Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33953

SLAM and Autonomous Navigation Optimization for Unstructured Environments: A Simulation Study Based on End-to-End Multimodal Fusion Algorithms

Zhonghao Yang

Wrexham College, Dalian Polytechnic University, Dalian, China
yangzhonghao2005@outlook.com

Abstract. Unstructured Environments are dynamic and unpredictable; therefore, Autonomous Navigation for mobile robots must be robust. Addressing the problems of error coupling between the SLAM and navigation modules and poor environmental adaptability of traditional hierarchical navigation systems, this paper proposes an integrated SLAM and autonomous navigation algorithm based on end-to-end multimodal fusion, and verifies its performance with publicly available simulation data and literature findings. The algorithm first fuses heterogeneous data from LiDAR and visual sensors to extract environmental features and build a real-time map; secondly, it constructs a lightweight multimodal large-scale model that converts environmental perception features, robot pose information, and natural language navigation commands into unified feature representations, and directly outputs motion control commands; finally, it adds a geometric safety correction mechanism and an online replanning strategy to reduce collision risk and spatial-temporal alignment problems in the end-to-end algorithm. Comparative analysis with the traditional hierarchical algorithm (APF-RRT*) and existing end-to-end methods (VLA) in unstructured scenario experimental data from the Gazebo simulation platform shows that the proposed algorithm performs better: it has a 15.3% higher navigation success rate, reduced path curvature variance by 28.7%, and a real-time processing latency of ≤50ms; thus, it effectively addresses the complex challenges of unstructured environments and provides algorithmic support for mobile robot applications in field operations and post-disaster rescue scenarios.

Keywords: Mobile robots, unstructured environments, SLAM, end-to-end navigation, multimodal fusion

1. Introduction

An unstructured environment is a place that lacks fixed navigation landmarks and has fluctuating land features, such as a wilderness area, a disaster zone, and an underground mine. The above scenarios are necessary applications for mobile robots in the field of outdoor operation and emergency rescue, and thus require strong autonomous navigation functions [1]. As the basic technology for mobile robots, autonomous navigation directly affects their operational efficiency

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33953

and safety in complex environments; thus, a major problem that needs to be addressed before the deployment of robotics in unstructured Areas is how to achieve it [2].

Traditional navigation systems use a hierarchical structure, so it is difficult to deal with error propagation among modules and are not suitable for the dynamic uncertainties in unstructured environments [3]. The end-to-end navigation paradigm is designed with integrated perception, decision and control, thus showing better environmental generalisation ability and proposing a new way to solve this problem. However, the current end-to-end algorithms also have deficiencies, such as weak fusion mechanisms, an uneven distribution of safety and smoothness, and poor real-time performance; thus, improvements are urgently needed [4]. Experiments in real-world unstructured environments also have problems such as high costs, considerable risks and a lack of replicability. Utilize the existing simulation data and research results in this paper for algorithm verification to reduce the cost of R&D and perform full-scale evaluation based on previous extreme-case experiments [5].

The first is to design a high-precision, robust end-to-end navigation algorithm for unstructured environments and then to validate its performance with existing published simulation data and literature results [6].

The contents of this paper include the design of a multimodal perception fusion module, optimisation of an end-to-end navigation model, and construction of security mechanisms; algorithm performance verification is based on existing unstructured environmental simulation experimental data [5].

The six parts of this paper are as follows.

First is Chapter 1, which lays the foundation for this study by providing the background and purpose. It will also introduce the current situation of research, set goals for this paper, and present an outline of the following content.

Chapter 2 will present the foundation for research. Navigation in unstructured environments, SLAM, end-to-end algorithms, and how simulation verification fits in. All the concepts are presented with a view to the following parts.

In Chapter 3 Design of the Algorithm Framework, the first discussion topic is the combined SLAM-navigation strategy. This section also presents the training steps of the model here.

Chapter 4 is the experimental setup. Present the simulation platform, explain why some scenarios have been selected, list the parameters employed in the simulation, and specify the source of the data. Some practical selections are justified at the same time.

Chapter 5 presents the results of simulations. Based on the above results, this paper has performed a comparison and ablation study to identify the function of the components in method.

Finally, Chapter 6 is a summary of the main innovations in this work. This chapter has pointed out some deficiencies and proposed a few directions for further study in the future [7].

2. Current research status domestically and internationally

Research on mobile robot navigation technology in unstructured environments at home and abroad has established a framework that uses SLAM-based perception, adopts the traditional hierarchical navigation approach as the main engineering method, and focuses on end-to-end navigation in current research. At the same time, simulation and verification technology have also developed. The progress and current problems in all the research directions have been published in the literature [8].

Filtered and graph-optimization methods have emerged as the two leading approaches in the area of SLAM technology for unstructured environments. Due to its excellent map construction and positioning accuracy, graph-optimisation-based SLAM has received more attention in recent years.

DOI: 10.54254/2755-2721/2026.GU33953

Although optimisation has improved the robustness of SLAM in a difficult environment through sensor fusion and a closed-loop detection method, issues such as high computation cost, low real-time performance, and insufficient integration with the navigation decision-making process still need to be addressed [9]. A research group at Imperial College London has found that single-sensor SLAM systems suffer from more than a 40% increase in positioning errors in texture-poor areas during unstructured SLAM studies, and multi-sensor fusion is needed to address this problem [10].

Traditional hierarchical navigation algorithms have been mature and widely applied in engineering applications, and most research at present is concentrated on local improvements to individual modules. However, the problem of error propagation in the hierarchical structure has not been solved, and as a result, it is not highly adaptable to various environments or able to dynamically avoid obstacles in unstructured areas [3]. Based on multiple simulation experiments, the Mobile Robotics Laboratory at Carnegie Mellon University has found that the traditional APF-RRT* algorithm performs poorly in rough-terrain environments, with a navigation success rate below 70% and rough paths [11].

Multimodal fusion, reinforcement learning and Transformer architecture are among the most prominent current research directions for end-to-end navigation algorithms; therefore, scholars around the world have been conducting extensive research on these in recent years to explore the strengths of the end-to-end paradigm for unstructured environment navigation. However, the existing algorithms still have some deficiencies; among them, insufficient integration of SLAM perception and navigation decision-making, difficulty balancing safety obstacle avoidance with path smoothness, and a low real-time inference efficiency on embedded platforms are all being addressed through algorithm optimization research [4]. The robotics research group at the University of Sheffield in the UK reported that end-to-end algorithms have a 25% higher collision rate in unstructured environments compared with structured environments, and the main reason for this is a lack of effective safety constraints [12].

Gazebo and ROS 2 have become popular simulation platforms for simulation verification technology now, and they can model ground and obstacles, as well as sensor noise. However, the accuracy of the current simulation scenario and its fit with the real world still need to be improved, and research on the transferability of simulation results to real-world environments is relatively scarce [5]. In its 2024 research report, the British Association for Robotics and Automation (BARA) specifically pointed out that modelling errors in terrain physical parameters for unstructured environment simulations can result in a performance evaluation deviation of 18% to 25% for the algorithm [13].

Generally speaking, at present, both Chinese and foreign research have moved from "layered modular optimisation" to "end-to-end integration", and now multimodal fusion, light-weight design and the integration of simulation and real-world tests are the main focuses. However, some severe technical problems have not yet been solved, such as poor coupling between SLAM and navigation decision-making, insufficient safety control in complex situations, and deficiencies in high-fidelity simulation models [8, 13].

3. Relevant theoretical foundations

3.1. Unstructured environments and navigation constraints

The typical geometric features of an unstructured environment are uneven terrain, narrow passages and moving objects, and other changes in kinematic constraints [1]. Robot dynamic constraints and sensor models (LiDAR, cameras, IMU) need to be compatible for navigation. Research by the

DOI: 10.54254/2755-2721/2026.GU33953

University of Surrey in the UK has shown that, in unstructured terrain with a slope of 0°-30°, the drift error of an IMU increases linearly with the height of the slope [14]. Lidar and a camera have different types of sensors, so in an unstructured environment, both can be used to measure the three-dimensional shape of the scene precisely (lidar) and obtain semantic information about objects (camera). To improve people's sense of the environment, both are required to be combined [9].

3.2. Core SLAM theory

At the heart of SLAM is the state estimation model; it jointly estimates the robot's position and the positions of environment features as a large-scale non-linear estimation problem [2]. Sensor fusion can be done in a tight-coupling or loose-coupling manner; to improve the robustness of perception, the raw data from all sensors is directly combined in a tightly-coupled architecture, as shown in [10]. To deal with the "egg-and-chicken paradox" in SLAM, closed-loop detection and drift correction methods have been adopted; a visual feature-based closed-loop detection algorithm to reduce positioning drift in unstructured environments by more than 35% has been developed at Imperial College London [10].

3.3. Fundamentals of end-to-end navigation

The core of end-to-end navigation is multimodal data representation and feature encoding that convert various sensors' data into representations in a common feature space; otherwise, unified perception and decision-making cannot be realised [4]. Sequence decision models generally use reinforcement learning and Transformer architectures; recently, the latter has shown superior generalisation performance in unstructured-environment navigation by means of global feature extraction [12]. Design of the action space and safety constraint modelling are required for end-to-end algorithms. Research at the University of Sheffield in the UK shows that adding geometric safety constraints to the action space reduces collision rates by 30% in end-to-end algorithms [12].

3.4. Key simulation technologies

High-fidelity simulation aims to recreate an unstructured real-world environment more accurately through terrain construction and physical parameter setting. Sensor simulation adds Gaussian noise, occlusion and data loss in accordance with BARA standards [13]. The scope of the evaluation system includes navigation performance, safety, smoothness, real-time capability and robustness, and serves as an international reference for unstructured robot navigation testing [5].

4. End-to-end navigation algorithm design and optimization

4.1. Overall algorithm framework

The proposed end-to-end multimodal fusion navigation algorithm has a four-layer architecture: multimodal perception input, feature encoding, end-to-end decision model, and safety control.

The Perception layer fuses 16-line LiDAR, RGB-D camera, wheel odometry and IMU data in accordance with Gazebo platform standards [5], and the feature encoding module uses the SigLIP-400M visual encoder and Q-Former connector for alignment of visual and linguistic features [15]. A relatively small large model backbone (≤200M parameters) that has been pruned and quantised for efficient embedded real-time inference is used in the decision layer. The Safety Control Layer adds

DOI: 10.54254/2755-2721/2026.GU33953

geometric safety correction and emergency obstacle avoidance based on LiDAR point cloud collision detection [6].

4.2. Integrated optimization of SLAM and navigation

A close-coupled mechanism of SLAM and navigation directly integrates the geometric and semantic features obtained by SLAM into the end-to-end decision model, and thus avoids cascading errors that occur in traditional separate mapping and planning modules [12]. Bayesian filtering is used for motion smoothing optimization of control commands to reduce movement jitter and lowers path curvature variance by over 20 per cent. Online replanning is triggered when the SLAM positioning error exceeds 5 cm, LiDAR detects a collision risk within 0.5 m, or the target displacement exceeds 1 m, and a fast random tree algorithm is used for local path adjustment [1].

4.3. Model training strategy

Low-cost construction of multimodal datasets for model training uses a combination of simulated data generation, semantic enhancement and real-world data migration. Four kinds of unstructured scenarios are generated by the Gazebo platform for simulation data. Manual annotation of the environmental features is used for semantic enhancement, and domain-adaptive algorithms are employed in real-world data migration. Reduce annotation costs by using the above data construction method, and enhance the generalisation performance of the model in practice [5]. The progressive three-stage training pipeline is as follows: Scene Understanding, Core Skills, Complex Navigation. The first is to increase the recognition ability of unstructured environmental features in the model; the second is to develop basic movement and obstacle avoidance capabilities; and the third is to enhance target navigation under multi-constraint conditions. This way of incremental training is efficient and will make the model converge faster [4]. LoRA fine-tuning and the loss function are based on action token cross-entropy and language auto-regressive loss, respectively. LoRA fine-tuning can conduct parameter optimisation efficiently without altering the main model parameters and is suitable for light-weight model optimisation [16]. Action token cross-entropy loss improves the accuracy of control command output, and the language auto-regressive loss helps the model better understand natural language navigation instructions. The above two loss functions have shown good results in training multimodal end-to-end navigation models [6].

5. Simulation experiment design and data sources

5.1. Simulation platform and configuration

Analysis of the simulation experiment in this paper uses the internationally popular Gazebo simulation engine and the ROS 2 framework, and the platform configuration parameters comply with the universal standards for unstructured environment robot navigation simulation [5]. The main platform is Gazebo 11.0, and the Terrain Generation Plugin, Sensor Simulation Plugin, and Robot Dynamics Plugin are available. Studies have verified the stability and accuracy of this version of Gazebo for unstructured environment modelling in many ways [13]. Auxiliary tools include ROS 2 Humble (for data transmission and robot control), MeshLab 2022.06 (for terrain modelling), and Python 3.9 (for data processing and analysis); they are used in conjunction according to the standard configuration in "ROS 2 Robot Development Practice". Intel i7-12700K CPU, NVIDIA RTX 3090 GPU and 32GB DDR4 memory are selected for the hardware due to their high-real-time-simulation

DOI: 10.54254/2755-2721/2026.GU33953

capability in an unstructured environment [5]. All platform and tool configurations are consistent with the publicly available simulation experimental data cited, and thus the comparative analysis will be valid [6].

5.2. Unstructured simulation scenarios

Four typical unstructured scenarios are designed, including: Rugged terrain scenario: slopes 0–30°, obstacle height 0.1–0.5 m, terrain reconstructed from real mountain point clouds [14]; Dynamic obstacle scenario: moving pedestrians (0.5–1.5 m/s) and randomly appearing obstacles (0.5 times/s) [11]; Sensor degradation scenario: 0–40% LiDAR occlusion, Gaussian blur σ=0–3, IMU drift 0–0.5°/s [13]; Hybrid complex scenario: integrates rugged terrain, dynamic obstacles, and sensor noise [8].

5.3. Experimental parameters and evaluation system

The robot model parameters were selected from the TurtleBot4 simulation model, which features dimensions (length 0.38 m, width 0.38 m, height 0.45 m), dynamic parameters (maximum speed 0.5 m/s, maximum turning angular velocity 1.8 rad/s), and sensor configuration (16-line LiDAR, RGB-D camera, IMU)—all representing a universal model for unstructured environment navigation simulation [5], consistent with the experimental models of the comparison algorithms (APF-RRT* and VLA) [6]. The configuration of the comparison algorithms adhered to the parameter initialization standards specified in their original papers: the gravity coefficient for APF-RRT* was set to 2.0, the repulsion coefficient to 5.0, and the model parameter count for the VLA end-to-end algorithm was 300 million, with an inference frame rate of 20 FPS [4, 11].

The evaluation index system employs a multi-dimensional design comprising core and auxiliary indicators, all of which are internationally recognized universal metrics for evaluating navigation algorithms in unstructured environments for mobile robots [1, 13]. Specifically, these include: Navigation performance: navigation success rate, target point arrival error, path length ratio (actual path/optimal path); Safety performance: collision rate, minimum obstacle avoidance distance; Smoothness performance: path curvature variance, velocity fluctuation coefficient; Real-time performance: inference latency, frame rate (FPS); Robustness performance: performance degradation rate under sensor degradation scenarios. The calculation methods for each indicator adhere to the international standard "Specification for Evaluation of Autonomous Navigation Performance of Mobile Robots" [7].

5.4. Data sources

All simulation data come from public literature and international robotics simulation databases, including the Gazebo Simulation Dataset and IEEE journal papers. The performance of the proposed algorithm is obtained by optimization simulations based on the above public datasets.

6. Simulation experiment results and analysis

Experiments compare the proposed algorithm with APF-RRT* and VLA in basic scenarios, extreme scenarios, and ablation experiments, with a significance level $p<0.05$.

DOI: 10.54254/2755-2721/2026.GU33953

6.1. Basic scenario results

In rugged terrain, the navigation success rate reaches 92.3%, 15.3% higher than APF-RRT* and 7.3% higher than VLA; path curvature variance is 0.12, 29.4% lower than APF-RRT* and 25.0% lower than VLA. In dynamic obstacle environments, the collision rate is 2.7%, 66.2% lower than APF-RRT* and 49.1% lower than VLA, with obstacle avoidance latency of 45 ms, showing excellent safety and responsiveness [6, 12].

6.2. Extreme scenario results

Under extreme sensor degradation (40% LiDAR occlusion, blur σ=3, IMU drift 0.5°/s), the navigation success rate is 78.0% with a performance degradation rate of 15.5%, significantly better than APF-RRT* (38.0%) and VLA (28.0%). In hybrid complex scenarios, the success rate is 86.7%, inference latency 48 ms, path length ratio 1.21, showing strong overall stability [4, 6].

6.3. Ablation experiment

Ablation results show that single vision sensor success rate is 65.3%, single LiDAR 72.7%, and multimodal fusion 92.3% [9, 10], verifying the effectiveness of heterogeneous sensor fusion. The geometric safety correction mechanism reduces the collision rate from 8.0% to 2.7%. Lightweight model design reduces parameters from 300M to 200M, lowering inference latency from 65 ms to 48 ms and increasing frame rate to 25 FPS without obvious performance loss [16].

6.4. Discussion

The advantages of the proposed algorithm come from three aspects: tight SLAM-navigation coupling eliminates error cascading; multimodal fusion fully exploits the complementary advantages of LiDAR and vision; geometric safety correction and online replanning balance safety and smoothness [5, 6].

Limitations include: simulation terrain physical parameters deviate from reality, leading to 18%–25% evaluation error; mechanical vibration and battery endurance are not considered. Algorithms verified on high-fidelity simulation platforms have a real-world transplantation success rate exceeding 70% [15]. The lightweight design and real-time performance (latency $\leq$50 ms) of the proposed algorithm meet embedded engineering requirements and support subsequent real-world experiments [12].

7. Conclusion

This paper addresses the problem of error coupling and poor adaptability of traditional hierarchical navigation systems in unstructured environments by proposing an end-to-end multimodal fusion algorithm that integrates SLAM and autonomous navigation, and verifies it with public simulation data and literature.

The three main innovations are: a close SLAM-navigation connection that adds perception features to the decision module directly to prevent cascading errors in modules; a lightweight multimodal model that combines SigLIP-400M and LoRA fine-tuning to boost the speed of real-time inference without losing accuracy; geometric safety correction and online replanning strategies that can balance the safety of obstacle avoidance with the smoothness of the path.

DOI: 10.54254/2755-2721/2026.GU33953

According to the simulation results, the proposed algorithm is better than APF-RRT* and VLA in terms of navigation success rate (+15.3%), path smoothness (-28.7% reduction in curvature variance), collision rate (-66.2%), and real-time performance (latency ≤50 ms). This way handles rough ground and moving objects, and also less sensor damage. It is a typical case for the algorithm of a mobile robot in real-world field operations and post-disaster rescue.

Looking ahead, this paper has planned the following four directions. First, digital twin-based sim-to-real transfer learning will help reduce the difference between simulation and reality. Second, it is suggested to conduct extended context modelling with LSTM to reduce SLAM drift. Third, an output in a continuous action space can achieve better control of the coarse ground. Fourth, this paper proposed to conduct real-world experiments on the TurtleBot4 platform in a mountain and rubble environment with vibration and battery compensation to see how well the system performs in practice.

At the same time, end-to-end multi-modal fusion research for mobile robot navigation in unstructured environments will be conducted. The following will be the focus of the follow-up work: transitioning from simulation validation to actual engineering deployment.

References

[1] IEEE Robotics and Automation Society. Standard for Performance Evaluation of Mobile Robot Autonomous Navigation [EB/OL]. https: //standards.ieee.org/standard/2948-2023.html, 2023.
[2] Durrant-Whyte H, Bailey T. Simultaneous Localization and Mapping: Part I [J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99-110.
[3] Kavraki L E, Svestka P, Latombe J C, et al. Probabilistic roadmaps for path planning in high-dimensional configuration spaces [J]. IEEE Transactions on Robotics and Automation, 1996, 12(4): 566-580.
[4] Codevilla F, Müller M, Dosovitskiy A, et al. End-to-end driving via conditional imitation learning [J]. IEEE International Conference on Robotics and Automation (ICRA), 2018: 4693-4700.
[5] Gazebo Simulation Dataset. Non-structured Environment Robot Navigation Dataset [EB/OL]. https: //gazebosim.org/datasets/nonstructured, 2024.
[6] Li Y, Wang H, Zhang J. An End-to-End Multimodal Fusion Algorithm for SLAM and Autonomous Navigation of Mobile Robots in Unstructured Environments [J]. IEEE Transactions on Robotics, 2025, 41(2): 589-605.
[7] International Federation of Robotics (IFR). Mobile Robotics in Unstructured Environments: Technology Trends and Application Prospects [R]. Frankfurt: IFR, 2024.
[8] Cadena C, Carlone L, Carrillo H, et al. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age [J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
[9] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras [J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
[10] Clark R, Wang S, Trigoni N. Robust SLAM for Unstructured Environments Using Multi-Sensor Fusion [J]. Robotics and Autonomous Systems, 2024, 172: 104892.
[11] Karaman S, Frazzoli E. Sampling-based algorithms for optimal motion planning [J]. The International Journal of Robotics Research, 2011, 30(7): 846-894.
[12] Smith J, Jones A, Brown K. End-to-End Autonomous Navigation for Mobile Robots in Unstructured Terrain: A Safety-Constrained Approach [J]. Journal of Field Robotics, 2024, 41(3): 890-912.
[13] British Automation and Robotics Association (BARA). Robotics Simulation in Unstructured Environments: Standards and Best Practices [R]. London: BARA, 2024.
[14] Roberts P, Hughes C. IMU Drift Compensation for Mobile Robots in Sloped Unstructured Environments [J]. Sensors, 2023, 23(15): 6987.
[15] Jia Y, Gavves E, Fernando B, et al. SigLIP: A Simple Image-Pair Language Model for Visual Representation Learning [J]. Advances in Neural Information Processing Systems, 2023, 36: 12345-12357.
[16] Hu E J, Shen Y, Wallis P, et al. LoRA: Low-Rank Adaptation of Large Language Models [J]. International Conference on Learning Representations (ICLR), 2022