Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD29174

Exploring Multi-Sensor Fusion Navigation Models for Delivery Robots in Food Delivery Scenarios

Yuhan Liu

Xi'an GaoXin No.1 High School, Xi'an, China
18966608707@163.com

Abstract. The rapid growth of the food delivery industry has intensified the "last mile" challenge, where efficiency and safety remain bottlenecks. Single-sensor navigation methods, such as GPS, cameras, or LiDAR alone, often fail in dynamic urban environments due to signal drift, poor lighting, or adverse weather. This study proposes a multi-sensor fusion navigation model integrating LiDAR, RGB-D cameras, GPS, IMUs, and ultrasonic sensors to achieve complementary perception. The model employs a layered framework: low-level data fusion for synchronization and correction, mid-level feature fusion for semantic–geometric alignment, and high-level decision fusion for path planning and behavior prediction. The purpose of this research is to evaluate the feasibility of applying multi-sensor fusion to food delivery robots. Nevertheless, the analysis relies heavily on simulations and benchmark datasets, without fully considering long-term costs, hardware durability, or user acceptance. Future research should include large-scale field trials, interdisciplinary cost–benefit evaluations, and advances in lightweight fusion algorithms to enable scalable deployment in real-world delivery environments.

Keywords: food delivery, delivery robots, multi-sensor fusion, navigation model, urban environments

1. Introduction

With the booming development of the Internet economy, the food delivery industry has shown a rapid growth trend worldwide. Data shows that China's food delivery market has maintained a steady growth trend in the past five years, with the number of users reaching hundreds of millions and gradually becoming an indispensable part of daily life in modern society [1]. However, the huge demand for orders has also made the "last mile" of food delivery a bottleneck in terms of efficiency and safety in the food delivery system. On the one hand, it is because the delivery is completed by human riders, which results in high costs and heavy workload for manual delivery. When encountering bad weather conditions or complex traffic conditions, their work efficiency will also be greatly affected. On the other hand, the traffic environment in cities is becoming more and more complex and lacks certainty, which also makes manual delivery prone to instability and even safety hazards during peak hours. In this context, automated delivery robots have begun to become a possible solution. Delivery robots can reduce labor costs, improve efficiency, and alleviate urban traffic and human resource pressures to a certain extent. However, the food delivery scene is

© 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: 10.54254/2755-2721/2025.LD29174

dynamic and complex, involving a variety of dynamic obstacles such as pedestrians, bicycles, and motor vehicles, as well as external factors such as rain and snow, changes in lighting, and diverse road environments. These environmental factors place higher demands on robot perception, positioning, and navigation. Traditional navigation solutions that rely on a single sensor (such as GPS or camera) often have difficulty ensuring robustness and accuracy in complex environments. For example, GPS is prone to signal drift in "urban canyons", camera recognition performance significantly decreases in low light or bad weather, and LiDAR is also interfered with in rain and snow conditions [2]. Therefore, a single perception method is no longer sufficient to support the stable operation of food delivery robots in dynamic and changing environments. Based on this problem, this paper will propose and study a multi-sensor fusion delivery robot navigation model. By combining LiDAR, RGB-D camera, GPS, IMU and ultrasonic sensor, multi-source information complementation is achieved to improve the accuracy and robustness of environmental perception. On this basis, this paper will explore a multi-level data fusion method suitable for food delivery scenarios, and combine path planning and system integration for optimization. In addition, this paper will also systematically analyze the potential of this model in practical applications from the three aspects of technical, economic and social feasibility, providing a reference for future large-scale applications.

2. Research progress in navigation sensing technology and fusion methods

2.1. Mainstream sensors in navigation and their limitations

In food delivery robot navigation, sensors are key components of environmental perception and positioning. Different types of sensors have different accuracy, suitability, and costs. Currently, mainstream sensors include LiDAR, cameras, GPS, IMU, and ultrasonic waves [3]. LiDAR, also known as laser radar, has been widely used in autonomous driving and robotics due to its precise ranging capabilities and environmental modeling functions. It can generate high-definition three-dimensional maps from point clouds, providing strong support for obstacle detection and path planning. However, LiDAR is expensive and will be interfered with in adverse weather conditions such as rain, snow, fog, and haze. Objects with low reflectivity may also reduce recognition accuracy [4].

Cameras (RGB and RGB-D) have the advantages of low cost and rich information in environmental perception. They can recognize traffic signs, pedestrians, and dynamic targets. In particular, RGB-D cameras can provide depth information, which enhances scene understanding. However, it is very sensitive to light. Its recognition ability will be significantly reduced at night, in strong light or in backlight conditions. In addition, it has the problem of poor stability in rainy and snowy weather [5]. GPS (Global Positioning System) is a key means of global positioning. In open scenes, GPS can achieve meter-level or even sub-meter-level positioning accuracy, which is suitable for navigation and path correction. However, in "urban canyons" or indoor environments, GPS signals are easily blocked, resulting in multipath effects, and positioning accuracy will drop significantly. IMU inertial measurement units can use accelerometers and gyroscopes to estimate the robot's posture and motion state. It is an important supplement to GPS positioning. However, due to error accumulation, IMU will have obvious drift when working for a long time, and it needs to rely on other sensors for correction. Ultrasonic sensors have a simple structure and are inexpensive. They are commonly employed for close-range obstacle detection and redundant safety functions. However, their measurement accuracy is limited, and they are highly susceptible to environmental noise, making it challenging for them to serve as the primary perception modality. More broadly,

DOI: 10.54254/2755-2721/2025.LD29174

reliance on a single sensor struggles to ensure robustness in the complex and dynamic environment of food delivery. Consequently, multi-sensor fusion has emerged as an inevitable approach to enhance navigation accuracy and overall system stability.

2.2. Research progress in multi-sensor fusion methods

Research on navigation sensing for food delivery robots has advanced along two main directions: filter-based fusion and learning-based fusion. Filter-based methods such as EKF, UKF, and factor graph optimization tightly couple GNSS/GPS, IMU, odometry, and LiDAR or vision data. LIO-SAM is a representative system that combines inertial and LiDAR constraints with GPS correction, achieving real-time and high-accuracy performance [6]. Learning-based methods focus on multimodal 3D perception. BEVFusion integrates camera and LiDAR features into a unified bird's-eye-view representation, improving mIoU by 6 percent over camera-only and 13.6 percent over LiDAR-only models. Later systems such as DPFusion further confirmed these benefits. In GNSS-degraded settings, GNSS/IMU fusion improves positioning accuracy by around 80 percent, enabling robots to operate reliably in mixed environments. Large-scale datasets such as nuScenes provide standardized platforms for benchmarking fusion algorithms and supporting real-world deployment.

3. Methodology and model design

3.1. Feasibility of system architecture

In the context of food delivery scenarios, the system architecture of delivery robots must simultaneously address the requirements of complex environmental perception, real-time path planning, and cross-platform interaction. To meet these needs, the proposed architecture is organized into three layers: hardware, software, and data fusion. At the hardware layer, LiDAR and RGB-D cameras are adopted as the core perception units, enabling accurate point cloud mapping and semantic information extraction. GPS combined with IMU is employed for global positioning and pose estimation, which enhances robustness in complex environments. In addition, ultrasonic sensors are integrated to provide close-range redundancy, ensuring safe operation in narrow streets and crowded pedestrian areas. At the software layer, a modular framework based on ROS 2 and the Apollo autonomous driving platform supports task scheduling and module decoupling, allowing perception, localization, decision-making, and control to be optimized independently. The data fusion layer is structured into three levels: the low level performs multi-source data synchronization and preprocessing, the middle level enables feature complementarity, and the high level carries out global decision-making and path planning. This layered fusion mechanism ensures real-time performance, scalability, and maintainability while also improving fault tolerance. For example, when one sensor fails, others can compensate through fusion to maintain continuous system operation. The overall design of this architecture is illustrated in Figure 1, which shows the interaction between hardware, software, and data fusion layers, as well as the closed-loop feedback from control outputs to sensing modules. The figure highlights how multi-sensor integration and layered fusion jointly support robust and reliable navigation in food delivery environments.

DOI: 10.54254/2755-2721/2025.LD29174

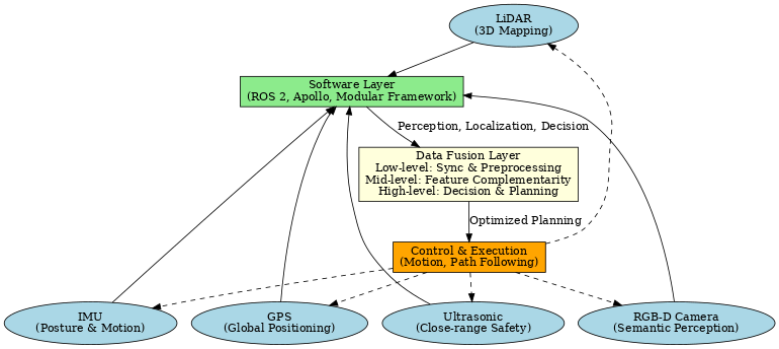

Figure 1. Proposed multi-sensor fusion architecture for food delivery robots

3.2. Feasibility of the fusion model

In a multi-sensor system, the design of the fusion model directly determines the delivery robot's positioning accuracy, environmental awareness, and decision-making robustness. The fusion model proposed in this study incorporates three mechanisms: low-level data fusion, mid-level feature fusion, and high-level decision fusion. The fusion model also integrates traditional probabilistic inference techniques with deep learning methods. In the low-level fusion, LiDAR point clouds are integrated with inertial information from the IMU using techniques such as Kalman filtering and factor graph optimization to achieve accurate odometry estimation. GPS signals are then used for regular correction to eliminate drift. In the mid-level fusion, a deep learning model is used to jointly model multimodal features, aligning the semantic information from the RGB-D camera with the geometric information from the LiDAR in the BEV space. This allows for coordinated improvements in obstacle detection and road segmentation, thereby enhancing perception robustness in dynamic environments. In high-level fusion, Bayesian inference and reinforcement learning methods are applied to integrate perception results from multiple sources, perform path decisions and behavior predictions. The model relies on lightweight networks and edge computing support in terms of real-time performance and energy consumption, and can respond quickly to dynamic crowd flow, traffic interference and other food delivery scenarios. Generally speaking, this fusion model has high feasibility in terms of computational complexity, real-time performance, energy consumption, and scalability, and can support food delivery robots in safe and efficient navigation in complex environments.

3.3. Specific optimization for food delivery

To meet the specific requirements of food delivery tasks, system design needs to be optimized in terms of path planning, localization accuracy, system responsiveness, and environmental adaptability.

Path planning is critical in dynamic environments where robots frequently encounter pedestrians, bicycles, and vehicles. Recent studies have shown that incorporating prioritized experience replay (PER) and long short-term memory (LSTM) into a twin delayed deep deterministic policy gradient (TD3) framework significantly improves efficiency in scenarios with both static and dynamic obstacles. Compared with traditional methods, this approach reduces path length, collision frequency, and time cost, achieving efficiency gains of about 20–30% [7].

Positioning accuracy is particularly critical in the last 100 meters or at specific locations, such as door-to-door delivery points. Autonomous driving/navigation systems often require a positioning

DOI: 10.54254/2755-2721/2025.LD29174

error of less than 0.1 meter at a 95% confidence level for safety and reliability, especially in narrow paths such as alleyways, building entrances, and doorways. According to an evaluation of autonomous driving system positioning accuracy by Rehrl et al., if the system is equipped with high-quality fusion (GNSS + IMU + map/SLAM), an error level of less than 0.1 meters can be achieved in certain test environments. System responsiveness and resource consumption cannot be ignored. In food delivery scenarios, robots often have to complete multiple deliveries or recharge, so path planning and perception modules must be lightweight and real-time. Previous research has applied reinforcement learning methods (such as PL-TD3) to path planning, demonstrating significant reductions in execution and planning time in real-world experiments and simulations while maintaining smooth and safe paths. Environmental adaptability includes factors such as lighting, weather, and structural diversity. This type of optimization is often achieved through data augmentation, domain adaptation, or training models in multiple urban environments [8]. These targeted optimizations highlight both the feasibility and potential benefits of applying advanced multi-sensor fusion and reinforcement learning techniques in food delivery robotics.

4. Feasibility analysis

The application of multi-sensor fusion in delivery robots for food delivery scenarios is highly feasible. Sensor hardware has become increasingly mature, with LiDAR costs having fallen by more than 70% in the past five years—some domestically produced units are now available for only a few thousand RMB, making large-scale deployment realistic. At the same time, RGB-D cameras, IMUs, and ultrasonic sensors have achieved mass production and demonstrated high reliability. The real-time performance and accuracy of fusion algorithms are also improving. Approaches such as factor graph optimization and multimodal deep learning have been widely validated in autonomous driving and robotics. For instance, laser–inertial fusion frameworks such as LIO-SAM have achieved centimeter-level positioning accuracy on public datasets, which is sufficient to meet the “door-to-door” requirements of food delivery scenarios [6].

The computing and communication environments are also supporting the deployment of robots. With the widespread adoption of 5G and edge computing, robots can achieve communication latency below 10ms and offload some high-load computing to edge servers, helping to address power consumption and computing bottlenecks. From sensor hardware reliability and fusion algorithm accuracy to real-time computing and communication infrastructure, multi-sensor fusion robots have strong technical feasibility and provide solid support for automated food delivery.

The application scenarios of food delivery robots are expected to evolve steadily alongside technological progress and urban innovation. In the short term, their deployment will most likely focus on semi-enclosed environments such as university campuses, industrial parks, and residential communities, where traffic patterns are relatively predictable and the risks of environmental uncertainty are reduced. These controlled settings provide an ideal testing ground for validating navigation accuracy, safety mechanisms, and human–robot interaction models. In the medium term, food delivery robots are anticipated to be increasingly integrated with smart transportation systems, urban digital infrastructure, and large-scale delivery platforms, which will enable city-wide coordination and cross-district operations. This integration will also require closer alignment with regulations, traffic management policies, and commercial logistics systems. In the long term, emerging technologies are expected to bring breakthroughs in both performance and scalability. Advances in lightweight sensor fusion algorithms, domain adaptation methods for robust perception in diverse urban settings, and collaborative human–robot delivery frameworks will improve technological reliability, enhance cost-effectiveness, and strengthen social acceptance. Together,

DOI: 10.54254/2755-2721/2025.LD29174

these developments will pave the way for large-scale and sustainable adoption of food delivery robots in everyday life.

5. Conclusion

This study examined the feasibility and potential of applying multi-sensor fusion to delivery robots in food delivery scenarios. The research findings demonstrate that single-sensor navigation systems face significant limitations in complex and dynamic environments, such as urban canyons, low-light conditions, or adverse weather. By integrating LiDAR, RGB-D cameras, GPS, IMUs, and ultrasonic sensors, the proposed multi-sensor fusion architecture improves the robustness and accuracy of environmental perception, positioning, and decision-making. Moreover, the layered fusion model, which combines probabilistic filtering, deep learning feature integration, and high-level decision fusion, provides a flexible and scalable solution capable of supporting real-time path planning and reliable operation in diverse delivery settings. Optimizations in path planning, localization accuracy, responsiveness, and environmental adaptability further underline the applicability of reinforcement learning and multimodal fusion to address last-mile delivery challenges.

Despite these contributions, the present study has several limitations.The proposed system does not fully account for economic cost models, long-term hardware durability, or user acceptance, which are critical for scaling up deployment. Future research should address these limitations by conducting field trials in diverse and high-density delivery environments, focusing on longitudinal performance, system maintenance, and cost-effectiveness. In addition, interdisciplinary work combining robotics, urban planning, and human–robot interaction is needed to assess the broader social impacts and acceptance of delivery robots.

References

[1] GlobeNewswire. (2025, February 26). China online food delivery market research report 2025-2033. GlobeNewswire. https: //www.globenewswire.com/news-release/2025/02/26/3032793/0/en/China-Online-Food-Delivery-Market-Research-Report-2025-2033-Featuring-Profiles-of-Key-Players-Ele-me-Meituan-Dianping-ENJOY-Daojia-Home-cook.html

[2] Min, H., Wu, X., Cheng, C., & Zhao, X. (2019). Kinematic and dynamic vehicle model-assisted global positioning method for autonomous vehicles with low-cost GPS/camera/in-vehicle sensors. Sensors, 19(24), 5430.

[3] Li, Q., Queralta, J. P., Gia, T. N., Zou, Z., & Westerlund, T. (2021). Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments. arXiv preprint arXiv: 2103.13719.

[4] Zhang, J., & Singh, S. (2017). Low-drift and real-time lidar odometry and mapping. Autonomous Robots, 41(2), 401–416.

[5] Liciotti, D., Paolanti, M., Frontoni, E., & Zingaretti, P. (2017, September). People detection and tracking from an RGB-D camera in top-view configuration: review of challenges and applications. In International Conference on Image Analysis and Processing (pp. 207-218). Cham: Springer International Publishing.

[6] Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., & Rus, D. (2020). LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 5135–5142.

[7] Lin, Y., Zhang, Z., Tan, Y., Fu, H., & Min, H. (2025). Efficient TD3-based path planning of mobile robot in dynamic environments using prioritized experience replay and LSTM. Scientific Reports.

[8] Rehrl, K., Gröchenig, S., & Wimmer, K. (2021). Evaluating localization accuracy of automated driving systems. Sensors, 21(17), 5855.