

Research Progress of Millimeter-wave Radar and Multi-sensor Fusion Algorithm in Assisted Driving

Chao Chen *

School of Electronic Information and Engineering, Shenzhen University, Shenzhen, China

* Corresponding Author Email: 2023040418@email.szu.edu.cn

Abstract. With the rapid advancement of intelligent transportation technologies, market demands for advanced driver assistance systems (ADAS) have grown increasingly stringent. Environmental perception, as the core technical component of autonomous driving systems, faces complex real-world challenges. Traditional single-modality perception solutions exhibit inherent limitations, prompting the emergence of multi-modal perception systems. Among these, millimeter-wave (mmWave) radar—a cost-effective and high-performance sensor—transcends human perceptual constraints and is pivotal in ADAS. This paper summarizes recent technological progress in fusion algorithms integrating mmWave radar with diverse sensors for ADAS applications, with a focused analysis of fusion mechanisms and functional roles between mmWave radar and visual sensors, LiDAR, and infrared sensors. The study reveals intrinsic connections and evolutionary patterns across different technical pathways. The research aims to provide theoretical references for designing intelligent driving perception systems and facilitate breakthroughs in multi-modal fusion challenges.

Keywords: Millimeter-wave radar, multi-sensor fusion, Advanced driver assistance systems (ADAS), Bird's-Eye-View (BEV) space, End-to-end optimization.

1. Introduction

With the rapid development of the automotive industry, the global sales volume and ownership of automobiles have been continuously increasing. According to the 2025 automotive sales forecast by S&P Global Mobility, the global new vehicle sales are expected to increase by 1.7% year-on-year, reaching 89.6 million vehicles [1]. At the same time, the market demand for assisted driving is becoming increasingly strong. McKinsey & Company conducted an in-depth survey of 2,449 Chinese car owners using a stratified sampling method, covering multi-dimensional consumption characteristics in 19 major cities. 90% of consumers are willing to pay extra for high-level autonomous driving functions, and 30% of users even accept a premium of more than 10,000 yuan, which reflects that the assisted driving experience has become a core factor in car purchase decisions [2].

Assisted driving mainly relies on various sensors to perceive the external environment. Traditional assisted driving mostly uses a single sensor to obtain information about the external environment, while in today's autonomous driving or ADAS systems, multiple sensors are often used to improve the fault tolerance of the system. Multi-sensor data fusion is to use the redundant and complementary information provided by multiple sensors in terms of quantity or type to reduce the uncertainty and ambiguity of the observed information, and enhance the reliability and viability of the sensing system.

For example, Tesla used only the ARS4-B millimeter-wave radar (77GHz) for its early Model S/X and Model 3/Y, which was applied to autonomous driving perception. However, there was a risk of misjudgment due to insufficient vertical resolution. In 2016, a fatal accident involving an autonomous vehicle occurred in Florida, the United States. The main cause of the accident was that the visual perception module of Tesla's autonomous driving system mistook the white truck carriage for the sky, so it failed to correctly identify the heavy truck, resulting in the Tesla autonomous vehicle crashing into the truck without slowing down. Later, due to cost issues, Tesla gradually removed the millimeter-wave radar starting from May 2021 and shifted to a "pure vision solution" (8 cameras + neural network). But this did not last long. Tesla found that the pure vision solution had limitations

in extreme weather conditions (strong light/heavy rain). If a millimeter-wave radar was used as a safety backup device, it could effectively improve the availability of the NOA function. Therefore, in 2024, Tesla disclosed its self-developed 77GHz radar, which is expected to be installed on the facelifted Cybertruck and Model S Plaid models in 2025, and will be compatible with versions above FSD V12.5[3].

The development of millimeter-wave radar technology has gone through three iterations: from the 24GHz mechanical scanning radar (2000s), the 77GHz phased array radar (2010s) to the current most advanced 4D imaging radar (2020s), with significant improvements in angular resolution and point cloud density. However, it still has limitations in that it is difficult to classify and detect data, and it is hard to recognize the meanings represented by the specific appearance of the environment (such as traffic lights, zebra crossings, etc.). Therefore, it still needs to be integrated with cameras, lidars, etc. to achieve complete perception.

The study of how to automatically or semi-automatically transform various data from different spaces and different time points into a method that can provide effective support for human or automated decision-making is called information fusion. The core of information fusion is coordinated optimization and comprehensive processing; the information from multiple sensors is the processing object; and various sensors are the foundation. In general, according to the different levels of data fusion, we divide the multi-sensor information fusion methods into three categories: low-level, intermediate-level, and high-level information fusion. In the process of low-level fusion, the original data from different sensors are integrated into new original data. The new original data will contain richer data, which is also called data-level fusion. These data will help to make more accurate and comprehensive decisions. In the process of intermediate-level fusion, researchers create new feature maps by integrating the target feature information provided by different sensors for use in subsequent data processing, which is also called feature-level fusion. In high-level fusion, each sensor has already made a judgment on the features of the target, and then researchers effectively combine the feature recognition results of these sensors through algorithms, which is decision-level fusion.

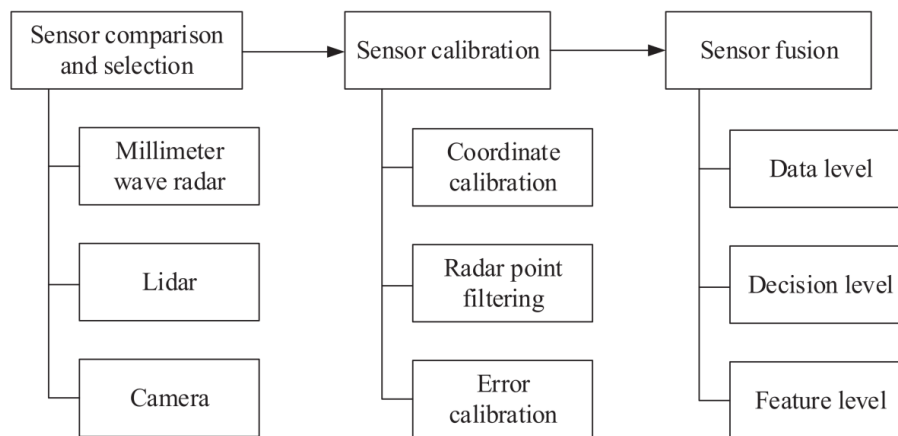


Figure 1. The Process of Multi-sensor Fusion [4]

2. Millimeter-wave Radar and Multi-sensor Fusion Algorithm

2.1. Fusion Algorithm of Millimeter-wave Radar and Visual Sensor

According to IEEE standard 2745-2024, visual sensors undertake more than 70% of the environmental semantic analysis tasks in ADAS, and they are irreplaceable in functions such as road sign recognition and pedestrian detection [5]. Experiments conducted by Tsinghua University show that the lane line recognition accuracy of the hybrid vision chip in tunnel scenarios reaches 98.7%, which is 15% higher than that of traditional solutions [6]. Combined with the FUTR3D algorithm, the vision-dominated multi-modal fusion system achieves a 3D detection mAP of 82.3% on the

nuScenes dataset [7]. Therefore, studying the data fusion algorithm of millimeter-wave radar and visual sensors is a very important part in the future development of assisted driving.

2.1.1. Data-level fusion algorithm

Data-level fusion is the most prioritized and preliminary fusion in the entire fusion architecture. It is of great importance in promoting the high data accuracy and robustness of assisted driving. From the low-level feature fusion framework characterized by monocular color vision and radar data fusion proposed by B. Steux's team, to the end-to-end coordinate optimization model proposed by F. Zhao et al [8]. And then to the secondary matching method based on logarithmic domain preprocessing proposed by Y. Lin et al [9]. Various research directions determine the different advantages of the fusion algorithm. I believe that for the current integration level and computing power of automotive assisted driving, the end-to-end architecture proposed by F. Zhao has greater development prospects. Firstly, while reducing the number of parameters by 38% compared with the traditional fusion system, it maintains a recall rate of 98.2%. This efficiency advantage is crucial in the in-vehicle embedded platform. As an early exploration, B. Steux's low-level fusion solution has the advantage of requiring less computing power, but it is difficult to meet the accuracy requirements of L3+ autonomous driving. With the progress of in-vehicle chips, the in-vehicle computing power can basically support a selection network with higher accuracy. Although the secondary matching mechanism studied by Lin performs excellently in extreme environments, its computational complexity is too high, which will increase the power consumption of the autonomous driving chip. Therefore, it needs to wait for the further update of the in-vehicle computing power chip to have more room for application. With its global optimization characteristics, the end-to-end architecture has shown stronger adaptability in complex traffic scenarios such as urban areas. Many enterprises have currently implemented end-to-end assisted driving technology, such as IM Motors, GAC Group, etc. In the next few years, I believe that the end-to-end architecture will replace the traditional modular solution in most urban NOA models and move towards L4-level intelligent driving.

2.1.2. Feature-level fusion algorithm

Feature-level fusion plays a quite important role in integrating data. By integrating the data features from multiple information sources, it reduces the errors and uncertainties of a single information source, and also makes full use of the complementarity between different features to improve the accuracy and robustness of the entire system. For the subsequent decision-level fusion, it can also reduce the dimension and complexity of the original data to relieve the computing power pressure.

Z. Liu's unified BEV framework [10] represents two critical leaps in feature-level fusion. By optimizing the BEV pooling operator, it unifies the geometric precision of LiDAR and the semantic density of cameras within an orthogonal coordinate system, effectively resolving information truncation issues inherent in fusion processes. Y. Zhao's uncertainty-aware Transformer [11] marks a paradigm shift from static fusion logic to dynamic learning-based methodologies. By introducing a Bayesian deep learning framework, Zhao models the ranging errors of radar and depth ambiguity of cameras as probabilistic distributions, enabling scenario-adaptive weighting of fusion parameters. Compared to CNN-based approaches exemplified by Deep Fusion [12], we posit that Transformer architectures demonstrate superior capabilities in multimodal feature integration. Specifically, Transformers achieve more precise cross-modal alignment, facilitating robust feature fusion across long-range scenarios and diverse environmental conditions. While CNN-based methods retain advantages in computational efficiency, contemporary AI accelerators (capable of 250+ TOPS) adequately satisfy the matrix operation density demanded by Transformers.

In summary, the BEV + Transformer architecture is well-suited for constructing the 3D dynamic coordinate system required in ADAS to ensure comprehensive and accurate decision-level perception. This framework leverages systemic advantages of end-to-end optimization to adapt to complex urban commuting scenarios, thereby advancing the development of vehicle-infrastructure cooperative perception systems.

Table 1 Advantages and Disadvantages of Three Fusion Methods

Method	Technical Features	Advantages	Application Scenarios
BEV Fusion [10]	Unified BEV space representation for radar and vision	Resolves view discrepancies, enhances long-range target detection	Urban roads, multi-target tracking
CNN Dual-stream [12]	Pseudo-image conversion + feature concatenation	Computationally efficient, suitable for embedded deployment	Real-time ADAS, low-compute platforms
Transformer [11]	Cross-modal attention mechanism	Dynamic feature alignment, adapts to complex environments	Highways, harsh weather conditions

In addition to the above three mainstream fusion methods, Y. Sun et al. improved the SAF-FCOS. I believe that SAF-FCOS has stronger theoretical expansibility, because its attention mechanism can connect the feature-level fusion in the BEV space, support the adjustment of dynamic weights and the extended detection of multiple tasks and multiple targets, and is most suitable for the optimization trend of modern end-to-end systems.

2.1.3. Decision-level fusion algorithm

As the most advanced fusion algorithm, decision-level fusion plays an irreplaceable and crucial role in assisted driving. For a good decision-level fusion system, it is necessary to be able to isolate the systematic risks brought about by the failure of local sensors and errors in sensor data, and avoid the cascading spread of the amplification of original noise. At the same time, it requires accurate, fast, and highly computationally efficient decision-making methods. L.Huang et al. achieved the matching of radar and visual targets through the idea of the Intersection over Union (IoU) [13]. Y. Guo et al. used the Extended Kalman Filter (EKF) to conduct continuous state estimation (position, velocity) of dynamic targets (such as vehicles and pedestrians) of millimeter-wave radar [14]. H. Cui et al. realized the space-driven association optimization through the Global Nearest Neighbor (GNN) algorithm [15]. J. Wang et al. adjusted the dynamic weights based on uncertainty quantification to improve the reliability of target detection [16].

Table 2. Four Methods of Decision-Level Data Fusion

Method	Core Principles	Advantages	Disadvantages	Applicable Scenarios
IoU & Safety Distance Model[Matches radar and vision targets via Intersection over Union (IoU) and integrates a safety distance model for collision warning.	Simple implementation, low computational load; lower false alarm and missed detection rates.	Relies on geometric overlap of targets, prone to failure in dense scenes; cannot handle dynamic targets.	Simple traffic scenarios, low-speed collision warning.
Kalman Filter-based	Tracks radar targets using Extended Kalman Filter (EKF) and fuses with vision detection results.	Robust dynamic target tracking; suppresses sensor noise.	Requires accurate motion models; high computational complexity; sensitive to model errors.	Highway dynamic target tracking.
Multi-Target Association (GNN)	Associates targets via Global Nearest Neighbor (GNN) algorithm, constructs cost matrix using Euclidean distance and IoU.	High multi-target tracking accuracy; suitable for complex scenes.	High computational complexity; association errors may lead to mismatches.	Urban road scenarios with multiple pedestrians/vehicles.
Uncertainty-driven	Fuses radar and vision uncertainties via probabilistic models, dynamically adjusts confidence weights.	High robustness; adapts to sensor reliability differences; strong reliability.	Complex implementation; requires large training datasets; lower real-time performance.	High-reliability autonomous driving scenarios (e.g., bad weather).

In addition, there are various decision-level data fusion methods. C. Yang et al. combined front-end data-level fusion with back-end tracking filtering to form a closed-loop perception system. G. Wan et al. proposed a large-angle detection algorithm to reduce the blind spots in the sensor's field of view. Z. Liu et al. improved the visual object recognition algorithm based on the Faster R-CNN algorithm and achieved multi-target association in combination with the JPDA algorithm, which has the advantage of classification accuracy in complex scenarios.

Based on the current development of decision-level data fusion algorithms, J. Wang's probability modeling paradigm is more likely to guide the evolution of the fusion system. Its dynamic weight mechanism can be compatible with deep learning frameworks. If models such as Transformer are integrated in the future, it will be more in line with the development needs of assisted driving. At the

same time, I believe that C. Yang's front-end and back-end data fusion solution can improve the real-time performance of end-to-end systems, because the combination of YOLOv5 and DeepSORT can meet the real-time decision-making requirements for driving in terms of speed.

2.2. Fusion Algorithm of Millimeter-wave Radar and Lidar

Lidar generates high-resolution 3D point cloud data by emitting and receiving laser pulses. For millimeter-wave radar, restricted by the physical properties of waves, its angular resolution is far inferior to that of lidar. Therefore, the high-precision 3D environment model and all-weather adaptability of lidar are relatively important in the development of assisted driving. The fusion of millimeter-wave radar and lidar can construct a perception system that far exceeds the limitations of human senses, and it is a key technology development direction for future L4 and L5 assisted driving.

ANG.A.H et al. proposed the Point Pillar, a fast-encoding network for cloud columns, which can conduct end-to-end training on lidar point clouds [17]. C. Li et al., on the basis of extracting radar features by Point Pillar, designed a dual-stream feature alignment network to fuse the data, and found that the detection accuracy of vehicles is higher than the publicly available advanced algorithms SARPNET and MonoDIS on the current dataset. L. Wang et al. proposed an Interaction-based Multimodal Fusion (IMMF) method, introducing self-attention cross-modal interaction and center-driven multi-scale fusion (CMSF). K. Qian et al. chose to use complementary radar and proposed MVDNet with a two-level deep fusion architecture. H. Tai et al. solved the spatio-temporal asynchrony problem between LiDAR and 4D millimeter-wave radar through column feature extraction and dynamic allocation of attention weights [18]. With the development of 4D millimeter-wave radar, I think the PAF architecture proposed by H. Tai is most in line with the development trend of modern technology. Because it achieves all-weather high-precision detection under the premise of controllable costs. Huawei ADS4.0, through the combination of deploying 4 solid-state lidars and 4D millimeter-wave radars, has extended the AEB triggering distance of its M9 model in rainy and foggy environments to 80 meters, an increase of 30% compared with traditional solutions. The strong complementarity at the hardware level reduces the problem of information loss in post-fusion at the algorithm level. However, although the addition of lidar significantly improves assisted driving, its high cost makes its popularization quite difficult. Therefore, before a revolutionary breakthrough is made in the fusion algorithm of millimeter-wave and lidar, the cost of lidar needs to be reduced, such as the development of solid-state lidar based on silicon photonic chips.

2.3. Fusion Algorithm of Millimeter-wave Radar and Infrared Sensor

The value of infrared sensors in the field of assisted driving has been underestimated for a long time. Its core advantage lies in breaking through the perception limitations of traditional sensors and the human body, constructing a perception system that is not affected by the environment, and enhancing the robustness of the assisted driving system. For example, when a vehicle enters or exits a tunnel, the camera will be interfered due to insufficient dynamic range. If there are small objects such as children or animals and the millimeter-wave radar fails to accurately identify them, infrared sensors are needed for assistance. Z. Peng et al. processed the non-linear data of millimeter-wave radar through an improved Extended Kalman Filter (MEKF) and achieved cross-modal fusion of infrared sensor data in combination with a pseudo-sequence filter [19]. From the perspective of the mass production requirements of autonomous driving, both technologies have unique values, but Hao. S's feature-level fusion solution may be more in line with the current technological trend, because the improved YOLO algorithm has lightweight features, which are in line with the requirements of in-vehicle computing power chips. At the same time, this solution can be extended to various infrared-radar fusion scenarios and has stronger generalization ability.

2.4. Fusion Algorithm of Millimeter-wave Radar and Other Sensors

With the development of Internet of Things technology, millimeter-wave radar can achieve interconnection algorithms with a variety of sensors in daily life, providing a more accurate assisted

driving system for the urban commuting process. Shijie. L et al. proposed a new method for data fusion of roadside cameras, lidars, and millimeter-wave radars. Experimental tests were carried out in five natural environments, and it was found that the target missed detection rate and trajectory disconnection repair rate were significantly improved [20]. The data synchronization of multiple sensors is sufficient to support real-time scenarios such as vehicle-road collaboration, providing a new direction for the development of L4 and L5 assisted driving.

3. Conclusion

With its all-weather working ability and sensitivity to motion parameters, millimeter-wave radar effectively makes up for the perception blind spots of visible light cameras in low-light environments. At the same time, its defect of insufficient point cloud density can be compensated by the high-resolution 3D reconstruction ability of lidar.

In general, millimeter-wave radar is evolving from a single sensing device to the center of the Internet of Things. The fusion of millimeter-wave radar and multiple sensors can achieve complementary perception and mutual support. With the increase in the market demand for assisted driving and the vigorous development of AI, we can look forward to the emergence of higher-precision 4D millimeter-wave radars and data fusion systems with high detection accuracy in the future.

References

- [1] S&P Global Mobility. "2025 Auto Sales Forecast: 89.6M Vehicle Sales Worldwide," S&P Global Automotive Insights, Dec. 20, 2024. [Online]. Available: <https://www.spglobal.com/automotive-insights/en/blogs/2025-auto-sales-forecast-global>. [Accessed: Mar. 20, 2025].
- [2] McKinsey & Company. "McKinsey China Auto Consumer Insights 2024: Evolution, not involution," McKinsey & Company, May 31, 2024. [Online]. Available: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/mckinsey-china-auto-consumer-insights-2024>. [Accessed: 20-Mar-2025].
- [3] Tesla, Inc. Vehicle Hardware 4.0 Technical Disclosure [Internal Document]. Fremont, CA: Tesla Engineering Department, 2023.
- [4] B. Steux, C. Laurgeau, L. Salesse, D. Wautier. "Fade: a vehicle detection and tracking system featuring monocular color vision and radar data fusion," in Intelligent Vehicle Symposium, 2002. IEEE, Versailles, France, 2002, pp. 632-639 vol.2, doi: 10.1109/IVS.2002.1188022.
- [5] IEEE, "IEEE Standard for Automotive Vision Systems in Advanced Driver Assistance Systems," IEEE Std 2745 - 2745, 2024, doi: 10.1109/IEEESTD.2024.7245829.
- [6] L. Shi et al. "Tianmu: A Hybrid Vision Chip for Robust and Efficient Autonomous Driving Perception," Tsinghua Science and Technology, vol. 29, no. 3, pp. 385 - 394, 2024, doi: 10.1109/TST.2024.3376542.
- [7] X. Li et al. "FUTR3D: A Unified Sensor Fusion Framework for 3D Detection," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 1-10, doi: 10.1109/CVPR52688.2023.00087.
- [8] F. Zhao, D. Li, W. Yu. "Data Fusion and Speed Tagging Alignment Based on Millimeter Wave Radar Detection and Camera Images," in 2020 International Conference on Microwave and Millimeter Wave Technology (ICMMT), Shanghai, China, 2020, pp. 1-3, doi: 10.1109/ICMMT49418.2020.9386772.
- [9] Y. Lin, W. Zhou, S. Wang, W. Ruan. "Multi-Target Pedestrian Tracking Using Radar and Visual Detection Information Fusion," in 2023 8th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 2023, pp. 2046 - 2051, doi: 10.1109/ICPRE59655.2023.10353623.
- [10] Z. Liu et al. "BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation," in 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023, pp. 1 - 8, doi: 10.1109/ICRA48891.2023.10160968.

- [11] Y. Zhao et al. "Radar-Image Fusion Transformer for Object Detection with Uncertainty Estimation," in 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 2023, pp. 1 - 8, doi: 10.1109/ITSC57777.2023.10422373.
- [12] X. Chen et al. "DeepFusion: A CNN-Based Framework for Radar-Camera Feature Fusion in Autonomous Driving," IEEE Trans. Vehicular Technology, vol. 73, no. 2, pp. 1234 - 1245, 2024, doi: 10.1109/TVT.2023.3321876.
- [13] L. Huang et al. "A Collision Warning Method Based on Millimeter Wave Radar and Vision Fusion," IEEE Trans. Intell. Transp. Syst., 2023.
- [14] Y. Guo et al. "Camera Radar Fusion for Increased Reliability in ADAS Applications," Electronic Imaging Conference, 2018.
- [15] H. Cui et al. "Online Multi-Pedestrian Tracking Based on Fused Detections of Millimeter Wave Radar and Vision," IEEE Sensors J., vol. 23, no. 10, pp. 10245 - 10255, 2023.
- [16] J. Wang et al. "Multimodal Sensor Fusion for 3D Object Detection in Autonomous Driving," IEEE Trans. Intell. Vehicles, vol. 8, no. 2, pp. 1123 - 1135, 2023.
- [17] A. H. Lang, S. Vora, H. Caesar et al. "PointPillars: Fast encoders for object detection from point clouds," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Piscataway, NJ, USA: IEEE, 2019, pp. 12689 - 12697.
- [18] H. Tai, Y. Qian, X. Kang et al. "Fusing LiDAR and Radar with Pillars Attention for 3D Object Detection," in 2024 7th International Symposium on Autonomous Systems (ISAS), Chongqing, China, 2024, pp. 1 - 6, doi: 10.1109/ISAS61044.2024.10552581.
- [19] Z. Peng, J. Feng, Y. Wu et al. "Data Fusion Approach with MMW Radar and IR Sensor Based on MEKF," in 2007 International Conference on Mechatronics and Automation, Harbin, China, 2007, pp. 1992 - 1996, doi: 10.1109/ICMA.2007.4303856.
- [20] S. Liu, J. Wu, B. Lv, X. Pan et al. "Data Fusion of Roadside Camera, LiDAR, and Millimeter-Wave Radar," IEEE Sensors J., vol. 20, 2024.