed-loop integration of active sensing and topology adaptation, small geometric features that would be lost under passive observation are preserved, resulting in more robust traversable area estimation.

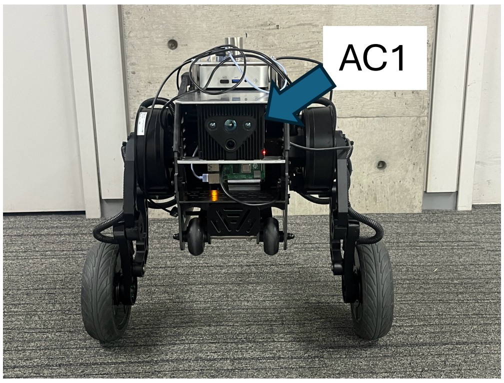

FIGURE 2. The experimental platform. The two-wheeled biped robot DIABLO equipped with the AC1 multimodal sensor unit.

4 Experiments

4.1 Hardware Configuration

As shown in Fig. 2, experiments were conducted using the two-wheeled biped robot DIABLO (Direct Drive Tech). The robot is equipped with a multimodal sensor unit, AC1 (Robosense), which integrates a 3D LiDAR, an RGB camera, and an IMU. In this study, environment modeling and traversable area estimation were performed primarily using 3D point cloud data obtained from the LiDAR sensor.

For computation, a PC equipped with an Intel Core i7 CPU, 32 GB RAM, and an NVIDIA RTX 4060 GPU running Ubuntu 22.04 was used. The PC was connected to the robot' s control board (Raspberry Pi) via wired Ethernet, enabling real-time execution of the proposed algorithm and transmission of control commands.

4.2 Experimental Environment

To evaluate the overall performance of the proposed framework, experiments were conducted in real outdoor environments, such as tiled open spaces in front of building entrances, as shown in Fig. 3. Small protrusions and minor

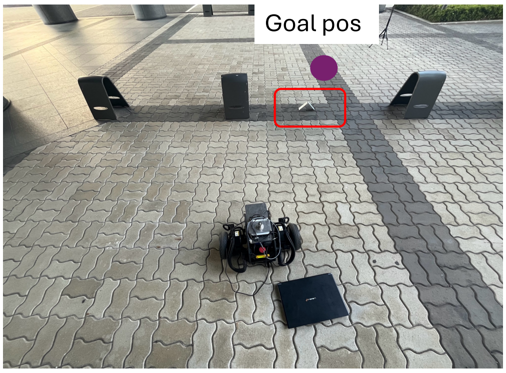

FIGURE 3. Outdoor experimental environment used for evaluation. Experiments were conducted in a tiled open space in front of a building entrance.

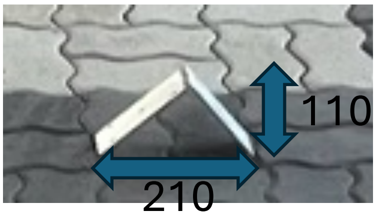

FIGURE 4. Example of small protrusions used for evaluation. The proposed method aims to detect such minor geometric irregularities.

obstacles, as illustrated in Fig. 4, were intentionally placed in the environment. The experiments were designed to evaluate: (1) whether attention nodes were appropriately generated through qualitative visualization, and (2) whether the improvement in geometric representation was achieved without a significant increase in the number of nodes.

4.3 Experimental Results

Fig 5 illustrates the process from attention node generation to attention-based local learning.

Fig 5(a) shows a grid-based map with a resolution of 20 mm, where obstacles are defined as regions with a height of 50 mm or higher. Small obstacles cause unstable classifications due to measurement noise, resulting in flickering obstacle detection.

Fig 5(b) presents the result after applying a correction process.

Fig 5(c) shows the initial state of the GNG. Red nodes indicate non-traversable regions, while green nodes represent traversable regions. Around small obstacles, non-traversable nodes (red) are not immediately generated; instead, orange attention nodes appear, indicating the detection of local geometric uncertainty.

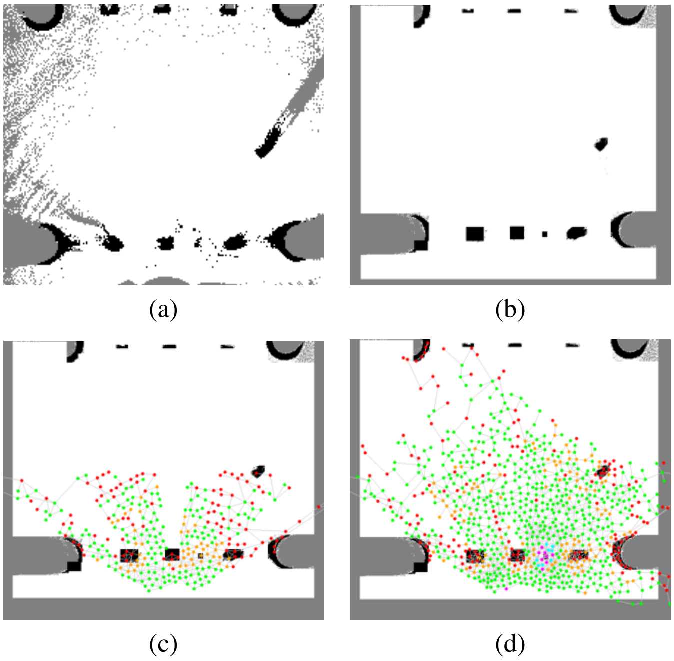

FIGURE 5. Process of attention node generation and local refinement. (a) Grid-based map (20 mm resolution, obstacles defined as height $\geq$ 50 mm). (b) Modified grid result. (c) Initial GNG state (red: non-traversable, green: traversable, orange: attention nodes). (d) Final result after local learning (cyan: non-traversable, light blue: traversable).

TABLE 1. Comparison of average number of GNG nodes with and without attention-based local learning.

Method	Avg. Nodes	Increase (%)
Without Attention	365	—
With Attention	448	22.7

Fig 5(d) shows the result during attention-based local learning. In this stage, light blue nodes represent traversable regions, and cyan nodes represent non-traversable regions. After refinement, small obstacles are correctly classified, and subtle geometric irregularities are stably preserved in the environment map.

Table 1 compares the average number of GNG nodes with and without attention-based local learning.

To quantitatively evaluate the network expansion, the node increase ratio is defined as

$$R_{\text{inc}} = \frac{N_{\text{att}} - N_{\text{base}}}{N_{\text{base}}} \times 100\ (\%), \tag{7}$$

where N_{base} denotes the average number of nodes without attention, and N_{att} denotes the average number of nodes with attention-based local learning.

Substituting $N_{\mathrm{base}} = 365$ and $N_{\mathrm{att}} = 448$ yields

$$R_{\mathrm{inc}} = 22.7\%. \tag{8}$$

Although geometric refinement was achieved around small obstacles, the increase in node count remains moderate. Since the refinement is performed locally around attention nodes, the overall network size does not grow excessively. Therefore, the proposed attention mechanism achieves a favorable balance between geometric accuracy and computational efficiency.

5 Conclusion and Future Work

This paper proposed an active traversable area estimation framework that cyclically integrates environment modeling and action generation. By combining planar deviation-based attention node detection with viewpoint control, the proposed method utilizes the robot's embodiment to actively preserve fine geometric features that are typically lost under uniform downsampling. Real-world experiments demonstrated that embodiment-aware active perception improves the robustness of small obstacle detection and traversable area estimation.

Nevertheless, several challenges remain for practical deployment. First, a more robust attention decision mechanism is required to mitigate false detections caused by sensor noise and localization errors. Explicit modeling of uncertainty should be incorporated into the attention generation process.

Second, viewpoint selection should be reformulated as an optimization problem. In the current implementation, the pitch angle during attention was fixed at 15 degrees; however, the quality of acquired geometric information strongly depends on factors such as target distance and local surface structure. Future work should determine observation poses that maximize information acquisition efficiency.

Finally, the proposed active perception framework should be integrated with higher-level navigation modules, including path planning and object detection, to achieve a fully autonomous robotic system capable of reliable operation in real-world environments.

Acknowledgements

This research was supported by JST [Moonshot Research and Development Program], Grant Number [JPMJMS2034]

References

[1] Panagiotis Papadakis. Terrain traversability analysis methods for unmanned ground vehicles: A survey. *Engineering Applications of Artificial Intelligence*, 26(4):1373–1385, 2013.

[2] Tianrui Guan et al. Ga-nav: Efficient terrain segmentation for robot navigation in unstructured outdoor environments. *IEEE Robotics and Automation Letters*, 7(3):8138–8145, 2022.

[3] Wei Zhou et al. Automated evaluation of semantic segmentation robustness for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):1951–1963, 2019.

[4] Thierry Siméon. Motion planning for a non-holonomic mobile robot on 3-dimensional terrains. In *Workshop on Geometric Reasoning for Perception and Action*, pages 38–50. Springer, 1991.

[5] Panagiotis Papadakis and Fiora Pirri. 3d mobility learning and regression of articulated, tracked robotic vehicles by physics-based optimization. In *Virtual Reality Interaction and Physical Simulation, Eurographics*, 2012.

[6] Bernd Fritzke. A growing neural gas network learns topologies. In *Advances in Neural Information Processing Systems*, volume 7, 1994.

[7] Yuichiro Toda et al. Growing neural gas with different topologies for 3d space perception. *Applied Sciences*, 12(3):1705, 2022.

[8] Azhar Aulia Saputra et al. Dynamic density topological structure generation for real-time ladder affordance detection. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3439–3444, 2019.

[9] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[10] Koki Ozasa, Yuichiro Toda, and Takayuki Matsuno. Growing neural gas based traversability clustering for an autonomous robot. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE, 2023.