Research on Environmental Perception Technology in Unmanned Delivery Vehicles

ZHOU Qiu
(Tianjin Yingzeng Trading Co., Ltd., Tianjin 300499,China)

Abstract:This paper comprehensively reviews the mainstream environmental perception technology systems for autonomous delivery vehicles, systematically analyzing the working principles, technical architectures, performance parameters, application scenarios, and advantages and disadvantages of four primary sensing technologies: LiDAR, visual cameras, millimeter-wave radar, and ultrasonic sensors. It delves into core algorithms, fusion architectures, and implementation processes of multi-sensor fused perception. Drawing on real-world deployment cases from leading companies such as JD.com, Meituan, and Nuro, and supported by authoritative 2025 industry statistics and market analysis in autonomous delivery, the study systematically summarizes current technological bottlenecks, scenario adaptation challenges, and engineering implementation gaps in environmental perception for autonomous delivery vehicles. Targeted optimization strategies are proposed, along with predictions on future technological evolution and industry development trends. The research provides theoretical foundations and practical references for optimizing perception systems, facilitating engineering deployment, and advancing industry-standardization efforts in autonomous delivery vehicles, while also offering reliable support for upgrading intelligent logistics equipment and advancing applied engineering research.
Key words:Unmanned delivery vehicle; environmental perception; multi-sensor fusion; LiDAR; visual perception; millimeter-wave radar; smart logistics

1 INTRODUCTION

1.1 Research Background

Under the comprehensive advancement of digital economy, smart logistics, and new infrastructure strategies, China's express logistics industry has entered a period of rapid development, with express volume consistently ranking first globally for multiple consecutive years. According to the "2025 Blue Book on the Development of China's Smart Logistics Industry," China's total express volume surpassed 130 billion pieces in 2025, with daily last-mile delivery orders exceeding 350 million. Traditional manual delivery models face multiple challenges, including rising labor costs, inconsistent delivery efficiency, insufficient coverage at the final delivery point, and operational limitations under adverse weather conditions—making intelligent transformation of last-mile logistics an urgent necessity. As low-speed autonomous intelligent devices, unmanned delivery vehicles offer core advantages such as round-the-clock operation, low operating costs, high delivery efficiency, and zero human intervention, precisely matching the requirements of last-mile delivery scenarios. They effectively address the shortcomings of manual delivery and have become a key breakthrough in the transformation and upgrading of smart logistics.

The autonomous driving system of unmanned delivery vehicles primarily consists of three core modules: environmental perception, decision-making and planning, and motion control. Among these, the environmental perception module serves as the vehicle's "eyes," forming the foundational basis for decision-making and motion control. Urban last-mile delivery environments are characterized by complex road conditions, dense pedestrian traffic, mixed non-motorized vehicles, random obstacles, and variable lighting and weather conditions, placing extremely high demands on the environmental perception capabilities of unmanned delivery vehicles. Individual sensors are limited by physical constraints and cannot simultaneously meet the requirements of all-weather, full-scenario, and high-precision perception. For example, cameras are easily affected by lighting and inclement weather; LiDAR is relatively expensive and lacks speed perception capability; millimeter-wave radars have limited resolution; and ultrasonic sensors have short detection ranges.

In this context, integrating multiple types of sensors to build a multi-modal environmental perception system has become a central direction for technological evolution in unmanned delivery vehicles. In-depth research into the working mechanisms, application characteristics, and fusion algorithms of various sensing technologies, along with analysis of industry implementation cases and market conditions, helps identify technical bottlenecks and optimization pathways. This research holds significant engineering value and industry relevance for enhancing the safety, stability, and scenario adaptability of unmanned delivery vehicles, promoting intelligent, standardized, and scalable deployment of last-mile logistics. Moreover, closely aligned with practical applications in intelligent logistics and autonomous driving engineering, this study provides systematic theoretical and practical support for mid-level engineering professionals engaged in intelligent equipment research, engineering optimization, and real-world deployment.

1.2 Current Research Status at Home and Abroad

1.2.1 Current situation of overseas research

Overseas research on perception technologies for unmanned delivery vehicles started earlier, primarily represented by the United States and Europe. The technological approach emphasizes high precision, high reliability, and full-scenario adaptability, focusing on implementing L4-level low-speed autonomous driving perception systems. Nuro, a leading company in the global unmanned delivery sector, pioneered a multi-sensor fusion architecture combining solid-state LiDAR, high-definition imaging millimeter-wave radar, and vision cameras. By abandoning traditional mechanical LiDAR and adopting compact solid-state LiDAR, it significantly improved system stability and environmental adaptability, enabling precise detection of small targets at long distances even in adverse weather conditions—ideal for urban open-road delivery scenarios. European companies, meanwhile, focus on safety and standardization in perception technology, emphasizing optimization of sensor anti-interference algorithms and scenario-adaptation models. Tailoring to the characteristics of narrow streets and slow-moving traffic in European cities, they have enhanced short-range obstacle detection and low-speed collision avoidance accuracy, establishing a mature standardization framework for perception technology deployment. Overall, foreign unmanned delivery vehicle perception technologies excel in high sensor hardware precision, mature fusion algorithms, and strong scenario adaptability. However, challenges remain, including high equipment costs, insufficient adaptation to local environments, and difficulties in large-scale implementation.[1-3]

1.2.2 Domestic Research Status

China's domestic unmanned delivery vehicle industry, supported by a vast last-mile delivery market, has achieved rapid iteration and development, establishing a localized technological path characterized by cost-effectiveness, scenario adaptability, fast iteration, and large-scale deployment. Leading companies such as JD.com, Meituan, and Cainiao have deeply engaged in complex local environments—including urban communities, industrial parks, and universities—continuously refining their perception technology systems to develop solutions tailored to China's road conditions. JD's sixth-generation autonomous delivery vehicle integrates a perception large model with fused point-cloud vision technology, expanding its sensing range 19-fold and improving perception performance threefold. Meituan's unmanned vehicles employ a "one primary, two auxiliary" LiDAR configuration combined with multi-camera and millimeter-wave radar fusion, achieving a 98% success rate in navigating the complex traffic conditions of Shenzhen's urban villages.

Domestic universities and research institutions have been conducting technical research in parallel, focusing on breakthroughs in core technologies such as multi-sensor fusion algorithms, semantic perception in complex scenarios, and low-computation power edge deployment. These efforts have effectively reduced hardware costs for perception systems and improved algorithm adaptability and implementation. However, overall, domestic perception technology for unmanned delivery vehicles still faces shortcomings: insufficient robustness in perceiving complex and extreme scenarios, limited accuracy in time and space synchronization across multiple sensors, numerous blind spots in aging environments, and reliance on imported high-end perception chips and core sensors—factors that hinder the industry's large-scale, high-quality development.[4,5]

2 Overall Overview of Environment Perception Technology for Autonomous Delivery Vehicles

2.1 Core Definition and Functions of the Environmental Perception System

The environment perception system for unmanned delivery vehicles is a core onboard system that relies on various vehicle-mounted sensors, data acquisition modules, and algorithm processing units to real-time capture environmental information around the vehicle, enabling environment modeling, object detection, semantic recognition, and state prediction. It forms the fundamental basis for autonomous navigation, obstacle avoidance, intelligent braking, and path adjustment in unmanned delivery vehicles. Unlike high-speed autonomous vehicles, unmanned delivery vehicles are classified as low-speed autonomous devices (maximum speed ≤30 km/h), primarily operating in urban low-speed roads, enclosed campuses, and narrow community pathways. Their perception systems emphasize four key capabilities: short-range high-precision sensing, small-object recognition, prediction of complex dynamic obstacles, and stable operation under all weather conditions.

The core functions of the environment perception system for unmanned delivery vehicles mainly consist of four components: first, environmental information acquisition, which uses various sensors to collect real-time multidimensional data such as surrounding road conditions, obstacles, pedestrians, non-motorized vehicles, traffic signs, and pavement conditions; second, object detection and classification, which employs algorithms to identify and distinguish between static obstacles (curbs, bollards, guardrails, parked vehicles) and dynamic obstacles (pedestrians, electric scooters, moving vehicles), enabling target classification, distance measurement, and speed estimation; third, scene semantic parsing, which recognizes semantic elements such as lane markings, traffic lights, intersections, crosswalks, and restricted zones, thereby constructing a real-time environmental semantic model; fourth, risk prediction and data output, which analyzes potential environmental risks in real time and transmits perception data to the decision-making and planning module, providing data support for vehicle maneuvers such as obstacle avoidance, deceleration, stopping, and lane changing.[6]

2.2 Overall Architecture of the Environmental Perception System

The environment perception system of unmanned delivery vehicles adopts a four-layer progressive architecture: "hardware perception layer—data preprocessing layer—algorithm fusion layer—decision output layer." This overall structure features clear logic and distinct hierarchical layers, enabling a full closed-loop process for environmental information—from acquisition and processing to fusion and output.

The first layer is the hardware perception layer, serving as the foundational component of the system. It integrates four core sensing devices: LiDAR, high-definition cameras, millimeter-wave radar, and ultrasonic sensors. These components comprehensively capture 3D point cloud data, image pixel data, speed and distance measurements, and close-range obstacle information around the vehicle, achieving seamless 360-degree coverage with no blind spots. The second layer is the data preprocessing layer, which employs filtering, calibration, and noise reduction algorithms to eliminate noise, redundancy, and anomalies from raw data. It also performs time synchronization and spatial alignment across multi-source data streams, thereby enhancing the accuracy of the original data. The third layer is the algorithm fusion layer, leveraging multi-sensor fusion algorithms to integrate the strengths of various sensors and compensate for the limitations of individual sensing modalities. This layer enables high-precision target detection, semantic segmentation, and trajectory prediction. The fourth layer is the decision output layer, which transmits the fused, high-accuracy perception data to the autonomous driving decision module, supporting key autonomous operations such as path planning, obstacle avoidance control, and speed regulation.[7]

2.3 Classification and Adaptation Standards for Environmental Perception Technologies

Perception technologies can be broadly categorized into passive and active perception. Passive perception relies primarily on visual cameras, utilizing ambient light to capture image data. It offers semantic recognition capabilities, low cost, and high resolution, but has weak anti-interference performance. Active perception includes LiDAR, millimeter-wave radar, and ultrasonic sensors, which actively emit detection signals to gather environmental data. This approach is unaffected by external lighting conditions, features strong anti-interference capability, and is suitable for all-weather operation scenarios.[8]

Considering the industry requirements of unmanned delivery vehicles—low-speed operation, complex environments, and controllable costs—a clear set of perception technology adaptation standards has emerged: first, high-precision obstacle detection at short distances (0 – 50 meters), ensuring safe navigation in narrow roads and close-range obstacle avoidance; second, real-time speed measurement and trajectory prediction for dynamic targets, enabling adaptation to chaotic pedestrian and non-motorized traffic; third, stable perception under adverse weather conditions such as rain, snow, bright sunlight, and nighttime, guaranteeing round-the-clock operations; fourth, low computational power and low power consumption, compatible with embedded on-board terminal deployment; fifth, low cost, compact size, and high reliability, meeting the demands of large-scale mass production and practical deployment. All types of perception technologies must be selected and optimized according to these criteria.

3 Research on the Principles and Applications of Single Environment Perception Technology for Unmanned Delivery Vehicles

3.1 LiDAR perception technology

3.1.1 Working principle

LiDAR is a core high-precision active perception sensor for autonomous delivery vehicles, operating primarily on the time-of-flight (ToF) principle. It actively emits high-frequency laser beams through a laser transmitter; when these beams encounter surrounding obstacles, they reflect back and are captured by a receiver. By calculating the time difference between emission and reception of the laser signals and incorporating the speed of light, the sensor accurately determines the distance to the obstacle. Combined with the laser's emission angle, it constructs a three-dimensional point cloud model of the surrounding environment, enabling 3D environmental reconstruction, obstacle localization, and contour recognition. [9]

Its fundamental distance measurement formula is:

$$L = \frac{1}{2} c \times \Delta t \tag{1}$$

In the formula: L is the distance between the sensor and the obstacle (m); c is the speed of light, taken as 3×10 m/s; Δt is the time difference between laser emission and reception (s).

By emitting laser beams through high-frequency scanning, LiDAR can rapidly capture a massive amount of three-dimensional coordinate points, generating high-density point cloud data that accurately reconstructs the outline, dimensions, and spatial position of obstacles. With extremely high ranging accuracy and spatial resolution, it serves as the core hardware for high-precision environmental modeling in unmanned delivery vehicles.

3.1.2 Technical Classification and Architectural Parameters

Currently, LiDAR sensors used in delivery robots are primarily categorized into three types: mechanical LiDAR, semi-solid-state LiDAR, and solid-state LiDAR, each suited to different positioning requirements and cost levels of unmanned delivery vehicles. Mechanical LiDARs, typically 16-line or 32-line models, offer wide scanning angles, high point cloud density, and mature technology, making them widely used in mid-to-high-end delivery robots. Semi-solid-state LiDARs are smaller in size and more stable, offering a balanced combination of precision and cost-effectiveness. Solid-state LiDARs, which have no moving mechanical parts, feature low failure rates, low power consumption, and suitability for mass production—making them the future mainstream direction for industry evolution. Nuro's new generation of autonomous vehicles has already fully adopted solid-state LiDAR to replace traditional mechanical sensors.

Key performance parameters of commonly used LiDAR sensors in delivery robots include: ranging distance from 0.05 to 200 meters, ranging accuracy of ±2 cm, horizontal scanning angle of 360° /120° , vertical scanning angle of 15° –30° , and scanning frequency of 10–20 Hz—fully meeting the needs of short-range, high-precision perception for low-speed autonomous vehicles. The prevailing industry configuration is a "primary radar + auxiliary radars" setup. For example, Meituan’s delivery robots use one primary and two auxiliary LiDARs: the primary sensor handles high-precision mapping of the forward core area, while the auxiliary sensors cover blind spots on the sides and rear, enabling full 360-degree, seamless perception coverage around the vehicle.[10]

3.1.3 Technical strengths and weaknesses

Lidar's core advantages include: first, extremely high ranging accuracy with centimeter-level perception precision, enabling accurate detection of fine targets such as curbs, stone blocks, and low obstacles; second, immunity to lighting interference, ensuring consistent performance in both day and night conditions, and adaptability to complex lighting scenarios including nighttime, backlit, and shaded environments; third, the ability to construct three-dimensional environmental models, accurately reconstructing spatial terrain and obstacle contours to support path planning and precise obstacle avoidance; fourth, strong resistance to ambient light interference, offering significantly greater stability than pure vision-based perception.

Key shortcomings: First, the hardware cost is relatively high, significantly increasing the mass production cost of autonomous vehicles. Second, traditional mechanical LiDARs contain rotating components, making them susceptible to wind, sand, and vibration during prolonged outdoor operation, resulting in a higher failure rate. Third, they cannot directly obtain obstacle information such as speed, texture, color, or semantics, requiring integration with other sensors. Fourth, under extreme weather conditions like rain, snow, or heavy fog, laser beams are easily scattered by moisture, leading to a slight decline in perception accuracy.

3.1.4 Applicable Scenes

LiDAR is primarily designed for high-precision perception scenarios in autonomous delivery vehicles, including accurate modeling of urban open roads, obstacle avoidance on narrow paths, detection of low obstacles, reconstruction of complex intersection environments, and static terrain perception. It serves as the core hardware component of the perception system for high-end autonomous delivery vehicles.

3.2 Camera-based visual perception technology

3.2.1 Working Principle

The visual camera is the most fundamental and cost-effective passive sensing hardware for unmanned delivery vehicles, based on the core principles of optical imaging and digital image processing. It captures natural light signals from the environment through an optical lens, converts these light signals into digital electrical signals via an image sensor (CMOS/CCD), and generates high-definition image video frames. Relying on deep learning vision algorithms (such as CNN, YOLO, and SegNet), it performs image preprocessing, feature extraction, object detection, semantic segmentation, and classification recognition, enabling precise identification of traffic signs, traffic lights, pedestrians, vehicles, and lane markings.

The core feature extraction formula for visual perception is:

$$F(x,y)=\sigma(W\times I(x,y)+b) \quad (2)$$

In the equation: $F(x,y)$ is the image feature output matrix; $I(x,y)$ is the original image pixel matrix; W is the convolutional weight parameter; b is the bias parameter; σ is the activation function.

Through multiple layers of convolution, pooling, and activation operations, shallow texture features and deep semantic features can be progressively extracted from images, enabling precise multi-object recognition in complex scenes. This forms the core component of semantic perception for unmanned delivery vehicles.[11]

3.2.2 Technical Classification and Architectural Parameters

The visual cameras mounted on delivery robots are primarily categorized into three types: monocular, binocular, and panoramic cameras. Monocular cameras are the most cost-effective and structurally simple, mainly used for recognizing traffic signs, traffic lights, and lane markings. Binocular cameras utilize the principle of disparity to enable distance measurement, compensating for the lack of depth perception in monocular systems, making them suitable for detecting nearby obstacles. Panoramic cameras provide 360-degree surround-view imaging, eliminating blind spots around the vehicle and ideal for complex, narrow environments such as residential communities and campuses.

Industry-standard configuration parameters include resolutions of 1080P/4K, frame rates of 30 - 60 fps, horizontal viewing angles of 90° - 180°, night vision, and wide dynamic range imaging capabilities, effectively suppressing glare and backlight interference to ensure reliable imaging performance both day and night. Both JD.com and Meituan's autonomous delivery vehicles employ multi-camera configurations, integrating front-, side-, rear-, and surround-view cameras to achieve comprehensive visual perception coverage.[12]

3.2.3 Technical strengths and weaknesses

Key advantages of visual perception: first, extremely low hardware cost, compact size, easy integration, and suitability for mass production; second, strong semantic recognition capability, enabling precise identification of traffic lights, road signs, lane markings, and categories of pedestrians and non-motorized vehicles—making it essential for semantic understanding in autonomous driving scenarios; third, high image resolution allows capturing fine texture details, suitable for target classification in complex environments; fourth, mature technology with rapid algorithm iteration and excellent adaptability.

Key shortcomings: First, it is highly dependent on ambient lighting, resulting in significantly degraded image quality under strong light, backlit conditions, nighttime, or adverse weather such as rain, snow, and fog, leading to frequent recognition failures and misidentifications. Second, the depth estimation accuracy of monocular cameras is low, failing to meet high-precision obstacle avoidance requirements. Third, the system has weak anti-interference capability for dynamic targets, making it vulnerable to occlusions and shadows, which can cause missed detections or false positives.

3.2.4 Applicable Scenes

Visual perception technology is primarily designed for semantic recognition scenarios in unmanned delivery vehicles, including traffic signal detection, road sign recognition, lane line detection, object classification, and scene semantic parsing. It is a standard sensing technology for all unmanned delivery vehicles and is typically used as the primary sensing hardware in combination with radar systems.

3.3 Millimeter-wave radar perception technology

3.3.1 Working principle

Millimeter-wave radar is the core hardware for all-weather dynamic perception in unmanned delivery vehicles, functioning as an active sensing device based on frequency-modulated continuous wave (FMCW) detection principles. It emits high-frequency millimeter-wave electromagnetic signals through its antenna; when these waves encounter obstacles, they reflect back. The radar receives the echo signals and precisely calculates the distance, relative velocity, and angular position of the obstacles by analyzing the frequency and phase differences between the transmitted and received signals, enabling real-time monitoring and trajectory tracking of moving targets.

The core formula for speed measurement using millimeter-wave radar is:

$$v = \frac{c \times \Delta f}{2f_0} \tag{3}$$

where v is the target relative velocity (m/s), c is the speed of light, Δf is the frequency difference between the transmitted and echo signals (Hz), and f_0 is the radar's center operating frequency (Hz).

Millimeter-wave radar inherently possesses velocity sensing capability, enabling real-time detection of dynamic obstacles' velocity changes and prediction of their motion trajectories, thereby compensating for LiDAR's lack of velocity measurement and low accuracy in visual velocity estimation.

3.3.2 Technical Classification and Architectural Parameters

Most unmanned delivery vehicles are equipped with two types of millimeter-wave radars: 24GHz and 77GHz. The 24GHz radar offers low cost, moderate detection range, and strong anti-interference capability, making it ideal for short-range obstacle monitoring and blind-spot warnings. In contrast, the 77GHz radar features wider bandwidth, higher resolution, and greater velocity measurement accuracy, enabling precise perception of distant dynamic targets and suitable for high-speed dynamic obstacle avoidance on open roads. Nuro's new generation of autonomous vehicles is equipped with high-definition imaging millimeter-wave radars, capable of accurately detecting small targets at long distances even in adverse weather conditions, significantly enhancing safety in complex environments.

Key performance parameters: Detection range of 0.2 - 150 meters, ranging accuracy of ±0.5 meters, speed measurement accuracy of ±0.2 m/s, horizontal detection angle of ±60°, and stable operation under all weather conditions without being affected by lighting, rain, snow, fog, or sandstorms.[13]

3.3.3 Technical strengths and weaknesses

Millimeter-wave radar offers several core advantages: first, it excels in all-weather operation, unaffected by lighting conditions, rain, snow, fog, or sandstorms, providing the best environmental robustness; second, it can directly and accurately measure velocity, enabling real-time tracking of dynamic targets' movements, making it ideal for scenarios involving pedestrians and non-motorized vehicles moving unpredictably;

third, its hardware is compact, low-power, moderately priced, and highly stable, making it well-suited for long-term in-vehicle deployment; fourth, it has penetration capability, allowing detection of partially occluded targets, with outstanding anti-interference performance.

Key shortcomings: First, low spatial resolution prevents the identification of obstacle contours, textures, and semantic information, making it difficult to distinguish target categories. Second, detection accuracy is limited at close range and is easily affected by ground clutter interference. Third, the system has weak capability in detecting small static obstacles, failing to meet the requirements for high-precision terrain perception.

3.3.4 Application Scenarios

Millimeter-wave radar is primarily suited for dynamic risk perception scenarios in unmanned delivery vehicles, including speed measurement and trajectory prediction of moving pedestrians, non-motorized vehicles, and other vehicles, all-weather monitoring under adverse conditions, blind spot risk warnings, and dynamic obstacle detection during high-speed driving.。[14]

3.4 Ultrasonic Sensor Perception Technology

3.4.1 Working Principle

The ultrasonic sensor is the core hardware for short-range auxiliary perception in unmanned delivery vehicles, classified as an active sensing device. Its fundamental principle relies on ultrasonic distance measurement. The sensor transmits high-frequency ultrasonic signals through its probe; when these waves encounter an obstacle, they reflect back and are received by the probe. By calculating the time difference between the emission and reception of the ultrasonic signal and combining it with the speed of sound propagation, the system determines the short-range distance between the sensor and the obstacle, enabling detection of nearby obstacles and collision avoidance warnings. [15]

The core distance calculation formula is:

$$S = \frac{1}{2} v_s \times \Delta t \tag{4}$$

where S is the detection distance (m); v_s is the ultrasonic wave propagation velocity at room temperature, taken as 340 m/s; Δt is the round-trip time difference of the ultrasonic wave (s).

3.4.2 Technical Specifications and Deployment Method

Ultrasonic sensors offer advantages such as ultra-low cost, high stability, and no blind spots, with a detection range of 0.02 – 5 meters and a distance measurement accuracy of ±1 cm. They respond quickly and can precisely detect small objects at close range, including low obstacles, walls, curbs, and steps. The industry-standard deployment method involves a full-vehicle surround layout, with 8 to 12 ultrasonic sensors evenly distributed on the front, rear, and both sides of the vehicle. This setup enables comprehensive, short-range collision avoidance sensing without blind spots, serving as a reliable backup to high-end perception hardware.

3.4.3 Technical strengths and weaknesses

Core Advantages: First, extremely low cost, simple structure, low failure rate, and maintenance-free operation; second, exceptionally high detection accuracy at short distances with no blind spots, effectively compensating for the near-range perception limitations of radar and cameras; third, strong all-weather stability unaffected by lighting or weather conditions; fourth, fast response time, suitable for emergency close-range collision avoidance scenarios.

Core Limitations: First, very short detection range, only suitable for short-range sensing, incapable of medium- to long-range environmental detection; second, susceptible to environmental airflow and noise interference, resulting in significant errors at longer distances; third, lacks speed measurement, semantic recognition, and 3D modeling capabilities, enabling only basic distance measurement functions.[16]

3.4.4 Applicable Scenarios

Ultrasonic sensors serve as auxiliary sensing devices, ideal for short-range, low-speed collision avoidance in unmanned delivery vehicles, including narrow road overtaking, parking alignment, close-range obstacle detection, and blind spot protection. They are essential low-cost backup sensing hardware for unmanned delivery vehicles.

4 Research on Multi-Sensor Fusion Perception Technology for Unmanned Delivery Vehicles

4.1 Core Concept and Necessity of Multi-Sensor Fusion Technology

As revealed by the analysis of the aforementioned single-sensing technologies, each type of sensor has inherent technical limitations, making a single sensing approach unsuitable for complex urban last-mile delivery scenarios: pure vision-based systems are vulnerable to interference and have high failure risks in adverse weather; LiDAR lacks semantic understanding and speed measurement capabilities; millimeter-wave radar provides no contour or semantic information; and ultrasonic sensors have limited detection ranges. Multi-sensor fusion perception technology overcomes these limitations by integrating the multidimensional strengths of LiDAR, visual cameras, millimeter-wave radar, and ultrasonic sensors through hardware 协同calibration and algorithmic fusion processing. This enables comprehensive perception capabilities including "high-precision 3D modeling, accurate semantic recognition, all-weather dynamic speed estimation, and close-range blind-spot-free fallback," effectively eliminating the shortcomings of individual sensing technologies. It significantly enhances the perception robustness and accuracy of unmanned delivery vehicles in complex road conditions, harsh weather, and dynamic environments, representing the only optimal solution for current unmanned delivery vehicle perception technology.[17]

4.2 Levels of Multi-Sensor Fusion and Core Algorithms

Based on data processing levels, multi-sensor fusion in unmanned delivery vehicles is categorized into three tiers: data-level fusion, feature-level fusion, and decision-level fusion. Each tier supports different algorithms and application scenarios, forming a complete fusion technology framework.

4.2.1 Data-Level Fusion

Data layer fusion refers to the integration of raw data at the foundational level, directly synchronizing, aligning, denoising, calibrating, and fusing raw point clouds, images, distance, and velocity data collected from various sensors. This approach preserves all original data characteristics, achieving the highest perception accuracy. The core adaptation algorithm is the Kalman filter, which effectively removes noise from raw data and enables spatiotemporal synchronization and calibration of multi-source data.

Kalman filter core iterative formula:

$$\begin{aligned}
&\text{状态预测：} \hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \\
&\text{协方差预测：} P_{k|k-1} = AP_{k-1|k-1}A^T + Q \\
&\text{增益计算：} K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \\
&\text{状态更新：} \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1}) \\
&\text{协方差更新：} P_{k|k} = (I - K_kH)P_{k|k-1}
\end{aligned} \tag{5}$$

In the equation: x^ is the system state value; P is the covariance matrix; K is the filter gain; A, B, and H are system parameter matrices; Q and R are noise matrices; z is the observation value. This algorithm enables real-time filtering and precise calibration of raw data from multiple sensor sources, providing a high-quality data foundation for higher-level fusion.[18,19]

4.2.2 Feature-level Fusion

Feature-level fusion is a mid-level integration of feature information, which extracts features from various sensor data separately and then fuses the multidimensional feature information, balancing accuracy and computational speed. It is the most commonly used fusion level in unmanned delivery vehicles. Visual algorithms extract image semantic features, LiDAR captures three-dimensional spatial features, millimeter-wave radar extracts velocity features, and ultrasonic sensors obtain short-range distance features. A weighted fusion algorithm associates and reconstructs these features, enabling precise target classification, localization, and speed estimation.

4.2.3 Decision-level Fusion

The decision-making layer integrates into a high-level result fusion, where various sensors independently perform target detection and risk assessment, outputting their respective decisions. These are then fused through the D-S evidence theory algorithm to generate a final perception and decision outcome. This approach features low computational load and high fault tolerance, primarily used in risk mitigation and emergency obstacle avoidance scenarios.

4.3 Multi-sensor Fusion Architecture and Implementation Process

This paper integrates mainstream industry technologies to establish a "four-dimensional integrated" multi-sensor fusion architecture tailored for unmanned delivery vehicles—comprising "LiDAR 3D modeling + visual semantic recognition + millimeter-wave dynamic speed measurement + ultrasonic close-range backup."

The overall implementation process consists of five steps: First, synchronized data acquisition from multiple sensors, with four types of sensors collecting multidimensional environmental data in parallel while ensuring hardware clock synchronization. Second, raw data preprocessing, including noise reduction via Kalman filtering, spatiotemporal calibration, and removal of abnormal data. Third, hierarchical feature extraction, separately extracting spatial, semantic, velocity, and short-range distance features. Fourth, multi-feature fusion analysis, integrating multidimensional features using weighted fusion algorithms to achieve target detection, scene modeling, and risk prediction. Fifth, perception result output, transmitting the fused high-precision environmental data to the decision and control module to support autonomous driving and obstacle avoidance operations.

4.4 Core Advantages of Fusion Technology

Multi-sensor fusion technology effectively overcomes the scenario limitations of single-sensor perception systems, offering three core advantages: First, comprehensive perception capability, covering long and short distances, moving and stationary targets, and all-day, all-weather conditions. Second, high-precision perception, integrating 3D space, semantics, and velocity information to enable accurate target localization, classification, and speed estimation. Third, high robustness—when one sensor fails, others can compensate and ensure continuous perception, significantly enhancing vehicle safety and operational stability, making it fully suitable for complex urban last-mile delivery environments.[20,21]

5 Engineering Application Case Study of Perception Technology in Unmanned Delivery Vehicles

5.1 JD.com's Unmanned Delivery Vehicle Perception Technology Case

As a leading enterprise in China's unmanned delivery industry, JD.com has been deeply committed to end-of-line unmanned delivery technology for many years. Its sixth-generation intelligent unmanned delivery vehicle is an industry benchmark product, featuring a technical solution centered on "large-scale perception models and deep integration of multiple sensors," designed to adapt to complex delivery environments across diverse scenarios such as urban open roads, narrow community streets, rural roads, and wholesale markets.

In terms of hardware configuration, the JD sixth-generation unmanned vehicle is equipped with a three-dimensional perception system. It features high-precision lidar mounted on the roof, multiple high-definition stereo cameras and panoramic surround-view cameras distributed across the body, and standard millimeter-wave radars and surround ultrasonic sensors on the front, rear, and sides—forming a four-layer perception architecture that relies primarily on lidar, supplements with vision, supports with radar, and provides protection via ultrasonic sensors. Technologically, the vehicle integrates a self-developed perception large model and point-cloud visual fusion technology, enabling deep integration of point-cloud and image data. This expands the perception range by 19 times compared to traditional solutions and triples overall perception performance. Additionally, it innovatively adopts a "lightweight map + strong perception" approach, eliminating reliance on high-definition maps and instead using onboard perception systems to achieve real-time mapping and environmental adaptation, significantly enhancing navigation capabilities in complex settings.

Engineering deployment results: To date, JD's unmanned delivery vehicles have been deployed at scale across more than 30 cities, over 200 parks, and communities nationwide, completing over 100 million last-mile deliveries. In challenging conditions—including narrow city streets, nighttime operations, strong backlighting or glare, light rain, and light snow—the perception accuracy reaches 99.2%, with a significant improvement in passage success rates under complex road conditions. Equipment failure rates have decreased by 40% year-on-year, effectively addressing industry challenges such as perception failure and operational lag in complex environments. The system now enables stable, round-the-clock, all-scenario delivery operations.[22]

5.2 Meituan's Autonomous Delivery Vehicle Perception Technology Application Case

Meituan's unmanned delivery vehicles primarily target high-frequency short-distance delivery scenarios in urban communities, villages within cities, and commercial districts. To address the extremely complex conditions of Chinese urban villages—such as narrow roads, chaotic obstacles, dense pedestrian and non-motorized traffic, and irregular road conditions—the company has developed a customized multi-sensor fusion perception solution.

The hardware architecture adopts the industry-standard "1 primary, 2 auxiliary" LiDAR layout. The main LiDAR is responsible for high-precision 3D modeling of the core 120-degree forward driving area, while two auxiliary LiDARs cover blind spots on the vehicle's sides and rear. Combined with 11 high-definition cameras, the system enables full-scenario semantic recognition. High-precision millimeter-wave radars are integrated to detect dynamic object speeds and predict trajectories, and ultrasonic sensors are deployed around the vehicle body to provide fallback protection during close-range navigation in narrow spaces. Through feature-level multi-sensor fusion algorithms, the system achieves real-time calibration by integrating 3D point cloud data, image semantics, and velocity information, significantly enhancing detection capabilities for low obstacles, temporary roadside clutter, and sudden pedestrian crossings.

Engineering implementation results: Meituan's autonomous vehicles have been fully deployed in urban villages and old residential communities across first-tier cities such as Shenzhen, Guangzhou, and Shanghai. They achieve a 98% passage success rate in complex urban village environments, precisely adapting to unstructured road conditions and densely populated dynamic 人流 scenarios, demonstrating exceptional stability day and night. By the end of 2025, Meituan's autonomous delivery service had surpassed 800 million cumulative orders, validating the scalability, adaptability, and reliability of its perception technology. It has become a benchmark solution for autonomous delivery perception in China’s complex urban environments.。

5.3 Nuro's Autonomous Delivery Vehicle Perception Technology Application Case in the United States

Nuro is a pioneer in the global low-speed autonomous delivery sector, focusing on lightweight unmanned delivery in open urban environments overseas. Its technology emphasizes high precision, high reliability, and adaptability to harsh weather conditions, with perception capabilities leading globally. Nuro's new generation of autonomous delivery vehicles abandons traditional mechanical LiDAR systems in favor of compact solid-state LiDARs, eliminating potential mechanical failure risks and enhancing long-term stability for outdoor operations. Additionally, they are equipped with high-definition imaging millimeter-wave radars, enabling long-range, high-precision detection of small objects and ensuring stable perception of minute obstacles even in extreme weather such as heavy rain, dense fog, or sandstorms.

The vehicle’s perception system employs a deeply integrated solution combining solid-state LiDAR, imaging millimeter-wave radar, and high-definition vision. Through proprietary fusion algorithms, it achieves millisecond-level synchronization of multi-source data, with key optimizations in predicting fast-moving dynamic targets, identifying distant obstacles, and resisting interference in extreme conditions. This allows for high-precision perception across all scenarios without relying on ultrasonic sensors, delivering industry-leading robustness and safety in perception performance.。

Engineering implementation results: Nuro's autonomous vehicles have obtained open-road autonomous operation permits in multiple U.S. states, enabling legal operation on urban public roads. They achieve stable delivery operations around the clock and in all weather conditions, with a 35% improvement in perception accuracy over traditional solutions under extreme or adverse weather conditions. The equipment downtime rate is below 0.5%, fully demonstrating the practical value of high-precision multi-sensor fusion technology and providing a reference direction for the evolution of advanced sensing solutions in the industry.[23]

6 Industry statistics and market development analysis

6.1 Overall market size of the industry

With the support of smart logistics policies and continuous advancements in sensing technologies, China's unmanned delivery vehicle industry has entered a period of rapid large-scale deployment. According to authoritative data from the "Blue Book on the Development of Low-Speed Autonomous Driving Industry (2025 - 2026)," China's unmanned delivery vehicle market surpassed 10 billion yuan in 2024, growing by 50% year-on-year to reach 15 billion yuan in 2025. In 2025, cumulative outdoor unmanned delivery vehicle shipments in China exceeded 39,000 units, with 27,000 newly delivered throughout the year—marking a significant leap compared to previous years. Industry institutions forecast that by 2030, China's unmanned delivery vehicle market will surpass 41 billion yuan, maintaining an annual compound growth rate above 35%, indicating substantial development potential.

Globally, in 2025, L4-level unmanned delivery vehicle shipments reached 30,000 units, with China accounting for over 40%, making it the world's largest market for both consumption and R&D of unmanned delivery vehicles. The three core application scenarios are last-mile express delivery, supermarket retail, and campus logistics, which account for 52%, 28%, and 15% respectively, while the remaining 5% covers niche areas such as healthcare and sanitation.

6.2 Analysis of the Perception Technology Market Landscape

The current industry has developed three mainstream sensing technology approaches, each tailored to different market segments: the first is the high-end fully integrated approach, represented by Nuro and JD's premium models, which employ a fusion solution combining solid-state LiDAR, high-definition millimeter-wave radar, and multi-camera vision systems. This approach offers high sensing accuracy and strong stability, suitable for open-road scenarios and targeting the high-end commercial market. The second is the mid-range cost-effective approach, exemplified by Meituan’s mainstream models, using a combination of mechanical LiDAR, stereo vision, and standard millimeter-wave radar. This balances precision with cost efficiency, making it ideal for complex urban community environments and currently the dominant mass-production solution in the industry. The third is the low-end pure vision approach, used in small-scale campus delivery robots, relying on multi-camera and ultrasonic sensor configurations. With extremely low costs, this solution suits simple, enclosed environments and targets lower-tier, underserved markets.

In terms of market share, as of 2025, the domestic autonomous delivery vehicle landscape has begun consolidating, with Neolix, Jiu Shi Intelligence, JD, and Meituan emerging as core leading players. Together, these four companies account for over 90% of the market. All have successfully achieved large-scale deployment of multi-sensor fusion perception technologies. Single-sensor solutions—pure vision or pure radar—are gradually being phased out of the mainstream, and multi-sensor fusion has become an essential technical requirement across the industry.

6.3 Current Status of Sensing Technology Industrialization

With continuous iteration of domestically produced sensing hardware and open-source optimization of fusion algorithms, the cost of perception systems for autonomous delivery vehicles has steadily declined, significantly accelerating industrialization and deployment. The localization rate of LiDAR sensors rose from 40% in 2022 to 75% in 2025, while overall sensor hardware costs dropped by 52% year-on-year, laying a solid foundation for large-scale production. Meanwhile, perception algorithms continue to be lightweighted and optimized for embedded deployment, enabling compatibility with low-computational-power onboard terminals. This addresses the traditional challenges of high computational demands and difficult implementation associated with fusion algorithms, greatly enhancing engineering practicality.

Currently, the industry has established an industrialization trend characterized by "domestically sourced hardware, lightweight algorithms, standardized fusion, and full-scenario coverage." Sensing technology is no longer a bottleneck for autonomous delivery vehicle deployment. Instead, adapting to complex environments, resisting interference in extreme weather conditions, and achieving a balance between low cost and high precision have become the key focus areas for ongoing technological evolution.[24]

7 Core Challenges in Environment Perception Technology for Autonomous Delivery Vehicles

7.1 Insufficient Robustness in Perceiving Complex and Extreme Scenarios

Current perception technologies for autonomous delivery vehicles perform stably under normal conditions but still exhibit significant shortcomings in highly complex and extreme scenarios. First, adaptability to extreme weather is poor: during heavy rain, dense fog, snowstorms, or severe sandstorms, LiDAR point cloud noise increases dramatically, visual imaging becomes blurred, and millimeter-wave radar suffers from clutter interference, leading to a sharp decline in target recognition accuracy. Second, visual semantic recognition is severely affected by complex lighting conditions—such as strong backlighting, flickering streetlights at night, or alternating shadows from trees—resulting in frequent false positives and missed detections. Third, the system struggles with detecting small, scattered objects; fine obstacles such as loose stones, debris, low barriers, or overhead cables are often overlooked, creating blind spots that pose safety risks.

7.2 Limited Synchronization Accuracy in Multi-Sensor Fusion

Multi-sensor fusion relies on precise time synchronization and spatial calibration. However, current engineering implementations still suffer from synchronization errors. Differences in sampling frequencies and response delays among sensors lead to millisecond-level timing discrepancies in data acquisition from LiDAR, cameras, and millimeter-wave radars, causing misalignment when fusing multi-source data. Additionally, long-term vehicle vibrations and outdoor variations in temperature and humidity can cause sensor mounting shifts, resulting in spatial calibration drift. This ultimately reduces fused perception accuracy, leading to target localization errors and identification confusion, thereby compromising driving stability.

7.3 Difficulty Balancing Cost and Precision

High-precision perception hardware—such as solid-state LiDAR and high-definition imaging millimeter-wave radars—is expensive, significantly increasing production costs for autonomous vehicles and hindering industry-wide adoption. In contrast, low-cost sensors offer poor precision and reliability, making them unsuitable for complex urban environments. The industry currently faces a dilemma: "high precision at high cost, low cost with low precision." How to improve perception accuracy through algorithm optimization and architectural iteration while keeping hardware costs under control remains a key challenge for industrial deployment.

7.4 Weak Algorithm Generalization Capability

Current mainstream perception fusion algorithms are primarily trained on conventional road datasets, resulting in insufficient training samples for complex localized scenarios in China—such as narrow alleys in urban villages, temporary road construction, disordered non-motorized traffic, and irregular obstacles—leading to poor algorithm generalization. These systems often fail to recognize unfamiliar environments or unconventional obstacles, making them unsuitable for the diverse and complex last-mile delivery conditions in China, with algorithm iteration lagging behind real-world deployment demands.

7.5 Insufficient Hardware Reliability and Durability

Unmanned delivery vehicles operate long-term in outdoor, open-air environments exposed to extreme temperatures, humidity, wind, sand, and rain. Traditional perception sensors have limited waterproofing, dustproofing, and shock resistance, leading to hardware aging, performance degradation, and frequent failures over time. Equipment durability and adaptability to outdoor conditions require further improvement, increasing maintenance costs and downtime losses.

8 Technical Optimization Strategies and Future Development Trends

8.1 Targeted Technical Optimization Strategies

8.1.1 Enhancing Perception Algorithms for Complex Scenarios

To address challenges such as extreme weather, complex lighting conditions, and small object detection, a localized dataset for complex scenarios should be constructed, enriched with samples of rain, snow, fog, strong light, shadows, and cluttered obstacles. The structure of deep learning models should be optimized. Large-scale models can be introduced to enhance scene semantic understanding, improving recognition and prediction capabilities for irregular obstacles and unknown environments, thereby strengthening algorithm generalization and robustness.

8.1.2 Improving Multi-source Data Synchronization and Calibration Accuracy

Establish a unified hardware clock synchronization mechanism and optimize timestamp calibration algorithms to eliminate timing discrepancies among multiple sensors. Design flexible anti-vibration mounting structures suitable for vehicle vibration conditions, perform periodic spatial calibration of sensors to correct installation offset errors, and refine Kalman filter fusion algorithms to strengthen anomaly detection, thus enhancing accuracy and stability in multi-source data fusion.

8.1.3 Building a Low-cost, High-precision Perception Architecture

Replace imported hardware with domestically produced sensors to reduce procurement costs. Optimize sensor layout by eliminating redundant components and adopting a differentiated configuration of "high-precision primary sensors + low-cost auxiliary sensors." Leverage algorithm lightweighting, intelligent data completion, and virtual sensing technologies to compensate for precision limitations of low-cost hardware, achieving an optimal balance between cost and performance.

8.1.4 Strengthening Outdoor Durability of Hardware

Select automotive-grade sensors with waterproof, dustproof, and shock-resistant properties, and optimize sensor sealing and heat dissipation designs. Implement real-time hardware health monitoring modules to detect precision degradation and faults early, enabling proactive maintenance, extending hardware lifespan, and reducing operational and maintenance costs.

8.2 Future Technological Development Trends

8.2.1 Integrated and Standardized Perception Fusion

Future unmanned delivery vehicles will move beyond single-sensor perception solutions. Deep integration of multiple sensors will become industry standard, gradually establishing unified standards for hardware layout, data fusion, and algorithm processing. This modular, standardized, and universal approach will significantly reduce R&D and mass production costs, accelerating industry-wide scale adoption.

8.2.2 Deep Empowerment by Perception Large Models

Perception large models will play a pivotal role in future development, providing powerful capabilities in scene understanding, context reasoning, and dynamic adaptation. They will enable more accurate and reliable perception in complex and unpredictable environments, driving innovation across autonomous delivery systems.

8.2.3 Hardware lightweighting, solid-state evolution, and domestication

LiDAR will fully transition from mechanical to solid-state sensors, achieving miniaturization, low power consumption, and high stability; millimeter-wave radar will upgrade toward high-definition imaging and higher resolution; vision cameras will deliver high dynamic range and ultra-clear night vision. Meanwhile, core perception hardware will be entirely replaced with domestically produced alternatives, significantly reducing equipment costs and accelerating the widespread adoption of unmanned delivery vehicles.[25]

8.2.4 Vehicle-Infrastructure Cooperative Perception Fusion Upgrad

Future unmanned delivery vehicles will no longer rely solely on autonomous onboard sensing. Instead, they will deeply integrate with smart roads, roadside perception devices, and cloud-based data platforms to achieve full-domain collaborative perception through "onboard sensing + roadside sensing + cloud-based dispatching." This integration enables early access to long-range traffic conditions and risk warnings, effectively eliminating blind spots and prediction delays in single-vehicle perception, thereby greatly enhancing driving safety and traffic efficiency.

8.2.5 Low-power edge-side intelligent deployment

By leveraging algorithm distillation, lightweight pruning, and computational optimization, high-precision fusion algorithms and large-model algorithms are efficiently deployed on low-computing-power in-vehicle terminals. This reduces the power consumption of perception systems, adapts to the power supply constraints and limited computing capabilities of autonomous vehicles, and enhances device endurance and long-term stable operation.

9 Conclusion

This paper focuses on environment perception technology for unmanned delivery vehicles, systematically organizing the overall framework of such technology. It thoroughly analyzes the working principles, core formulas, performance parameters, advantages and disadvantages, and application scenarios of four single-sensor perception technologies: LiDAR, visual cameras, millimeter-wave radar, and ultrasonic sensors. Particular emphasis is placed on the layered architecture, core algorithms, implementation processes, and technical advantages of multi-sensor fusion perception. Drawing on real-world engineering deployment cases from three leading companies—JD.com, Meituan, and Nuro—and supported by the latest 2025 industry market statistics, the study comprehensively examines the current state of engineering applications, market landscape, and industrialization trends in unmanned delivery vehicle perception technology. It precisely identifies key bottlenecks in existing technological applications, proposes targeted optimization strategies, and forecasts future directions of technological evolution.

The research concludes that due to physical limitations, each individual perception technology has inherent shortcomings in scenario adaptability and cannot meet the demands of all-weather, high-precision, and high-safety perception required for complex urban last-mile deliveries. Multi-sensor fusion technology, by integrating the multidimensional strengths of various sensing hardware, enables comprehensive coverage in spatial awareness, semantic understanding, velocity estimation, and close-range detection, representing the optimal solution at present and a core trend in the industry. The sector has now entered a phase of large-scale deployment, with the primary focus of perception technology iteration shifting from "having or lacking perception" to "precise, intelligent, low-cost, and full-domain perception." Key breakthroughs in the future will hinge on adapting to complex environments, algorithm generalization capability, balancing cost and precision, and achieving integration between vehicles and infrastructure.

With continuous advancements in artificial intelligence, sensor hardware, big data, and vehicle-infrastructure coordination, the environment perception systems of unmanned delivery vehicles will rapidly evolve toward integration, intelligence, standardization, full-domain coverage, and low cost. This will fundamentally resolve challenges in perceiving complex environments, significantly enhance the safety, stability, and efficiency of autonomous delivery operations, and continuously drive the intelligent and unmanned transformation of last-mile logistics. It will provide solid technical support for high-quality development in smart logistics and digital city construction. Furthermore, the findings of this study offer systematic theoretical foundations and practical references for mid-level engineering professionals engaged in R&D, optimization, and deployment of intelligent unmanned equipment, demonstrating significant value in both engineering applications and academic research.□

REFERENCES

[1] Ge Xinjian. Research on Navigation and Obstacle Avoidance Technology for Unmanned Delivery Vehicles in Campus Environment [D]. Jinan: Shandong Jianzhu University, 2024.

[2] Ge Xinjian. A Review of Intelligent Driving Technology in Outdoor Environments [J]. Computer Measurement & Control, 2024, 32(10): 1-8.

[3] Li Yule, Zhang Xiaoyu, Liu Na. Research on an Improved Scheme for Intelligent Campus Logistics Management Based on Unmanned Delivery Vehicles [J]. Logistics Technology, 2025, 48(13): 57-61.

[4] Research on Environmental Perception Algorithms for Low-Speed Unmanned Logistics Vehicles Based on LiDAR [D]. 2020.

[5] Research on the Design of Unmanned Delivery Vehicles Based on Context Awareness [D]. Nanjing: Southeast University, 2021.

[6] Vermesan O, Bahr R, Sand H E, et al. Advancing edge AI perception platforms and sensor fusion for last-mile delivery autonomous vehicles[M]//AI and Edge Computing for Autonomous Vehicles. London: Taylor & Francis, 2026.

[7] Du H, Ren L, Wang Y D, et al. Advancements in perception system with multi-sensor fusion for embodied agents[J]. Information Fusion, 2025, 115: 102637.

[8] Sensor systems for autonomous vehicles: functionality and reliability challenges in adverse environmental conditions[J]. Measurement, 2025, 242: 116538.

[9] The automation concept of the U-Shift II vehicle[J]. SAE Technical Papers, 2025, 1: 2025-01-7023.

[10] LogisticsVISTA: 3D terminal delivery services with UAVs, UGVs and USVs based on foundation models and scenarios engineering[C]//2024 IEEE Intelligent Vehicles Symposium (IV). Macau: IEEE, 2024.

[11] In-Depth Analysis of the Core Technology System for Autonomous Delivery Vehicles: Full-Chain Technological Breakthroughs from Perception to Energy[EB/OL]. (2025-08-15)[2026-07-01]. https://four-faith.com/jishuyingyong/1653.html.

[12] Upstream Technology Analysis of the Autonomous Delivery Vehicle Industry Chain[EB/OL]. (2025-08-01)[2026-07-01]. https://www.automotiveworld.cn.

[13] Unmanned Vehicles Toward Embodied Intelligence: Perception Enhancement and End-to-End Decision-Making [J]. New Electronics Technology Magazine, 2025, (12): 1 - 10.

[14] Liu Yang, Lei Ting. Design of an Intelligent Multifunctional Campus Unmanned Logistics Vehicle [J]. Mechanical & Electrical Products Development and Innovation, 2024, 37(5): 53 - 57.

[15] Zhang Yuzheng. Exploring the Role of Artificial Intelligence in Modern Logistics Management Models [J]. Logistics Technology, 2022, 45(10): 105 - 107.

[16] Xue Xiaowei. Research on the Application of Artificial Intelligence in Modern Logistics Systems [J]. China Logistics and Procurement, 2024, (9): 106 - 107.

[17] Autonomous Vehicles and Drones: 5G Accelerates the Arrival of Intelligent Logistics Era [J]. Logistics Technology, 2019, 42(7): 1 - 10.

[18] Yang Ping, Tu Guangming. Exploring Innovative Directions for Modern Logistics Management in the Age of Artificial Intelligence [J]. Logistics Engineering and Management, 2021, 43(3): 68 - 70.

[19] Intelligent Mobile Robot Systems Facilitate Smart Upgrading of Factory Logistics [J]. Automation Expo, 2020, 37(7): 36 - 40.

[20] Xiao Gangchao, Zhu Xiaoming, Li Zubiao, et al. Research and Application of a New Flexible Distribution Service Model for Huaihua Tobacco Logistics [J]. China Market, 2022, (31): 173 - 177.

[21] Yao Changqing. A Process-Based Strategy for Flexible Logistics Design in Tobacco Distribution Centers [J]. Chongqing and the World (Academic Edition), 2015, 32(12): 138 - 141.

[22] Ma Libing. Artificial Intelligence Drives Transformation of Modern Logistics Enterprises—Taking Warehouse-Based Logistics Companies as Examples [J]. China Storage and Transportation, 2019, (9): 142 - 144.

[23] Li B, Liu S S, Tang J, et al. Autonomous last-mile delivery vehicles in complex traffic environments[J/OL]. arXiv preprint, 2020, arXiv:2001.08096. https://arxiv.org/abs/2001.08096.

[24] Meituan Unmanned Delivery Department. Visual Global Localization Technology Based on Feature Points[EB/OL]. (2019-04-26). https://cloud.tencent.com/developer/article/1418690.

[25] Tsinghua University, Huazhong University of Science and Technology, Beijing SanKuai Online Technology Co., Ltd. (Meituan), Beijing Jingdong Ganshi Technology Co., Ltd. (JD.com), et al. Intelligent Perception and Navigation Control Technology for Heterogeneous Autonomous Unmanned Systems in Complex Multi-Domain Environments [R]. Beijing: China Association of Automation Scientific and Technological Achievements Appraisal Conference,2024.

Author Introduction: Zhou Qiu, male, engineer, currently engaged in automation engineering design and sales.

Received: 2026-04-09
Revised: 2026-05-18
Accepted: 2026-05-20