Scene-Aware 2D-3D Hybrid Mapping for Computationally Efficient Robot Navigation

Muhammad Imad, Soo Young Shin*
imadsafi08@kumoh.ac.kr, wdragon@kumoh.ac.kr*
Department of IT Convergence Engineering
Kumoh National Institute of Technology, Gumi, South Korea

***Abstract*—Autonomous mobile robots require accurate environmental representations for efficient path planning. While 2D maps are computationally lightweight, they fail to capture elevation information crucial for navigating stairs or ramps. Full 3D maps, though rich in detail, introduce significant memory and processing burdens. To address this, we propose a scene-aware 2D-3D hybrid mapping (SA-2D3D) framework that leverages deep scene recognition to dynamically switch between 2D and 3D mapping modes. Unlike prior methods relying on ultrasonic sensors to detect elevation, our approach utilizes lightweight convolutional neural networks (CNNs) trained for scene classification, enabling semantic understanding of transitional zones like staircases, corridors, and ramps. Experimental results across both simulation and real-world deployments demonstrate that our method reduces mapping time by 49% and memory usage by 92% compared to full 3D mapping, while maintaining semantic fidelity and navigational integrity.**

***Index Terms*—Hybrid Mapping, Scene Recognition, Autonomous Navigation, 2D-3D Switching, SLAM**

I. Introduction

Autonomous robots depend heavily on efficient spatial representations for navigation and path planning. Traditional 2D occupancy maps, tahough computationally efficient, lack elevation details and are insufficient in environments with staircases or ramps. Full 3D mapping approaches such as OctoMap [1] or Normal Distribution Transform [2] offer rich spatial representations but at the cost of high memory consumption and longer processing times.

To mitigate this trade-off, previous works have explored hybrid representations, such as the 2D-3Dh mapping approach [3], which switches to 3D only when elevation is detected using hardware-based ultrasonic sensors. However, such methods require precise sensor tuning and lack semantic understanding of the environment.

In this paper, we propose an enhanced method that utilizes a lightweight scene recognition model to classify the environment context (e.g., ramp, staircase, corridor) and dynamically switch between 2D and 3D mapping. This semantic trigger replaces traditional sensor-based thresholds, enabling more robust, context-aware mapping in complex indoor and semi-structured environments.

II. Methodology

A. Mapping Backbone

Our method builds upon Rao-Blackwellized Particle Filter (RBPF) SLAM to maintain a pose-graph and generate a 2D occupancy grid map. When elevation-related scenes are recognized, a parallel 3D mapping module using OctoMap is activated, creating voxel-based 3D segments overlayed on the 2D structure.

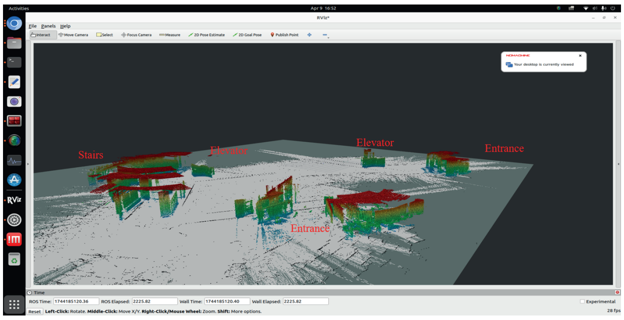

Fig. 1. Switching between 2D and 3D mapping for computationally efficient robot navigation

B. Scene-Aware Switching

We trained a MobileNetV3-based classifier on a custom dataset of indoor environments labeled into classes such as *corridor*, *staircase*, *ramp*, and *room*. When the scene classifier detects a transition to an elevation-relevant class (e.g., staircase or ramp), the system switches to 3D mapping mode. This semantic-driven approach removes reliance on range-based sensors and enhances adaptability to unseen environments.

C. Switching Strategy

2D Mode: Default mapping using LiDAR-based occupancy grids for flat scenes.
3D Mode: Triggered when classifier confidence exceeds a threshold ($\geq$ 0.8) for elevation-related classes. 3D point clouds from LiDAR are accumulated into a voxel grid using OctoMap.

III. Experimental Results

A. Setup

We tested our system on a Scout Mini UGV equipped with a 128-channel Ouster LiDAR and a front-mounted camera. Environments included indoor labs, office corridors with stairs, and semi-outdoor ramps. Baseline comparisons were made against full 3D OctoMap SLAM and the original ultrasonic-based 2D-3Dh approach.

B. Metrics

Time Efficiency: Our method reduced average mapping time from 22.6 min (3D) to 11.5 min.
Memory Usage: Average map size dropped from 450 KB (3D) to 36 KB.
Accuracy: Scene switching precision was 94.7% on the test dataset.

IV. Conclusion

We introduced a scene-aware hybrid mapping technique that uses semantic perception to control 2D-3D switching during SLAM. Our approach provides significant computational savings without compromising environmental understanding. This framework is adaptable to various platforms and scales well with complex environments where elevation changes are common.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2024-00437190) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation, 50%).
This research was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1A6A1A03024003, 50%).

References

[1] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, vol. 34, no. 3, pp. 189–206, 2013.
[2] J. P. Saarinen, H. Andreasson, T. Stoyanov, and A. J. Lilienthal, "3D normal distributions transform occupancy maps: An efficient representation for mapping in dynamic environments," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1627–1644, 2013.
[3] R. K. Megalingam, S. Tantravahi, H. S. Tammana, and H. S. R. Puram, "2D-3D hybrid mapping for path planning in autonomous robots," *Int. J. of Intelligent Robotics and Applications*, vol. 7, no. 3, pp. 291–303, 2023.