International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

Autonomous Multi-floor Logistics: A Comprehensive Survey on Vision-Guided SLAM and Navigation Frameworks for Multi-floor Indoor Delivery Robots

Prabhas K S[1], Sathwik Padmanabha[2], Vedanta M S[3], Harshith V[4]

[1,2,3,4]Department of Computer Science (AI & ML), Vidyavardhaka College of Engineering, Mysuru, India

Abstract
The Autonomous Mobile Robots (AMRs) are increasingly being deployed in hotels, offices and college complexes, where they deliver objects from one point to another. However, extending navigation to multi-floor environments introduces new challenges such as vertical movement, multi-level mapping and dynamic human interaction. SLAM is one of the most essential and fundamental technologies that allows autonomous bots to create a map of the indoor environment and estimate their location in it. This survey consolidates recent progress in different types of SLAM algorithms: W-VSLAM, ORB-SLAM2/3, RI-LIO, RGB- D based SLAM, sensor-fusion techniques: LVI-SAM, RI-LIO, FAST-LIO2, FAST-LIVO, LVI-fusion, and multi-sensor fusion SLAM systems designed for indoor multi-floor navigation. It further examines mapping, localization, path planning, and autonomous elevator interaction strategies, identifies limitations, and highlights opportunities for future work.

Keywords: Autonomous mobile robots, SLAM, multi-floor navigation, sensor fusion, eleva- tor interaction, LiDAR-inertial odometry, visual-inertial odometry, 3D mapping, autonomous delivery, indoor localization, path planning

1 Introduction

Indoor delivery robots are no longer science fiction—you can already find them moving through hospitals, offices, and university halls, quietly carrying supplies, meals, or mail. But most of these robots have one big limitation: they can only move on a single floor. The moment an elevator or staircase shows up, they stop. That's where multi-floor navigation becomes the next big challenge. A robot that can travel between floors needs to see and understand complex 3D spaces, handle elevators, and stay aware of people moving around it. Traditional Simultaneous Localization and Mapping (SLAM) systems work well on flat ground, but things get messy when the environment changes vertically. Vision-guided SLAM—powered by cameras, LiDAR, and inertial sensors—offers a way forward. It helps robots map their surroundings, plan safe paths, and react to obstacles in real time. When combined with navigation frameworks like ROS 2, it opens the door to truly autonomous delivery systems that can operate across multiple floors with little or no human help. This survey examines how recent research efforts address the challenges of multi-floor indoor navigation. It reviews state-of-the-art advancements in mapping, localization, path planning, and autonomous elevator interaction, all directed toward developing reliable and scalable multi-floor delivery robots. By synthesizing these approaches, the paper identifies current strengths, existing limitations, and key opportunities for future innovation in vision-guided SLAM and navigation frameworks.

2 Literature Review

2.1 SLAM

2.1.1 LiDAR's Foundational Role in SLAM

Initial SLAM frameworks employed 2D laser scan matching with probabilistic pose estimation to build occupancy grid maps. These systems achieved real-time indoor navigation and obstacle avoidance through iterative scan alignment. While simple and stable in structured environments, they were constrained by planar assumptions. Their inability to handle elevation changes or sparse features limited scalability to 3D environments.

Later frameworks enhanced localization by integrating LiDAR intensity information with geometric features. Cylindrical-projection descriptors improved loop closure detection and map alignment in textureless environments. This method strengthened distinctiveness and robustness in feature-scarce areas but increased computational cost and noise sensitivity due to intensity-based feature extraction.

RLI-SLAM introduced a tightly coupled fusion of LiDAR, UWB, and IMU data using factor graph optimization. This multimodal system maintained accurate localization under rapid motion and GPS-denied conditions. It exhibited strong robustness against motion distortion and occlusion but required precise sensor synchronization, adding implementation complexity. Deep neural networks were introduced for loop closure detection and 6-DoF pose estimation directly from LiDAR point clouds. Frameworks like LCDNet improved long-term map consistency by learning feature representations from revisited scenes. Despite adaptability to unstructured environments, their reliance on large datasets and high computational demand constrained

real-time deployment.

FAST-LIO2 employed an iterated Kalman filter (IKF) for direct LiDAR–IMU fusion, avoiding explicit feature extraction. The approach achieved high-speed state estimation and reduced drift using continuous-time LiDAR integration. It delivered robust performance across environments but remained sensitive to calibration errors and LiDAR noise.

Research on cost-effective SLAM adapted Google Cartographer for affordable LiDAR sensors through improved pose graph optimization and scan matching. This approach enabled reliable mapping for educational and small-scale robotics applications. While democratizing SLAM accessibility, it compromised on mapping fidelity and range precision compared to premium systems.

2.1.2 Visual SLAM-Based Systems

ORB-SLAM3 marks a major step forward in feature-based SLAM, combining monocular, stereo, RGB-D, and visual–inertial inputs in a unified framework. Using ORB feature extraction and bundle adjustment, it delivers precise pose estimation and robust loop closure. Its ability to handle multiple maps enables long-term consistency and relocalization across sessions. Although highly accurate and versatile, it requires significant computation and precise sensor synchronization for optimal performance.

Deep-learning based loop closure and CNN-based global descriptors enhance place recognition in repetitive indoor settings, reducing false matches and drift. RGB-D integration and motion segmentation help differentiate static and moving elements in crowded spaces, though they may suffer in poor lighting or with low-quality depth sensors.

LR-SLAM refines visual–inertial SLAM by incorporating both point and line features, improv- ing pose estimation in structured and low-texture scenes. W-VSLAM combines visual–inertial data with wheel odometry to minimize drift in mobile platforms.

Improvements to ORB-SLAM include accelerated preprocessing with quasi-physical sampling and enhanced KD-tree matching, enabling dense 3D reconstruction with sensors like KINECT and maintaining real-time performance. Hybrid systems that integrate object detection (e.g., YOLOv5) to filter dynamic features have also been shown to increase robustness in dynamic scenes.

2.1.3 Multi-Sensor Fusion SLAM Systems

LiDAR–visual fusion approaches have focused on improving data reliability and 3D reconstruction quality by rejecting outliers and boundary-aware rendering. LVI-Fusion provides a tightly integrated solution combining LiDAR, camera, and IMU data using precise synchronization and semantic segmentation to reject dynamic objects. LVI-SAM adopts factor-graph-based tightly-coupled LiDAR, visual, and inertial odometry. FAST-LIVO and FAST-LIVO2 aim for computational efficiency by directly aligning sparse LiDAR points with image and IMU data. LIVER focuses on extreme and low-light conditions using visual enhancement and deep models. RI-LIO augments LiDAR–inertial systems with reflectivity image analysis for improved rotational accuracy and reduced long-term drift.

2.2 Sensor Fusion

This study advances grid-based SLAM by refining Rao-Blackwellized Particle Filters (RBPF). It introduces a novel proposal distribution that effectively integrates robot motion data with the latest sensor observations, significantly reducing state estimation uncertainty. Furthermore, an adaptive resampling mechanism is employed to limit particle degeneration, allowing resampling only when necessary. These improvements not only enhance map accuracy but also reduce com- putational overhead by requiring far fewer particles compared to earlier RBPF approaches. This work presents a monocular visual–inertial SLAM framework optimized for real-time augmented reality (AR) on mobile devices. The system fuses camera and IMU data through a visual–inertial odometry (VIO) framework, supplemented by a fast optical-flow-based visual odometry (VO) module for enhanced responsiveness. A distinctive adaptive execution module dynamically transitions between VIO and fast VO modes based on IMU data variation, achieving an optimal trade-off between computational efficiency and localization accuracy in resource-constrained AR environments.

2.3 Navigation

VINS-Mono introduces a tightly coupled visual–inertial estimation framework utilizing a single monocular camera and a low-cost IMU. The system achieves precise motion estimation by jointly optimizing visual features and inertial constraints, enabling robust performance across various applications such as UAV navigation and mobile robotics. It also supports loop closure and map reuse, ensuring global consistency and long-term localization reliability. This study proposes the Multi-State Constraint Equivariant Filter (MSCEqF), a novel filtering approach for vision-aided inertial navigation that maintains robustness even under poor initialization. Unlike conventional filters, MSCEqF leverages system symmetries encompassing all state components—including IMU biases and camera calibration parameters—resulting in improved consistency, convergence, and resilience to estimation drift in complex navigation scenarios. Forster et al. introduce an on-manifold preintegration framework that significantly improves the computational efficiency of visual–inertial odometry (VIO). By compactly representing a sequence of inertial measurements between keyframes as a single preintegrated constraint, the method reduces processing time while preserving accuracy. Its manifold-based representation also properly handles nonlinear rotational dynamics, leading to superior performance in real-time state estimation tasks. NF-iSAM introduces a non-Gaussian inference framework for SLAM that captures the full posterior distribution of robot states and environmental features, even under highly uncertain or ambiguous conditions. By integrating normalizing flow neural networks into the factor graph optimization process, the approach models complex, multi-modal distributions more accurately than traditional Gaussian-based methods, resulting in enhanced robustness and adaptability to real-world noise and outliers. Direct Sparse Odometry (DSO) is a direct visual odometry method that estimates camera motion and scene geometry directly from image intensity values, bypassing the need for feature extraction or matching. It employs a photometric error minimization strategy over a sparse set of high-gradient pixels, achieving exceptional accuracy and robustness, particularly in low-texture or illumination-varying environments.

2.4 Autonomous Mobile Robots

Recent advancements in autonomous mobile robotics have spanned domains from agricultural automation to industrial testing and urban delivery, each contributing uniquely to navigation, safety, and user acceptance. Tobias et al. (2024) present a modular agricultural robot architecture that leverages off-the-shelf components to reduce development time while enabling effective mapping and navigation in structured, corridor-like farm environments. However, their system remains limited in handling unstructured terrains or complex path planning beyond simple row following. Manikandan and Kaliyaperumal propose a hardware–software framework incorporating a linear actuator-based steering mechanism and a curve-adaptive Pure Pursuit algorithm for both on-road and off-road navigation. While promising, their work offers limited validation in scenarios with dense or dynamic obstacles. Kaiser et al. explore the societal dimension of automation by analyzing consumer acceptance of Autonomous Delivery Robots (ADRs) in last-mile logistics, distinguishing between package and meal delivery contexts. Although they provide valuable behavioral insights, the study leaves open questions about how effort expectancy influences adoption. Wu et al. advance testing methodologies through a Vision-Language-Model (VLM)- driven simulation framework (RVSG) that replicates diverse human interactions to evaluate the functional and safety robustness of industrial AMRs. Despite improving test coverage, the method cannot fully capture the unpredictability of real human behavior in real-world settings. Collectively, these works highlight interdisciplinary progress across design, navigation, human–robot interaction, and testing frameworks, while underscoring the ongoing challenges of unstructured navigation, behavioral uncertainty, and comprehensive performance validation in autonomous mobile robotics.

3 Related Work

3.1 SLAM

Among the reviewed frameworks, FAST-LIO2 represents one of the most balanced and high-performance LiDAR-based SLAM systems. By employing a direct LiDAR–IMU fusion strategy through an iterated Kalman filter (IKF), it eliminates the need for explicit feature extraction, significantly reducing computational latency while maintaining high localization accuracy. Its continuous-time LiDAR integration ensures robust pose estimation and minimal drift across both indoor and outdoor environments. This approach effectively addresses the trade-off between accuracy and real-time operation seen in earlier SLAM systems. Despite its reliance on precise sensor calibration, FAST-LIO2 remains a benchmark solution for achieving efficient, reliable, and real-time LiDAR-inertial odometry in modern autonomous navigation.

3.2 Navigation

VINS-Mono demonstrates exceptional precision and robustness by tightly coupling monocular vision with inertial measurements, ensuring consistent localization and long-term map reuse for mobile and aerial robotics. The MSCEqF filter further strengthens state estimation stability by leveraging full-system symmetries, maintaining reliable performance even under poor ini- tializations. On-Manifold Preintegration improves computational efficiency in visual–inertial odometry by compactly integrating IMU data while accurately handling nonlinear rotational dynamics, enabling real-time operation without compromising accuracy. Meanwhile, NF-iSAM advances SLAM inference by incorporating normalizing flow networks to capture non-Gaussian uncertainties, enhancing adaptability in complex or dynamic environments. Finally, Direct Sparse Odometry (DSO) achieves high accuracy through photometric optimization on raw image intensities, delivering reliable motion estimation under challenging visual conditions. Collectively, these approaches mark a shift toward more precise, efficient, and resilient SLAM frameworks, with VINS-Mono and On-Manifold Preintegration offering the most balanced trade-off between accuracy and real-time performance.

3.3 Autonomous Mobile Robots

Recent advancements in autonomous mobile systems span agricultural, industrial, and urban domains. A cost-efficient agricultural platform demonstrates robust localization, navigation, and mapping in corridor-like environments such as vineyards, showcasing scalability for precision farming. Research on Autonomous Delivery Robots (ADRs) highlights increasing societal acceptance for last-mile logistics and explores their potential in emergency and healthcare operations. In industrial settings, the RVSG Vision–Language Model (VLM) framework enhances safety validation by simulating complex human–robot interactions for Autonomous Mobile Robots (AMRs). Additionally, a hybrid on/off-road navigation system employing a Linear Actuator steering mechanism and Curve-aware Pure Pursuit (C-PP) algorithm improves trajectory tracking and control across varied terrains. Collectively, these studies underscore the role of multi-sensor fusion, adaptive control, and human-centered design in advancing reliable and efficient autonomous mobility.

3.4 Sensor Fusion

Recent advancements in Simultaneous Localization and Mapping (SLAM) research have focused on enhancing the efficiency and precision of grid-based and visual–inertial systems. A refined Rao-Blackwellized Particle Filter (RBPF) approach introduces an improved proposal distribution that effectively combines robot motion and sensor data to minimize state estimation uncertainty, supported by an adaptive resampling strategy that reduces particle degeneration and compu- tational load. Complementing this, a monocular visual–inertial SLAM framework has been developed for real-time augmented reality (AR) applications on mobile platforms, integrating camera and IMU inputs via a visual–inertial odometry (VIO) scheme. The framework's adaptive execution module intelligently alternates between VIO and optical-flow-based visual odometry (VO) modes, achieving an optimal balance between accuracy and computational efficiency in resource-constrained AR environments.

4 Proposed System

Our proposed system consists of two major components: the Delivery Bot and the mobile application that manages the operation for placing the deliveries. Together, these components provide end-to-end autonomous deliveries within multi-floor apartment buildings by using intelligent navigation, obstacle avoidance, mapping and interaction with the elevator.

4.1 Delivery Bot

To achieve reliable state estimation, the delivery robot employs a deeply integrated sensor fusion framework, initiated by an Extended Kalman Filter (EKF). The EKF provides a high- frequency ego-motion estimate by fusing the proprioceptive sensors (wheel odometry and IMU). This estimate serves as a robust motion prior for the primary graph-based SLAM algorithm, LIO-SAM (LiDAR-Inertial Odometry via Smoothing and Mapping), which leverages the 360° LiDAR data to build precise, geometrically-consistent maps. To correct long-term drift and handle perceptual ambiguity, a parallel visual place recognition module leverages the forward-facing stereo camera—using ORB feature extraction and a DBoW2 vocabulary—to provide high-confidence loop closure and relocalization constraints to the LIO-SAM graph. This localization-aware foundation enables a five-phase operational workflow, encompassing Mission Setup, elevator interaction, intra-floor navigation, and drop-off Confirmation.

4.1.1 Mission Setup and Navigation

When the delivery request is placed, the delivery bot receives the sender's information through the app. It begins the mission from a particular docking station, first the mission path is set to go to the dispatcher. The intra-floor navigation process is handled through a tightly coupled visual–inertial SLAM framework implemented using ORB-SLAM3. The system employs keypoint- based feature extraction via Oriented FAST and Rotated BRIEF (ORB) descriptors, allowing robust frame-to-frame tracking under varying illumination and texture conditions. Each camera frame contributes to a pose graph, optimized using nonlinear least-squares minimization (g2o) to reduce the overall reprojection error across keyframes. To enhance localization robustness, visual odometry is fused with Inertial Measurement Unit (IMU) and wheel odometry data through an Extended Kalman Filter (EKF) or factor graph–based fusion using frameworks such as GTSAM. This fusion ensures accurate pose estimation during short-term visual loss, such as when entering low-light or reflective areas near the elevator zone. The 3D point cloud map generated by ORB-SLAM3 is projected into a 2D occupancy grid, which is subsequently fused with LiDAR-based mapping using a Normal Distributions Transform (NDT) registration. The resulting hybrid map serves as the input for the global planner, which computes an optimal route using A*, Theta*, or D* Lite algorithms, depending on environmental constraints and computation time. The local planner—typically implemented via Dynamic Window Approach (DWA) or Timed Elastic Band (TEB) controllers—generates smooth velocity profiles while maintaining obstacle avoidance through a layered costmap architecture. Dynamic obstacle tracking is achieved using spatio-temporal costmap updates, allowing the robot to dynamically replan its path in real time. The navigation stack continuously publishes pose estimates, velocity commands, and costmap updates over the ROS 2 DDS middleware to ensure deterministic communication and modular scalability across navigation nodes.

4.1.2 Elevator Interaction

The elevator interaction module enables autonomous vertical mobility by integrating hardware interfacing and decision-making within a dedicated ROS 2 node. It communicates with the building's Elevator Control Unit (ECU) through an IoT bridge using MQTT or RESTful APIs. Upon reaching the elevator lobby, the robot retrieves the real-time state of all elevators and evaluates a multi-objective cost function that considers floor proximity, arrival time, and occupancy. The elevator with the lowest cost is selected for boarding. Visual alignment is achieved using AprilTag or ArUco markers, refined through the Iterative Closest Point (ICP) algorithm with LiDAR data for sub-5 cm accuracy. The robot then sends a boarding command and docks securely within the elevator compartment. During transit, SLAM is suspended, and IMU-based dead reckoning maintains orientation. On arrival, ORB-SLAM3's relocalization restores the 6-DoF pose, allowing seamless reinitialization of navigation for continued intra-floor movement.

4.1.3 Delivery Completion and Handover

Upon reaching the delivery zone, the robot refines its localization using ORB-SLAM3's relocaliza- tion module, aligning real-time ORB features with stored keyframes and optimizing poses via g2o. LiDAR-based GICP registration ensures centimeter-level docking accuracy by matching the local point cloud to the static map. The local trajectory planner (TEB or DWA) generates smooth, time-optimal velocity profiles for final approach, governed by a finite state machine managing approach, alignment, verification, handover, and departure. AprilTag detection with PID-based control enables precise docking. Delivery is authorized only after OTP-based verification: a cloud-generated OTP is validated using a SHA-256 token within the ROS 2 authentication node. Once verified, a payload-release flag triggers servo actuation, logging mission data to the control node. Post-handover, the robot initiates a Return-to-Base routine, where a hybrid A* planner computes an energy-efficient route minimizing distance and battery cost, ensuring autonomous return and data synchronization.

4.2 Application System

The mobile application links users and the bot. Sender inputs location and recipient credentials; recipient confirms floor and room. The backend initializes delivery tasks and dispatches the nearest available bot. The bot queries the ECU, updates maps, and streams telemetry over ROS 2–MQTT bridges. OTP-based verification ensures secure handover; mission data syncs to the cloud.

5 Challenges and Issues in Multi-level Indoor Navigation

Achieving reliable autonomy in multi-floor and multi-building indoor environments introduces a spectrum of spatial, computational, and communication-level challenges.

5.1 Inter-robot Coordination and Cross-floor Task Orchestration

The second challenge domain concerns inter-robot coordination and cross-floor task orchestration. When multiple autonomous delivery units operate concurrently, ROS 2 DDS middleware can experience latency jitter, message queuing delays, and inconsistent Quality of Service (QoS) parameters, leading to task desynchronization. Implementing a real-time distributed task allocation framework requires deterministic scheduling under asynchronous communication and network congestion, especially in elevator handovers or docking zones shared across floors. Furthermore, the absence of a unified spatio-temporal coordination protocol for multi-agent navigation causes difficulties in maintaining global state awareness, which can result in race conditions or deadlocks in dense multi-floor environments.

5.2 Dynamic Path Reconfiguration and Environment Modeling

A third significant challenge arises in dynamic path reconfiguration and environment modeling. Multi-level indoor structures exhibit high environmental entropy—frequent human interference, moving obstacles, and dynamically changing layouts—causing traditional DWA (Dynamic Window Approach) or TEB (Timed Elastic Band) planners to degrade under uncertainty. The integration of semantic mapping using CNN-based perception pipelines can enhance contextual awareness but introduces high computational overhead and inference latency. Maintaining real- time costmap updates across multiple levels requires efficient spatial partitioning and layer-based occupancy grid merging, which becomes computationally intensive when performed on edge- limited onboard processors. Moreover, vertical transition zones (e.g., elevators, escalators) require specialized motion primitives and finite state machine (FSM) logic for state synchronization and collision-free maneuvering.

5.3 System-level Robustness and Energy-aware Control

Finally, system-level robustness and energy-aware control remain persistent challenges in multi- level autonomy. Precision docking and payload transfer demand sub-centimeter accuracy from ultrasonic and vision-based proximity sensors, which are highly susceptible to multipath inter- ference and lighting variability. Mechanical misalignments during lift-based payload transfers introduce calibration drift in the kinematic chain, necessitating self-correcting control loops or PID-adaptive controllers for alignment compensation. Additionally, energy-aware mission planning is constrained by limited onboard computation and power, making predictive battery scheduling and fault-tolerant recovery mechanisms essential to prevent mission dropouts. Inte- gration of modular payload systems further complicates sensor calibration, requiring real-time inertial frame re-registration upon module reattachment.

5.4 Summary

Collectively, these challenges highlight the need for multi-modal sensor fusion architectures, low-latency edge computing nodes, and federated SLAM frameworks that can handle spatial hierarchy, distributed map consistency, and real-time cooperative autonomy across interconnected smart building environments.

6 Future Scope

Future developments will enhance scalability, multi-floor autonomy, and cross-building logistics through layered navigation. A multi-floor delivery mechanism enables the payload compartment to transfer between floors while the base stays docked, using a modular lift-transfer algorithm with RFID identification, visual features, and ultrasonic sensing for alignment and docking.

Docking stations synchronize with a central controller via ROS 2 DDS for task scheduling, battery monitoring, and positional handover. The elevator or transfer module operates via an FSM controlling queueing, lift occupancy detection, and inter-floor path confirmation, ensuring reliable vertical logistics.

A hierarchical multi-agent network includes an Outdoor Delivery Bot (ODB) and Indoor Delivery Bots (IDBs). The ODB uses GPS-aided SLAM with RTK-GNSS, Hybrid-A* planning, and LiDAR+IMU+vision fusion. At building hubs, it hands off via encrypted Wi-Fi/5G with QR/RFID-based authentication.

IDBs use AMCL, costmaps, and DWA/TEB planning for corridor navigation. Future research explores cloud-based fleet coordination using distributed SLAM, federated learning, edge computing near docking stations, and multi-robot task allocation (Hungarian/auction- based), enabling scalable indoor-outdoor logistics.

7 Conclusion

This literature review examines recent advancements in SLAM-based navigation frameworks that support autonomous indoor delivery systems, particularly those operating across multiple floors. State-of-the-art solutions such as ORB-SLAM3, FAST-LIO2, and hybrid LiDAR–visual–inertial systems have improved localization, mapping accuracy, and trajectory tracking in cluttered indoor environments. With the integration of deep learning–based perception, semantic segmentation, and dynamic object detection, modern SLAM systems can better differentiate static and moving obstacles, thereby enhancing situational awareness and map stability.

Despite these advancements, significant research gaps persist. Most existing works focus on single-floor, planar navigation and lack robust mechanisms for vertical mobility, elevator integration, or reliable localization across building levels without GPS. Current systems also struggle with real-time multi-sensor fusion under lighting or texture variations, leading to degraded performance. Additionally, computational overhead and power consumption pose challenges for resource-constrained delivery robots deployed in large buildings.

Future work must emphasize lightweight, adaptive, and multi-modal SLAM architectures capable of long-term operation in dynamic, multi-floor spaces. Integrating semantic reasoning with geometric understanding may enable more efficient planning and decision-making. Address- ing these challenges will be key to realizing scalable indoor delivery solutions for smart buildings, campuses, and healthcare environments.

References

1. Piao, J.-C., Kim, S.-D. (2017). Adaptive Monocular Visual–Inertial SLAM for Real-Time AR Applications. *Sensors*, 17(11), 2567.
2. Zhi, C., Xiumin, S. (2019). Research on Cartographer Algorithm based on Low Cost Lidar. *IJERT*, 8(10).
3. Zhang, X., Zhang, H., Qian, C., Li, B., Cao, Y. (2023). A LiDAR-intensity SLAM and loop closure detection method using cylindrical-projection shape context descriptor. *Journal of Advanced Research*, 153, 103419.
4. Xiang, Z., Guo, J., Meng, J., et al. (2024). Accurate localization of indoor high similarity scenes using visual SLAM combined with loop closure detection algorithm. *PLOS ONE*, 19(12), e0312358.
5. Qin, T., Li, P., Shen, S. (2018). VINS-Mono: A Robust and Versatile Monocular Visual- Inertial State Estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020.
6. Kaiser, R., De Benedetto, S., Planing, P., Müller, P. (2024). What Will the Delivery Robots Bring Us Tomorrow?
7. Huang, Q., Pu, C., Khosoussi, K., et al. (2023). Incremental non-Gaussian inference for SLAM using normalizing flows. *IEEE Transactions on Robotics*, 39(2), 1458–1475.
8. Saat, S., Rashid, W.N.A., Tumari, M.Z.M., Saealal, M.S. (2020). HECTORSLAM 2D MAPPING FOR SIMULTANEOUS LOCALIZATION AND MAPPING (SLAM). *Journal of Physics: Conference Series*, 1529(4), 042032.
9. Luo, D., Huang, Y., Huang, X., Miao, M., Gao, X. (2024). W-VSLAM: A Visual Mapping Algorithm for Indoor Inspection Robots. *Sensors*, 24(17), 5662.
10. Cattaneo, D., Vaghi, M., Valada, A. (2022). LCDNet: Deep Loop Closure Detection and Point Cloud Registration for LiDAR SLAM. In *2022 IEEE ICRA*, 1–7. IEEE.
11. Jiang, H., Cang, N., Lin, Y., Guo, D., Zhang, W. (2024). LR-SLAM: Visual Inertial SLAM System with Redundant Line Feature Elimination. *Journal of Intelligent & Robotic Systems*, 110(169).
12. Wen, T., Fang, Y., Lu, B., et al. (2024). LIVER: A Tightly Coupled LiDAR-Inertial-Visual State Estimator With High Robustness for Underground Environments. *IEEE RA-L*, 9(3), 2399–2406.
13. Liu, Z., Li, Z., Liu, A., et al. (2024). LVI-Fusion: A Robust Lidar-Visual-Inertial SLAM Scheme. *Remote Sensing*, 16(9), 1524.

14. Shan, T., Englot, B., Ratti, C., Rus, D. (2021). LVI-SAM: Tightly-coupled Lidar-Visual- Inertial Odometry via Smoothing and Mapping. In *2021 IEEE ICRA*, 5692–5698.
15. Xu, W., Cai, Y., He, D., Lin, J., Zhang, F. (2021). FAST-LIO2: Fast Direct LiDAR-inertial Odometry. In *2021 IEEE/RSJ IROS*, 1–8.
16. Zheng, C., Zhu, Q., Xu, W., et al. (2022). FAST-LIVO: Fast and Tightly-Coupled Sparse- Direct LiDAR-Inertial-Visual Odometry. In *2022 IEEE/RSJ IROS*, 1–8.
17. Geng, M., Jia, Y., Qian, X., et al. (2023). Research on Multi-Sensor Fusion SLAM Algorithm Based on Outlier Rejection and Boundary Rendering. In *2023 ICCC*, 1–6. IEEE.
18. Zhang, Y., Tian, Y., Wang, W., et al. (2023). RI-LIO: Reflectivity Image Assisted Tightly- Coupled LiDAR-Inertial Odometry. *IEEE RA-L*, 8(3), 1802–1809.
19. Kuang, H., Wang, X., Liu, X., Ma, X., Li, R. (2017). An Improved Robot's Localization and Mapping Method Based on ORB-SLAM. In *2017 10th ISCID*.
20. Campos, C., Elvira, R., Rodrıguez, J.J.G., Montiel, J.M.M., Tardós, J.D. (2021). ORB- SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM. *IEEE Transactions on Robotics*, 37(6), 1874–1890.
21. Xin, R., Guo, N., Ma, X., Liu, G., Feng, Z. (2024). RLI-SLAM: Fast Robust Ranging- LiDAR-Inertial Tightly-Coupled Localization and Mapping. *Sensors*, 24(17), 5672.
22. Ebadi, K., Bernreiter, L., Biggie, H., et al. (2022). Present and Future of SLAM in Extreme Underground Environments. *IEEE RA-L*, 7(2), 3338–3345.
23. Engel, J., Koltun, V., Cremers, D. (2018). Direct Sparse Odometry. *IEEE TPAMI*, 40(3), 611–625.
24. Jiao, Y., Wang, Z. (2019). Stereo vision SLAM system based on ORB. In *IAEAC 2019*.
25. Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D. (2017). On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1), 1–21.
26. Fornasier, A., van Goor, P., Allak, E., Mahony, R., Weiss, S. (2024). MSCEqF: A Multi State Constraint Equivariant Filter for Vision-Aided Inertial Navigation. *IEEE RA-L*, 9(1), 731–738.
27. Tobias, J., Lim, S.H., Duke, M., McGuinness, B., Au, C.K. (2024). Implementation of an autonomous mobile platform for agricultural tasks in corridor-like environments.
28. Wu, J., Lu, C., Arrieta, A., Ali, S., Peyrucain, T. Vision Language Model-based Testing of Industrial Autonomous Mobile Robots.
29. Manikandan, N.S., Kaliyaperumal, G. (2023). Hardware and Software Design for Mobile Robot Navigation over On-road and Off-road Curvature Paths.
30. Li, J., Dai, J., Su, Z., Zhu, C. (2024). RGB-D Based Visual SLAM Algorithm for Indoor Crowd Environment. *Journal of Intelligent & Robotic Systems*, 110(27).
31. Zhou, W., Meng, X., Liu, Z., Zhang, X., Wang, S., Song, Y. (2024). Enhancing ORB-SLAM3 Pose Estimation in Dynamic Scenes with YOLOv5 Object Detection. In *2024 ICIPMC*.