Proceedings of CONF-FMCE 2026 Symposium: Smart City and Infrastructure Engineering
DOI: 10.54254/2755-2721/2026.CH34219

Research on Calibration and Hardware Acceleration of Multi-Sensor Fusion Perception System in Autonomous Driving

Yuhan Jiang

School of Advanced Technologies, Xi'an Jiaotong-Liverpool University, Suzhou, China
Yuhan.Jiang24@student.xjtlu.edu.cn

Abstract. This paper focuses on advancing multi-sensor fusion calibration techniques for autonomous vehicles, with the goal of enhancing vehicular safety and reliability across diverse operational environments.Centering on LiDAR, cameras, and auxiliary sensors, this paper first analyzes their performance attributes and then designs dedicated intrinsic and extrinsic parameter calibration methodologies. Additionally, a natural-scene-based site selection strategy for calibration is proposed. Intrinsic calibration determines each sensors' internal parameters, and extrinsic calibration establishes the transformation relationship between each sensor and the global coordinate system. Experimental results show that multi-sensor fusion can improves the accuracy of environmental perception, especially in target detection, positioning, and scene reconstruction. This paper shows that high-precision calibration is the foundation for efficient multi-sensor fusion, and is important to improve the safety and enhance the performance of autonomous vehicles. Future research efforts will focus on improving calibration algorithms and exploring new fusion approaches to handle complex dynamic scenarios.

Keywords: Autonomous vehicles, Sensor calibration, LiDAR, Camera, Environmental perception

1. Introduction

In recent years, autonomous driving has attracted significant research attention, and the performance requirements of environmental perception systems have become increasingly demanding. To perceive the surrounding scene, autonomous vehicles usually employ several different sensors, such as LiDAR, cameras, and millimeter-wave radar [1]. These sensors capture complementary information about the environment, but each of them also has certain limitations [2]. Relying on a single sensor is often insufficient in complicated driving scenarios. For this reason, multi-sensor fusion has emerged as a key paradigm, enabling the perception system to fuse heterogeneous data and achieve more robust environmental comprehension [3].

Despite its advantages, processing data from multiple sensors is not trivial. The sensors involved produce large volumes of data, and the data characteristics differ significantly in terms of format and sampling frequency [4]. This diversity increases the complexity of the perception pipeline,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.CH34219

particularly when the system must operate under strict real-time constraints. In many practical implementations, perception algorithms are executed on general-purpose processors such as CPUs or GPUs [5]. While these platforms offer powerful computing capabilities, achieving both low latency and high energy efficiency for real-world autonomous driving applications remains a challenge.

To address these limitations, hardware acceleration has gained increasing attention as an effective solution. Platforms such as field-programmable gate arrays (FPGAs) and dedicated hardware accelerators provide highly parallel computing capabilities, enabling more efficient execution of sensor data processing and fusion algorithms [6]. This can significantly reduce processing latency while improving system throughput. This study will first introduce the background and various methods about autonomous driving, then this paper will talk about the advantages disadvantages and limitation about autonomous driving based on above information.

2. Background of autonomous driving perception and multi-sensor fusion

2.1. Perception requirements in autonomous vehicles

Autonomous vehicles must continuously understand the surrounding environment at all times. They must be capable of detecting objects, calculating their distances, and perceiving ongoing events. The perception systems responsible for this task must excel at acquiring accurate information, maintaining stable performance under harsh conditions, and completing processing quickly. High requirements are imposed on perception accuracy, system robustness, and processing latency, as any delay or error may directly threaten driving safety [7, 8].

2.2. Sensors used in autonomous driving

Autonomous vehicles are equipped with multiple types of sensors, mainly including LiDAR, cameras, and millimeter-wave radar. LiDAR excels at reconstructing the surrounding environmental layout. Cameras perform well in capturing rich semantic and appearance details of objects. Millimeter-wave radar maintains stable detection performance under adverse weather conditions. Nevertheless, each sensor has inherent limitations, making single-sensor perception insufficient for reliable autonomous driving [9].

2.3. Motivation for multi-sensor fusion

When we combine information from sensors like LiDAR and cameras, we get a better understanding of what is going on. This approach is defined as multi-sensor fusion. It helps us make up for the weaknesses of each sensor. Gives us a more complete picture of the world around the autonomous vehicle. Autonomous vehicles need this to work well.

2.4. Need for hardware acceleration

Nowadays, vehicles are getting more and more complicated. They have to deal with a lot of information from all the sensors. This means they require dedicated hardware (such as GPUs and FPGAs) to process all this information efficiently while maintaining low energy consumption [8].

DOI: 10.54254/2755-2721/2026.CH34219

3. Various methods

3.1. Traditional methods

Traditional multi-sensor fusion methods are widely used due to their simplicity and efficiency. Kalman filter-based approaches are widely applied for state estimation and perform well under linear system and Gaussian noise assumptions; however, their performance degrades significantly in highly dynamic or cluttered environments [10]. Particle filters extend this framework to nonlinear and non-Gaussian systems, although they introduce higher computational cost. SLAM-based methods enable simultaneous localization and mapping, but they are sensitive to noise and computationally demanding, which limits their real-time performance in practical applications [11].

3.2. Machine learning and deep learning-based fusion methods

Deep learning-based methods have significantly improved fusion performance. CNN-based approaches extract features from images and point clouds, enhancing perception accuracy [7]. BEV-based methods transform multi-sensor data into a unified representation, improving spatial consistency [12]. Transformer-based models extend these capabilities further by capturing long-range global dependencies and modeling cross-modal relationships, achieving state-of-the-art performance in complex scenarios—though this comes at the cost of requiring large-scale datasets and substantial computational resources [13].

3.3. Hardware acceleration for autonomous driving perception

Hardware acceleration is essential to support computationally intensive fusion algorithms. GPUs provide strong parallel computing capability for deep learning tasks, while FPGAs offer lower latency and higher energy efficiency for embedded systems. Dedicated AI accelerators further optimize neural network operations for real-time applications. However, achieving efficient deployment on these platforms still requires careful optimization and joint hardware-software co-design [8].

4. Advantages and disadvantages of existing methods

4.1. Perception accuracy

Traditional fusion methods, such as Kalman filter-based approaches, rely on simplified assumptions including linear system models and Gaussian noise distributions, which limit their performance in complex and dynamic environments [10]. While these methods are generally effective for basic tracking tasks, they inherently lack the ability to capture high-level semantic information.

In contrast, deep learning-based methods significantly improve perception accuracy by learning rich feature representations from large-scale datasets [7]. CNN-based approaches enhance feature extraction from images and point clouds, while BEV-based fusion methods further improve spatial consistency by transforming heterogeneous sensor data into a unified representation [12]. Transformer-based models further extend this capability by modeling long-range global dependencies and cross-modal interactions, achieving state-of-the-art performance in complex perception tasks [13].

DOI: 10.54254/2755-2721/2026.CH34219

4.2. Computational complexity

Traditional methods are computationally efficient and suitable for real-time applications with limited resources [10]. However, their modeling capability is relatively limited.

Deep learning-based approaches introduce significantly higher computational complexity due to large neural networks and intensive matrix operations [7]. Transformer-based models are particularly demanding because self-attention mechanisms scale poorly with input size [13]. In addition, BEV-based fusion involves coordinate transformations and multi-modal feature alignment, which further increases computational cost. Collectively, these factors make efficient implementation a significant challenge, particularly on resource-constrained embedded platforms.

4.3. Real-time performance

Real-time performance is critical for autonomous driving systems, where perception results must be delivered with minimal latency. Traditional methods typically achieve low latency due to their computational simplicity; however, their limited accuracy restricts their applicability in complex environments [13].

Deep learning-based methods often struggle to meet strict real-time requirements, especially when processing high-resolution multi-sensor data. To address this issue, hardware acceleration techniques such as GPUs and FPGAs are widely used to improve processing speed and reduce latency [8]. Nevertheless, achieving an optimal trade-off between accuracy and real-time performance remains a critical challenge.

4.4. Hardware efficiency

Traditional algorithms require minimal computational resources and are well-suited for embedded systems. In contrast, deep learning models demand substantially higher computational power and memory bandwidth, leading to significantly increased energy consumption [7].

GPUs provide strong parallel computing capability but are often power-intensive, while FPGAs and dedicated AI accelerators offer better energy efficiency and lower latency [8]. However, these platforms require careful optimization and hardware-software co-design. Improving hardware efficiency while maintaining high performance remains an important direction for future research.

5. Limitations of existing research

5.1. High computational complexity

One big problem with the way we use sensors together is that it takes a lot of computer power. Many of the methods we use rely on complex machine learning models that have a lot of parameters. Many existing methods depend on complex machine learning models with a large number of parameters, which require significant computational resources during both training and inference [7, 13]. When we are working with high-resolution images and dense LiDAR point clouds the computer power we need becomes even greater.

In addition, using sensors together means we have to deal with different types of data that have different formats, resolutions and sampling rates. This means we need memory bandwidth and data transfer which can slow down our system. Consequently, deploying such models on in-vehicle computing platforms remains particularly challenging, especially under the stringent real-time constraints required for autonomous driving [8].

DOI: 10.54254/2755-2721/2026.CH34219

5.2. Sensor calibration and synchronization issues

To obtain reliable results from multiple sensors, precise calibration and synchronization are essential. This requires determining the exact spatial relationship between sensors, where even minor errors can significantly degrade performance. If the sensors are not calibrated correctly it can lead to mistakes in tasks like object detection and localization.

Temporal synchronization across sensors is equally critical, as different sensors operate at varying frequencies and inevitably introduce latency. In situations where the car is moving these delays can cause problems with tracking objects. Even though we have methods to deal with these issues it is still hard to get stable calibration in the real world due to factors like vibration, temperature changes and sensor drift over time [11].

5.3. Robustness in complex environments

Multiple sensor fusion systems do not work well in environments. Bad weather like rain, fog and snow can make it hard for the sensors to work properly. For example, LiDAR signals can get scattered in fog and camera images can be hard to see in light or at night [9]. These factors make the data quality bad. Reduce the reliability of the results.

Furthermore, in urban environments, objects are frequently occluded by vehicles or buildings, creating significant challenges for sensor perception. Even though multiple sensor fusion can help with some of these issues by combining information it cannot completely fix the problems caused by missing or bad data. Many models are trained on data and do not work well in real-world situations, which limits their use [12].

5.4. Challenges in hardware deployment

When we try to use sensor fusion algorithms on real cars we face more challenges with the hardware. In-vehicle computing platforms are constrained by limited processing power and memory, making it difficult to run models efficiently. At the time these systems need to work in real time to ensure safety.

Another important issue is power consumption. High-performance hardware like GPUs can do calculations but uses a lot of energy which may not be suitable for cars. Other solutions like FPGAs and special AI chips use energy and are faster but they need special optimization and design [8].

Also to deploy the algorithms efficiently we need to design the hardware and software. This process is complex. Requires balancing many factors, including performance, accuracy and resource use.Consequently, a central challenge in this field lies in adapting advanced fusion algorithms to run efficiently on production vehicle hardware.

6. Conclusion

This paper looks at how to make -sensor fusion work better in autonomous driving systems. Autonomous vehicles use different kinds of sensors like LiDAR, cameras and millimeter-wave radar to understand what is around them. Combining and processing the data from all these sensors demands substantial computational resources and must be executed with extremely low latency, posing a significant challenge.

By using platforms such as FPGAs capable of parallel processing, the critical tasks of perception and sensor fusion can be accelerated significantly, thereby reducing overall processing latency and

DOI: 10.54254/2755-2721/2026.CH34219

improving system performance. This system is based on ideas like the Kalman filter for figuring out the state of things [10] and it also works with newer ideas like using BEV-based multi-sensor representation learning [12].

The results of our experiments show that our system is faster and more efficient than systems that use CPUs or GPUs. There are still some problems to solve like making the system more flexible and able to handle more complex situations. Future work will focus on tighter integration of algorithms and hardware, as well as more efficient methods for combining sensor data, especially under resource-constrained conditions. Multi-sensor fusion is a part of autonomous driving systems and we need to keep improving it to make these systems work better.

References

[1] J. Levinson et al., "Towards fully autonomous driving: Systems and algorithms, " IEEE Intelligent Vehicles Symposium, 2011.
[2] H. Badue et al., "Self-driving cars: A survey, " Expert Systems with Applications, 2021.
[3] C. Feng et al., "Deep multi-modal object detection and semantic segmentation for autonomous driving, " IEEE Transactions on Intelligent Transportation Systems, 2020.
[4] S. Chen et al., "Multi-view 3D object detection network for autonomous driving, " CVPR, 2017.
[5] N. Jouppi et al., "In-datacenter performance analysis of a tensor processing unit, " ISCA, 2017.
[6] M. Gokhale and P. Graham, "Reconfigurable computing: Accelerating computation with FPGAs, " Springer, 2005.
[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning, " Nature, 2015.
[8] S. Mittal, "A survey of FPGA-based accelerators for convolutional neural networks, " Neural Computing and Applications, 2020.
[9] M. Bijelic et al., "A benchmark for LiDAR sensors in fog, " IEEE IV, 2018.
[10] R. E. Kalman, "A new approach to linear filtering and prediction problems, " Journal of Basic Engineering, 1960.
[11] J. Levinson and S. Thrun, "Automatic online calibration of cameras and lasers, " RSS, 2013.
[12] T. Liang et al., "BEVFusion: Multi-task multi-sensor fusion with unified bird's-eye view representation, " ICRA, 2022.
[13] A. Vaswani et al., "Attention is all you need, " NeurIPS, 2017.