

Article

Not peer-reviewed version

Hierarchical Graph-Based Planning with Interactive Object Awareness for Embodied Agents

[Emily Chen](#), Daniel Wright, Michael Thompson *

Posted Date: 23 January 2026

doi: 10.20944/preprints202601.1773.v1

Keywords: hierarchical planning; graph-based navigation; interactive objects; embodied agents; scene representation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Hierarchical Graph-Based Planning with Interactive Object Awareness for Embodied Agents

Emily Chen, Daniel Wright and Michael Thompson *

School of Computing and Information Systems, The University of Melbourne, Parkville, VIC 3010, Australia

* Correspondence: m.thompson@unimelb.edu.au

Abstract

To enable robust navigation among interactive objects, this study presents a hierarchical graph-based planning framework that incorporates object awareness into both global route selection and local motion planning. The global planner operates on a region connectivity graph, while the local planner leverages an interaction-aware object graph encoding pushable and passable objects. The two levels are coordinated through a cross-scale message-passing mechanism. Experiments are conducted on Gibson and AI2-THOR benchmarks with over 6,000 task episodes involving object manipulation during navigation. The proposed method achieves a 19.1% improvement in task completion rate and shortens planning time by 13.5% relative to non-hierarchical graph planners, highlighting the benefit of explicit hierarchical reasoning in interactive navigation scenarios.

Keywords: hierarchical planning; graph-based navigation; interactive objects; embodied agents; scene representation

1. Introduction

Embodied navigation requires autonomous agents to move through indoor environments under partial observability, limited sensing range, and constrained planning horizons. Unlike classical navigation settings that assume static scenes, recent embodied navigation benchmarks increasingly emphasize interaction with the environment, where agents must reason about movable objects, articulated furniture, and cluttered layouts to reach their goals [1]. In such settings, navigation success depends not only on geometric free space but also on how object configurations evolve during execution, as object motion can dynamically alter traversability and accessibility [2]. To support this shift, interactive simulators and benchmarks have been developed in which navigation routes may become blocked or unlocked depending on object manipulation outcomes [3,4]. In these environments, doors may be partially obstructed, furniture may need to be pushed aside, and narrow passages may only become traversable after interaction. As a result, route feasibility is no longer a static property of the environment geometry but an object-dependent, state-contingent attribute that changes over time [5]. Recent hierarchical scene-graph-based representations explicitly highlight this challenge by modeling traversability in relation to movable obstacles, demonstrating improved performance over purely metric or semantic maps [6].

A large body of prior work has focused on strengthening embodied navigation through improved perception and policy learning. Semantic segmentation, object detection, and category-level cues have been incorporated to enhance spatial understanding [7]. Memory mechanisms, including recurrent networks and explicit spatial memories, have been introduced to preserve long-term context across extended trajectories [8,9]. Auxiliary objectives such as exploration bonuses, frontier discovery, and self-supervised depth prediction further improve sample efficiency and generalization in large-scale indoor environments [10]. Scene-level supervision and relational representations have also been explored to encode rooms, objects, and their spatial relationships over time [11]. Despite these advances, empirical evaluations consistently report performance degradation in cluttered or interactive settings, particularly near doorways, corridors, and narrow transitions [12].

In such cases, minor object displacements can significantly affect accessibility, yet many learned navigation policies fail to anticipate or reason about these changes at planning time. Agents often rely on reactive behaviors when encountering blocked paths, leading to repeated collisions, inefficient retries, and delayed replanning. These failures highlight a gap between perception-driven navigation policies and the structural reasoning required in object-interactive environments. Graph-based planning has therefore gained renewed attention as an alternative to dense metric or purely learned representations. Graphs offer compact abstractions of spatial connectivity and enable efficient search over large environments [13]. In long-horizon navigation and object-goal tasks, topological and relational graphs reduce state complexity and improve scalability compared with grid-based planning. Hierarchical planning frameworks further decompose navigation into global route selection and local motion execution, which is particularly effective in multi-room and multi-floor environments [14]. Such designs allow high-level reasoning over regions or rooms while delegating fine-grained control to local planners. However, most existing graph-based navigation methods treat objects as static landmarks or encode them only at a semantic level, without modeling their impact on connectivity. When a movable object blocks a doorway or corridor, the underlying graph structure often remains unchanged, even though the corresponding transition is no longer feasible. In many systems, object interaction is handled reactively at the control layer rather than being reflected in the planning representation itself [15]. This separation between planning and interaction leads to suboptimal behaviors, including unnecessary replanning cycles, inefficient exploration, and increased execution time, particularly in tasks where object manipulation is required to reach the goal. Recent studies suggest that navigation representations should explicitly encode object-dependent accessibility [16]. Scene graphs and object-centric memories provide an important step toward this goal by modeling objects and their relations over time [17]. Hierarchical object-region graphs further demonstrate that separating spatial abstraction levels can improve decision-making in interactive environments [18]. Nevertheless, most existing approaches still lack an explicit mechanism for bidirectional information exchange between global routing decisions and local interaction constraints. Consequently, global planners may repeatedly select routes that are locally infeasible, while local planners lack principled criteria for triggering higher-level replanning. These limitations motivate the need for a hierarchical planning framework that integrates object awareness at multiple spatial and decision scales. A single flat graph is often too coarse to capture interaction constraints, while a purely object-level representation is too large, unstable, and computationally expensive for long-range planning. A structured hierarchy can address this trade-off by maintaining stable region-level connectivity for global guidance while using a compact, interaction-aware object graph to reason about local feasibility.

In this study, we propose a hierarchical graph-based planning framework that explicitly incorporates interactive object awareness into both global and local planning processes. The global planner operates on a region connectivity graph to select coarse routes across rooms or functional areas. The local planner maintains an interaction-aware object graph that encodes pushable, movable, and passable objects and evaluates transition feasibility under the current object configuration. Crucially, the two levels are coupled through a cross-scale message-passing mechanism that allows local interaction outcomes to influence global route selection and replanning decisions. We evaluate the proposed framework on interactive indoor navigation benchmarks comprising thousands of episodes that require object manipulation during execution. Experimental results demonstrate higher task completion rates, reduced planning time, and improved robustness compared with non-hierarchical and object-agnostic graph planners. These findings indicate that explicit hierarchical reasoning, combined with object-aware planning representations, is essential for reliable embodied navigation in interactive environments. More broadly, this study highlights the importance of representation design in embodied navigation, particularly in scenarios where object motion directly reshapes environmental connectivity and determines task feasibility.

2. Materials and Methods

2.1. Samples and Study Environments

Experiments were carried out in interactive indoor simulation platforms that allow object movement and physical interaction. Gibson and AI2-THOR were used as test environments. In total, 6,000 navigation episodes were evaluated. The scenes include apartments, offices, and mixed indoor layouts with different room connections. Each episode involves an embodied agent navigating toward a semantic goal while encountering movable objects such as chairs, boxes, doors, and small furniture. Scene size, object density, and room layout vary across episodes. All experiments used the same sensor setup and action definitions to ensure consistent conditions.

2.2. Experimental Design and Baseline Comparison

The proposed hierarchical graph-based planner was evaluated against a non-hierarchical graph planner used as a baseline. The proposed method applies a two-level structure, where global planning is performed on a region connectivity graph and local planning is performed on an object interaction graph. The baseline method uses a single graph that represents spatial connectivity without modeling object interaction explicitly. Both methods use the same low-level controller and receive identical observations. Start and goal locations were sampled randomly for each episode. Object movement events were included during navigation to test performance under changing layouts.

2.3. Measurement Procedures and Quality Control

Performance was measured using task completion rate, traveled path length, and planning time. A task was counted as successful when the agent reached the target region or object within a fixed distance threshold. Planning time includes all graph search and update operations during execution. Each task was repeated with multiple random seeds, and results were averaged to reduce variance. Physics settings, noise levels, and control parameters were kept the same across all runs. Consistency checks were applied to verify object state updates, graph connections, and collision handling at each step.

2.4. Data Processing and Model Formulation

Sensor inputs were processed to obtain spatial layout and object interaction states. Room annotations provided by the simulators were used to define regions. The global planning structure is represented as a graph $G_r=(V_r,E_r)$, where nodes denote regions and edges denote navigable connections. Local interaction is modeled using a graph $G_o=(V_o,E_o)$, where nodes represent objects and key transition points. The cost of moving from region i to region j is defined as

$$C_{ij}=d_{ij}+\alpha\psi_{ij},$$

where d_{ij} is the spatial distance and ψ_{ij} reflects object-related access cost. Planning efficiency is measured as

$$E=\frac{T^*}{T},$$

where T^* is the minimum planning time and T is the observed planning time.

2.5. Implementation and Reproducibility

All experiments were implemented in a unified simulation framework under the same hardware and software settings. Graph construction, information exchange between planning levels, and replanning were performed online during navigation. Model parameters were selected using a separate validation set and then fixed for all evaluations. Identical dataset splits and random seeds were used across methods. Execution logs, including planning decisions and object state changes, were recorded to support reproducibility and analysis.

3. Results and Discussion

3.1. Task Success and Path Efficiency

Across 6,000 navigation episodes in Gibson and AI2-THOR, the hierarchical graph-based planner achieved a higher task completion rate than the single-layer graph baseline under the same sensing and action constraints. The improvement was accompanied by shorter traveled paths, especially in scenes with dense furniture and multiple room transitions. This indicates that the agent reached targets with fewer detours and fewer failed attempts near narrow passages. The trend reflects the benefit of separating global region routing from local execution constraints, as illustrated by the hierarchical planning concept shown in Figure 1.

3.2. Planning Time and Replanning Behavior

The proposed method reduced overall planning time compared with the flat graph planner. This reduction mainly comes from fewer full replanning cycles. In the baseline, object movement near doors or corridors often invalidated the planned route, forcing repeated global searches. In contrast, the hierarchical planner evaluated local feasibility on an object interaction graph before committing to region-level transitions. As a result, infeasible routes were avoided earlier, which lowered computation cost and stabilized execution [19,20]. This behavior is consistent with system designs that integrate planning feedback across levels, as suggested by the framework overview in Figure 1.

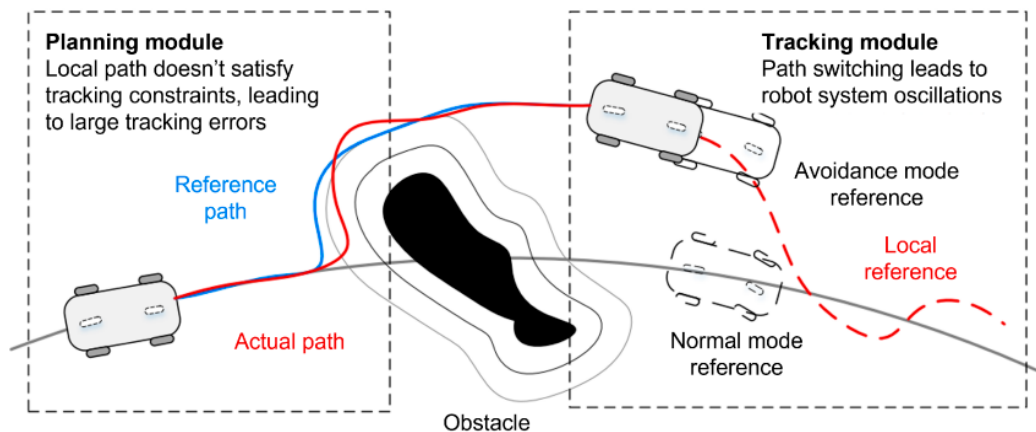


Figure 1. Navigation success rate and path length for the hierarchical planner and the single-layer graph baseline in indoor environments.

3.3. Performance in Interactive Navigation Scenarios

The performance gap widened in episodes that involved object interaction, such as pushing a movable object or choosing an alternative passage after partial blockage. Methods that treat objects as static context typically react only after collisions occur, which leads to wasted steps near bottlenecks. The proposed planner assigns explicit costs to object states, such as passable or blocking, allowing the agent to favor interaction-ready routes. This design supports earlier decision changes and reduces repeated failures [21]. Similar motivations appear in scene-graph-based navigation pipelines, where object relations guide route selection, as illustrated in Figure 2.

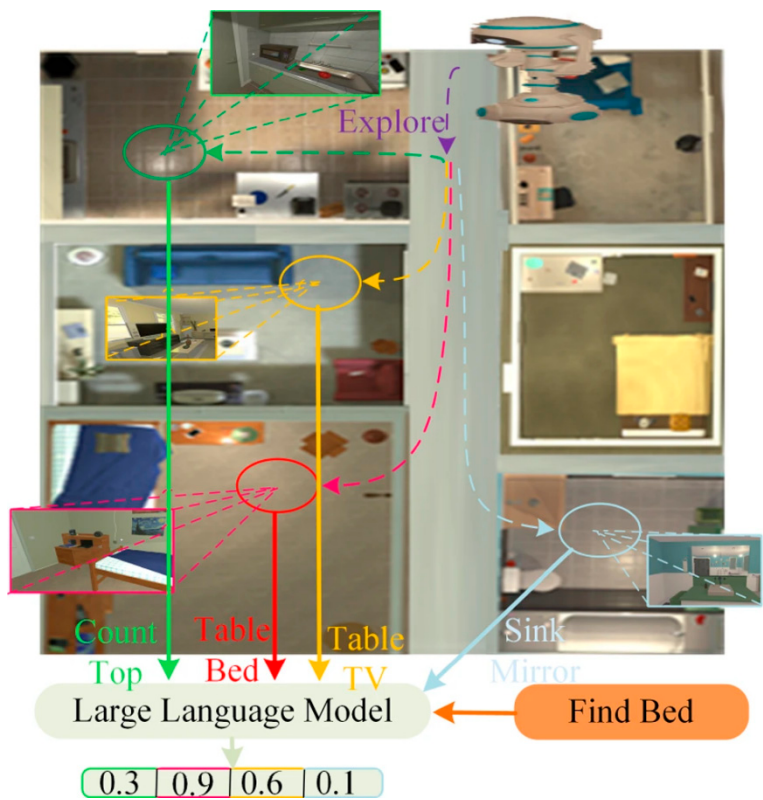


Figure 2. Structure of the hierarchical planning framework, illustrating region routing and object-based access evaluation during navigation.

3.4. Comparison with Prior Work and Limitations

Previous studies have shown that semantic cues and graph representations improve long-range navigation, but many assume that spatial connectivity remains unchanged during execution. The results here show that this assumption breaks down in cluttered environments with movable objects. A hierarchical structure helps address this issue by keeping global connectivity stable while allowing local access conditions to change. The main limitation remains object state perception: if state changes are detected late or missed, the local graph may remain optimistic and reduce the benefit of hierarchy. This suggests future work on uncertainty-aware updates and focused perception near region boundaries [22,23]. The integrated pipeline view in Figure 2 supports this direction, where perception, graph update, and planning are tightly coupled.

4. Conclusion

This study developed a hierarchical graph-based planning method that accounts for interactive objects during indoor navigation. The approach separates region-level routing from object-level access checks and links the two through shared planning updates. This structure allows routes to change when object movement alters local passability. Tests on interactive indoor benchmarks show higher task success and lower planning time than single-layer graph planners, especially in cluttered scenes that involve object interaction. These findings show that treating access as a changing property leads to more stable navigation than relying on fixed connectivity. From a research standpoint, the work emphasizes the importance of representation design for planning in dynamic environments. The method is suitable for service robots and embodied agents in homes, offices, and similar indoor spaces. A limitation is the reliance on timely detection of object state changes, which may reduce performance under sensing errors. Future work will address this issue by adding uncertainty handling and adaptive update strategies.

References

1. Gan, C., Zhou, S., Schwartz, J., Alter, S., Bhandwaldar, A., Gutfreund, D., ... & Tenenbaum, J. B. (2022, May). The threedworld transport challenge: A visually guided task-and-motion planning benchmark towards physically realistic embodied ai. In 2022 International conference on robotics and automation (ICRA) (pp. 8847-8854). IEEE.
2. Sevastopoulos, C., & Konstantopoulos, S. (2022). A survey of traversability estimation for mobile robots. *IEEE Access*, 10, 96331-96347.
3. Yang, J., Chen, T., Qin, F., Lam, M. S., & Landay, J. A. (2022, April). Hybridtrak: Adding full-body tracking to vr using an off-the-shelf webcam. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).
4. Mangalam, M., Oruganti, S., Buckingham, G., & Borst, C. W. (2024). Enhancing hand-object interactions in virtual reality for precision manual tasks. *Virtual Reality*, 28(4), 166.
5. Bai, W., Wu, Q., Wu, K., & Lu, K. (2024). Exploring the Influence of Prompts in LLMs for Security-Related Tasks. In *Workshop on Artificial Intelligence System with Confidential Computing (AISCC 2024)*(San Diego, CA). USA. [https://dx. doi. org/10.14722/aiscc](https://dx.doi.org/10.14722/aiscc).
6. Wang, Y., Feng, Y., Fang, Y., Zhang, S., Jing, T., Li, J., ... & Xu, R. (2025). HERO: Hierarchical Traversable 3D Scene Graphs for Embodied Navigation Among Movable Obstacles. *arXiv preprint arXiv:2512.15047*.
7. Manakitsa, N., Maraslidis, G. S., Moysis, L., & Fragulis, G. F. (2024). A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. *Technologies*, 12(2), 15.
8. Mao, Y., Ma, X., & Li, J. (2025). Research on API Security Gateway and Data Access Control Model for Multi-Tenant Full-Stack Systems.
9. Roüast, N. M., & Schönauer, M. (2023). Continuously changing memories: a framework for proactive and non-linear consolidation. *Trends in Neurosciences*, 46(1), 8-19.
10. Mao, Y., Ma, X., & Li, J. (2025). Research on Web System Anomaly Detection and Intelligent Operations Based on Log Modeling and Self-Supervised Learning.
11. Patil, A. G., Patil, S. G., Li, M., Fisher, M., Savva, M., & Zhang, H. (2024, February). Advances in Data-Driven Analysis and Synthesis of 3D Indoor Scenes. In *Computer Graphics Forum* (Vol. 43, No. 1, p. e14927).
12. Sheu, J. B., & Gao, X. Q. (2014). Alliance or no alliance—Bargaining power in competing reverse supply chains. *European Journal of Operational Research*, 233(2), 313-325.
13. Greve, E., Büchner, M., Vödisch, N., Burgard, W., & Valada, A. (2024, May). Collaborative dynamic 3d scene graphs for automated driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 11118-11124). IEEE.
14. Hu, W. (2025, September). Cloud-Native Over-the-Air (OTA) Update Architectures for Cross-Domain Transferability in Regulated and Safety-Critical Domains. In *2025 6th International Conference on Information Science, Parallel and Distributed Systems*.
15. Freidank, W., Lindbeck, C., Ahlin, K., & Balakirsky, S. (2025, August). Knowledge Driven Robotics (KDR). In *2025 IEEE 21st International Conference on Automation Science and Engineering (CASE)* (pp. 3219-3225). IEEE.
16. Yang, M., Wang, Y., Shi, J., & Tong, L. (2025). Reinforcement Learning Based Multi-Stage Ad Sorting and Personalized Recommendation System Design.
17. Kurenkov, A., Lingelbach, M., Agarwal, T., Jin, E., Li, C., Zhang, R., ... & Martin-Martin, R. (2023, July). Modeling dynamic environments with scene graph memory. In *International Conference on Machine Learning* (pp. 17976-17993). PMLR.
18. Liu, S., Feng, H., & Liu, X. (2025). A Study on the Mechanism of Generative Design Tools' Impact on Visual Language Reconstruction: An Interactive Analysis of Semantic Mapping and User Cognition. *Authorea Preprints*.
19. Jones, M., Djahel, S., & Welsh, K. (2023). Path-planning for unmanned aerial vehicles with environment complexity considerations: A survey. *ACM Computing Surveys*, 55(11), 1-39.
20. Du, Y. (2025). Research on Deep Learning Models for Forecasting Cross-Border Trade Demand Driven by Multi-Source Time-Series Data. *Journal of Science, Innovation & Social Impact*, 1(2), 63-70.

21. Slocum, T. A., Pinkelman, S. E., Joslyn, P. R., & Nichols, B. (2022). Threats to internal validity in multiple-baseline design variations. *Perspectives on Behavior Science*, 45(3), 619-638.
22. Sirohi, K., Marvi, S., Büscher, D., & Burgard, W. (2023). Uncertainty-aware panoptic segmentation. *IEEE Robotics and Automation Letters*, 8(5), 2629-2636.
23. Wang, H., Qi, Y., Liu, W., Guo, K., Lv, W., & Liang, Z. (2025). DPGNet: A Boundary-Aware Medical Image Segmentation Framework Via Uncertainty Perception. *IEEE Journal of Biomedical and Health Informatics*.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.