World Electric Vehicle Journal

MDPI

Article

A Review of Environmental Perception Technology Based on Multi-Sensor Information Fusion in Autonomous Driving

Boquan Yang, Jixiong Li * and Ting Zeng

School of Mechanical and Electrical Engineering and Automation, Foshan University, Foshan 528000, China; 2112351106@stu.fosu.edu.cn (B.Y.); 2112451103@stu.fosu.edu.cn (T.Z.)
* Correspondence: lijx@fosu.edu.cn

Abstract: Environmental perception is a key technology for autonomous driving, enabling vehicles to analyze and interpret their surroundings in real time to ensure safe navigation and decision-making. Multi-sensor information fusion, which integrates data from different sensors, has become an important approach to overcome the limitations of individual sensors. Each sensor has unique advantages. However, its own limitations, such as sensitivity to lighting, weather, and range, require fusion methods to provide a more comprehensive and accurate understanding of the environment. This paper describes multi-sensor information fusion techniques for autonomous driving environmental perception. Various fusion levels, including data-level, feature-level, and decision-level fusion, are explored, highlighting how these methods can improve the accuracy and reliability of perception tasks such as object detection, tracking, localization, and scene segmentation. In addition, this paper explores the critical role of sensor calibration, focusing on methods to align data in a unified reference frame to improve fusion results. Finally, this paper discusses recent advances, especially the application of machine learning in sensor fusion, and highlights the challenges and future research directions required to further enhance the environmental perception of autonomous systems. This study provides a comprehensive review of multi-sensor fusion technology and deeply analyzes the advantages and challenges of different fusion methods, providing a valuable reference and guidance for the field of autonomous driving.

Keywords: autonomous driving; multi-sensor information fusion; environmental perception; machine learning

Academic Editor: Grzegorz Sierpiński

Received: 30 October 2024
Revised: 19 December 2024
Accepted: 26 December 2024
Published: 2 January 2025

Citation: Yang, B.; Li, J.; Zeng, T. A Review of Environmental Perception Technology Based on Multi-Sensor Information Fusion in Autonomous Driving. *World Electr. Veh. J.* **2025**, *16*, 20. https://doi.org/10.3390/wevj16010020

1. Introduction

In recent years, autonomous vehicles have shown great potential to improve the efficiency of intelligent transportation systems, enhance road safety, and optimize energy consumption [1]. Significant progress in sensor technology, artificial intelligence (AI), and vehicle control systems has driven rapid development in the field of autonomous driving. These technological advances have enabled autonomous vehicles (AVs) to operate at higher levels of autonomy and safety. However, with the rapid development of autonomous driving control and regulation technologies, higher requirements have been placed on perception systems, requiring more complex and reliable systems. The development of autonomous driving control systems is a key factor in enhancing the overall capabilities of Avs [2]. Advanced technologies such as model predictive control (MPC) [3], robust path tracking [4], and deep reinforcement learning (DRL) [5] are increasingly being applied to autonomous vehicle systems. The goal of these control technologies is to ensure that

the vehicle can still achieve accurate trajectory tracking, stable operation, and reasonable decision-making in dynamic and unpredictable environments. With the continuous advancement of control systems, the requirements for autonomous vehicle perception systems are becoming more and more stringent. The perception system is responsible for providing accurate and real-time environmental data to the control system, which is the basis for the accurate operation of the control system. The close interaction between the control system and the perception system is crucial because the control algorithm is highly dependent on accurate and up-to-date environmental information to adjust its decisions in real time. The rapid development of autonomous driving control technology has led to higher and higher expectations for perception systems, hoping that they can provide high-precision real-time data under various driving conditions.

The perception system in autonomous driving is one of the key technologies to achieve autonomous driving capabilities. It perceives the surrounding situation by analyzing and understanding the environment around the vehicle, making subsequent decisions and planning driving paths. Autonomous vehicles rely on advanced perception systems to obtain accurate and comprehensive information about the surrounding environment. Therefore, the research and development of perception systems are crucial for promoting autonomous driving technology. By continuously improving the accuracy, stability, and adaptability of the perception system, safer and more efficient intelligent driving cars can be achieved, bringing huge potential benefits to society. Multi-sensor information fusion comprehensively utilizes the information obtained by different sensors, avoiding the perception limitations and uncertainties of a single sensor, forming a more comprehensive perception and recognition of the environment or target, and improving the system's external environment perception ability. In recent years, multi-sensor information fusion technology has been fully applied to fault detection [6], remote sensing technology [7], robotics technology [8], simultaneous localization and mapping (SLAM) [9], health detection [10], and advanced driver assistance systems (ADAS) [11].

Environmental perception systems provide a comprehensive perception network for autonomous vehicles by equipping them with multiple sensors, such as lidar, cameras, millimeter-wave radar, GNSS/IMU, etc. This paper focuses on how multi-sensor information fusion technology can enhance environmental perception tasks. Cameras can provide rich visual information, such as color, texture, and contour, and are widely used for object detection and tracking. Millimeter-wave radar provides data on the distance and speed of objects, which has a wide range of applications in obstacle detection [12], pedestrian [13], and vehicle [14] recognition. GNSS and IMU provide global position information and inertial information for autonomous vehicles to determine their location. This allows the vehicle to update its position in real time on the high-precision map. Lidar can provide accurate distance measurement and three-dimensional spatial information, and is mainly used for positioning, obstacle detection, and environmental reconstruction. Since 3D data provides richer and more accurate environmental information than 2D data, 3D lidar plays an increasingly important role in autonomous driving systems. It can maximize the reconstruction of the real traffic environment to obtain the details of the target by combining sensors such as cameras and lidar.

However, during the fusion process, the targets of multiple sensors are not in the same coordinate system, and the data transmission rates of different sensors are also different. Therefore, the multi-sensor information must be synchronized and aligned to a common spatiotemporal reference frame. In addition, there are differences in the physical information form and sensors, which requires calibration of the external parameters of multiple sensors to determine the final position and posture. In the process of sensor fusion, the current research methods are different, and the information fusion, fusion

methods, and fusion algorithms used by multiple sensors are also different. From the perspective of fusion methods, there are many combinations of sensors, including lidar–radar, camera–lidar, radar–camera, camera–lidar–radar, etc. According to the different forms of information processing by information fusion, the methods can be divided into data-level fusion, feature-level fusion, and decision-level fusion. In the process of data fusion, different fusion levels are usually adopted based on different abstraction levels of sensor data.

This paper will explore the application of multi-sensor information fusion in environmental perception, focusing on object detection and tracking, localization, mapping, and scene segmentation. By integrating data from different sensors, autonomous driving systems can more accurately identify and track pedestrians, vehicles, and obstacles, thereby improving safety. Multi-sensor fusion can also improve real-time positioning and mapping, ensuring reliable navigation in complex environments. This paper is organized as follows: Section 2 discusses the types, characteristics, and applications of sensors used in autonomous vehicles. Section 3 reviews sensor calibration methods. Section 4 studies multi-sensor fusion methods and algorithms, and Section 5 concludes and provides suggestions for future research.

2. Sensors in Environmental Perception Systems

The environmental data that autonomous vehicles can make clear decisions depend largely on the data obtained by sensors. The type and performance of sensors in autonomous vehicles directly determine the quantity and quality of external information collected by the system. In addition to using vehicle-to-vehicle communications such as V2V (vehicle to vehicle) and V2I (vehicle to infrastructure), autonomous vehicles also analyze and integrate various sensor data to perceive the external environment. Currently, the commonly used sensor types include radar, camera (including RGB-D, infrared camera), ultrasonic, lidar, and GNSS/IMU. Because the detection capabilities and reliability of various sensors in different environments are limited, the fusion of multi-sensor data can improve the detection and recognition accuracy. Table 1 summarizes the advantages and disadvantages of the above sensors and their detection ranges, indicating that different sensors have obvious differences in working characteristics. At the same time, by fusing multi-sensor data, the perception capabilities of AD vehicles can be improved in all aspects, effectively ensuring the safety of drivers [15]. Therefore, this section will mainly discuss the characteristics and advantages and disadvantages of these sensors.

Table 1. Comparison of different sensors [15].

Type	Advantage	Disadvantage	Max Working Distance
MMW Radar	(1) long working distance; (2) applicable to radial speed; (3) suitable for all weather conditions	(1) unapplicable for static objects; (2) frequent false alarms	5–200 m
Camera	(1) excellent recognition; (2) available lateral speed; (3) available color distributions	(1) large amount of calculation; (2) light interference; (3) vulnerable to weather disturbances	250 m
LiDAR	(1) wide field of view (2) wide range resolution (3) high angular resolution	(1) unable to tolerate bad weather (2) expensive	200 m
Ultrasonic	(1) cheap	(1) low resolution (2) not suitable for high speed	2 m
DSRC	(1) suitable for high speed (up to 150 km/h) (2) relatively mature technology (3) low latency (0.2 s)	(1) low data rate (2) small coverage area	100–300 m

Table 1. *Cont.*

Type	Advantage	Disadvantage	Max Working Distance
LTE-V2X	(1) long working distance (2) relatively high data transfer rates (up to 300 mbps)	(1) high latency over long distances (>1 s) (2) not suitable for time-critical events	Up to 20 km
5G-V2X	(1) ultra-high data transfer rate (2) low latency (<80 s) (3) high bandwidth (4) suitable for high speed (up to 500 km/h)	(1) immature application	100–300 m

2.1. Millimeter-Wave Radar

Millimeter-wave (MMW) radar uses electromagnetic waves in the millimeter-wavelength range (usually between 30 GHz and 300 GHz) to detect objects and measure their distance, speed, and relative position. The radar transmits electromagnetic waves, which are reflected from objects in the environment. The radar system then measures the time delay and frequency shift of the reflected waves and uses this data to calculate the distance and speed of the object. The frequency shift occurs due to the Doppler effect, allowing the radar to track the speed of moving objects. MMW radar works by transmitting a continuous or pulsed signal and receiving the reflected signal from the object. It can be divided into two types: frequency modulated continuous wave (FMCW) and pulse Doppler radar. FMCW radar is particularly useful in autonomous driving because it is able to measure distance and relative speed at the same time. The main types of MMW radar currently used in autonomous vehicles are as follows: (1) FMCW radar, which is widely used in autonomous driving systems because of its ability to provide distance and speed measurements. It works by modulating the frequency of the transmitted signal and measuring the frequency difference of the returned signal, which allows accurate estimation of distance and speed. (2) Pulse Doppler radar uses pulses of electromagnetic energy to detect objects. Although the resolution may be lower than that of FMCW radar, it is particularly suitable for long-range detection. (3) Short-range, medium-range, and long-range radars are designed to provide different levels of coverage. Short-range radar provides high-resolution data to detect close-range objects. Long-range radar detects objects at a greater distance, such as the vehicle ahead.

One of the advantages of incorporating MMW radar into autonomous driving is that it is not affected by light and weather conditions. It can work in darkness and detects snow, rain, fog, or dust almost equally well. Long-range radar can detect up to 250 m in very adverse conditions where other sensors cannot operate [16]. MMW radar can accurately measure the speed of moving objects through the Doppler effect, making it ideal for adaptive cruise control (ACC) and collision avoidance systems.

However, it also has some disadvantages and difficulties. The low resolution cannot distinguish objects that are closer or makes it difficult to detect small objects in the environment. Although it is very good at detecting large objects and measuring distance and speed, it lacks the detailed visual information provided by cameras, such as object shape, texture and color. In addition, MMW radar systems are susceptible to interference from other radar sensors, especially in urban environments with many radar vehicles. MMW radar is sensitive to target reflectivity. This is because different object materials have different abilities to reflect radar waves. Metallic objects can significantly enhance radar signals, which helps to identify targets such as vehicles, but it can also make small metal objects on the road, such as discarded cans, appear larger than they actually are, while some others (such as wood) hardly reflect radar waves. This situation may lead to false detections and missed detections.

A millimeter-wave radar is an important sensor for autonomous driving systems, and it excels in object detection, speed measurement, and coping with complex environmental conditions. The all-weather performance and long-range detection capabilities of millimeter-wave radar make it an indispensable part of applications such as adaptive cruise control, obstacle detection, and pedestrian safety. However, due to its low angular resolution and limited details, complementary sensors such as cameras and lidar are needed to achieve a fully integrated and reliable perception system.

2.2. Camera

Cameras are one of the most widely used sensors in autonomous driving, providing critical visual information for the vehicle's perception system. They capture two-dimensional (2D) images of the surrounding environment and then process them to detect objects, lanes, road signs, and traffic lights. Cameras rely on optical principles to capture light reflected from objects and convert it into electrical signals that create digital images. Cameras used in autonomous driving include RGB and special variants such as infrared cameras, fisheye cameras, etc., which can provide additional depth, infrared, or visual information under special conditions. It can provide the necessary visual information for perception, navigation and decision-making, including color, texture and shape features. Cameras can provide raw visual data and work together with sensors such as lidar, radar, and ultrasound to provide the vehicle with a full understanding of the surrounding environment, thereby improving driving safety. Object detection and recognition, lane detection and tracking, traffic sign recognition, and pedestrian detection are the main applications of cameras. Based on deep learning methods, deep neural networks have revolutionized computer vision tasks, achieving more accurate and efficient object detection, classification, and scene understanding. Advanced deep learning models, such as convolutional neural networks (CNNs), can achieve more accurate and efficient object detection, classification, semantic segmentation, and scene understanding from camera data. These models can learn complex features and patterns directly from raw pixel data, significantly improving perception performance. The most common models of camera-based perception in autonomous driving are the (1) stereo vision system, which uses two or more cameras placed at different angles to simulate human binocular vision. By comparing the images of each camera, the stereo vision system can estimate the depth and distance to the object, thereby achieving 3D reconstruction of the environment. (2) Monocular vision systems rely on a single camera and typically use machine learning algorithms to infer depth and 3D information from 2D images. While monocular vision systems are more cost-effective, they are not as accurate as stereo vision systems. (3) Infrared cameras are used to enhance vision in low-light environments and provide a thermal image of the surrounding environment. Infrared cameras are particularly useful for detecting objects with higher body temperatures, such as pedestrians or animals, in the dark.

However, cameras are susceptible to lighting conditions, and reliability and accuracy decrease when lighting conditions change (e.g., vehicles exiting tunnels, shadows, low light). Cameras are also easily affected by weather conditions such as rain, snow, or fog. Water droplets on the lens blur the image and fog reduces visibility, making it difficult to detect objects at a distance. Unlike lidar, which provides accurate 3D spatial data, cameras typically provide 2D data. Although techniques such as stereo vision and monocular depth estimation can infer depth, their accuracy is not as good as lidar, especially in complex scenes. Camera data requires a lot of computing resources to process, especially when using deep learning models for real-time object detection and classification. This increases the computational burden of the entire system. By fusing camera data with data from other sensors (radar, lidar) to confirm the observation results and improve

reliability, the limitations of a single sensor can be compensated. Using deep learning methods, convolutional neural networks (CNNs) trained on large-scale datasets can adapt to different lighting environment conditions and improve object detection capabilities.

2.3. LiDAR

As a distance measurement technology [17], LiDAR works by measuring the time interval between the emitted laser pulse and the reflected light scattered by the surrounding targets, thereby achieving accurate distance calculation. The round-trip delay of the LiDAR signal is called the time of flight (TOF), which can be obtained by modulating the frequency, phase, intensity, etc., of the emitted light and measuring the time it takes for the receiver to detect the modulation pattern [18–20]. LiDAR systems can be divided into laser ranging systems and scanning systems [21]. Rangefinders that measure the distance to an object using a laser beam are called laser rangefinders. The way they work depends on the type of signal modulation used in the laser beam. Their time of flight (TOF) can be measured using pulsed lasers, and these are called direct detection laser rangefinders. The laser signal can also be frequency modulated continuous wave (FMCW), which can indirectly measure distance and speed through the Doppler effect. These are called coherent detection laser rangefinders. TOF LiDARs dominate the current automotive LiDAR market due to their simple structure and signal processing methods. However, due to eye safety requirements, the limited transmit power limits the potential to increase their maximum range. Their return signals may also be interfered by strong sunlight or other TOF LiDAR beams. FMCW LiDAR continuously transmits frequency modulated laser signals to the target, continuously illuminating the object with less transmission power, thus meeting eye safety requirements and opening the possibility of using more power to expand its field of view (FOV). Compared with cameras, LiDAR is insensitive to light and provides practical and accurate 3D perception capabilities during the day and night. It can provide high-resolution and real-time 3D point clouds of the environment [22], with the aim of obtaining the shape and distance of surrounding vehicles and pedestrians as well as road geographic information, while facilitating object detection and classification. Multi-line LiDAR continuously transmits laser beams through the transmitter, and the receiver collects the scattered light of the target as a point cloud image, which helps perceive and identify pedestrians and vehicles.

However, multi-thread LiDAR systems in autonomous driving bring challenges such as data synchronization, resource management, real-time performance, concurrency issues, scalability, and fault tolerance. Effective management of these issues is crucial to improving the reliability of autonomous driving systems. LiDAR can sense the surrounding environment in real time and form high-definition 3D graphics [23]. It has the advantages of fast response, long detection distance, and high accuracy. The main performance of LiDAR point cloud in autonomous driving is (1) real-time environment perception and processing for scene understanding and target detection [24]; (2) generation and construction of high-definition maps and city models for reliable positioning and construction. LiDAR also has unique advantages in ranging. Reference [25] proposed an obstacle detection and tracking method based on three-dimensional light detection and ranging (LiDAR) to obtain the motion state of obstacles in real traffic scenes. The experimental results show that the method has good performance in real urban scenes and has high reliability. For the problem that adjacent obstacles are difficult to distinguish and distant obstacles are easily detected as multiple targets, reference [26] proposed an obstacle detection and tracking method based on multiple LiDARs. The average detection accuracy of this method was 97.53%, and the average tracking accuracy was 95.1%. The results showed that it was superior to other methods in obstacle detection and tracking.

Although LiDAR is superior to other sensors in ranging accuracy and 3D perception, its performance is poor in harsh environments such as fog, snow, or rain. Heterogeneous sensor data fusion can eliminate information redundancy and loss, and provide reliable, stable, and efficient environmental perception capabilities, but the cost of the system will increase accordingly.

2.4. Global Navigation Satellite System (GNSS)/Inertial Measurement Unit (IMU)

The Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU) are key components in autonomous driving systems, helping to determine the position, speed, and orientation of the vehicle. GNSS is a satellite-based navigation system that provides precise positioning and timing information anywhere on Earth. By triangulating signals from multiple satellites, a GNSS receiver can calculate the vehicle's position (latitude, longitude, and altitude) and speed. GNSS is widely used for global positioning and mapping, especially in open areas such as highways. In autonomous driving, it is one of the most commonly used vehicle positioning sensors. GNSS provides precise global positioning, which enables accurate positioning in open environments such as highways or rural roads. Differential GPS (DGPS), Real-Time Kinematic GPS (RTK-GPS), and Precise Point Positioning (PPP) technologies can be used to improve the accuracy of GNSS. The differential global positioning system (DGPS) operation consists of a reference station and a rover. Both stations use GPS receivers to receive positioning data from GPS satellites, use the positioning data collected by the reference station to calculate the positioning error, and transmit the error correction to the rover to improve the positioning accuracy [27]. RTK can achieve real-time centimeter-level positioning accuracy through double-difference ambiguity resolution (AR) [28]. Compared with differential techniques, precise point positioning (PPP) has a significant advantage in that it can accurately determine the position of a GNSS rover receiver by using external corrections from the Internet or dedicated correction satellites [29]. While both DGPS and RTK can achieve high positioning accuracy, GPS signal interruptions in urban environments and tunnels remain a challenge for these sensors.

The IMU can provide information about the vehicle's attitude, velocity, and direction. The IMU helps determine vehicle motion between GNSS updates and ensures continuous tracking of the vehicle's path, especially in areas where GNSS signals may be temporarily lost, such as tunnels and dense urban areas. The IMU provides real-time feedback on the vehicle's motion, which is critical for continued tracking, especially in dynamic environments. As the positioning error accumulates during the vehicle's driving, the IMU signal error accumulation will drift over time, so the IMU needs to continuously correct the estimated position. The data fusion of GNSS and IMU can achieve vehicle state estimation and ensure a continuous positioning process. GNSS and IMU data are often fused with other sensors such as LiDAR and radar to improve positioning accuracy and build a complete picture of the vehicle's surroundings. By fusing these data sources, autonomous vehicles are able to navigate in a variety of environments, even without direct GNSS signals. There are currently several models for combining GNSS and IMU data. (1) GNSS/IMU integration: combining GNSS data with IMU information can create a dead reckoning system that continuously estimates the vehicle's position when GNSS signals are unavailable. In this model, the IMU uses accelerometers and gyroscopes to track vehicle motion and orientation to fill in the gaps in GNSS data. (2) Tightly coupled system: in this, GNSS and IMU data are fused and processed in real time. This can improve the accuracy of vehicle positioning, especially in challenging environments where GNSS signals are weak or unstable. (3) Loosely coupled system: in this system, GNSS and IMU operate independently and merge the data in the post-processing stage. Compared with the tightly coupled system, this method is less computationally intensive, but less accurate.

To mitigate the limitations of GNSS and IMU, these sensors are often fused with other perception technologies, such as LiDAR, radar, and cameras. GNSS/IMU systems are often integrated with LiDAR and radar sensors to enhance the vehicle's perception of its surroundings. LiDAR provides a high-precision 3D map of the environment around the vehicle, while radar provides reliable object detection in adverse weather conditions. The integration of these sensors improves positioning, obstacle detection, and path planning. Cameras are used in conjunction with GNSS/IMUs to help detect road signs, lane markings, and other visual cues that support positioning. By fusing visual data and GNSS/IMU information, autonomous driving systems can achieve higher vehicle positioning and decision-making accuracy. The integration of GNSS and IMU is critical for the precise positioning and navigation of autonomous vehicles. While GNSS provides reliable global positioning services, IMU ensures continuous tracking when GNSS signals are missing or unstable. GNSS/IMU data is fused with sensors such as LiDAR, radar, and cameras to ensure that autonomous vehicles can operate reliably in a variety of environments.

In general, the fusion of multi-source heterogeneous sensor data will improve the perception capability and range of autonomous vehicles, but the resulting computational pressure needs to be resolved. The combination of V2V, V2I, and cloud computing will reduce the computational pressure of vehicles processing massive data.

3. Multi-Sensor Information Fusion

Autonomous vehicles travel in dynamic and unpredictable environments, and a single type of sensor cannot provide the comprehensive data required for safe and efficient navigation. Each sensor has its own unique advantages and limitations. Multi-sensor fusion can increase redundancy and reduce the possibility of errors due to sensor-specific limitations. By combining data from multiple sensors, the system obtains overlapping coverage, ensuring robust perception while improving the accuracy and reliability of the data. Fusion sensor systems provide more reliable and detailed perception of the environment, which helps make real-time decisions in complex scenarios such as urban navigation and highway driving. By gaining a more comprehensive understanding of the environment, fusion systems can reduce risks by improving object detection, obstacle avoidance, and path planning, enabling autonomous vehicles to adapt to different scenarios. However, the fusion of multiple sensors is also challenging. Since the fusion algorithm requires a lot of computing resources, this increases system complexity and energy consumption. The fusion process requires precise alignment of sensor data in time and space, which is difficult to achieve in real-time applications. Similarly, differences or noise in sensor data can lead to information conflicts, requiring complex algorithms to resolve inconsistencies. The integration of multiple sensors brings additional hardware and software complexity, making the system more expensive and complex to maintain. By addressing these challenges, multi-sensor fusion systems can provide a reliable foundation for advanced autonomous vehicle perception and decision-making. Therefore, multi-sensor fusion is essential for autonomous driving, enabling vehicles to operate safely and efficiently in various environments. This section will focus on the calibration and fusion methods in multi-sensor fusion.

3.1. Multi-Sensor Calibration

The calibration of multiple sensors is crucial for accurate data fusion in autonomous driving systems. The calibration task is to ensure accurate alignment of multiple sensors on a vehicle for effective environmental perception and understanding. There are two types of multi-sensor calibration [30]: intrinsic calibration and extrinsic calibration. Intrinsic calibration focuses on the internal parameters of individual sensors, correcting their inherent

distortions and inaccuracies. Intrinsic calibration addresses sensor-specific parameters and is performed before external calibration. Most intrinsic parameters are provided or calculated by the manufacturer, so calibration focuses mainly on extrinsic parameters. In order to adjust the relative position and orientation of multiple sensors to ensure that their data can be accurately combined, the sensors need to be calibrated externally. External calibration is a rigid transformation (or Euclidean transformation) that maps points from one 3D coordinate system to another 3D system, such as mapping points from a 3D world or 3D LiDAR coordinate system to a 3D camera coordinate system. External calibration estimates the position and orientation of the sensor in three orthogonal axes (also known as six degrees of freedom, 6DoF) of the 3D space relative to an external reference frame [31–33]. This section reviews the extrinsic calibration methods for multi-sensor systems, including emerging methods currently used in research.

There are three different types of fusion levels in sensor fusion: data-level fusion, feature-level fusion, and decision-level fusion [34]. Data-level fusion and feature-level fusion are early fusion, while decision-level fusion is late fusion. Early fusion refers to the integration of data series at the feature level [35]. Unlike early fusion, late fusion is handled by each sensor independently for classification or recognition. Since the integration of data series is done at the semantic level, synchronization and calibration are required in the early fusion stage, but not in the late fusion stage. In early fusion, the input sources from different sensors need to be regularized. The goal is to ensure that different types of sensors can achieve the same goal in the same coordinate system. The purpose of multimodal sensor calibration is to determine how to transform data from different sources into a common reference system required for early sensor fusion [36]. Target-based calibration and targetless calibration are two traditional calibration methods. Next, we will introduce them respectively and discuss their advantages and disadvantages. Finally, we will give the emerging methods currently under research.

3.1.1. Target-Based Calibration Method

The target-based calibration method uses a specially designed calibration target. Figure 1 is a general flow chart of target calibration. The calibration plate is a sensor calibration workpiece used in the calibration target. Its surface is usually an object plane or approximate plane with a specific set of shapes, patterns or features. In the external calibration of LiDAR and camera, it is used as a reference object to measure the correspondence between the image or point cloud observed by the sensor and the actual geometric shape, such as chessboard pattern [37], polygonal flat plate [38], nearly orthogonal multi-plane chessboard [39], arbitrary trihedron [40], plane target with holes [41], spherical pattern [42], etc., to calibrate a multi-sensor system. Typical features such as chessboards have the ability to accurately detect corners and intersections, and they provide high flexibility [43–45]. Most of the research focuses on the external calibration of LiDAR-camera, with chessboard patterns as its main target. Reference [46] proposed an external calibration method to calibrate the direction and position of 2D LiDAR and camera through multi-angle chessboard pattern poses. The calibration parameters are estimated by solving a nonlinear optimization problem. Reference [47] proposed an extrinsic calibration method for 2D LiDAR and stereo vision cameras based on 3D reconstruction of a chessboard and inverted plane geometric constraints between two sensor views, and applied a nonlinear optimization algorithm based on geometric constraints to solve the extrinsic parameters. For feature matching, multiple calibration board views are required to determine the feature correspondence in order to establish a set constraint between the two sensors and estimate the relative transformation. Reference [48] proposed a calibration method based on point-to-plane constraints. The key to this method is that the laser points on the chessboard are obtained from

the initially positioned LiDAR and the same LiDAR rotated vertically from the original position. Additional constraints are constructed and the basic constraints are strengthened, thereby solving the problem of LiDAR sensitivity to posture. References [49,50] proposed an optimal extrinsic calibration algorithm between a binocular stereo vision system and a 2D LiDAR under the Mahalanobis distance constraint. Instead of repeating the same calibration process from the LiDAR to each camera, they calibrated the multi-sensor system based on the point-to-plane geometric constraints of the 3D reconstruction of the chessboard calibration board. Reference [51] proposed an improved calibration method for the joint calibration of 2D LiDAR and color camera, estimating the 2D homography by using point-to-line constraints established by a triangle plate. They proposed a data preprocessing method to improve the measurement error of the LiDAR. Reference [52] used point calibration targets to locate the target and LiDAR coordinates in the image, that is, the parameter-constrained point-to-point correspondence between the two. Specifically, the target in front of the sensor is repeatedly moved downward until the LiDAR captures the target to obtain the correspondence. Most calibration targets are composed of alternating black and white color blocks with different regions, such as chessboard and ArUco [53,54]. The outer edges of some calibration targets have obvious geometric features, which greatly facilitates the detection of features. However, in order to overcome the problem that the traditional chessboard may cause uneven illumination of the calibration target under strong annular illumination, resulting in the failure of ordinary chessboard detection. Reference [55] proposed a calibration method based on the Charuco plate, which can solve the problems caused by the traditional calibration plate and can also be used to detect other saddle points. To date, there are many external calibration tools for LiDAR and cameras. These calibration targets are usually designed based on chessboard patterns. Compared with 2D LiDAR, 3D LiDAR can obtain more information, such as the normal vector of a point [56]. The calibration of 3D LiDAR and camera is a 6-degree-of-freedom (DOF) problem. Similar to the calibration of a 2D LRF and camera, the chessboard provides three constraints for each observation, but this will affect the calibration accuracy. To compare the effects of different calibration targets, some typical external calibration methods for 3D LiDAR and camera are given in Table 2. Existing calibration tools can only solve the calibration of paired sensors with at most two sensing modes. References [57,58] proposed an external calibration tool for radar, camera and LiDAR, and bound it to the Robot Operating System (ROS). Three configurations are proposed to estimate the pose of the sensor by simultaneously detecting multiple calibration plate positions, as shown in Figure 2. (1) Minimum connected pose estimation (MCPE): in the MCPE configuration, all sensors are calibrated in pairs relative to a selected reference sensor. The advantage of this method is that it is computationally simple because only the reference sensor is involved, so the computational effort is small. However, if the calibration of the reference sensor is biased, it will affect the calibration results of all other sensors. Therefore, choosing the right reference sensor is crucial to the success of MCPE. (2) Fully connected pose estimation (FCPE): the FCPE configuration performs pairwise calibration between all sensors without specifying a specific reference sensor. This method is similar to loop closure optimization in SLAM and can improve the robustness of calibration because it does not rely on the accuracy of a single reference sensor. However, as the number of sensors increases, the number of transformation matrices that need to be estimated increases, and additional loop closure constraints need to be added, which may increase the computational complexity. (3) Pose and structure estimation (PSE): the PSE configuration estimates the poses of all sensors and the pose of the calibration plate at the same time. This method is similar to bundle adjustment because it estimates the pose of the sensor and the pose of the calibration plate at the same time. The advantage of this method is that it can avoid the use of heterogeneous

error functions (pixel and Euclidean distance) and use a unified error function instead. However, the optimization of this approach is more complex and thus takes longer to compute. In addition, loop closure constraints are not explicitly enforced, which may affect the accuracy of the calibration. They compared the joint optimization results of PSE, MCPE, and FCPE depending on multiple variables, such as the number of calibration plate locations and the choice of MCPE reference sensor. The results show that the joint optimization of FCPE outperforms MCPE and PSE when more than five calibration plate locations are used. However, the increase in the number of calibration plate locations has a significant impact on the computation time. Reference [59] introduced L2V2T2Calib and open-sourced it as a toolbox to unify extrinsic calibration between different lidars, visual cameras, and thermal cameras. All sensors can use the same calibration target (a four-circular hole plate). A general method for automatically detecting planar calibration targets is proposed, using the concept of template matching to improve calibration efficiency.

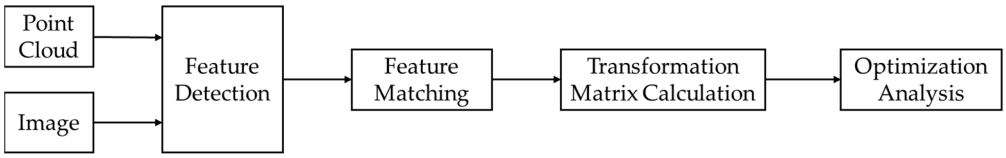

Figure 1. Target calibration general flow chart.

Table 2. Some representative 3D LiDAR and camera calibration methods.

Calibration Plate	Features	Year	Literature
Checker board	Expanding from 2D LiDAR to 3D LiDAR	2010	[60]
	Introducing unit normal vector uncertainty	2012	[56]
	Fit a chessboard model to the chessboard point cloud; optimize using the Levenberg–Marquardt method	2017	[61]
	By combining 3D line and plane correspondences, the number of poses is reduced to one	2018	[37]
	Intensity clustering method based on Gaussian mixture model (GMM)	2022	[62]
Polygonal flat plate	Utilizes 2D–3D correspondence	2014	[38]
Design—flat panel	AprilTag is placed at the intersection of two reflective cross stripes	2022	[63]
Nearly orthogonal Multiplane chessboard	Solve nonlinear constraint adjustment problems using sequential quadratic programming (SQP) methods	2022	[39]
Any trihedron	Prevalent in structured environments	2013	[40]
Plane target with a hole	There is a triangular hole on the plane; establish 3D–3D correspondence	2012	[64]
Spheric	Available for different modality devices; the center of the sphere is calculated separately from the LiDAR point cloud and the image	2020	[65]
Ordinary box	Only a simple cardboard box of known dimensions is needed; Can be used to calibrate camera–LiDAR and LiDAR–LiDAR	2017	[66]
	Directly solved through the E-PnP algorithm	2018	[67]
	A method for fitting a plane using points, independent of edge points	2018	[68]
ArUco marking	Use polygon corners as corresponding points to avoid errors	2021	[53]
	Calibration of multiple scenes at different distances	2022	[69]
Panoramic infrastructure	Robust calibration using data from a single frame	2021	[70]

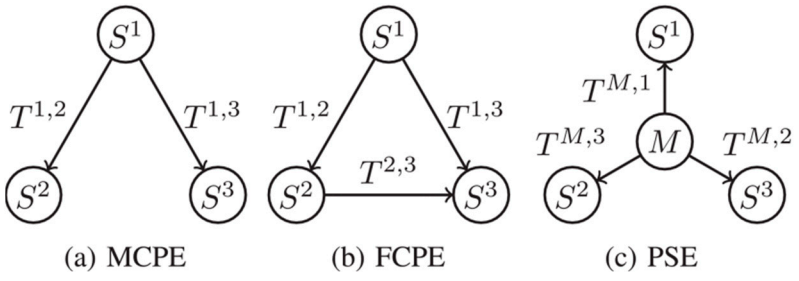

Figure 2. Optimal configuration for joint calibration [57,58].

Although the target-based calibration method has been more widely adopted, it also has certain limitations. First, the target-based calibration method requires setting a specific

calibration target; however, the target board has the defect of relatively complex calculation. Secondly, the vehicle will cause changes in hardware, software and the surrounding environment during long-term driving, which will eventually lead to the misalignment of the sensor system, increase the error of the calibration result, or even fail. Therefore, it is crucial to improve the reliability and accuracy of the calibration method of the autonomous driving fusion sensor system.

3.1.2. Targetless-Based Calibration Method

Targetless calibration can be based on targetless external calibration, which uses the estimated motion of each sensor or uses features in the perceived environment for calibration. However, using features of the perceived environment requires multimodal sensors to extract the same features in the environment and is sensitive to the calibration environment [71,72]. Its overall calibration process is similar to that of the target-based calibration method. In general, the target-based calibration structure is also applicable to this section. Since it relies on real objects in natural scenes, it is more general and flexible and suitable for various scenarios and applications. Specifically, targetless calibration methods can be divided into feature-based, mutual information-based, and motion-based methods. Given that most current research focuses on LiDAR–camera external calibration, this section mainly outlines targetless calibration methods based on LiDAR–camera.

(1) Feature-based methods

Feature-based methods directly extract features from the environment for matching and external parameter estimation. Specifically, these features can be divided into: geometric, semantic and motion features. Geometric features are composed of some geometric features in the environment, such as points, edges, etc. Semantic features are semantic perceptions of specific objects in a given environment, such as skylines, cars, and telephone poles. Motion features describe the properties of moving objects, including posture, velocity and acceleration. In reference [73], it first manually selects point correspondences from laser–camera acquisition in natural scenes, then uses the PnP algorithm to estimate external parameters, and then uses the Levenberg–Marquardt algorithm to solve the nonlinear optimization problem. Reference [74] proposes a key point-based 2D–3D pose estimation network for the real-time calibration of camera–LiDAR. A trainable point weighting layer is used to find more obvious local features for easy matching, such as tree trunks, traffic signs and road edges. The network can extract sparse key points and give corresponding weights for posture estimation, which further improves the robustness of the network. In addition to point features, edge features are another type of geometric features widely used in LiDAR–camera external calibration. Edge features in point clouds and images contain useful environmental geometric information. Reference [75] proposed a calibration method based on local features by jointly registering edge feature points between LiDAR and camera data. They do not require geometric primitives or any other type of environmental structure. They use a combination of P3P and MSAC to reduce the error of external parameters, and finally use a Kalman filter to smooth the final calibration. In reference [76], pixel-level accuracy is achieved by aligning natural edge features in LiDAR–camera. They proposed an algorithm based on point cloud voxel cutting and plane fitting that can accurately and reliably extract LiDAR edges. Its robustness and consistency are verified in various natural scenes. In the process of selecting points and edges, the features need to be translated and rotated, and distortion and other problems will occur in this process. Therefore, it is necessary to introduce feature operators. The scale-invariant feature transform (SIFT) method proposed in the literature [77] converts an image into a large number of local feature vectors, each of which is invariant to image translation, scaling and rotation, and is invariant to illumination changes and affine or 3D projection parts. This is

a popular operator. Sped-up robust features (SURFs) [78,79] are a faster method for SIFT to extract interest point features. When point features are unstable and insufficient to describe environmental features, edge features become particularly important. The commonly used method is to use depth discontinuity to extract edge information from LiDAR point clouds. The main principle is to set a depth difference threshold between adjacent points. Points whose depth changes exceed the threshold will be retained. On the contrary, they will be filtered out, and then these edges can be identified from the points. This idea has been widely used in various point cloud edge extraction methods. The literature [80] uses the classic Canny method to extract point features from a single-frame LiDAR and extracts edge contours from point clouds by setting a depth threshold. The literature [81] optimizes external parameters by minimizing the difference between the LiDAR depth map and the camera depth map. Similar to point cloud edge feature extraction, image edge extraction in cameras adds various operators, such as Sobel operator, Canny operator [82], and LSD operator [83]. Different semantic feature information in the environment, such as skyline, lanes, and poles, can be used to calibrate the LiDAR–camera. Reference [84] proposed a skyline-based LiDAR–camera automatic registration method. This method extracts skyline pixels from images and skyline points from point clouds, respectively, and uses a strong optimization method to search for the best matching parameters between them. Reference [85] proposed a calibration-independent track-to-track association method. This method selects trajectories in the same time series as features based on time information. Reference [86] proposed a line feature-based automatic calibration and refinement method. They extracted straight-line features from road lanes and poles in images and point clouds. These features not only provide sufficient spatial constraints to robustly estimate accurate initial calibration, but also provide rich semantic information for further refinement. In existing research, we can perform external calibration based on various detection and tracking algorithms by receiving the estimated trajectory of the object through the LiDAR and camera. It is worth noting that the two trajectories should match as closely as possible. In the literature [87], they proposed a multi-sensor calibration method based on the Gaussian process (GP) to estimate the trajectory of moving objects. This method uses the GP regression method to estimate the object trajectory and applies it to object tracking. In addition, the obtained time delay is used to estimate the external parameters based on the 3D point correspondence relationship.

(2) Information-based approach

Information-based methods mainly estimate external parameters by maximizing the similarity transformation between the LiDAR and the camera, which is achieved through various information metrics. The information-based LiDAR–camera calibration method can be summarized in three steps: first, the 3D–2D projection of the LiDAR points, that is, projecting the 3D LiDAR point set into the image. Second, the statistical similarity measurement, which uses the similarity distribution between some sensor data obtained from the LiDAR and the camera to measure the statistical similarity between the 2D projected image and the camera image. Finally, the statistical correlation measure is optimized, which is usually a non-convex function and requires an optimization method to reach the global optimum. It is worth noting that the data obtained from the LiDAR and the camera have several properties that have similar distributions. For example, LiDAR data points with high reflectivity usually correspond to bright areas in the image, while data points with low reflectivity correspond to dark areas [88]. Therefore, the similarity between reflectivity and image intensity can be used to test the similarity between the LiDAR and camera data. In addition, the gradient information in the lidar point cloud data and the camera image can also be used to measure the similarity between the two [89]. Some commonly used point cloud and image attribute pairs include reflectivity

and grayscale intensity, gradient amplitude and direction, 3D semantic label–2D semantic label, and 3D–2D attribute pair combinations. Reflectivity and grayscale intensity represent the return intensity of the LiDAR point cloud and the pixel intensity in the grayscale image, respectively. When the LiDAR and camera observe the environment at the same time, there will be statistical similarities between them, and the same surface properties of the object determine the properties of both. Similarly, other attribute pairs such as reflectivity and hue [90] and reflectivity and visible wavelength [91] also depend on the same surface properties of the object. Since they are relative values inferred from the image and not directly measured physical quantities, they may be affected by environmental conditions and lighting changes. Using 3D semantic label–2D semantic label attribute pairs can solve the above problems well. The semantic labels of the 3D LiDAR point cloud correspond to the 2D image pixels of the camera, and this semantic information can be used for data association [92]. Some studies have found that the use of a combination of multiple features can improve the robustness of algorithms in different environments. Reference [93] uses a combination of 3D–2D attribute pairs to measure similarity and assigns appropriate weights to them. These attribute pairs can be a combination of the above attribute pairs, such as reflectivity, surface discovery, gradient information, etc. Next, we need to use various statistical correlation metrics to measure the statistical similarity between attribute pairs, where larger metrics have better correspondence. The mutual information-based method is currently the most commonly used method. Reference [94] reviews the commonly used statistical dependency metrics in existing information-based methods. The mutual information-based method is currently the most commonly used method. Finally, we need an optimization algorithm to solve non-convex functions to achieve global optimality. Commonly used optimization algorithms include the Barzilai–Borwein steepest descent method [95], the Nelder–Mead downhill simplex method [96], the Levenberg–Marquardt algorithm [97], the particle swarm optimization (PSO) [98], the BFGS quasi-Newton method [99], and the Boundary-Optimized Quadratic Approximation (BOBYQA) algorithm [100].

This section describes an information-based targetless external calibration method that infers the external pose of a camera or sensor by analyzing feature acquisition and relationships in image or sensor data. Compared with targeted calibration methods, it does not require the setting of a dedicated calibration plate or target, and has a variety of attribute pairs or attribute pair combinations to choose from, is easy to calculate, and has a certain degree of flexibility. However, its accuracy depends on data quality, and some factors such as reflectivity and grayscale intensity are more dependent on the environment, which constitutes its limitations.

(3) Motion-based methods

The motion-based targetless extrinsic calibration method uses the motion of sensors mounted on a moving vehicle to estimate extrinsic parameters. Currently, this method mainly involves finding the correspondence between the trajectories generated by the sensors. Some methods attempt to find the correspondence between trajectories through odometry technology or IMU and GNSS measurements. According to existing literature, motion-based methods can be mainly divided into hand–eye calibration and calibration based on 3D structure estimation. The method of changing the position and orientation of the sensor and calibrating using the motion observed by each sensor is called hand–eye calibration [101]. The hand–eye calibration problem is mainly used in robot vision. The traditional hand-eye calibration is extended to the calibration problem of LiDAR and camera for wider application. Solving the homogeneous matrix equation $AX = XB$ [102], the extrinsic parameters between sensors can be obtained. Hand–eye calibration can be roughly divided into three stages: estimating the motion of each sensor, estimating extrinsic

parameters, and refining extrinsic parameters. The state transformation matrix of LiDAR and camera is estimated by considering the rotation and translation between adjacent frames. The extrinsic parameters are estimated by the motion state of the sensor. For the motion estimation of LiDAR, the literature [103] combines LiDAR odometry and ICP algorithm to obtain the transformation matrix of LiDAR as accurately as possible. For camera motion estimation, reference [104] uses a combination of motion structure (SFM) and visual odometry to calculate the camera transformation matrix. However, both the ICP algorithm and SFM are currently the more commonly used methods for estimating LiDAR and camera motion. Reference [105] summarizes the current hand–eye based LiDAR–camera external calibration methods, as shown in Table 3.

Table 3. LiDAR–camera calibration method based on hand–eye calibration.

Literature	Motion Estimation Method (LiDAR–Image)	Rotation Parameter Representation	Refinement Method
[106]	ICP and SFM	angular axis	edge-to-edge
[107]	ICP and Visual Odometry	angular axis	color matching
[103]	LiDAR Odometry and Visual Odometry	angular axis	intensity matching
[81]	ICP and Visual Odometry	quaternions	edge alignment
[108]	ICP and Visual Odometry	lie algebra	depth matching and edge alignment
[104]	LiDAR Odometry and Visual Odometry	rotation matrix	3D–2D point matching

Unlike the hand–eye calibration method, the 3D structure-based method analyzes the 3D structure of the surrounding environment from the image. Among them, SFM is one of the most commonly used methods [109]. SFM can estimate the 3D model from overlapping 2D image sequences [110]. It has a wide range of applications in 3D modeling, visual SLAM, augmented reality, etc. Specifically, when estimating 3D structure based on the SFM method, the camera is installed on a moving vehicle and a series of images are captured while the vehicle is moving, thereby converting the LiDAR–camera external calibration problem into a registration task in the 3D domain. Reference [111] describes a method for registering a panoramic image sequence to a LiDAR point cloud using a non-rigid version of ICP containing a bundle adjustment framework. The registration is then improved by integrating the SIFT correspondence from the image to the reflectivity data into the bundle adjustment. Reference [112] proposes a procedure for automatically combining and jointly registering images and LiDAR data. This method uses the SFM reconstruction method to calculate high-precision image orientation and sparse point clouds, so that the 3D–3D correspondence can be accurately determined. Reference [113] uses scene sequence information of vehicle motion to obtain initial extrinsic parameters. This method estimates the LiDAR motion by registering 3D point clouds through the ICP algorithm, estimates the 3D structure from a 2D image sequence that may be coupled with local motion signals using the SFM algorithm, and then projects the 3D LIDAR points onto the 2D image plane using the initialization parameters. By searching for the edges of the image and LiDAR, the SIFT feature points between the two sensors are calculated respectively. According to the distribution and confidence of the points, the registered SIFT feature points are selected and added to the objective function. Finally, the optimal parameters are obtained by combining the optimization algorithm. However, the SFM method may result in sparse point clouds when converting 2D images to 3D point clouds. In this case, the matching rate will decrease, and the use of the ICP algorithm will increase the error. To solve this problem, reference [114] proposed an automatic registration method based on semantic features extracted from panoramic images and point clouds. The method first estimates the precise rotation parameters between the panoramic camera and the laser scanner using GPS and IMU-assisted motion structure (SFM). Then, Faster-RCNN is used to extract vehicles in the panoramic image as candidate primitives. Finally, the translation between the panoramic

camera and the LiDAR is refined by maximizing the overlapping area of corresponding primitive pairs based on particle swarm optimization (PSO). Reference [115] uses SFM to generate point clouds from image sequences recorded by a moving vehicle, connects the image domain and 3D space, and uses this as the basis for registration, and performs object-level alignment between the LiDAR and the generated point cloud. Methods based on 3D structure estimation can recover the three-dimensional form of a scene from a collection of images without relying on precise camera internal and external parameters or detailed prior information about the scene. However, they face the challenge of high computational cost when processing large-scale scenes. Although effective reconstruction methods such as SFM are proposed in current research, various measures to increase the number of point clouds are also required.

3.1.3. Deep Learning-Based Methods

Currently, deep learning is widely used in various applications, including the external calibration of multiple sensors. Among them, the external calibration method based on deep learning is an emerging research method that can calibrate the LiDAR camera system without using calibration objects, mileage information or mutual information. In current research, many researchers use training data to train neural network models to estimate the relationship between camera parameters and input images. End-to-end methods can simplify the calibration steps using neural network models. They use these models to learn useful features, and in this case, there is no need to manually define features. It uses a neural network model to process the input camera image and LiDAR point cloud data, and then directly outputs external parameters to achieve the optimal calibration parameters by minimizing the corresponding loss function. Reference [116] proposed RegNet, which is the first deep convolutional neural network (CNN) for inferring 6 degrees of freedom (DOF) external calibration between multimodal sensors. It integrates three traditional calibration steps (feature extraction, feature matching and global regression) into the convolutional neural network. Based on RegNet, reference [117] proposed a deep learning-based online calibration method for visual sensors and depth sensors. They first merged the LiDAR point cloud and depth image into one point cloud, and considered that the entire point cloud was generated by a virtual depth sensor and then calibrated the virtual sensor with the camera. RegNet can complete the calibration without manual intervention and has higher calibration accuracy than traditional methods. It can provide stable initial estimates and continuous online correction of external parameters. However, the performance of the network is limited by its structural design and capabilities. The overly simple feature extraction and matching network fails to fully consider the geometry of the point cloud. When the internal parameters of the camera change, the trained model needs to be fine-tuned, which may not fully capture the features and relationships between complex sensors, thus affecting the accuracy. Reference [118] proposed CalibNet, which is the first geometrically supervised deep learning method. They use a new architecture based on 3D space transformer, which learns to solve the underlying physical problem by using geometric and photometric consistency to guide the learning process. Reference [119] proposed CalibRCNN to infer the 6-degree-of-freedom (DOF) rigid body transformation between 3D LiDAR and 2D camera. It not only uses LSTM network to extract temporal features between 3D point clouds and RGB images of consecutive frames, but also uses geometric loss and photometric loss obtained by inter-frame constraints to refine the calibration accuracy of predicted transformation parameters. In the reference [120], they proposed a novel CalibDNN for accurate calibration between multimodal sensors. It is a simple system with a single model and a single iteration, which considers transformation loss and geometric loss to maximize the consistency of multimodal data. And it is applied

to challenging datasets with complex and diverse scenes. Reference [121] proposed an online LiDAR camera self-calibration network (LCCNet). Unlike other learning-based methods, they constructed a cost volume between RGB features and depth features for feature matching instead of directly connecting them. In addition to the smooth L1 loss between the predicted calibration and the ground truth, an additional point cloud distance loss is proposed. Reference [122] proposed a new adaptive LiDAR–camera calibration method ATOP, which implements a cascade process from attention to optimization. In the attention stage, they proposed a cross-modal object matching network (CMON) to find overlapping FOVs between the camera and LiDAR and match 2D objects in the image with their 3D versions in the point cloud. In the optimization stage, the center and vertices of each matched object are collected to construct 2D–3D point pairs. They adopted two cascade PSO-based methods: Point-PSO and Pose-PSO for pose initialization and refinement in the optimization stage. The method does not require the estimation of the initial pose and does not rely on a specific calibration target. However, the various networks mentioned above solve the external calibration as a regression task without considering the geometric constraints involved. Reference [123] proposed a new end-to-end external calibration method called DXQ-Net, which applies a differentiable pose estimator module to estimate external parameters and constrain 2D–3D correspondences with uncertainty during training. Deep learning-based methods use neural networks to extract latent features from LiDAR and camera data. With a large amount of training data, these methods can obtain suitable feature extraction results. However, the accuracy of this method is inevitably limited by the size of the training dataset and the framework of the deep learning network.

3.2. Multi-Sensor Information Fusion Method

In the process of multi-source heterogeneous sensor fusion, different methods represent different levels of thinking about the raw data in the fusion stage. Since the fusion of multiple sensors adopts different fusion methods at different data abstraction levels, different fusion algorithms are deployed. According to the level of information fusion, information fusion can be divided into data-level fusion, feature-level fusion, and decision-level fusion. Table 4 summarizes the different fusion methods. Next, they will be explained one by one.

Table 4. Summary of the fusion methods.

Fusion Level	Features	Advantages	Disadvantages
data level	combine raw data from multiple sensors	high precision and rich data	the computational complexity is high and precise alignment is required
feature level	fuse features such as edges, shapes or textures from sensors	balance accuracy and efficiency, retain useful information	need to extract features, which may cause information loss
decision-making level	combining each sensor decision	modular, highly resistant to sensor failures, easy to use in real time	information loss, reduced detail, conflicting decisions

3.2.1. Data-Level Fusion

Data-level fusion refers to the fusion process of directly fusing the raw perception data of different sensors and then further processing the fused data. It is a low-level fusion process. This method is most commonly used for multi-source image enhancement, where combining the raw data can present the environment more richly and comprehensively, especially for applications such as remote sensing and image enhancement. In autonomous driving, sensors such as LiDAR and millimeter-wave radar (MMW–radar) generate data at different resolutions and sampling rates. Therefore, spatial and temporal alignment of sensor data is essential to ensure that objects detected by each sensor correspond accurately

across all sensors. This means that the data must be integrated into a unified coordinate system to seamlessly fuse them. For example, compared to camera images, LiDAR provides high spatial resolution but has limited horizontal and vertical resolution. On the other hand, MMW–radar has advantages in bad weather but lacks instant imaging capabilities due to its long wavelength. Data-level fusion achieves fusion by aligning the frames of these different data sources and then integrating them into a comprehensive representation. In recent years, some research has focused on radar imaging [124,125]; however, this is not sufficient to distinguish multiple objects in complex scenes. Multi-source data fusion can improve image clarity and target detection capabilities. Generating a raster map based on radar or LiDAR data and then fusing it with an optical image can also be considered a data-level fusion method. Generally speaking, in the research on radar or laser radar (LiDAR) and camera fusion, data-level fusion methods can be divided into two research directions. One is to use the obstacle detection results of radar or laser radar to generate a raster map. The other is to use the optical image as a real sample and generate radar or laser radar images through a generative adversarial network (GAN) [126,127].

In the process of autonomous driving, multi-source heterogeneous pixel-level fusion usually relies on the resolvable units or generated images of radar and laser radar (LiDAR). This process aims to extract environmental features and target parameters from the fused data for further decision making. Data-level fusion methods can directly combine sensor data without deep information processing. Although this method can achieve the maximum fusion of multi-source data, the redundancy between data also leads to reduced fusion efficiency.

3.2.2. Feature-Level Fusion

Feature-level fusion refers to the fusion of multi-source heterogeneous sensor data after feature extraction. It is a mid-level fusion that first extracts features from the raw sensor data, such as edges, corners, shapes, and motion patterns, and then merges these extracted features to form a single feature set that can be further processed by machine learning or decision algorithms. Since the same target can be extracted from different sensors in different directions, better target detection and recognition can be achieved. Target parameter extraction and data feature extraction are two extraction methods in autonomous driving systems. Target parameter extraction includes extracting information such as the size, distance, direction, speed, and acceleration of the target from preprocessed data. Many studies use radar or LiDAR to extract the location features of the target and assist image recognition by generating a region of interest (ROI). This region directly converts the target position detected by the radar into an image, thereby forming a specific area. Data feature extraction is to extract the shape, edge, texture, time–frequency characteristics, and color distribution of the target from the image or processed data for classification and recognition. In order to reduce the large amount of computation caused by the large number of regions of interest (ROIs) generated in the image that may be included in the computer vision, many studies use radar and LiDAR to first extract the range and azimuth information of the target, and then integrate its location information with the image data to produce fewer regions of interest (ROIs) and computational complexity. Finally, a pre-trained model is used to further identify these regions and accurately classify the target categories. Many studies apply machine learning methods to further perception tasks after extracting the ROI. Traditional machine learning usually requires the extraction of standard features such as Haar operators, HOG operators, and gray-level co-occurrence matrices (GLCMs). These features are then classified using SVM [128], Adaboost, and other methods. In recent studies, neural networks such as YOLO, CNN, and ANN are often used to achieve target classification and recognition.

Feature-level fusion requires a certain degree of information extraction from the raw data and the integration of unrelated dimensional features or parameters from multiple sensor data. These high-dimensional features provide stronger discrimination capabilities in target recognition, thereby improving fusion efficiency and overcoming the inherent limitations of a single sensor. In recent years, due to the direct use of existing visual pattern recognition neural network architectures, research on the combination of multi-sensor features has been insufficient. Most studies focus on implementing feature-level fusion strategies through target parameter extraction methods.

3.2.3. Decision-Level Fusion

Decision-level fusion is a high-level fusion. Each sensor does not merge the raw data and features, but processes its own data independently first, and then fuses the final decision of each sensor through the decision fusion method, so as to make a decision or prediction on the detected object or environment. Common methods of decision fusion include probability-based methods, deep learning methods and fuzzy subset hypothesis methods. Reference [129] uses decision-level fusion of radar signals and LiDAR point cloud data, and then uses a nonlinear Kalman filter method to detect obstacles and state tracking. Reference [130] proposes a two-sensor decision fusion method based on evidence reasoning rules for object classification. This method uses a Logistic model to calculate the reliability of the sensor based on the difference in classification decisions within a certain time span, and calculates the adaptive weight of the sensor based on the coefficient of variation to adapt to different environments. They used the Nuscenes dataset and the Waymo dataset to conduct a comparative experimental study on object classification decision fusion in autonomous driving. The results show that when there are only two sensors, the proposed method can effectively improve the fusion accuracy. Reference [131] proposed an attention-based fusion neural network (AFnet) model that can decouple data correlation through the encoder without considering traditional constraints and make full use of the nonlinear fitting ability of deep learning. Experimental results in NuScenes and Carla show that under the decision-level fusion framework, AFnet shows excellent performance in dealing with complex fusion problems caused by vehicle occlusion and overlap, achieving a state-of-the-art fusion matching accuracy of 99.11%. Reference [132] proposed a framework that combines fuzzy logic and neural networks. The framework combines Kalman filtering and an adaptive filtering algorithm (i.e., ANFIS) to construct an effective data fusion method for target tracking systems. The fuzzy adaptive fusion algorithm is an effective tool to keep the actual quality of the residual covariance consistent with its assumed value. ANFIS has good absorption and prediction capabilities, which makes it an effective tool for dealing with empirical defects in any system. This method uses the learning and prediction capabilities of ANFIS to adjust the confidence of the sensor through the training data set, thereby improving the accuracy of target tracking.

Decision-level fusion combines multiple decisions made by different sensors. The final fusion effect depends on the performance of the fusion strategy. Through this level of information fusion, the final decision is directly generated. Since the data of each sensor areprocessed independently before fusion, the system does not need to process a large amount of raw data or features, thereby reducing the overall computational burden. Decision-level fusion allows for modularity, where sensors operate independently using dedicated algorithms suitable for their specific data types. This makes it easier to integrate new sensors or update individual sensor algorithms without affecting the entire system. However, since only the final decision is fused, many of the detailed information available in the raw data or features are lost. This may limit the system's ability to make a nuanced interpretation of the environment. It can also be challenging to handle conflicting and

redundant data from multiple sensors. In some studies, it is combined with other fusion methods to take advantage of the advantages of each level while mitigating its limitations.

3.3. Multi-Sensor Information Fusion Algorithm

Multi-sensor information fusion involves integrating data from multiple sensors to improve the accuracy, reliability, and robustness of environmental perception in autonomous driving. To this end, various methods and algorithms have been developed, each with its own advantages and disadvantages. Since there is currently no completely unified algorithm that can adapt to all scenarios, it is necessary to select appropriate algorithms according to different application scenarios. This section discusses several key algorithms used in multi-sensor information fusion, including Kalman filter, Bayesian estimation, D-S evidence theory, and deep learning methods, and then compares and analyzes these methods. Table 5 summarizes the advantages and disadvantages of commonly used fusion algorithms and their application scenarios.

Table 5. Common fusion algorithms.

Fusion Algorithm	Advantages	Disadvantages	Application
Kalman Filter	effectively handle Gaussian noise; real-time processing capability	nonlinear systems perform poorly	vehicle tracking and navigation
Bayesian Estimation	able to integrate prior knowledge; suitable for redundant data	requires probabilities to be independent; may have convergence issues	data fusion and object recognition
D-S Evidence Theory	ability to handle uncertainty and inconsistent information	Evidence of an inability to effectively handle conflict	decision making and target tracking
Deep Learning	it has the ability of self-learning and has a higher accuracy when processing large data sets.	computationally demanding; lack of interpretability	object detection and classification

3.3.1. Kalman Filter

The Kalman filter is a recursive algorithm that uses input measurements from a mathematical process model to recursively estimate the current state of a system over time [133]. It is implemented in two steps: first, in the prediction phase, an estimate of the current state is given under uncertainty. Then, after the measurement is obtained, the previous estimate is updated by a weighted average. The Kalman filter fuses multi-sensor data information with dynamic environmental information. When the system and sensor noise are both Gaussian white noise in a linear dynamic model, the Kalman filter can provide a statistically optimal estimate of the fused information. Therefore, the Kalman filter can provide an optimal estimate for statistical multi-sensor system information fusion, and its recursive nature can eliminate the large amount of storage and computation required for information processing. Today, the Kalman filter is widely used in multi-target tracking and state estimation (such as position, velocity, and direction) in multi-sensor systems. The Kalman filter uses multi-sensor information such as LiDAR, camera, IMU, etc., to locate the vehicle and build a map. However, the Kalman filter can only accurately estimate linear systems and it is difficult to achieve optimal estimates for nonlinear systems. Since the motion process of the vehicle is nonlinear, many studies have adopted variants of Kalman filtering, such as extended Kalman filtering and unscented Kalman filtering, to linearize the nonlinear problem, and select the optimal algorithm according to different linearity models.

3.3.2. Bayesian Estimation

Bayesian estimation is a method that represents various uncertain information provided by multiple sensors as probabilities, and processes them using the Bayesian conditional probability formula in probability theory [134]. Bayesian estimation is a common method for multi-sensor low-level redundant data fusion, and its information is described

as the probability distribution of sensor information with Gaussian noise uncertainty added. Bayesian estimation is based on minimizing risk cost as a model and can update the hypothesis likelihood function given a prior likelihood estimate and additional observations. Bayesian estimation is applicable to redundant data, but it requires probability independence and requires prior probability and conditional probability. Particle filtering is a Bayesian algorithm that can handle nonlinear and non-Gaussian estimation problems. Particle filtering is based on the point mass ("particle") representation of probability density. It is achieved by sampling a certain number of discrete samples (particles) from a suitable probability density function and using the probability density (or probability) of the sample points as the corresponding weights [135]. Using these samples and corresponding weights, the posterior probability density can be approximately estimated, thereby achieving state estimation. When the sample size is large enough, the discrete particle estimation method will approach the posterior probability density of any distribution with high accuracy. In autonomous vehicles, particle filters can fuse multiple sensors for positioning, tracking, and map matching.

3.3.3. D-S Evidence Theory

The D-S evidence theory is a generalization of Bayesian estimation. The D-S evidence theory is not hindered by incomplete models or a lack of prior information. The Bayesian theory requires the definition or assumption of prior probabilities and can only assign evidence to one hypothesis. In the D-S evidence theory, evidence is assigned only based on known information and no assumptions are made. Evidence can be assigned to multiple hypotheses, which constitute a proposition [136]. In multi-sensor information fusion, the D-S evidence theory is structurally divided into three stages. The first stage is the target synthesis, which integrates the observations of multiple sensors into a single total output. The second stage is inference, which obtains and infers observations and expands them into target reports. The third stage is updating, because of the random errors of various sensors, a set of time-independent continuous reports from the same sensor is more reliable. It can provide a faster and more accurate method for ignorant multi-sensor fusion. However, its disadvantage is that it cannot effectively handle contradictory evidence, and the calculation is often more complicated than other methods. The D-S combination rule is sensitive, and sometimes some slight changes in the underlying probability distribution may lead to significant changes in the results.

3.3.4. Deep Learning

Deep learning can be considered an improvement in neural networks, and its core concept is artificial neural networks. Artificial neural networks are a type of non-programmed, adaptive, brain-like information processing. Their essence is to achieve parallel distributed information processing functions through network transformation and dynamic behavior, and to imitate the information processing of the human brain and nerves to varying degrees and levels [137]. Artificial neural networks have high fault tolerance, robustness, and self-organization in information processing. In the process of realizing multi-sensor information fusion, it is necessary to first determine the neural network model, network topology, and learning rules according to the system requirements and the form of fusion. CNN and RNN are two commonly used algorithms for deep learning sensor fusion in autonomous driving. Figure 3 summarizes these two algorithms and their variants. In the absence of a functional model of the sensor fusion system, artificial neural networks can be trained through a large amount of test data to obtain the network structure and mapping relationship, which is suitable for complex multi-sensor information fusion scenarios.

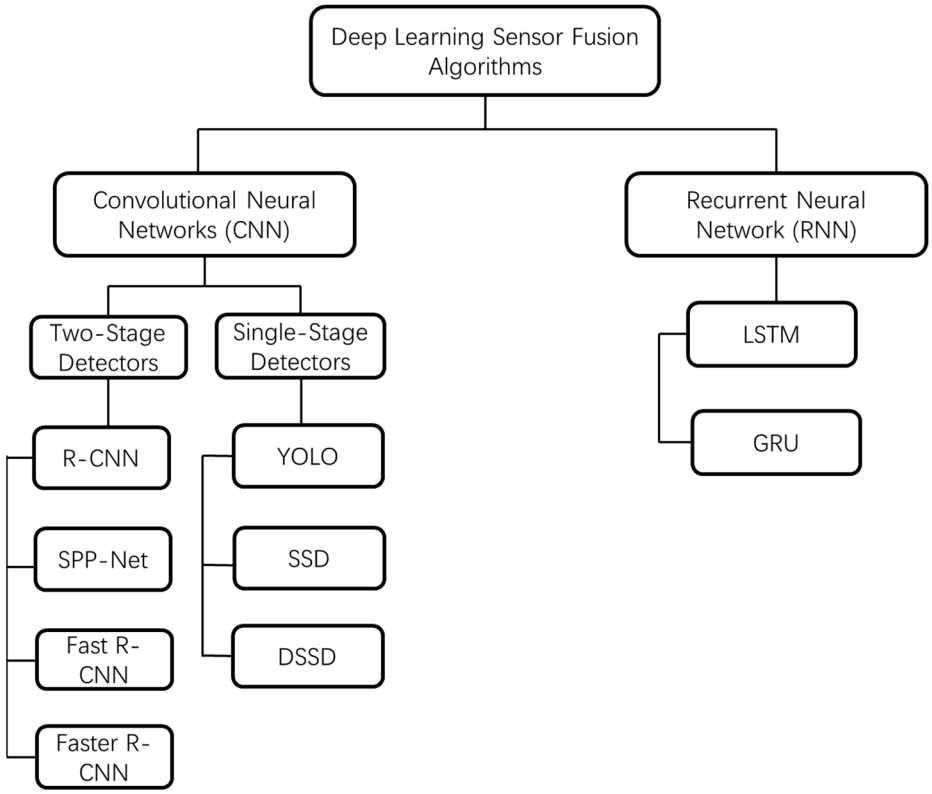

Figure 3. Two commonly used deep learning sensor fusion algorithms in autonomous driving [138].

Among the existing various sensor information fusion technologies, different information fusion algorithms have their own shortcomings and limitations. Therefore, in future research, we should focus on organically combining different algorithms to form new fusion algorithms and further improve the performance of the fusion system. At present, there are some algorithm fusion methods, such as combining fuzzy logic, artificial neural networks, wavelet transform, and Kalman filtering.

4. Application of Multi-Sensor Information Fusion in Environmental Perception

During normal driving, the driver needs to constantly observe the surrounding traffic conditions. In addition to observing traffic lights, it is more important to analyze and predict the behavioral intentions of pedestrians and other vehicles, and make responses in advance based on personal experience to avoid potential dangers that may occur during driving. In order to achieve this goal, autonomous driving needs to monitor changes in the surrounding environment in real time, and track and analyze the movement intentions and behavioral trends of the target, so that correct decisions can be made in advance to avoid dangers. Multi-sensor information fusion can make up for the situation where a single sensor is easily disturbed in a complex traffic environment and causes information distortion. Multi-sensor fusion plays an important role in improving the safety, reliability and stability of autonomous driving systems. By integrating data information from sensors such as cameras, lidars, and radars, vehicles can perceive and understand the surrounding environment more comprehensively and accurately, and apply specific fusion strategies in the process of perceiving the surrounding environment to ensure the reliability of the final result. This section will discuss the main applications of multi-sensor information fusion in autonomous driving environment perception.

4.1. Object Detection and Tracking

The perception system of an autonomous vehicle needs to understand all objects in the surrounding environment that may interact with it in real time. For example, lane line detection should also detect other vehicles, pedestrians, obstacles, and traffic signs. In addition to detection, the system should also be able to track detected objects in the spatial domain in the time domain. The main goal is to predict possible events based on the predicted motion of other vehicles, people, or obstacles [139]. This ensures the safety of the autonomous vehicle, passengers, and other road users. In a complex traffic environment, due to the presence of interference factors, it is difficult for a single sensor to guarantee the accuracy of detection. This usually involves information fusion of multiple sensors. Different sensor data information complements each other, which improves the credibility of the data and provides a more comprehensive and accurate understanding of the surrounding environment. Many papers focus on the fusion of the main sensors in autonomous driving (cameras, LiDAR, radar), some focus on traditional fusion methods using state estimators (Kalman filters, particle filters), and others use machine learning fusion methods (such as deep neural networks). In addition, some papers focus on single target detection and tracking with multi-sensor fusion, while others focus on multi-target detection and tracking problems with multi-sensor fusion. In [140], a new fusion extended Kalman filter (fusion-EKF) was proposed, which combined error bound (EB) and homography estimation to align different coordinate systems of radar and camera sensors. This study aims to fuse data from heterogeneous sensors such as mmWave radar and camera to improve the reliability and accuracy of tracking and detection in advanced driver assistance systems (ADAS). Sensor fusion and association are performed in the fusion EKF using homography estimation method (HEM), timeline alignment and region search. Reliable detection and cross-validated target tracking are achieved. At the same time, they introduced the concept of error bound (EB) to define the approximate region of sensor data, thereby enhancing the accuracy of data association in fusion-EKF. Experimental results show that the proposed fusion system can achieve a distance accuracy of 0.29 m and an angle accuracy of 0.013 rad in real time. Therefore, the proposed fusion system is effective, reliable and computationally efficient for real-time motion fusion applications. In [141], a fusion method of millimeter wave (MMW) radar and camera vision is proposed for pedestrian tracking. They used an unscented Kalman filter (UKF) for radar tracking. After detecting and locating pedestrian targets from vision using YOLOv5 and DeepSORT, an unscented Kalman filter (UKF) was used for radar tracking. The detection results of the two sensors were unified into a polar coordinate system with the sensor as the pole. Then, the target information obtained by visual processing was associated with the target information obtained by the radar module. Finally, the error covariance of each module was applied to achieve fusion tracking. In [142], a new spatial attention fusion (SAF) method for obstacle detection using millimeter-wave radar and visual sensors was proposed. The method is that the SAF block generates an attention weight matrix that can refine visual features based on radar data. SAF is implemented within the FCOS (fully convolutional single-stage object detection) framework to enhance the model's ability to detect small and long-distance obstacles. In [143], a radar–camera fusion method for multi-target detection and tracking in intelligent transportation systems (ITS) was proposed. This method uses a position inference algorithm and an improved EKF method to fuse radar and camera to detect and track pedestrians and vehicles. Reference [144] proposes a new deep learning method for multi-object tracking (MOT) that integrates data from millimeter-wave radar and camera sensors to improve the accuracy and robustness of autonomous driving. The method uses a bidirectional long short-term memory (Bi-LSTM) network to integrate long-term temporal information and improve motion prediction. In addition, an appearance

feature model inspired by FaceNet is adopted to ensure consistent tracking by associating objects between different frames. Reference [145] solves the problem of object detection in severe weather conditions by fusing cameras, LiDAR and radar using an adaptive deep fusion architecture. Reference [146] reviews the latest technologies for multi-object detection (MOD) and multi-object tracking (MOT) using deep neural networks (DNNs) for camera, LiDAR and radar sensor fusion. Most autonomous driving systems rely on 3D perception because it can provide depth information and 3D structure information around the vehicle. The precise 3D point cloud data provided by LiDAR is fused with visual information to achieve 3D object detection and tracking. Reference [147] developed a new method called DeepFusions to enhance 3D object detection by fusing deep features from LiDAR and camera sensors. The method consists of two key methods (InverseAug and LearnableAlign), which can help camera images to be effectively aligned with LiDAR point clouds at marginal computational cost (i.e., only one cross-attention layer). Reference [148] proposed a novel camera–LiDAR fusion 3D MOT framework called CAMO-MOT, which uses both camera and LiDAR data and significantly reduces tracking failures caused by occlusion and false detection. Reference [149] proposed a joint multi-object detection and tracking (JMODT) system based on end-to-end camera–LiDAR fusion, which performs joint detection and tracking by using 2D and 3D measurements for parallel object detection and association. Reference [150] proposes a new simultaneous detection and tracking baseline algorithm MotionTrack with multimodal sensor input in an autonomous driving environment. Current advances in deep learning will lead to more powerful object detection and tracking models, and with the application of mobile edge computing [151], sensor data can be processed faster, reducing latency and improving real-time performance.

4.2. Positioning and Mapping

Autonomous driving needs to accurately determine the position and orientation of the environment in which it is located, and create or update the map of the environment in real time. This allows the vehicle to identify and avoid surrounding obstacles, while achieving safe navigation. In addition, the positioning system should be sufficiently robust and accurate to cope with various complex environments and severe weather conditions. Generally, in autonomous driving systems, it is a common practice to increase the overall positioning and mapping performance by fusing two or more sensors. Reference [152] proposed a camera–radar sensor fusion framework based on vehicle component (rear corner) detection and positioning to improve the stability of vehicle positioning. The main idea of this method is to enhance the azimuth accuracy of radar information by detecting and locating the rear corner of the target vehicle from the image. The proposed method can effectively track the corner points of the vehicle using only one visible corner even in occluded scenes. GPS may experience signal attenuation in indoor spaces and urban canyons, and IMU may drift over time due to error accumulation when integrating acceleration to determine velocity and position. Reference [153] uses an unscented Kalman filter (UKF) Bayesian filter to fuse GPS and IMU data. Reference [154] uses the error state Kalman filter (ESKF) and the piecewise Rauch–Tung–Striebel (RTS) smoothing algorithm to fuse GNSS and IMU to improve the positioning accuracy and robustness of GNSS and IMU sensors. The proposed ESKF-RTS method provides more accurate and reliable positioning for autonomous vehicles, especially in environments where GNSS signals are unreliable. Reference [155] proposes a fusion algorithm that combines the Kalman filter (KF) with a new INS/GPS neural network framework (GI-NN) to assist INS in reducing the accumulated navigation error during GPS signal loss. Reference [156] studies the information fusion of GPS and INS sensors using the Kalman filter. In order to improve the performance of the Kalman filter, a robust method using the Mahalanobis distance is adopted. In order to

compensate for the problem of GPS data interruption, an artificial neural network is used to assist GPS\INS information fusion. Visual sensors are important elements in autonomous driving positioning and mapping systems. Reference [157] proposes an improved multi-sensor fusion positioning system based on GNSS/LiDAR/Vision/IMU. With semi-tight coupling and graph optimization. The system tightly couples the raw observation data of LiDAR, vision and IMU, while adding the positioning information of GNSS as a loosely coupled component to reduce the cumulative error. They use factor graph optimization methods to integrate data from different sensors and improve the accuracy and reliability of positioning through nonlinear optimization. In order to solve the limitations of SLAM algorithms in map drift, the literature [158] proposed a sensor fusion SLAM and positioning method based on LiDAR for offline mapping and online positioning of autonomous driving vehicles, integrating LiDAR with other sensors such as GNSS, IMU and vehicle status data to improve map accuracy and positioning performance, even when satellite signals are unavailable or unreliable. This type of fusion will have many advantages by combining the power of accurate LiDAR depth estimation and camera tracking capabilities. However, the fusion of LIDAR and visual SLAM will produce large cumulative errors in high-speed motion. Therefore, low-cost and high-performance inertial sensing units have become the first choice to make up for this defect [159]. Reference [160] proposed the first LiDAR–inertial–vision fusion SLAM system, LVI-SLAM. The system uses a tightly coupled framework to fuse the measurements of all three sensors, addressing the challenges of noisy point clouds, fast sensor movement, and tunnel environments. Reference [161] proposed mVIL-Fusion, a new SLAM system designed to fuse data from a monocular camera, an IMU (inertial measurement unit), and a 3D rotating LiDAR. The system aims to overcome the limitations of traditional SLAM methods, such as time synchronization issues and poor mapping performance in highly variable environments. Reference [162] proposed an effective positioning algorithm for autonomous vehicles based on vision–LiDAR–IMU fusion. In order to solve the problem of excessive back-end optimization calculations, a balanced selection strategy is adopted to reduce the computational complexity, focusing on key frames and sliding windows to optimize pose estimation. In dealing with large-scale drift, a loop detection algorithm based on iterative closest point (ICP) is proposed to improve long-term positioning accuracy. Common positioning and mapping techniques, as shown in Table 6.

Table 6. Comparison of localization and mapping technologies in terms of accuracy, cost, computational load, sources of external influences, and data storage size [138].

Method	Accuracy	Cost	Computational Load	External Effect	Data Size
GPS/IMU	low	medium	low	signal outage	low
GPS/INS/LiDAR/Camera	high	medium	medium	map accuracy	high
SLAM	high	low	high	illumination	high
Visual Odometry	medium	low	high	illumination	high
Map-Based Matching	very High	medium	very high	map change	very high

4.3. Scene Segmentation

Scene segmentation based on multi-sensor fusion is to accurately classify different parts of the environment through different sensor data, so as to accurately segment and classify different areas in the environment, enhance the understanding of the environment of autonomous driving vehicles, and improve the detection and positioning capabilities of objects. According to different environments, scenes can be divided into structured scenes (urban roads, parking lots, etc.) and unstructured scenes (rural towns, wilderness, etc.). Since unstructured scenes present higher complexity, most of them lack clear lane lines or are affected by irregular light changes. Therefore, structured scenes are easier to handle

than unstructured scenes. Existing large-scale open-source datasets are mainly concentrated in structured scenes, some of which contain some unstructured scenes. Therefore, many current studies focus on structured scene segmentation. Reference [163] proposed a Perception-Aware Multi-Sensor Fusion (PMF) scheme to effectively fuse the perception information from RGB images and point clouds. By fusing the spatiotemporal depth information of point clouds and the appearance information of RGB images, PMF can solve the segmentation problems of poor lighting conditions and sparse point clouds. Experiments show that PMF is robust to complex outdoor scenes. Reference [164] uses a bird's eye view (BEV) generated from a LiDAR point cloud combined with camera image segmentation to achieve more accurate lane marking detection. The DeepLabV3+ network image segmentation method is first used to segment the image captured by the camera, and then fused with the LiDAR data to generate a more accurate 3D spatial understanding of the lane marking. The segmentation network based on the DeepLabV3+ architecture is combined with a long short-term memory (LSTM) module to utilize temporal information and improve segmentation accuracy. Reference [165] proposes a new camera–LiDAR fusion model for lane line segmentation, and proposes a multimodal network modeling method based on information theory to optimize the fusion strategy and improve the robustness of lane segmentation. The optimal fusion network obtained by the proposed method achieves a lane line accuracy of more than 85% and an overall accuracy of more than 98.7%. Reference [166] proposes an entropy-based adaptive entropy multimodal fusion method for the night lane segmentation problem. This method uses attention to capture the spatial relationship between modalities and illumination distribution, and adaptively fuses them. Through the proposed lane feature enhancement module, they enhance lane features globally and locally, improving the network's ability to capture lanes. Since it is not feasible to directly apply existing methods to scenes of different structural types for robust autonomous driving systems, some research on unstructured scene segmentation has emerged in recent years. Reference [167] proposed an effective multimodal network called M2F2-Net for free space detection in unstructured off-road scenes. The network uses a multimodal cross fusion (MCF) fusion module to fuse the features of RGB images and surface normal maps (compiled from LiDAR point clouds). Reference [168] uses a grouped attention network (GA-Nav) to classify drivable and obstacle areas in RGB images, and then the Patchwork++ algorithm segments the LiDAR point cloud into ground and non-ground areas. Finally, a late fusion method is proposed to better fuse the two results to classify the drivable area. The fusion model successfully corrects the misclassification of the camera-based system by integrating LiDAR data. For example, bushes that were misclassified as drivable by the camera are correctly segmented as obstacles after fusion. Reference [169] proposed a multi-sensor fusion network combined with surface normals (SN) for unstructured scene segmentation. The network effectively combines 3D information from LiDAR and high-resolution color and texture features from RGB cameras. It fuses point cloud representations (based on point and range views) with reweighted RGB images of different scales. Surface normal features are extracted from LiDAR data to reweight the RGB image data to reduce the negative impact of inaccurate or unreliable visual data, especially in low-light conditions.

5. Conclusions and Future Research Suggestions

This paper deeply analyzes the environmental perception technology of multi-sensor information fusion in autonomous driving, and reveals the key role of making full use of sensor complementarity in improving the performance of autonomous driving systems. This paper focuses on the methods of heterogeneous sensor fusion, including the integration of lidar, camera, millimeter-wave radar, and GNSS/IMU, to effectively improve the

vehicle's perception of the surrounding environment and enhance the accuracy, stability and robustness of the system. Through a review of relevant literature, this paper summarizes the advantages of these sensors in applications such as data acquisition, target detection and tracking, positioning and map construction, and explores the key methods of calibration and data fusion in multi-sensor information fusion. This study finds that multi-sensor information fusion technology plays a key role in environmental perception of autonomous driving. Fusion technology significantly improves the accuracy of target detection and tracking, enhances the reliability of positioning and map construction, and improves the effect of scene segmentation. In particular, the introduction of deep learning technology has opened up new ways for multi-sensor information fusion, and the perception performance has been significantly improved by learning complex data features and patterns. In addition, this paper also explores the application of the Kalman filter, Bayesian estimation, D-S evidence theory, and deep learning algorithm in multi-sensor information fusion, and analyzes their role in specific tasks such as target detection and tracking, positioning and mapping, and scene segmentation.

Looking to the future, the development direction of autonomous driving technology is to optimize sensor fusion algorithms, thereby improving the ability of autonomous driving systems to cope with complex and dynamic environments and improving real-time decision-making under various driving conditions. Improving sensor performance, especially performance assurance in bad weather and complex environments, will be the key to ensuring the safety and reliability of autonomous driving. The seamless integration of perception systems and vehicle control systems is essential for the safety of autonomous vehicles, especially in high-speed or high-risk driving scenarios. In addition, the integration of edge computing and cloud computing can effectively solve the problem of limited on-board computing resources and achieve real-time and efficient data processing. The integration of smart cities and vehicle-to-everything (V2X) can enhance the interaction and data exchange between autonomous vehicles and the surrounding environment, thereby achieving more comprehensive environmental perception. By deeply exploring these research directions, the environmental perception ability of autonomous vehicles will be raised to a new level, promoting the transformation and widespread application of autonomous driving technology. This study emphasizes the importance of multi-sensor information fusion in autonomous driving and points out the direction of future research. Through the analysis of this article, we believe that multi-sensor information fusion technology is the key to achieving safe and reliable autonomous driving, and with the continuous advancement of technology, this field will continue to develop and improve.

Author Contributions: Formal analysis, J.L.; Writing—original draft, B.Y.; Writing—review & editing, T.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by the Guangdong Basic and Applied Basic Research Foundation (no. 2020A 1515110999).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. PraveenKumar, S.; Agyekum, E.B.; Kumar, A.; Velkin, V.I. Performance evaluation with low-cost aluminum reflectors and phase change material inte-grated to solar PV modules using natural air convection: An experimental investigation. *Energy* **2023**, *266*, 126415. [CrossRef]
2. Liang, J.; Yang, K.; Tan, C.; Wang, J.; Yin, G. Enhancing High-Speed Cruising Performance of Autonomous Vehicles through Integrated Deep Rein-forcement Learning Framework. *arXiv* **2024**, arXiv:2404.14713.

3. Liang, J.; Tian, Q.; Feng, J.; Pi, D.; Yin, G. A Polytopic model-based robust predictive control scheme for path tracking of autonomous vehicles. *IEEE Trans. Intell. Veh.* **2023**, *9*, 3928–3939. [CrossRef]
4. Peng, H.; Wang, W.; An, Q.; Xiang, C.; Li, L. Path tracking and direct yaw moment coordinated control based on robust MPC with the finite time horizon for autonomous independent-drive vehicles. *IEEE Trans. Veh. Technol.* **2020**, *69*, 6053–6066. [CrossRef]
5. Chen, I.-M.; Chan, C.-Y. Deep reinforcement learning based path tracking controller for autonomous vehicle. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2021**, *235*, 541–551. [CrossRef]
6. Zhang, H.; Deng, Y. Weighted belief function of sensor data fusion in engine fault diagnosis. *Soft Comput.* **2020**, *24*, 2329–2339. [CrossRef]
7. Luo, L.; Wang, X.; Guo, H.; Lasaponara, R.; Zong, X.; Masini, N.; Wang, G.; Shi, P.; Khatteli, H.; Chen, F.; et al. Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: A review of the century (1907–2017). *Remote Sens. Environ.* **2019**, *232*, 111280. [CrossRef]
8. Mao, X.; Li, W.; Lei, C.; Jin, J.; Duan, F.; Chen, S. A brain–robot interaction system by fusing human and machine intelligence. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 533–542. [CrossRef]
9. Tao, Y.; He, Y.; Ma, X.; Xu, H.; Hao, J.; Feng, J. SLAM Method Based on Multi-Sensor Information Fusion. In Proceedings of the 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 24–26 September 2021.
10. Gan, S.; Zhuang, Q.; Gong, B. Human-computer interaction based interface design of intelligent health detection using PCANet and multi-sensor information fusion. *Comput. Methods Programs Biomed.* **2022**, *216*, 106637. [CrossRef]
11. Jia, X.; Hu, Z.; Guan, H. A new multi-sensor platform for adaptive driving assistance system (ADAS). In Proceedings of the 2011 9th World Congress on Intelligent Control and Automation (WCICA 2011), Taipei, Taiwan, 21–25 June 2011.
12. Rosero, L.A.; Osório, F.S. Calibration and multi-sensor fusion for on-road obstacle detection. In Proceedings of the 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR), Curitiba, Brazil, 8–11 November 2017; IEEE: Piscataway, NJ, USA, 2017.
13. Etinger, A.; Balal, N.; Litvak, B.; Einat, M.; Kapilevich, B.; Pinhasi, Y. Non-imaging MM-wave FMCW sensor for pedestrian detection. *IEEE Sens. J.* **2013**, *14*, 1232–1237. [CrossRef]
14. Lee, S.; Yoon, Y.J.; Lee, J.E.; Kim, S.C. Human–vehicle classification using feature-based SVM in 77-GHz automotive FMCW radar. *IET Radar Sonar Navig.* **2017**, *11*, 1589–1596. [CrossRef]
15. Wang, Z.; Wu, Y.; Niu, Q. Multi-Sensor Fusion in Automated Driving: A Survey. *IEEE Access* **2020**, *8*, 2847–2868. [CrossRef]
16. Marti, E.D.; de Miguel, M.A.; Garcia, F.; Perez, J. A review of sensor technologies for perception in automated driving. *IEEE Intell. Transp. Syst. Mag.* **2019**, *11*, 94–108. [CrossRef]
17. Middleton, W.E.K.; Spilhaus, A.F. *Meteorological Instruments*; University of Toronto Press: Toronto, ON, Canada, 1941.
18. Amann, M.C.; Bosch, T.M.; Lescure, M.; Myllylae, R.A.; Rioux, M. Laser ranging: A critical review of unusual techniques for distance measurement. *Opt. Eng.* **2001**, *40*, 10–19.
19. Lum, D.J. Ultrafast time-of-flight 3D LiDAR. *Nat. Photonics* **2020**, *14*, 2–4. [CrossRef]
20. Behroozpour, B.; Sandborn, P.A.; Wu, M.C.; Boser, B.E. Lidar system architectures and circuits. *IEEE Commun. Mag.* **2017**, *55*, 135–142. [CrossRef]
21. Li, Y.; Ibanez-Guzman, J. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and per-ception systems. *IEEE Signal Process. Mag.* **2020**, *37*, 50–61. [CrossRef]
22. Roriz, R.; Cabral, J.; Gomes, T. Automotive LiDAR technology: A survey. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 6282–6297. [CrossRef]
23. Wang, P. Research on comparison of lidar and camera in autonomous driving. *J. Phys. Conf. Ser.* **2021**, *2093*, 012032. [CrossRef]
24. Yang, B.; Luo, W.; Urtasun, R. Pixor: Real-time 3D object detection from point clouds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7652–7660.
25. Xie, D.; Xu, Y.; Wang, R. Obstacle detection and tracking method for autonomous vehicle based on three-dimensional LiDAR. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881419831587. [CrossRef]
26. Li, J.; Zhang, Y.; Liu, X.; Zhang, X.; Bai, R. Obstacle detection and tracking algorithm based on multi-lidar fusion in urban environment. *IET Intell. Transp. Syst.* **2021**, *15*, 1372–1387. [CrossRef]
27. Chew, W.K.; Zakaria, M.A. Outdoor localisation for navigation tracking using differential global po-sitioning system estimation (DGPS): Positioning errors analysis. *Mekatronika* **2019**, *1*, 103–114. [CrossRef]
28. Li, X.; Huang, J.; Li, X.; Shen, Z.; Han, J.; Li, L.; Wang, B. Review of PPP–RTK: Achievements, challenges, and opportunities. *Satell. Navig.* **2022**, *3*, 28. [CrossRef]
29. Elsheikh, M.; Iqbal, U.; Noureldin, A.; Korenberg, M. The Implementation of Precise Point Positioning (PPP): A Comprehensive Review. *Sensors* **2023**, *23*, 8874. [CrossRef] [PubMed]
30. Kummerle, J.; Kuhner, T. Unified intrinsic and extrinsic camera and LiDAR calibration under uncertainties. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 6028–6034.

31. Yeong, D.J.; Velasco-Hernandez, G.; Barry, J.; Walsh, J. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors* **2021**, *21*, 2140. [CrossRef] [PubMed]
32. Lesson 3: Sensor Calibration—A Necessary Evil—Module 5: Putting It Together—An Autonomous Vehicle State Estimator | Coursera. Available online: https://www.coursera.org/lecture/state-estimation-localization-self-driving-cars/lesson-3-sensorcalibration-a-necessary-evil-jPb2Y (accessed on 15 June 2020).
33. Kwak, K.; Huber, D.F.; Badino, H.; Kanade, T. Extrinsic calibration of a single line scanning lidar and a camera. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 3283–3289. [CrossRef]
34. Hall, D.; Llinas, J. An introduction to multisensor data fusion. *Proc. IEEE* **1997**, *85*, 6–23. [CrossRef]
35. Kumar, P.; Gauba, H.; Roy, P.P.; Dogra, D.P. Coupled HMM-based multi-sensor data fusion for sign language recognition. *Pattern Recognit. Lett.* **2017**, *86*, 1–8. [CrossRef]
36. Qiu, Z.; Martínez-Sánchez, J.; Arias-Sánchez, P.; Rashdi, R. External multi-modal imaging sensor calibration for sensor fusion: A review. *Inf. Fusion* **2023**, *97*, 101806. [CrossRef]
37. Zhou, L.; Li, Z.; Kaess, M. Automatic Extrinsic Calibration of a Camera and a 3D LiDAR Using Line and Plane Correspondences. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 5562–5569. [CrossRef]
38. Park, Y.; Yun, S.; Won, C.S.; Cho, K.; Um, K.; Sim, S. Calibration between color camera and 3D LIDAR instruments with a polygonal planar board. *Sensors* **2014**, *14*, 5333–5353. [CrossRef] [PubMed]
39. Yoon, S.; Ju, S.; Nguyen, H.M.; Park, S.; Heo, J. Spatiotemporal Calibration of Camera-LiDAR Using Nonlinear Angular Constraints on Multiplanar Target. *IEEE Sens. J.* **2022**, *22*, 10995–11005. [CrossRef]
40. Gong, X.; Lin, Y.; Liu, J. 3D LIDAR-camera extrinsic calibration using an arbitrary trihedron. *Sensors* **2013**, *13*, 1902–1918. [CrossRef] [PubMed]
41. Velas, M.; Spanel, M.; Materna, Z.; Herout, A. Calibration of RGB camera with velodyne LiDAR. In Proceedings of the 22nd International Conference in Central Europeon Computer Graphics, Visualization and Computer Visionin Co-Operation with EUROGRAPHICS Association, Plzeň, Czech Republic, 2–5 June 2014; WSCG 2014 Communication Papers Proceedings. Václav Skala-UNION Agency: Plzeň, Czech Republic, 2014; pp. 135–144.
42. Kummerle, J.; Kuhner, T.; Lauer, M. Automatic calibration of multiple cameras and depth sensors with a spherical target. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Madrid, Spain, 1–5 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8.
43. Yang, R.; Cheng, S.; Chen, Y. Flexible and accurate implementation of a binocular structured light system. *Opt. Lasers Eng.* **2008**, *46*, 373–379. [CrossRef]
44. Rufli, M.; Scaramuzza, D.; Siegwart, R. Automatic detection of checkerboards on blurred and distorted images. In Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, 22–26 September 2008; pp. 3121–3126. [CrossRef]
45. Lee, S.-H.; Kim, T.-E.; Choi, J.-S. Correction of radial distortion using a planar checkerboard pattern and its image. *IEEE Trans. Consum. Electron.* **2009**, *55*, 27–33. [CrossRef]
46. Zhang, Q.; Pless, R. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566), Sendai, Japan, 28 September–2 October 2004; Volume 3, pp. 2301–2306. [CrossRef]
47. Li, Y.; Ruichek, Y.; Cappelle, C. 3D triangulation based extrinsic calibration between a stereo vision system and a LIDAR. In Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), Washington, DC, USA, 5–7 October 2011; pp. 797–802. [CrossRef]
48. Itami, F.; Yamazaki, T. An improved method for the calibration of a 2-D LiDAR With respect to a camera by using a checkerboard target. *IEEE Sens. J.* **2020**, *20*, 7906–7917. [CrossRef]
49. Li, Y.; Ruichek, Y.; Cappelle, C. Extrinsic calibration between a stereoscopic system and a LIDAR with sensor noise models. In Proceedings of the 2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Hamburg, Germany, 13–15 September 2012; pp. 484–489. [CrossRef]
50. Li, Y.; Ruichek, Y.; Cappelle, C. Optimal Extrinsic Calibration Between a Stereoscopic System and a LIDAR. *IEEE Trans. Instrum. Meas.* **2013**, *62*, 2258–2269. [CrossRef]
51. Chu, X.; Zhou, J.; Chen, L.; Xu, X. An improved method for calibration between a 2D LiDAR and a camera based on point-line correspondences. *J. Phys. Conf. Ser.* **2019**, *1267*, 012048. [CrossRef]
52. Itami, F.; Yamazaki, T. A simple calibration procedure for a 2D LiDAR with respect to a camera. *IEEE Sens. J.* **2019**, *19*, 7553–7564. [CrossRef]

53. Povendhan, A.P.; Yi, L.; Hayat, A.A.; Le, A.V.; Kai, K.L.J.; Ramalingam, B.; Elara, M.R. Multi-sensor Fusion Incorporating Adaptive Transformation for Reconfigurable Pavement Sweeping Robot. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 300–306. [CrossRef]
54. Sani, M.F.; Karimian, G. Automatic navigation and landing of an indoor AR. drone quadrotor using ArUco marker and inertial sensors. In Proceedings of the 2017 International Conference on Computer and Drone Applications (IConDA), Kuching, Malaysia, 9–11 November 2017; pp. 102–107. [CrossRef]
55. An, G.H.; Lee, S.; Seo, M.-W.; Yun, K.; Cheong, W.-S.; Kang, S.-J. Charuco board-based omnidirectional camera calibration method. *Electronics* **2018**, *7*, 421. [CrossRef]
56. Zhou, L.; Deng, Z. Extrinsic calibration of a camera and a lidar based on decoupling the rotation from the translation. In Proceedings of the 2012 IEEE Intelligent Vehicles Symposium (IV), Madrid, Spain, 3–7 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 642–648.
57. Domhof, J.; Kooij, J.F.; Gavrila, D.M. An extrinsic calibration tool for radar, camera and lidar. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019.
58. Domhof, J.; Kooij, J.F.; Gavrila, D.M. A joint extrinsic calibration tool for radar, camera and lidar. *IEEE Trans. Intell. Veh.* **2021**, *6*, 571–582. [CrossRef]
59. Zhang, J.; Liu, Y.; Wen, M.; Yue, Y.; Zhang, H.; Wang, D. L2V2T2 Calib: Automatic and Unified Extrinsic Calibration Toolbox for Different 3D LiDAR, Visual Camera and Thermal Camera. In Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 4–7 June 2023; IEEE: Piscataway, NJ, USA, 2023.
60. Pandey, G.; McBride, J.; Savarese, S.; Eustice, R. Extrinsic calibration of a 3D laser scanner and an omnidirectional camera. *IFAC Proc.* **2010**, *43*, 336–341. [CrossRef]
61. Wang, W.; Sakurada, K.; Kawaguchi, N. reflectance intensity assisted automatic and accurate extrinsic calibration of 3D LiDAR and panoramic camera using a printed chessboard. *Remote Sens.* **2017**, *9*, 851. [CrossRef]
62. Lai, Z.; Wang, Y.; Guo, S.; Meng, X.; Li, J.; Li, W.; Han, S. Laser reflectance feature assisted accurate extrinsic calibration for non-repetitive scanning LiDAR and camera systems. *Opt. Express* **2022**, *30*, 16242–16263. [CrossRef] [PubMed]
63. Grammatikopoulos, L.; Papanagnou, A.; Venianakis, A.; Kalisperakis, I.; Stentoumis, C. An effective camera-to-lidar spatiotemporal calibration based on a simple calibration target. *Sensors* **2022**, *22*, 5576. [CrossRef] [PubMed]
64. Ha, J.-E. Extrinsic calibration of a camera and laser range finder using a new calibration structure of a plane with a triangular hole. *Int. J. Control. Autom. Syst.* **2012**, *10*, 1240–1244. [CrossRef]
65. Tóth, T.; Pusztai, Z.; Hajder, L. Automatic LiDAR-camera calibration of extrinsic parameters using a spherical target. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; IEEE: Piscataway, NJ, USA, 2020.
66. Pusztai, Z.; Hajder, L. Accurate Calibration of LiDAR-Camera Systems Using Ordinary Boxes. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 394–402. [CrossRef]
67. Pusztai, Z.; Eichhardt, I.; Hajder, L. Accurate calibration of multi-lidar-multi-camera systems. *Sensors* **2018**, *18*, 2139. [CrossRef]
68. Chai, Z.; Sun, Y.; Xiong, Z. A novel method for lidar camera calibration by plane fitting. In Proceedings of the 2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Auckland, New Zealand, 9–12 July 2018; IEEE: Piscataway, NJ, USA, 2018.
69. Zamanakos, G.; Tsochatzidis, L.; Amanatiadis, A.; Pratikakis, I. A cooperative LiDAR-camera scheme for extrinsic calibration. In Proceedings of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 26–29 June 2022; IEEE: Piscataway, NJ, USA, 2022.
70. Fang, C.; Ding, S.; Dong, Z.; Li, H.; Zhu, S.; Tan, P. Single-shot is enough: Panoramic Infrastructure based calibration of multiple cameras and 3D LiDARs. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; IEEE: Piscataway, NJ, USA, 2017.
71. Persic, J.; Markovic, I.; Petrovic, I. Extrinsic 6DoF calibration of 3D LiDAR and radar. In Proceedings of the 2017 European Conference on Mobile Robots (ECMR), Paris, France, 6–8 September 2017; IEEE: Piscataway, NJ, USA, 2017.
72. Scholler, C.; Schnettler, M.; Krammer, A.; Hinz, G.; Bakovic, M.; Guzet, M.; Knoll, A. Targetless rotational auto-calibration of radar and camera for intelligent transportation systems. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; IEEE: Piscataway, NJ, USA, 2019.
73. Scaramuzza, D.; Harati, A.; Siegwart, R. Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes. In Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; IEEE: Piscataway, NJ, USA, 2007.
74. Ye, C.; Pan, H.; Gao, H. Keypoint-based LiDAR-camera online calibration with robust geometric network. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 2503011. [CrossRef]

75. Muñoz-Bañón, M.Á.; Candelas, F.A.; Torres, F. Candelas, and Fernando Torres. Targetless camera-LiDAR calibration in unstructured environments. *IEEE Access* **2020**, *8*, 143692–143705. [CrossRef]
76. Yuan, C.; Liu, X.; Hong, X.; Zhang, F. Pixel-level extrinsic self calibration of high resolution LiDAR and camera in targetless environments. *IEEE Robot. Autom. Lett.* **2021**, *6*, 7517–7524. [CrossRef]
77. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; IEEE: Piscataway, NJ, USA, 1999; Volume 2.
78. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006: Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006*; Proceedings, Part I 9; Springer: Berlin/Heidelberg, Germany, 2006.
79. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
80. Ma, H.; Liu, K.; Liu, J.; Qiu, H.; Xu, D.; Wang, Z.; Gong, X.; Yang, S. Simple and efficient registration of 3D point cloud and image data for an indoor mobile mapping system. *J. Opt. Soc. Am. A* **2021**, *38*, 579–586. [CrossRef]
81. Xu, H.; Lan, G.; Wu, S.; Hao, Q. Online Intelligent Calibration of Cameras and LiDARs for Autonomous Driving Systems. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; IEEE: Piscataway, NJ, USA, 2019.
82. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *6*, 679–698. [CrossRef]
83. Von Gioi, R.G.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: A line segment detector. *Image Process. Line* **2012**, *2*, 35–55. [CrossRef]
84. Zhu, N.; Jia, Y.; Ji, S. Registration of panoramic/fish-eye image sequence and LiDAR points using skyline features. *Sensors* **2018**, *18*, 1651. [CrossRef] [PubMed]
85. Peršić, J.; Petrović, L.; Marković, I.; Petrović, I. Online multi-sensor calibration based on moving object tracking. *Adv. Robot.* **2021**, *35*, 130–140. [CrossRef]
86. Ma, T.; Liu, Z.; Yan, G.; Li, Y. Crlf: Automatic calibration and refinement based on line feature for lidar and camera in road scenes. *arXiv* **2021**, arXiv:2103.04558.
87. Peršić, J.; Petrović, L.; Marković, I.; Petrović, I. Spatio-temporal multisensor calibration based on gaussian processes moving object tracking. *arXiv* **2019**, arXiv:1904.04187.
88. Pandey, G.; McBride, J.; Savarese, S.; Eustice, R. Automatic targetless extrinsic calibration of a 3d lidar and camera by maximizing mutual information. *Proc. AAAI Conf. Artif. Intell.* **2012**, *26*, 2053–2059. [CrossRef]
89. Taylor, Z.; Nieto, J. A mutual information approach to automatic calibration of camera and lidar in natural environments. In Proceedings of the Australian Conference on Robotics and Automation (ACRA), Melbourne, Australia, 4–6 December 2012.
90. Zhao, Y.; Wang, Y.; Tsai, Y. 2D-image to 3D-range registration in urban environments via scene categorization and combination of similarity measurements. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; IEEE: Piscataway, NJ, USA, 2016.
91. Pascoe, G.; Maddern, W.; Newman, P. Direct visual localisation and calibration for road vehicles in changing city environments. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015.
92. Jiang, P.; Osteen, P.; Saripalli, S. Semcal: Semantic lidar-camera calibration using neural mutual information estimator. In Proceedings of the 2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Karlsruhe, Germany, 23–25 September 2021; IEEE: Piscataway, NJ, USA, 2021.
93. Irie, K.; Sugiyama, M.; Tomono, M. Target-less camera-lidar extrinsic calibration using a bagged de-pendence estimator. In Proceedings of the 2016 IEEE International Conference on Automation Science and Engineering (CASE), Worth, TX, USA, 21–25 August 2016; IEEE: Piscataway, NJ, USA, 2016.
94. Li, X.; Xiao, Y.; Wang, B.; Ren, H.; Zhang, Y.; Ji, J. Automatic targetless LiDAR–camera calibration: A survey. *Artif. Intell. Rev.* **2023**, *56*, 9949–9987. [CrossRef]
95. Barzilai, J.; Borwein, J.M. Two-point step size gradient methods. *IMA J. Numer. Anal.* **1988**, *8*, 141–148. [CrossRef]
96. Nelder, J.A.; Mead, R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308–313. [CrossRef]
97. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168. [CrossRef]
98. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; IEEE: Piscataway, NJ, USA, 1995; Volume 4, pp. 1942–1948.
99. Kelley, C.T. *Iterative Methods for Optimization*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 1999.
100. Powell, M.J.D. *The BOBYQA Algorithm for Bound Constrained Optimization Without Derivatives*; Cambridge NA Report NA2009/06; University of Cambridge: Cambridge, UK, 2009; Volume 26, pp. 26–46.

101. Ishikawa, R.; Oishi, T.; Ikeuchi, K. Lidar and camera calibration using motions estimated by sensor fusion odometry. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 7342–7349.
102. Horaud, R.; Dornaika, F. Hand-eye calibration. *Int. J. Robot. Res.* **1995**, *14*, 195–210. [CrossRef]
103. Shi, C.; Huang, K.; Yu, Q.; Xiao, J.; Lu, H.; Xie, C. Extrinsic calibration and odometry for camera-LiDAR systems. *IEEE Access* **2019**, *7*, 120106–120116. [CrossRef]
104. Park, C.; Moghadam, P.; Kim, S.; Sridharan, S.; Fookes, C. Spatiotemporal camera-LiDAR calibration: A targetless and structureless approach. *IEEE Robot. Autom. Lett.* **2020**, *5*, 1556–1563. [CrossRef]
105. Liu, Z.; Chen, Z.; Wei, X.; Chen, W.; Wang, Y. External Extrinsic Calibration of Multi-Modal Imaging Sensors: A Review. *IEEE Access* **2023**, *11*, 110417–110441. [CrossRef]
106. Taylor, Z.; Nieto, J. Parameterless automatic extrinsic calibration of vehicle mounted lidar-camera systems. In Proceedings of the International Conference on Robotics and Automation: Long Term Autonomy Workshop, Hong Kong, China, 31 May–7 June 2014.
107. Taylor, Z.; Nieto, J. Motion-based calibration of multimodal sensor extrinsics and timing offset estimation. *IEEE Trans. Robot.* **2016**, *32*, 1215–1229. [CrossRef]
108. Liao, Q.; Liu, M. Extrinsic calibration of 3D range finder and camera without auxiliary object or human intervention. In Proceedings of the 2019 IEEE International Conference on Real-Time Computing and Robotics (RCAR), Irkutsk, Russia, 4–9 August 2019; IEEE: Piscataway, NJ, USA, 2019.
109. Ullman, S. The interpretation of structure from motion. *Proc. R. Soc. B* **1979**, *203*, 405–426. [CrossRef]
110. Iglhaut, J.; Cabo, C.; Puliti, S.; Piermattei, L.; O'Connor, J.; Rosette, J. Structure from motion photogrammetry in forestry: A review. *Curr. For. Rep.* **2019**, *5*, 155–168. [CrossRef]
111. Swart, A.; Broere, J.; Veltkamp, R.; Tan, R. Refined non-rigid registration of a panoramic image sequence to a LiDAR point cloud. In Proceedings of the Photogrammetric Image Analysis: ISPRS Conference, PIA 2011, Munich, Germany, 5–7 October 2011; Springer: Berlin/Heidelberg, Germany, 2011.
112. Moussa, W.; Abdel-Wahab, M.; Fritsch, D. Automatic fusion of digital images and laser scanner data for heritage preservation. In *Progress in Cultural Heritage Preservation: Proceedings of the 4th International Conference, EuroMed 2012, Limassol, Cyprus, 29 October–3 November 2012*; Proceedings 4; Springer: Berlin/Heidelberg, Germany, 2012.
113. Wang, L.; Xiao, Z.; Zhao, D.; Wu, T.; Dai, B. Automatic extrinsic calibration of monocular camera and LiDAR in natural scenes. In Proceedings of the 2018 IEEE International Conference on Information and Automation (ICIA), Wuyishan, China, 11–13 August 2018; IEEE: Piscataway, NJ, USA, 2018.
114. Li, J.; Yang, B.; Chen, C.; Huang, R.; Dong, Z.; Xiao, W. Automatic registration of panoramic image sequence and mobile laser scanning data using semantic features. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 41–57. [CrossRef]
115. Nagy, B.; Kovács, L.; Benedek, C. Online targetless end-to-end camera-LiDAR self-calibration. In Proceedings of the 2019 16th International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 27–31 May 2019; IEEE: Piscataway, NJ, USA, 2019.
116. Schneider, N.; Piewak, F.; Stiller, C.; Franke, U. RegNet: Multimodal sensor registration using deep neural networks. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; IEEE: Piscataway, NJ, USA, 2017.
117. Liu, H.; Liu, Y.; Gu, X.; Wu, Y.; Qu, F.; Huang, L. A deep-learning based multi-modality sensor calibration method for usv. In Proceedings of the 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), Xi'an, China, 13–16 September 2018; IEEE: Piscataway, NJ, USA, 2018.
118. Iyer, G.; Ram, R.K.; Murthy, J.K.; Krishna, K.M. CalibNet: Geometrically supervised extrinsic calibration using 3D spatial transformer networks. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; IEEE: Piscataway, NJ, USA, 2018.
119. Shi, J.; Zhu, Z.; Zhang, J.; Liu, R.; Wang, Z.; Chen, S.; Liu, H. Calibrcnn: Calibrating camera and lidar by recurrent convolutional neural network and geometric constraints. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; IEEE: Piscataway, NJ, USA, 2020.
120. Zhao, G.; Hu, J.; You, S.; Kuo, C.C.J. CalibDNN: Multimodal sensor calibration for perception using deep neural networks. In Proceedings of the Signal Processing, Sensor/Information Fusion, and Target Recognition XXX, Online, 12 April 2021; Volume 11756.
121. Lv, X.; Wang, B.; Dou, Z.; Ye, D.; Wang, S. LCCNet: LiDAR and camera self-calibration using cost volume network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021.
122. Sun, Y.; Li, J.; Wang, Y.; Xu, X.; Yang, X.; Sun, Z. ATOP: An attention-to-optimization approach for automatic LiDAR-camera calibration via cross-modal object matching. *IEEE Trans. Intell. Veh.* **2022**, *8*, 696–708. [CrossRef]

123. Jing, X.; Ding, X.; Xiong, R.; Deng, H.; Wang, Y. DXQ-Net: Differentiable lidar-camera extrinsic calibration using quality-aware flow. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; IEEE: Piscataway, NJ, USA, 2022.
124. Gisder, T.; Meinecke, M.M.; Biebl, E. Synthetic aperture radar towards automotive applications. In Proceedings of the 2019 20th International Radar Symposium (IRS), Ulm, Germany, 26–28 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–10.
125. Yao, S.; Guan, R.; Peng, Z.; Xu, C.; Shi, Y.; Yue, Y.; Gee Lim, E.; Seo, H.; Man, K.L.; Zhu, X.; et al. Radar perception in autonomous driving: Exploring different data representations. *arXiv* **2023**, arXiv:2312.04861.
126. Shi, X.; Zhou, F.; Yang, S.; Zhang, Z.; Su, T. Automatic target recognition for synthetic aperture radar images based on super-resolution generative adversarial network and deep convolutional neural network. *Remote Sens.* **2019**, *11*, 135. [CrossRef]
127. Paoletti, M.E.; Haut, J.M.; Ghamisi, P.; Yokoya, N.; Plaza, J.; Plaza, A. U-IMG2DSM: Unpaired simulation of digital surface models with generative adversarial networks. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1288–1292. [CrossRef]
128. Hyun, E.; Jin, Y. Doppler-spectrum feature-based human–vehicle classification scheme using machine learning for an fmcw radar sensor. *Sensors* **2020**, *20*, 2001. [CrossRef]
129. Shahian Jahromi, B.; Tulabandhula, T.; Cetin, S. Real-time hybrid multi-sensor fusion framework for perception in autonomous vehicles. *Sensors* **2019**, *19*, 4357. [CrossRef] [PubMed]
130. Ren, M.; He, P.; Zhou, J. Decision fusion of two sensors object classification based on the evidential reasoning rule. *Expert Syst. Appl.* **2022**, *210*, 118620. [CrossRef]
131. Xu, C.; Zhao, H.; Xie, H.; Gao, B. Multi-sensor Decision-level Fusion Network Based on Attention Mechanism for Object Detection. *IEEE Sens. J.* **2024**, *24*, 31466–31480. [CrossRef]
132. Petković, D. Adaptive neuro-fuzzy fusion of sensor data. *Infrared Phys. Technol.* **2014**, *67*, 222–228. [CrossRef]
133. Khodarahmi, M.; Maihami, V. A review on Kalman filter models. *Arch. Comput. Methods Eng.* **2023**, *30*, 727–747. [CrossRef]
134. Lu, P.; Dai, F. An overview of multi-sensor information fusion. In Proceedings of the 2021 6th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Oita, Japan, 25–27 November 2021; IEEE: Piscataway, NJ, USA, 2021; Volume 6.
135. Liu, Y.; Zhou, Y.; Hu, C.; Wu, Q. A review of multisensor information fusion technology. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; IEEE: Piscataway, NJ, USA, 2018.
136. Foley, B.G. A Dempster-Shafer Method for Multi-Sensor Fusion. Master's Thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, USA, 2012.
137. Wu, Y.-C.; Feng, J.-W. Development and application of artificial neural network. *Wirel. Pers. Commun.* **2018**, *102*, 1645–1656. [CrossRef]
138. Fayyad, J.; Jaradat, M.A.; Gruyer, D.; Najjaran, H. Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors* **2020**, *20*, 4220. [CrossRef] [PubMed]
139. Velasco-Hernandez, G.; Barry, J.; Walsh, J. Autonomous driving architectures, perception and data fusion: A review. In Proceedings of the 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 3–5 September 2020; IEEE: Piscataway, NJ, USA, 2020.
140. Zhang, R.; Cao, S. Extending reliability of mmWave radar tracking and detection via fusion with camera. *IEEE Access* **2019**, *7*, 137065–137079. [CrossRef]
141. Yang, C.; Huan, S.; Wu, L.; Weng, Q.; Xiong, W. Fusion of Millimeter-Wave Radar and Camera Vision for Pedestrian Tracking. In Proceedings of the 2023 5th International Conference on Communications, Information System and Computer Engineering (CISCE), Guangzhou, China, 14–16 April 2023; IEEE: Piscataway, NJ, USA, 2023.
142. Chang, S.; Zhang, Y.; Zhang, F.; Zhao, X.; Huang, S.; Feng, Z.; Wei, Z. Spatial attention fusion for obstacle detection using mmwave radar and vision sensor. *Sensors* **2020**, *20*, 956. [CrossRef]
143. Deng, J.; Zhu, B.; Chu, X.; Wang, L.; Lu, Z.; Hu, Z. Robust target detection, position deducing and tracking based on radar camera fusion in transportation scenarios. In Proceedings of the 2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring), Helsinki, Finland, 19–22 June 2022; IEEE: Piscataway, NJ, USA, 2022.
144. Cheng, L.; Sengupta, A.; Cao, S. Deep Learning-Based Robust Multi-Object Tracking via Fusion of mmWave Radar and Camera Sensors. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 17218–17233. [CrossRef]
145. Bijelic, M.; Gruber, T.; Mannan, F.; Kraus, F.; Ritter, W.; Dietmayer, K.; Heide, F. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020. [CrossRef]
146. Ravindran, R.; Santora, M.J.; Jamali, M.M. Multi-object detection and tracking, based on DNN, for autonomous vehicles: A review. *IEEE Sens. J.* **2021**, *21*, 5668–5677. [CrossRef]
147. Li, Y.; Yu, A.W.; Meng, T.; Caine, B.; Ngiam, J.; Peng, D.; Shen, J.; Lu, Y.; Zhou, D.; Le, Q.V.; et al. Deepfusion: Lidar-camera deep fusion for multi-modal 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.

148. Wang, L.; Zhang, X.; Qin, W.; Li, X.; Gao, J.; Yang, L.; Li, Z.; Li, J.; Zhu, L.; Wang, H.; et al. Camo-mot: Combined appearance-motion optimization for 3D multi-object tracking with camera-lidar fusion. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 11981–11996. [CrossRef]
149. Huang, K.; Hao, Q. Joint multi-object detection and tracking with camera-LiDAR fusion for autonomous driving. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; IEEE: Piscataway, NJ, USA, 2021.
150. Zhang, C.; Zhang, C.; Guo, Y.; Chen, L.; Happold, M. Motiontrack: End-to-end transformer-based multi-object tracking with lidar-camera fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023.
151. Feng, C.; Han, P.; Zhang, X.; Yang, B.; Liu, Y.; Guo, L. Computation offloading in mobile edge computing networks: A survey. *J. Netw. Comput. Appl.* **2022**, *202*, 103366. [CrossRef]
152. Kang, D.; Kum, D. Camera and Radar Sensor Fusion for Robust Vehicle Localization via Vehicle Part Localization. *IEEE Access* **2020**, *8*, 75223–75236. [CrossRef]
153. Alaba, S.Y. GPS-IMU Sensor Fusion for Reliable Autonomous Vehicle Position Estimation. *arXiv* **2024**, arXiv:2405.08119.
154. Yin, Y.; Zhang, J.; Guo, M.; Ning, X.; Wang, Y.; Lu, J. Sensor fusion of GNSS and IMU data for robust localization via smoothed error state kalman filter. *Sensors* **2023**, *23*, 3676. [CrossRef]
155. Liu, Y.; Luo, Q.; Zhou, Y. Deep learning-enabled fusion to bridge GPS outages for INS/GPS integrated navigation. *IEEE Sens. J.* **2022**, *22*, 8974–8985. [CrossRef]
156. Aslinezhad, M.; Malekijavan, A.; Abbasi, P. ANN-assisted robust GPS/INS information fusion to bridge GPS outage. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 129. [CrossRef]
157. Zhu, J.; Zhou, H.; Wang, Z.; Yang, S. Improved Multi-sensor Fusion Positioning System Based on GNSS/LiDAR/Vision/IMU with Semi-tightly Coupling and Graph Optimization in GNSS Challenging Environments. *IEEE Access* **2023**, *11*, 95711–95723. [CrossRef]
158. Dai, K.; Sun, B.; Wu, G.; Zhao, S.; Ma, F.; Zhang, Y.; Wu, J. Lidar-based sensor fusion slam and localization for autonomous driving vehicles in complex scenarios. *J. Imaging* **2023**, *9*, 52. [CrossRef]
159. Chen, W.; Zhou, C.; Shang, G.; Wang, X.; Li, Z.; Xu, C.; Hu, K. SLAM Overview: From Single Sensor to Heterogeneous Fusion. *Remote Sens.* **2022**, *14*, 6033. [CrossRef]
160. Liu, K. A robust and efficient lidar-inertial-visual fused simultaneous localization and mapping system with loop closure. In Proceedings of the 2022 12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Hangzhou, China, 27–31 July 2022; IEEE: Piscataway, NJ, USA, 2022.
161. Wang, Y.; Hongwei, M. mvil-fusion: Monocular visual-inertial-lidar simultaneous localization and mapping in chal-lenging environments. *IEEE Robot. Autom. Lett.* **2022**, *8*, 504–511. [CrossRef]
162. Cheng, J.; Zhang, L.; Chen, Q.; Fu, Z.; Du, L. High Precision and Robust Vehicle Localization Algorithm with Visual-LiDAR-IMU Fusion. *IEEE Trans. Veh. Technol.* **2024**, *73*, 11029–11043. [CrossRef]
163. Zhuang, Z.; Li, R.; Jia, K.; Wang, Q.; Li, Y.; Tan, M. Perception-aware multi-sensor fusion for 3D LiDAR semantic segmentation. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
164. Yin, R.; Cheng, Y.; Wu, H.; Song, Y.; Yu, B.; Niu, R. FusionLane: Multi-sensor fusion for lane marking semantic segmentation using deep neural networks. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 1543–1553. [CrossRef]
165. Zou, Z.; Zhang, X.; Liu, H.; Li, Z.; Hussain, A.; Li, J. A novel multimodal fusion network based on a joint-coding model for lane line segmentation. *Inf. Fusion* **2022**, *80*, 167–178. [CrossRef]
166. Zhang, X.; Yin, X.; Gao, X.; Qiu, T.; Wang, L.; Yu, G.; Wang, Y.; Zhang, G.; Li, J. Adaptive Entropy Multi-modal Fusion for Nighttime Lane Segmentation. *IEEE Trans. Intell. Veh.* **2024**, 1–13. [CrossRef]
167. Ye, H.; Mei, J.; Hu, Y. M2F2-Net: Multi-modal feature fusion for unstructured off-road freespace detection. In Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 4–7 June 2023; IEEE: Piscataway, NJ, USA, 2023.
168. Duraisamy, P.; Natarajan, S. Multi-Sensor Fusion Based Off-Road Drivable Region Detection and Its ROS Im-plementation. In Proceedings of the 2023 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 29–31 March 2023; IEEE: Piscataway, NJ, USA, 2023.
169. Feng, Y.; Li, X.; Ni, P.; Liu, X.; Jiang, T. Multi-sensor Fusion Network for Unstructured Scene Segmentation with Surface Normal Incorporated. *IEEE Sens. J.* **2024**, *24*, 13589–13603. [CrossRef]