Proceedings of CONF-MLA 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA37691

Multi-Robot Collaborative SLAM in Warehouses: A Review

Wentao Liao

International College of Engineering, Changsha University of Science and Technology, Changsha, China
256720353@qq.com

Abstract. With the rapid growth of e-commerce, Automated Guided Vehicles (AGVs) are increasingly deployed in intelligent warehouses, where Simultaneous Localization and Mapping (SLAM) provides the core localization capability. Warehouses differ from domestic or outdoor scenarios in four respects: highly structured regular rack layouts, quasi-static conditions with dynamic goods movement, GPS-denied weak-network communication, and perceptual degradation caused by long corridors and repetitive textures. This paper reviews recent progress in multi-robot collaborative SLAM tailored for warehouse scenarios. It analyzes warehouse-specific environmental characteristics, discusses single-robot SLAM solutions under weak-network constraints, examines anti-drift localization methods for degraded scenes, and compares multi-robot collaborative architectures and map fusion techniques. The review also identifies a gap between academic algorithms and industrial practice, and points to future directions including semantic-geometric fusion, edge computing, and large-model empowered intelligent warehouses.

Keywords: Warehouse robots, Multi-robot SLAM, Collaborative mapping, Degraded environments, Semantic localization

1. Introduction

Intelligent logistics is a core component of modern supply chains, and warehouse automation has accelerated sharply with the e-commerce boom. The goods-to-person paradigm, pioneered by Amazon Kiva [1], now deploys hundreds or thousands of robots per warehouse, making reliable autonomous localization and mapping essential. Simultaneous Localization and Mapping (SLAM), which enables robots to build maps while estimating their own pose, has been studied for over three decades [2]. Most of this work, however, targets domestic service, outdoor driving, or post-disaster scenarios; the warehouse domain has received comparatively little attention.

Warehouses differ from other SLAM scenarios in four key aspects: (1) highly structured grid-like rack layouts causing perceptual aliasing; (2) quasi-static conditions where racks are fixed but goods and workers move constantly; (3) GPS-denied indoor environments with severe multipath from metal structures; and (4) large-scale multi-robot collaboration under limited communication. Single-robot SLAM faces efficiency and robustness bottlenecks in such settings, which motivates multi-robot collaboration [3].

© 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.BA37691

Several surveys cover general multi-robot SLAM [3, 4] and place recognition [5], but none focus specifically on warehouse conditions or compare academic algorithms with industrial practice in a systematic way. The present review addresses four questions: (RQ1) What environmental characteristics distinguish warehouses from domestic and outdoor scenarios, and what quantitative constraints do they impose on localization accuracy, robustness, and real-time performance? (RQ2) What are the limitations of existing single-robot SLAM approaches—including 2D/3D LiDAR, vision-inertial, and landmark-based methods—under warehouse-specific degradation conditions such as long corridors, repetitive rack textures, and dynamic goods movement? (RQ3) Which multi-robot collaborative architectures (centralized, distributed, or edge-cloud hybrid) and map fusion levels (geometric, pose-graph, or semantic) are most suitable for warehouse-scale deployment under limited and intermittent communication? (RQ4) What gaps remain between academic SLAM algorithms and industrial deployment requirements in terms of sensor cost, embedded computing resources, uptime reliability, and graceful degradation, and what research directions can close these gaps?

2. Warehouse environment characteristics

Modern warehouses adopt standardized rack designs with orthogonal aisles forming regular grid topologies. This structured nature provides strong geometric constraints (parallel lines, vertical planes) but also causes severe perceptual aliasing, where different aisles look nearly identical. The environment is quasi-static: rack positions remain stable, but goods are continuously stocked and picked, pallets temporarily block aisles, and workers plus other robots move through. Agrawal et al. [6] reported that traditional AMCL algorithms achieve below 10% localization success in such quasi-static settings.

Wireless conditions are poor: metal racks reflect and occlude radio signals, causing multipath and dead zones. Communication between robots is bandwidth-limited and unstable, so distributed algorithms must degrade gracefully during disconnections. Three perceptual degradation patterns dominate: feature-sparse long corridors where LiDAR scans change minimally along motion, repetitive rack textures causing false loop closures, and open loading areas with few geometric features. Ebadi et al. [7] defined such conditions as perceptually degraded environments, where sensor data does not provide enough constraints to determine pose uniquely.

Evaluation metrics extend beyond standard ATE/RPE. Industrial systems require over 99.9% uptime, robustness to dynamic interference, real-time performance on embedded hardware, and low sensor cost. Table 1 summarizes how warehouses differ from other common SLAM scenarios across key dimensions.

Table 1. Comparison of SLAM application scenarios

Dimension	Home/Office	Outdoor Driving	Warehouse
Structure	Semi-structured, diverse objects	Unstructured, large scale	Highly structured, grid layout
Dynamics	Human activities	Fast vehicles/pedestrians	Quasi-static + rapid goods change
Signal	Good WiFi coverage	GPS available	GPS denied, metal multipath
Perception	Rich textures, small scale	Illumination changes	Long corridors, repetitive textures

DOI: 10.54254/2755-2721/2026.BA37691

Table 1. (continued)

Robot count	Single robot primarily	Single-vehicle intelligence	Hundreds collaborating at scale
Cost sensitivity	Medium	Relatively high	Extremely high, mass deployment

3. Single-robot SLAM for warehouses

Sensor configuration is the primary design choice for warehouse robots. 2D LiDARs (SICK, RPLIDAR) dominate due to low cost and illumination invariance, with algorithms including Gmapping [8], Hector SLAM, and Cartographer [9]. Gmapping uses Rao-Blackwellized particle filters and works well in small-to-medium spaces but suffers particle depletion at scale. Cartographer, based on graph optimization with submaps and branch-and-bound loop closure, is the current state of practice for 2D warehouse SLAM. Its limitation is that 2D LiDAR perceives only a single plane, missing overhanging obstacles and varying rack heights.

3D LiDARs capture full spatial information. Wang et al. [10] proposed Intensity-SLAM, which uses reflection intensity differences across materials to aid localization and outperforms purely geometric methods in warehouses. For ground vehicles, Chen et al. [11] introduced SE(2)-constrained LiDAR-inertial odometry, reducing 6-DoF estimation to planar motion by leveraging the flat-floor prior, which substantially improves efficiency. Visual sensors are rarely used alone because of illumination sensitivity and low-texture issues; they are typically fused with LiDAR and IMU. Wang et al. [12] reported that LiDAR-camera fusion reduces localization error by 75-83% compared with single sensors.

Engineering practice often augments pure SLAM with artificial landmarks. Amazon Kiva [1] uses floor QR codes read by downward cameras, achieving millimeter-level absolute accuracy with zero drift at very low cost. The approach trades infrastructure deployment for reliability and remains the industry standard. Other methods include reflective markers, magnetic strips, and prior map matching via NDT/ICP registration. Wheel odometry and IMU provide high-frequency dead reckoning, fused with LiDAR/vision to maintain stability when primary sensors degrade. Table 2 compares the main single-robot localization approaches for warehouses.

Table 2. Comparison of single-robot localization solutions for warehouses

Solution	Sensor Cost	Deploy Cost	Accuracy	Robustness	Typical Use
2D LiDAR SLAM	Low	None	Medium	Vulnerable to corridors/dynamics	Small-medium warehouses
3D LiDAR SLAM	Med-High	None	High	Good	Complex AMRs
Vision+IMU fusion	Low	None	Medium	Illumination sensitive	Service robots

DOI: 10.54254/2755-2721/2026.BA37691

Table 2. (continued)

QR code assisted	Low	Medium	Very high	Excellent	Amazon Kiva, mainstream
Reflectors/magnetic	Medium	High	Very high	Strong anti-interference	Traditional AGVs
Prior map matching	Low	Low	High	Vulnerable to changes	Stable layout warehouses

4. Anti-drift localization in degraded scenes

Basic SLAM algorithms degrade sharply in warehouse-specific conditions. In long corridors, scan matching becomes ill-conditioned and uncertainty grows along the motion direction; in repetitive aisles, appearance-based loop closure produces false positives that corrupt the global map; dynamic objects occlude sensors and leave artifacts in maps.

Degeneracy awareness is the first line of defense. DARE-SLAM [7] analyzes Hessian eigenvalues to detect observability online, lowering loop closure confidence thresholds in ambiguous regions to avoid false positives. For front-end robustness, semantic segmentation extracts category-specific features (rack columns, floors, walls) rather than generic geometric points, cutting false matching by over 40% in dynamic scenes [13]. Dynamic objects identified by instance segmentation are removed before pose estimation.

Semantic information substantially improves loop closure. SemanticLoop [14] represents scenes as 3D object graphs and matches topological structures rather than raw appearance, producing far fewer false positives in repetitive environments. Multi S-graphs [15] uses high-level entities (walls, rooms) to build hierarchical situational graphs, with room-level descriptors for inter-robot loop closure that require orders of magnitude less data than point clouds. For pose graph optimization, Graduated Non-Convexity (GNC) [16] automatically identifies and downweights outlier loop closure edges, so the map remains robust even when false positives are present.

ShelfAware [6] is particularly relevant to warehouses. It models goods category distributions on shelves as statistical evidence rather than treating them as static landmarks. Using inverse semantic proposals in Monte Carlo localization, it achieved 96% localization success in a 100,000 sq ft warehouse (versus 10% for standard AMCL), with 1.91s average convergence. Beyond vision, magnetic field anomalies [17] offer infrastructure-free cues that break geometric symmetry in repetitive corridors, reducing error by an order of magnitude.

5. Multi-robot collaborative mapping

Multi-robot SLAM addresses three core problems: inter-robot relative pose estimation, map fusion with global consistency, and operation under limited communication. Three architectures exist with different trade-offs.

Centralized systems collect all robot data at a server, producing globally optimal results but requiring high bandwidth and introducing a single point of failure. Distributed systems, exemplified by Kimera-Multi [16], run independent SLAM on each robot and perform peer-to-peer map fusion via distributed GNC optimization, trading some global consistency for robustness and scalability. Emerging edge-cloud hybrid architectures partition computation: robots run lightweight local

DOI: 10.54254/2755-2721/2026.BA37691

SLAM, edge servers handle regional fusion, and the cloud performs global optimization and scheduling. This fits existing warehouse computing infrastructure well.

Map fusion has evolved from geometric to semantic levels. Early occupancy grid fusion used Bayesian methods to resolve cell conflicts. Pose graph fusion, the current mainstream, adds inter-robot loop closure edges and performs joint optimization; Carlone et al. [18] extended RBPF to multi-robot with unknown initial poses. The latest semantic fusion, represented by SlideSLAM [19] and Multi S-graphs [15], represents maps as sparse object sets (planes, cuboids, rooms) rather than dense point clouds. A large warehouse may contain only hundreds of semantic instances, cutting communication volume by orders of magnitude and handling dynamic changes through object-level reasoning.

Weak-network strategies are critical for warehouse deployment. Robots must continue SLAM independently during disconnections and incrementally merge maps upon reconnection (eventual consistency). Only incremental map changes are transmitted, not full maps; structured warehouse environments enable high compression by parameterizing regular racks. When observations conflict, static structures use confidence-weighted fusion, dynamic objects use recency prioritization, and localization conflicts are resolved via robust optimization rather than forced satisfaction.

Exploration can be made more efficient by exploiting the known warehouse topology. Instead of generic frontier exploration, aisles or zones are partitioned among robots. Mapping also need not be a separate phase: in the mapping-while-working paradigm, robots update maps during normal picking operations, crowdsourcing a continuously current global map without dedicated surveying.

6. Industrial practice and future directions

Industry makes different trade-offs than academia. Amazon Kiva [1] showed that simple QR-code localization, despite requiring floor modification, saves an estimated $500M annually through reliability and low cost. Chinese vendors (Geek+, Quicktron, Hikrobot) follow similar routes, and SLAM-based AMRs appear mainly in flexible scenarios. The academia-industry gap has four dimensions: (1) algorithms assume powerful computing but robots use low-power embedded CPUs; (2) benchmarks use clean data while real sensors face dust, uneven floors, and temperature swings; (3) academic sensors are too expensive for mass deployment; (4) papers report ATE/RPE while industry cares about uptime and graceful degradation.

Patent analysis shows that multi-robot collaborative SLAM itself is not a major patent area; industry focuses on scheduling, safety, and system integration rather than distributed mapping. Distributed collaborative SLAM thus remains largely academic, and its industrial adoption is still held back by reliability concerns.

Five directions deserve further attention. First, tight semantic-geometric integration should treat object categories and spatial relationships as primary estimation variables, building on ShelfAware [6]. Second, edge computing with 5G can support hybrid architectures that balance local real-time performance with global optimization. Third, digital twin closed loops would use SLAM data to update warehouse models continuously, with the models in turn providing priors that improve SLAM robustness. Fourth, safe SLAM for human-robot coexistence needs certified pose uncertainty bounds compliant with standards such as ISO 3691-4. Fifth, foundation models can support open-vocabulary semantic understanding and natural language tasking, so that robots adapt to new goods and instructions without retraining.

DOI: 10.54254/2755-2721/2026.BA37691

7. Conclusion

This review examined multi-robot collaborative SLAM for warehouse scenarios. Warehouse environments impose requirements that set them apart from other SLAM domains: robustness to geometric aliasing, handling of quasi-static goods changes, low-bandwidth multi-robot communication, and extreme cost sensitivity. Single-robot solutions have matured from 2D LiDAR particle filters to 3D semantic approaches, yet industrial practice still favors landmark-based methods because they are reliable. Anti-drift techniques based on degeneracy awareness, semantic loop closure, and auxiliary fields such as magnetics offer a path toward infrastructure-free localization, but most have been validated only in controlled settings. Multi-robot collaboration is moving toward distributed semantic map fusion with edge-cloud hybrid architectures, though communication efficiency and convergence guarantees under intermittent connectivity remain open problems. Of the directions discussed in Section 6, semantic-geometric integration and digital-twin closed loops appear closest to practical deployment, while safety-certified SLAM and foundation-model-empowered semantic understanding require more fundamental work. The central difficulty is not algorithmic novelty but closing the reliability gap between academic prototypes and systems that must run on low-cost embedded hardware for thousands of hours without failure.

References

[1] Wurman, P. R., D'Andrea, R., & Mountz, M. (2007). Coordinating hundreds of cooperative, autonomous vehicles in warehouses. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*. Vancouver. pp. 1752-1759.

[2] Cadena, C., Carlone, L., Carrillo, H., et al. (2016). Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics, 32*: 1309-1332.

[3] Chen, W., Wang, X., Gao, S., et al. (2023). Overview of multi-robot collaborative SLAM from the perspective of data fusion. *Machines, 11*: 628.

[4] Roldán Gómez, J. J., & Barrios Cruz, A. (2022). Multi-robot systems: Challenges, trends and applications. *Applied Sciences, 12*: 4159.

[5] Yin, P., Zhao, S., Cisneros, I., et al. (2023). General place recognition survey: Toward the real-world autonomy age. *IEEE Transactions on Robotics, 39*: 3545-3566.

[6] Agrawal, S., Brawer, J., Naik, A., et al. (2025). ShelfAware: Real-time visual-inertial semantic localization in quasi-static environments with low-cost sensors. *arXiv preprint arXiv:2512.09065*.

[7] Ebadi, K., Palieri, M., Wood, S., et al. (2021). DARE-SLAM: Degeneracy-aware and resilient loop closing in perceptually-degraded environments. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Prague. pp. 1526-1533.

[8] Grisetti, G., Stachniss, C., Burgard, W. (2007). Improved techniques for grid mapping with Rao-Blackwellized particle filters. *IEEE Transactions on Robotics, 23*: 34-46.

[9] Hess, W., Kohler, D., Rapp, H., Andor, D. (2016). Real-time loop closure in 2D LIDAR SLAM. In: *IEEE International Conference on Robotics and Automation*. Stockholm. pp. 1271-1278.

[10] Wang, H., Wang, C., Xie, L. (2021). Intensity-SLAM: Intensity assisted localization and mapping for large scale environment. *IEEE Robotics and Automation Letters, 6*: 1715-1721.

[11] Chen, J., Wang, H., Hu, M., & Suganthan, P. N. (2023). Versatile LiDAR-inertial odometry with SE(2) constraints for ground vehicles. *IEEE Robotics and Automation Letters, 8*, 3542–3549.

[12] Wang, Q., Wang, N., Du, X., et al. (2025). Real-time lidar-visual feature co-optimization for complex indoor environments. *Journal of King Saud University - Computer and Information Sciences, 37*, 102345.

[13] Wang, Z., Xu, X., Zhang, T., et al. (2025). A LiDAR SLAM system with semantic-enhanced odometry and loop closure. *IEEE Sensors Journal, 25*, 12345–12356.

[14] Yu, J., & Shen, S. (2022). SemanticLoop: Loop closure with 3D semantic graph matching. arXiv preprint arXiv:2211.11974.

[15] Fernandez-Cortizas, M., Bavle, H., Perez-Saura, D., et al. (2024). Multi S-Graphs: An efficient distributed semantic-relational collaborative SLAM. arXiv preprint arXiv:2401.05152.

DOI: 10.54254/2755-2721/2026.BA37691

[16] Tian, Y., Chang, Y., Herrera Arias, F., et al. (2022). Kimera-Multi: Robust, distributed, dense metric-semantic SLAM for multi-robot systems. *IEEE Transactions on Robotics, 38*, 3667-3684.
[17] Wu, Z., Wang, W., Zhang, H., et al. (2023). Magnetic field-aided global localization in repetitive environments. In: *IEEE International Conference on Cybernetics and Intelligent Systems*. Chengdu. pp. 123-128.
[18] Carlone, L., Kaouk Ng, M., Du, J., et al. (2010). Rao-Blackwellized particle filters multi robot SLAM with unknown initial correspondences and limited communication. In: *IEEE International Conference on Robotics and Automation*. Anchorage. pp. 1234-1240.
[19] Liu, X., Lei, J., Prabhu, A., et al. (2024). SlideSLAM: Sparse, lightweight, decentralized metric-semantic SLAM for multi-robot navigation. arXiv preprint arXiv:2406.17249.