Agentic AI for Spatial Intelligence: A Comprehensive Survey

Author One, Author Two, Author Three
Department of Computer Science
University of Example City, Country
email@example.com

Abstract—This survey provides a comprehensive overview of the rapidly advancing intersection of Agentic Artificial Intelligence (AI) and Spatial Intelligence. While autonomous agents are demonstrating increasingly sophisticated capabilities, their ability to understand, reason about, and interact with the physical world remains a critical frontier. We address a significant gap in the literature by providing a unified taxonomy that connects agentic architectures with the diverse tasks of spatial reasoning. This paper reviews the foundational concepts of agentic AI, including memory, planning, and tool use, and systematically categorizes spatial intelligence tasks across navigation, scene understanding, manipulation, and geospatial analysis. We survey state-of-the-art methods, including embodied agents, multimodal large language models, and graph neural networks, and analyze the landscape of existing benchmarks, highlighting the need for more integrated evaluation frameworks. Furthermore, we explore key applications in critical infrastructure and outline the pressing technical, ethical, and safety challenges. By synthesizing these fragmented fields, we provide a roadmap for future research, aiming to accelerate the development of robust, safe, and effective spatially-aware autonomous systems.

Index Terms—Agentic AI, Autonomous Agents, Spatial Intelligence, Spatial Reasoning, Geospatial AI, Survey

I. INTRODUCTION

THE RISE of autonomous systems marks a pivotal moment in the evolution of Artificial Intelligence (AI). We are witnessing a paradigm shift from specialized models that excel at narrow tasks to goal-oriented, self-directed agents capable of complex decision-making in dynamic environments. This field, which we term **Agentic AI**, represents a significant leap towards creating machines that can operate with a higher degree of autonomy and intelligence. Concurrently, the ability for these agents to perceive, comprehend, and act within the physical world—a capability we define as **Spatial Intelligence**—has become a primary bottleneck and a critical area of research. The convergence of these two domains is essential for developing AI systems that can effectively and safely navigate real-world complexities, from autonomous vehicles and robotic assistants to large-scale urban planning and disaster response systems.

Despite the rapid progress in both agentic systems and spatial reasoning, the research landscape remains fragmented. Numerous surveys have independently covered topics such as Large Language Model (LLM) agents [1], embodied AI [2], and geospatial analysis [3]. However, a comprehensive

synthesis that bridges the architectural components of agentic AI with the functional requirements of spatial intelligence is notably absent. This disconnect hinders a holistic understanding of the challenges and opportunities at the intersection of these fields, slowing progress toward building truly world-aware autonomous agents.

This survey aims to fill this critical gap. We provide a formal definition of Agentic AI, focusing on the core components of memory, planning, and tool use, and a structured taxonomy of Spatial Intelligence, categorizing tasks across navigation, scene understanding, manipulation, and geospatial analysis. Our primary contributions are threefold:

1) A novel, unified taxonomy that connects agentic architectures with spatial intelligence tasks, providing a structured framework for understanding and categorizing research in this interdisciplinary area.
2) A comprehensive review of the state-of-the-art methods, evaluation benchmarks, and real-world applications, synthesizing findings from previously disparate fields.
3) A forward-looking analysis of the open challenges and a research roadmap to guide future work in developing more capable, robust, and safe spatially-aware agentic systems.

By providing this synthesis, we aim to create a foundational reference for researchers, developers, and policymakers, fostering a more integrated approach to building the next generation of autonomous intelligence.

REFERENCES

[1] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Luo, *et al.*, "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 3, pp. 1–43, 2024.
[2] D. Driess, F. Xia, M. S. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Fu, T. Murthy, *et al.*, "Palm-e: An embodied multimodal language model," *arXiv preprint arXiv:2303.03378*, 2023.
[3] Z. Jiang, Z. Li, Y. Xu, Y. Li, Z. Li, H. Li, M. Mendieta, C. Chen, and A. Plaza, "Crom: A remote sensing representation benchmark for cropping, reconstruction, and observation-based modeling," *arXiv preprint arXiv:2402.19300*, 2024.