

Advancements in Autonomous Driving Systems: Enhancing Environmental Perception with Lidar Integration and Innovative Algorithms

Chen Lu^{1,a,*}

¹*School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China*

a. u202110869@hust.edu.cn

**corresponding author*

Abstract: This paper discusses the intricacies of autonomous driving systems, focusing on the four core modules: perception, sensing, decision-making, and control. The sensing module integrates various sensors like LiDAR, cameras, millimeter-wave radars, IMU, and GPS to gather essential data. The paper emphasizes the role of LiDAR in 3D object detection and point cloud segmentation, which are crucial for environmental perception. It also addresses the challenge of fusing data from multiple sensors into a unified coordinate system for effective decision-making. This research presents several significant advances in environmental perception. First there is a single-stage 3D object detection algorithm using sparse convolutional neural networks for improved speed and accuracy. Moreover there is a two-stage 3D object detection method focuses on detecting difficult targets through target center prediction. This method employs a novel neural network architecture, AOAP-SCNN, for feature extraction. Furthermore there is a point cloud instance segmentation algorithm (a method that separates and identifies individual objects in 3D space) utilizes a front view approach and a DLA network with Tiny-PointNet for feature extraction. Finally there is also an automatic calibration algorithm for LiDAR and camera systems based on segmentation, designed for real-time operation on embedded platforms, which corrects the extrinsic matrix and improves the self-calibration system's practical value. Building upon these technical innovations, the paper provides a comprehensive analysis of intelligent vehicle technology including the history of development, the importance of environmental perception, and the application and limitations of LiDAR in intelligent vehicles.

Keywords: Autonomous Driving Systems, Environmental Perception, LiDAR.

1. Introduction

Autonomous driving represents the pinnacle of modern automotive innovation, promising a future where vehicles navigate with precision and safety without human intervention. At its core, this technology represents a sophisticated system that goes beyond traditional driving capabilities. It achieves this by integrating advanced sensors, algorithms, and computing power to handle real-time environmental interactions. This introduction sets the stage for a comprehensive exploration of the autonomous driving system, focusing on the critical aspects that enable vehicles to operate with a

high degree of autonomy. The autonomous driving system consists of three major modules that work in concert: environmental perception (including sensing), decision-making, and control. The perception module is paramount, as it involves the collection and interpretation of multimodal data from an array of sensors, including Light Detection and Ranging (LiDAR), cameras, millimeter-wave radars, Inertial Measurement Units (IMU), and Global Positioning System (GPS). These sensors feed a wealth of information into the system, allowing it to construct a detailed understanding of the vehicle's surroundings.

Among these sensors, LiDAR plays a crucial role by providing high-resolution, three-dimensional data about the environment. And central to this paper is the discussion of LiDAR, a sensor that provides high-resolution, three-dimensional data about the environment. LiDAR's capabilities extend beyond simple object detection, offering detailed semantic segmentation and instance segmentation of point clouds. This level of detail is crucial for the accurate estimation of object positions and postures, and for generating precise bounding boxes that aid in sensor fusion and system calibration.

However, the effectiveness of an autonomous driving system is not solely dependent on the quality of sensor data but also on the algorithms that process this data. This paper delves into the development of advanced algorithms that enhance the speed and accuracy of environmental perception through 3D object detection and point cloud segmentation. This paper introduces a single-stage 3D object detection algorithm based on sparse convolution, which offers a significant improvement over traditional methods. Additionally, this paper presents a two-stage detection algorithm that leverages a novel neural network architecture to detect challenging targets with greater accuracy.

The paper also addresses sensor calibration by proposing an automatic algorithm that optimizes the extrinsic parameter matrix for LiDAR-camera fusion. This algorithm is designed for real-time operation on embedded platforms, ensuring that the fusion of sensor data is as accurate as possible. In the realm of intelligent vehicles, environmental perception is not just a feature but a foundational element that enables advanced decision-making and control. This paper first examines the fundamental technology of environmental perception, then analyzes key sensor principles, and finally proposes innovative concepts in sensor fusion for obstacle detection. Before examining these technical aspects in detail, it is important to reflect on the historical development of intelligent vehicles, highlighting the milestones that have shaped the field and the ongoing challenges that researchers and engineers continue to address. The journey from concept to practical implementation is fraught with technical and logistical hurdles, but it is also a testament to the potential of autonomous driving to transform transportation as people know it. This paper aims to contribute to this evolving landscape by presenting new research findings and technical solutions that push the boundaries of what is possible in the world of autonomous driving.

2. Advancements in Autonomous Driving Systems

Autonomous driving is a sophisticated system that encompasses four major modules: perception, sensing, decision-making, and control. The sensing module, as depicted in Figure 1, is equipped with an array of sensors including LiDAR, cameras, millimeter-wave radars, IMU, and GPS. These sensors provide crucial multimodal information for autonomous driving systems. The information includes LiDAR point clouds, color images, infrared images, millimeter-wave radar data, vehicle acceleration, and position.

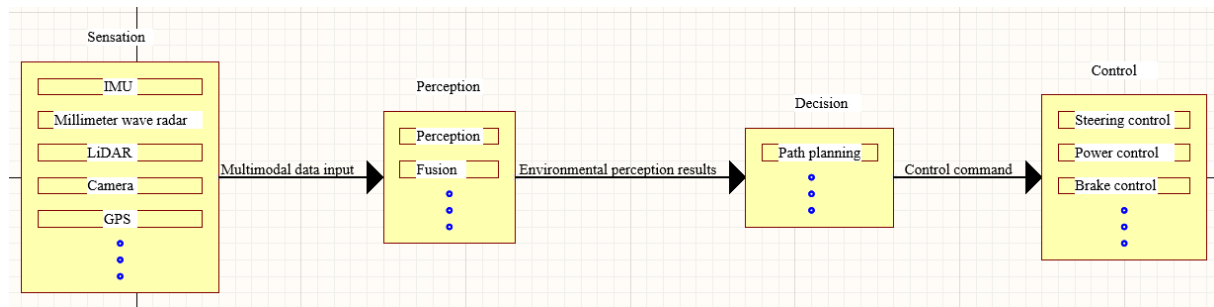


Figure 1: The structure of autonomous driving system [1]

LiDAR's 3D object detection capabilities allow for the identification, localization, and estimation of the posture of targets in the environment, thereby estimating the 3D bounding box of these targets, as shown in Figure 2. Furthermore, LiDAR's 3D point cloud semantic segmentation function enables the classification of each point in the scene point cloud. Point cloud instance segmentation distinguishes individual point clouds and provides refined target classification by excluding background points. This process generates irregular bounding boxes that enclose the target, which is beneficial for sensor fusion and calibration.

As an autonomous driving system is a multi-sensor system, each sensor perceives the environment in its own coordinate system. These perception results need to be transformed into the same coordinate system for result fusion. Typically, for the LiDAR subsystem result fusion algorithm, the perception results of the LiDAR are translated to the vehicle or camera coordinate system using the LiDAR relative camera extrinsic matrix. They are then fused and summarized with the perception results of other sensors before being sent to the decision module. This extrinsic matrix is derived from the calibration of the camera and LiDAR. The accuracy of this matrix directly impacts the fusion effect. Only the correct extrinsic matrix can accurately project the point cloud onto the camera image, allowing for the fusion of the image and point cloud detection results.

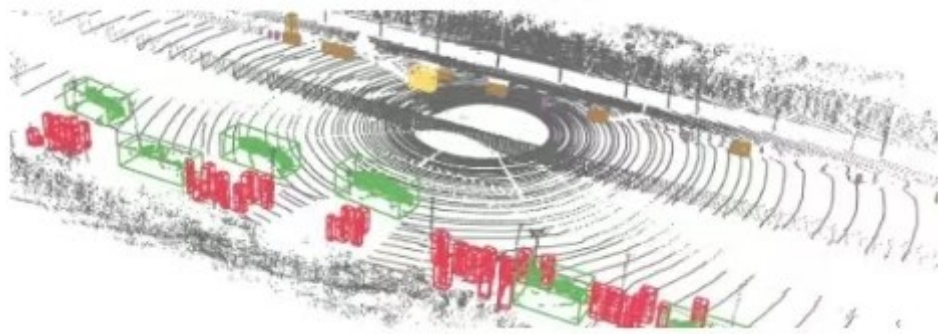


Figure 2: The figure of 3D environmental perception and calibration [1]

Environmental perception systems are vital components of autonomous driving. They collect multimodal data from various sensors, analyze and understand the surrounding environment, fuse the perception results, and forward them to the decision-making module. This serves as an essential basis for guiding vehicle travel. LiDAR, being one of the most critical sensors in the environmental perception system, provides point cloud information of the surrounding scene for vehicle navigation. LiDAR-based 3D detection and instance segmentation enable vehicles to better understand their surrounding environment. The perception results of LiDAR need to be integrated with the perception results from cameras and other sensors to compile the information on the targets. The calibration technology of multiple sensors is a significant factor affecting the fusion effect. Environmental

perception systems still face challenges in meeting autonomous driving requirements, particularly in terms of accuracy, fusion quality, and computing efficiency.

To enhance the speed and accuracy of environmental perception for autonomous driving, this paper conducts a theoretical analysis, algorithm design, technical implementation, and practical verification on key technologies of environmental perception, namely three-dimensional object detection, three-dimensional point cloud instance segmentation, and automatic calibration algorithms between cameras and LiDAR. The main research contents include:

(1) A single-stage 3D object detection algorithm is proposed, which utilizes sparse convolution to balance efficiency and accuracy through a fully sparse convolutional neural network. Instead of employing the anchor box detection method, it uses foreground segmentation, predicting multiple bounding boxes in different directions for each foreground point. Additionally, a new object bounding box encoding method is introduced, representing the object bounding box as two mutually perpendicular line segments passing through the foreground point, and indirectly calculating the object's bounding box by predicting the offset of the segment endpoints relative to the foreground point. Experiments on the KITTI dataset demonstrate that this proposed single-stage 3D object detection algorithm based on sparse convolution significantly improves both speed and accuracy compared to mainstream algorithms.

(2) A two-stage three-dimensional object detection algorithm based on sparse convolution is proposed to address the low detection accuracy of weak and small difficult targets in three-dimensional object detection. This method is designed based on the target center, predicting the position of the target center on the bird's-eye view, calculating the features on each target center, and finally predicting the bounding box of the target on each center. To efficiently calculate the features of the target center, an Assignable Output Active Point Sparse Convolutional Neural Network (AOAP-SCNN) is proposed. The AOAP-SCNN extracts multi-scale features from three different feature map scales for each target center. Compared with the single-stage three-dimensional object detection algorithm based on sparse convolution, this method significantly improves detection indicators for medium and difficult targets on the KITTI dataset and can process point clouds in real-time.

(3) A point cloud instance segmentation algorithm based on the front view is proposed to tackle the poor segmentation accuracy in fast instance segmentation algorithms. The algorithm restores the point cloud to the native range view (NRV) as the front view and then uses front view segmentation to achieve point cloud segmentation. A DLA network embedded with Tiny-PointNet is designed for the feature extraction network. This network uses a simplified PointNet network to extract local features of the point cloud in the front view, while the DLA network performs multi-scale feature extraction. After the features output by the two are superimposed and fused, the multi-task head network predicts the foreground score, center offset, target size, and target orientation for each pixel point on the fused features. Finally, a clustering algorithm is used to cluster the predicted target centers, with each category representing an instance. For training purposes, the point cloud instance segmentation dataset was derived from the KITTI 3D object detection dataset. The results show that the algorithm proposed in this paper is superior to contemporary algorithms in both segmentation accuracy and speed.

(4) A calibration algorithm between LiDAR and camera based on point cloud segmentation and image segmentation is proposed. Existing automatic calibration algorithms often suffer from large computational requirements and long processing times. To address these issues, this paper presents an automatic calibration system for cameras and LiDAR, implemented on the Hisilicon Hi3559AV100 embedded platform. The system enables real-time calibration across diverse driving scenarios and operating conditions. The method is based on point cloud and image segmentation, generating a three-dimensional conical frame enclosing the target point cloud and a bounding box of

the two-dimensional target segmentation mask. The algorithm generates 2D-3D point correspondences through virtual point mapping, then derives the camera-LiDAR extrinsic matrix using PnP. Additionally, a model transformation and inference method for the automatic calibration algorithm, specifically designed for the Hisilicon Hi3559AV100 embedded platform, is proposed. Simulation experiments and outdoor experiments on the KITTI dataset demonstrate that the automatic calibration algorithm proposed in this paper plays a role in correcting the extrinsic matrix of the camera and LiDAR, and the self-calibration system has application value.

Environmental perception in intelligent vehicles involves real-time data acquisition from multiple sensors, enabling various recognition tasks such as lane detection and obstacle identification. It is the foundation of intelligent vehicle decision-making and control and a key link in achieving intelligent driving. This paper systematically studies the environmental perception technology of intelligent vehicles, deeply analyzes the application principles and control strategies of several important sensors such as machine vision, millimeter-wave radar, LiDAR, and infrared sensors, and analyzes their advantages and disadvantages. Based on the application methods of sensors in different environments, a sensor information fusion-based obstacle detection concept is proposed, and the method of improving the environmental perception ability of intelligent vehicles through information fusion is elaborated in detail.

Intelligent vehicles represent an evolving concept in automotive technology, encompassing a comprehensive system with multiple advanced functions. Intelligent vehicles are built upon conventional vehicles by incorporating advanced information perception systems, sophisticated control systems, and reliable actuators. Intelligent information exchange is realized through onboard sensing systems and information terminals. Intelligent environmental perception enables automatic analysis of the vehicle's state, ultimately aiming to replace manual operation with autonomous driving capabilities.

The development of intelligent vehicles has a rich history, with significant milestones shaping the field. The concept of unmanned driving originated from the DARPA Grand Challenge launched by the U.S. Department of Defense, attracting many research institutions and universities such as Stanford University, Carnegie Mellon University, and FMC Corporation, greatly promoting the development of intelligent vehicles. Stanford University's approach combined multi-directional LiDARs for road surface detection with visual cameras for trajectory monitoring, and ultimately won the unmanned vehicle challenge championship in the shortest time [2,3]. Google started intelligent driving research in 2009 and quickly became one of the benchmarks in the field of unmanned vehicles [4]. In China, Tsinghua University's series unmanned vehicle THMR, designed with quasi-structured and unstructured road environments as the research background, has high decision-making and planning capabilities [5]. The Red Flag Q3 unmanned vehicle of the National University of Defense Technology, the JLUIV and DLIUV series intelligent vehicles of Jilin University [6,7]. And the intelligent vehicles developed by Shanghai Jiao Tong University, Chang'an University, and Hunan University, can also achieve positioning, navigation, obstacle avoidance, and following functions [8,9]. Among the key technologies in intelligent driving - perception, decision-making, and planning - environmental perception serves as the cornerstone, aiming to detect and recognize surrounding roads, vehicles, obstacles, traffic lights, etc., and is a key link in intelligent vehicles [10]. This paper analyzes the working principle, application background, and advantages and disadvantages in the application process of LiDAR sensors.

Despite its significant advantages, LiDAR technology faces challenges in intelligent vehicles primarily due to optical system sensitivity. However, the obstacle information detected by LiDAR can directly obtain the three-dimensional data description of the environment. Therefore, the research on LiDAR distance imaging technology has been valued in the field of intelligent vehicles and is widely used in obstacle detection, acquisition of three-dimensional environmental information, and

vehicle obstacle avoidance. LiDAR is an active sensor that detects targets by emitting laser beams [11]. When the emitted signal encounters a target, it generates a backscatter signal. Then, LiDAR begins to collect the target backscatter, and after signal processing by the target detection algorithm, information such as the target's distance and direction can be obtained. Based on laser beam configuration, LiDAR systems are classified into two types: front-mounted single-line LiDAR for fast 2D detection, and roof-mounted multi-line LiDAR for comprehensive 3D scanning. The first one is the single-line LiDAR placed in front of the vehicle, which can be used to detect objects in front of the vehicle and has a fast measurement speed, but can only generate two-dimensional data, and the other is the multi-line LiDAR placed on the roof of the vehicle, which can obtain targets by rotating scanning in 360°. While more beams provide more detailed target information, they also generate larger point clouds that demand greater processing and storage capabilities. LiDAR measures distance using Time of Flight (ToF) principles. This method can be categorized into two types: direct and indirect measurement. The direct method measures the time difference of the laser pulse traveling between the target and the radar. The indirect method calculates the phase difference between the received and emitted waveforms to infer distance information. This paper mainly introduces the principle of direct distance measurement. The relationship between the distance H (unit: meters) of the target object and the LiDAR and other influencing factors is shown in Equation 1.

$$H = \frac{c \times t_d}{2} \quad (1)$$

Where: c is the speed of light; t_d is the time difference between the emission time and reception time recorded by the emitter, unit: s. from this, the three-dimensional spatial coordinates of the target object can be calculated, allowing for a comprehensive understanding of its spatial relationship within the environment. Through the above three-dimensional spatial coordinate relationship, the distance relationship between the LiDAR and the target object can be calculated as shown in the following equations.

$$x = H \times \cos \theta \times \sin \alpha \quad (2)$$

$$y = H \times \cos \theta \times \cos \alpha \quad (3)$$

$$z = H \times \sin \theta \quad (4)$$

From this, the three-dimensional coordinate data of the target object can be obtained, achieving the detection of obstacles. The process of target detection using LiDAR often requires extensive datasets for algorithm training, which can be both time-consuming and costly to acquire and process. Currently, LiDAR applications primarily focus on offline high-precision mapping by combining 3D environmental data with vehicle positioning, while target recognition applications remain limited.

3. Conclusion

Autonomous driving represents the pinnacle of modern automotive innovation, promising a future where vehicles navigate with precision and safety without human intervention. At the heart of this technology lies a sophisticated system that transcends traditional driving by integrating advanced sensors, algorithms, and computing power to perceive, analyze, and act upon the dynamic environment in real-time. This introduction sets the stage for a comprehensive exploration of the autonomous driving system, focusing on the critical aspects that enable vehicles to operate with a high degree of autonomy.

The autonomous driving system is comprised of four major modules that work in concert: perception, sensing, decision-making, and control. The perception module is paramount, as it involves the collection and interpretation of multimodal data from an array of sensors, including Light Detection and Ranging (LiDAR), cameras, millimeter-wave radars, Inertial Measurement Units

(IMU), and Global Positioning System (GPS). These sensors feed a wealth of information into the system, allowing it to construct a detailed understanding of the vehicle's surroundings.

Central to this paper is the discussion of LiDAR, a sensor that provides high-resolution, three-dimensional data about the environment. LiDAR's capabilities extend beyond simple object detection, offering detailed semantic segmentation and instance segmentation of point clouds. This level of detail is crucial for the accurate estimation of object positions and postures, and for generating precise bounding boxes that aid in sensor fusion and system calibration.

However, the effectiveness of an autonomous driving system is not solely dependent on the quality of sensor data but also on the algorithms that process this data. This paper delves into the development of advanced algorithms for 3D object detection and point cloud segmentation, which are designed to enhance the speed and accuracy of environmental perception. This paper introduces a single-stage 3D object detection algorithm based on sparse convolution, which offers a significant improvement over traditional methods. Additionally, this paper presents a two-stage detection algorithm that leverages a novel neural network architecture to detect challenging targets with greater accuracy.

The paper also addresses the importance of sensor calibration, where this paper proposes an automatic calibration algorithm that corrects the extrinsic matrix between LiDAR and camera systems. This algorithm is designed for real-time operation on embedded platforms, ensuring that the fusion of sensor data is as accurate as possible. In the realm of intelligent vehicles, environmental perception is not just a feature but a foundational element that enables advanced decision-making and control. This paper systematically examines the technology behind environmental perception, analyzing the principles and strategies of key sensors and proposing new concepts in sensor fusion for obstacle detection.

References

- [1] Wang, B. (2022). *Key Technologies for 3D Environmental Perception System of Autonomous Driving Based on LiDAR* [Doctoral dissertation, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences]. <https://doi.org/10.3969/j.issn.1003-8639.2023.06.002>
- [2] Chen, Q., Ozguner, U., & Redmill, K. (2004). Ohio state university at the 2004 darpa grand challenge: Developing a completely autonomous vehicle. *IEEE Intelligent Systems*, 19(5), 8-11. <https://doi.org/10.1109/MIS.2004.48>
- [3] Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., ... & Mahoney, P. (2006). Stanley: The robot that won the DARPA Grand Challenge. *Journal of Field Robotics*, 23(9), 661-692. https://doi.org/10.1007/978-3-540-73429-1_1
- [4] Liu, C. (2016). *Changing the World: The Path of Google's Autonomous Vehicle Development*. *Auto Review*, 5(1), 94-99.
- [5] Wang, Y. (2017). *Design and Implementation of Intelligent Traffic Monitoring System on THMR-V Platform*. *Computer Measurement & Control*, 25(7), 106-109.
- [6] Chen, Z. (2007). Hongqi HQ3: The Combination of Intelligence and Safety. *Science and Technology Trend*, 11(1), 23-27.
- [7] Wang, R., You, F., Cui, G., & Zhang, H. (2004). Road recognition for high-speed intelligent vehicles based on computer vision. *Computer Engineering and Applications*, 26(26), 18-21.
- [8] Jahan, C., & Mandic, D. P. (2014). A class of quaternion Kalman filters. *IEEE Transactions on Neural Networks and Learning Systems*, 25(3), 533-544. <https://doi.org/10.1109/TNNLS.2013.2277540>
- [9] Jiang, Y., Xu, X., & Zhang, L. (2021). Heading tracking of 6WID/4WIS unmanned ground vehicles with variable wheelbase based on model free adaptive control. *Mechanical Systems and Signal Processing*, 159, 107715. <https://doi.org/10.1016/j.ymssp.2021.107715>
- [10] Meng, X., Zhang, C., & Su, C. (2019). A Review of Key Technologies for Autonomous Driving Systems. *Times Auto*, 17(1), 4-5.
- [11] Sun, P. (2019). *Research on the Key Technologies of Precise Perception of Driving Environment for Intelligent Vehicles in Urban Environment*, Chang'an University.