Recent Advances of Autonomous Vehicle Perception: A review

Ahmed ElKailany[1] and Soad Ali[2]
[1]Center of Research & Consultancy & Training, University of Tobruk
[2]Higher Institute of Science and Technology, Tobruk

Abstract

An autonomous vehicle (AV), also known as a self-driving car, is a vehicle that can sense its surrounding environment and navigate safely with little or no human input. The Society of Automotive Engineers (SAE) Level 5 AV can completely control the vehicle to improve road safety and reduce traffic congestion without operations from humans (SAE, 2014). For AVs to perform their tasks autonomously, they need to be able to perceive their position, and the environment and predict the possible movements/routes of environmental factors (Panda, S.S., 2021). In this paper, authors investigate novel approaches in the AV perception area, comparing results and techniques presented in State-of-the-Art literature. The collected research published between 2020 and 2024 introduces different approaches to enhance the accuracy performance of perception. This study helps to enhance the current understanding by highlighting major proposed research of the selected area. A comparison between sensor fusion algorithms, their advantages, disadvantages, applications, and fusion level was critiqued. A comparison of the algorithm's accuracy performance was discussed.

Keywords: Autonomous Vehicles, Sensor Fusion, Algorithms, keyword, keyword,keyword, keyword

Recent Advances of Autonomous Vehicle Perception: A Review

The field of study in autonomous vehicles is mainly focused on one of these areas: perception, localization, behavioral prediction, or planning of autonomous vehicles. (Kayin et.al, 2023) Although AVs are expected to reduce the density of transportation and eliminate human-caused accidents, conditions are challenging where visibility is limited, or road conditions and traffic congestion are sudden. AVs need sensing mechanisms for detecting their surroundings, i.e. objects, measure their distance, velocity, and other environmental characteristics. In the next section a brief review of the basic 5 sensors used in AVs

1. **LiDAR** is one of the core perception sensors in the autonomous driving field. The use of 3D-LiDAR on cars has not exceeded much more than a decade and has already demonstrated its indispensability in ADAS (advanced driver-assistance system) with high measurement accuracy and illumination-independent sensing capabilities (Thrun et al., 2006). This 3D laser scanning technology has some key attributes: measurement range, measurement accuracy, point density, scan speed and configuration ability, wavelength, robustness to environmental changes, form factor, and cost (Carballo et al., 2020).

2. **The automotive radar system** consists of a transmitter and a receiver. The transmitter sends out radio waves that hit an object (static or moving) and bounce back to the receiver, determining the object's distance, speed, and direction. Automotive radar typically operates at bands between 24 GHz and 77 GHz which are known as mm-wave frequencies, while some on-chip radar also operates at 122 GHz. Radar can be used in the detection of objects and obstacles like in the parking assistance system, and also in detecting positions, and speed relative to the leading vehicle as in the adaptive cruise control system (Patole et al., 2017). There is also an FMCW (Frequency Modulated Continuous Wave) form for radar

   where the frequency of the transmitted signal is continuously varied at a known rate which makes the difference between the transmitted and the reflected signal proportional to the time of flight. Besides the speed measurement advantage, FMCW radar shows superior range resolution and accuracy (Gao et al., 2021)
3. **Camera** is one of the widest-used sensors in perception tasks, while also one of the most vulnerable in adverse weather conditions. Adhered to the interior windshield, sometimes rear or other windows, dashcams (dashboard cameras) continuously record the surroundings of a vehicle with an angle as wide as 170◦ (Rexing, 2021). Numerous autonomous driving datasets started with dashcam recordings at an early stage while nowadays professional camera sets and fisheye lens cameras are being deployed for an even larger field of view (Yogamani et al., 2019).
4. **Ultrasonic sensors** are commonly installed on the bumpers and all over the car body serving as parking assisting sensors and blind spot monitors (Carullo and Parvis, 2001). The principle of ultrasonic sensors is pretty similar to radar, both measuring the distance by calculating the travel time of the emitted electromagnetic wave, only ultrasonic operates at ultrasound band, around 40 to 70 kHz. In consequence, the detecting range of ultrasonic sensors normally does not exceed 11 m (Frenzel, 2021), and that restricts the application of ultrasonic sensors to close-range purposes such as backup parking. Efforts have been done to extend the effective range of ultrasonic and make it fit for long-range detecting (Kamemura et al., 2008).
5. **The global navigation satellite system (GNSS)** is an international system ofmultiple constellations of satellites, including systems such as GPS (United States), GLONASS (Russia), BeiDou (China), Galileo (European Union), and

   other constellations and positioning systems. GNSS operates in the L-Band (1 to 2 GHz) which can pass through clouds and rain, with a minimum impact on the transmitted signal in terms of path attenuation. GNSS sensors include one or more antennas, reconfigurable GNSS receivers, processors, and memory. GNSS is often in combination with real-time kinematic positioning (RTK) systems using ground base stations to transmit correction data.

where the frequency of the transmitted signal is continuously varied at a known rate which makes the difference between the transmitted and the reflected signalproportional to the time of flight. Besides the speed measurement advantage, FMCW radar shows superior range resolution and accuracy (Gao et al., 2021)

Odometer and inertial navigation systems (INS) use dead reckoning to computeposition, velocity, and orientation without using external references. INS combines motion sensors (accelerometers), rotation sensors (gyroscopes), and alsomagnetic field sensors (magnetometers). For the advanced INS, fiber optic gyroscopes (FOG) are used: with no moving parts, and two laser beams propagating in opposite directions through very long fiber optic spools, the phase difference between the two beams is compared and it is proportional to the rate of rotation. The combination of the above, such as GNSS with INS (GNSS+INS) andother sensors, with an algorithm, is a common approach to improve positioning accuracy and reduce drift.

AVs perception main tasks are environmental perception and localization. Both sensorfusion and information fusion can be defined as the process of managing and handlingdata and information coming from several types of sources to improve some specific criteria and data aspects for decision tasks (Fayyad, et. al., 2021). Perception enhancements achieved by both sensor fusion solutions and algorithms, in the next section the state-of-the-art in sensor fusion and algorithms will be reviewed.

Literature Review

According to (Fayyad, et. al., 2021) AVs research applications are categorized as Pedestrian Detection, Road Detection, Vehicle Detection, Lane Detection, Visual Odometry,SLAM, Navigation, and Ego Position. The classical algorithms utilize data fusion for the development of applications that require to modelling and propagation of data imperfections(inaccuracy, uncertainty) are Statistical Methods, Probabilistic Methods, Knowledge-based Theory Methods, Evidence Reasoning Methods, and Interval Analysis Theory. Those classical algorithms are summarized in Table 1

Table 1 Classical approaches for sensor fusion algorithms

Probabilistic Methods	Statistical Methods	Knowledge-Based Methods	Evidence Reasoning Methods	Interval Analysis Methods
Bayesian Networks State-space Model KNN Least Square Estimation	Cross-Covariance Covariance Intersection	ANN Fuzzy Logic Genetic Algorithm Particle Swarm Ant Colony	Dempster-Shafer Recursive Operators Combination Rules	

Deep Learning algorithms sensor fusion algorithms discussion are mainly Based on CNN and RNN and the applications are SPP-Net, Spatial Pooling Network, You Only LookOnce (YOLO), Single-Shot Multibox Detector (SSD), Deconvolutional Single-Shot, Multibox Detector, Long-Short Term Memory (LSTM), and Gate Recurrent Unit (GRU).

In conclusion, algorithms empowered by deep learning algorithms can take advantageof data-driven knowledge discovery rather than physics-based models. enrich the field of autonomous vehicles. The recommendations focused on environmental perception, localization, and mapping, and how to further utilize deep learning algorithms to improve theperformance of sensor fusion networks.

In (Lai-Dang, et. al., 2023), the goal is to effectively combine data from cameras and LiDAR to capture both local and global contextual relationships. The proposed method outperforms previous approaches in driving and infraction scores on challenging benchmarks.

The paper categorizes existing fusion methods into detection-level fusion, point-level fusion,and proposal-level fusion. It introduces novel techniques to capture local and global relationships, addressing previous limitations. The methodology consists of spatial feature extraction, integration of encodings for interpretable features, and prediction of forward waypoints. The proposed method utilizes attention mechanisms in the Transformer architecture to aggregate sensor data features. Experimental results demonstrate the superiority of the proposed approach on CARLA benchmarks.

In (Jiang et. al., 2024) a novel approach called High-order Attention Mechanism Fusion Networks (HAMFNs) for accurate 3D object detection in autonomous driving was presented. The paper addresses the challenges in detecting 3D objects in autonomous driving scenarios and proposes a fusion network that combines LiDAR and camera data. The existingapproaches focus on basic architectural designs and fixed 3D bounding boxes, neglecting the exploration of feature interrelations and the varying scales of 3D objects.

HAMFNs leverage a high-order attention mechanism for image expression and multi-scale learning. The network incorporates high-order convolution layers for tensor filtering and discriminative representations of the holistic image. It also includes a multi-scale query module to capture the saliency properties of 3D objects.

The proposed method outperforms existing state-of-the-art methods in terms of mean Average Precision (mAP) on the nuScenes dataset, achieving a 0.7% increase. Additionally, HAMFNs are integrated into popular architectures like ResNet-50, ResNet-101, and ResNet- 152, resulting in improved performance with minimal parameter increase.

The paper highlights the limitations of traditional deep convolution networks in learning higher-order features and adapting to objects with large scale-wise variations. To address these issues, HAMFNs introduce a feature extraction network based on a high-order attention mechanism and an object detection approach based on multi-scale detection andscale linear regression.

The contributions of this work are summarized as follows:

1. Proposing a feature extraction network based on a high-order attention mechanismto capture higher-order features and utilize contextual information.

2. Introducing an object detection approach with multi-scale detection and scale linearregression to adapt to objects with varying scales.

3. Applying a high-order attention mechanism to 3D object detection in the fusion ofLiDAR and camera data.

4. Demonstrating superior performance compared to existing methods in 3D objectdetection on the nuScenes dataset.

Overall, the paper presents HAMFNs as an effective fusion network for 3D object detection in autonomous driving scenarios, addressing the limitations of existing approachesand achieving improved accuracy.

(Alaba et. al., 2024) provides a comprehensive review of the current state and advancements in 3D object detection for autonomous vehicles. The paper emphasizes the importance of accurate perception systems in autonomous driving to ensure safe and efficientdecision-making.

The paper highlights that while 2D object detection and classification have improvedwith the use of deep learning in computer vision, they lack depth information crucial for understanding driving environments. Therefore, 3D object detection becomes a cornerstone for autonomous driving and robotics, providing precise estimations of object locations and enhancing environmental comprehension.

To address the challenges in 3D object detection, researchers are exploring multimodal fusion techniques that combine information from multiple sensors such as cameras, radar, and LiDAR. The paper discusses the advantages and drawbacks of each sensor and emphasizes the necessity of equipping autonomous vehicles with diverse sensorsfor robust and reliable operation.

The paper categorizes multimodal fusion-based 3D object detection methods into datafusion, feature fusion, and decision fusion approaches. It also explores the advancements in deep learning models, including Convolutional Neural Networks (CNNs) and Transformer networks, for multimodal fusion. Furthermore, the paper discusses the challenges, open issues, and potential directions for future research in 3D object detection.

Overall, the paper provides a comprehensive overview of multimodal fusion-based 3Dobject detection methods, highlights the importance of diverse sensor integration, and identifies areas for further research in the field of autonomous vehicle perception.

In (Teng et. al., 2024), addresses the challenges of motion planning for autonomous mining trucks in open-pit mines. The authors propose a comprehensive paradigm for unmanned transportation in such environments. They introduce FusionPlanner, a multi-task motion planning algorithm that utilizes multi-sensor fusion to adapt both lateral and longitudinal control tasks for unmanned transportation. They also develop a benchmark called MiningNav to evaluate the trustworthiness and robustness of algorithms in open-pit mine transportation roads. Additionally, they present the Parallel Mining Simulator (PMS), ahigh-fidelity simulator designed specifically for open-pit mining scenarios.

The paper discusses the unique challenges of open-pit mining environments comparedto urban traffic scenarios, emphasizing the need for specialized solutions. Autonomous mining trucks offer improved safety and efficiency in transportation tasks, but deploying autonomous driving in open-pit mines faces obstacles due to complex operational conditions and adverse environmental factors.

To address these challenges, the authors propose FusionPlanner as an end-to-end motion planner designed specifically for mining trucks. The algorithm incorporates multi-sensor fusion to enhance control robustness and trustworthiness. They also introduce MiningNav, a novel benchmark tailored for evaluating motion planners in open-pit mines,and PMS, a simulation platform for data collection and algorithm validation.

The paper highlights the contributions of this research, including the proposed FusionPlanner algorithm, the MiningNav benchmark, and the PMS simulation platform. The empirical results demonstrate the effectiveness of FusionPlanner in reducing collisions and takeovers. The authors anticipate that their unmanned transportation paradigm will enhance the trustworthiness and robustness of mining trucks in continuous round-the-clock unmannedtransportation.

Overall, this paper presents a comprehensive approach to motion planning for autonomous mining trucks in open-pit mines, offering specialized algorithms, benchmarks,and simulation tools to address the unique challenges of this environment.

(Zhu et. al., 2024) proposes a testbed for collecting extensive data in the field of millimeterWave (mmWave) vehicular communication. The testbed aims to address the challenges faced by connected and automated vehicles (CAVs) in terms of high attenuationduring mmWave signal propagation and mobility management.

The authors highlight the potential of using multi-sensor fusion and artificial intelligence (AI) techniques to overcome the limitations of existing solutions that rely on pilot signals for channel estimation. By utilizing sensors such as LiDAR, cameras, and ultrasonic devices, the testbed can build a 3D map around the vehicle and estimate the signalpropagation path, eliminating the need for iterative pilot signal processes

The paper discusses the importance of mmWave communication in CAVs and its advantages over existing standards like DSRC. It also presents the challenges in applying mmWave technology to vehicular scenarios, particularly the need for directional transmissionand the time-consuming iterative process of obtaining precise propagation information.

To overcome these challenges, the authors propose a sensor fusion approach that leverages out-of-band information from LiDAR, cameras, and ultrasonic sensors to facilitatemmWave transmission. They demonstrate the effectiveness of this approach in improving performance, even in non-line-of-sight (NLoS) scenarios.

The paper introduces the HawkRover testbed, which utilizes commercially available equipment such as TP-Link Talon AD7200 routers and an RC car, making it a cost-effective solution for collecting mmWave wireless and multi-sensory data. The data collected from various sensors are time-synchronized, and a deep learning algorithm is applied to predict themmWave beamforming direction based on the sensor input.

The results of the study show promising performance in real-world data, indicating the potential of the constructed dataset and the deep learning algorithm for future mmWavevehicular networks.

Overall, the paper contributes to the field of vehicular communication by proposing alow-cost testbed that combines mmWave communication, sensor fusion, and deep learning techniques to enhance the performance of connected and automated vehicles.

(Abohassan et. al, 2024) explores the use of virtual simulations to analyze the interaction between autonomous vehicles (AVs) and their surrounding environment. The authors develop a framework to estimate the complexity of the environment by calculatingthe real-time data processing requirements for AVs to navigate effectively.

The study focuses on rural settings and examines the impact of various road featureson scene complexity and the occurrence of wildlife-vehicle collisions (WVCs). The VISTAsimulator is used to replicate the captured environment accurately. The authors dissect roadways into relevant road features (RRFs) and the full environment (FE) to study the influence of roadside features on scene complexity.

The results indicate that roadside features significantly increase environmental complexity by up to 400%. Increasing the width of a single lane on the road leads to a 12.3- 16.5% increase in processing requirements. Crest vertical curves cause a decrease in data rates due to occlusion challenges, with an average data loss of 4.2%. Sag curves, on the otherhand, increase complexity by 7%. Roadside occlusion in horizontal curves results in a severeloss of road information, leading to a decrease in data rate requirements by up to 19%. In terms of weather conditions, heavy rain increases the AV's processing demands by 240% compared to normal weather conditions.

The authors suggest that the findings of this study can be utilized by AV developersand government agencies to better design AVs and meet the necessary infrastructure requirements. The paper highlights the importance of virtual simulations and the use of LiDAR data to create digital twins of the physical environment for risk-free testing and validation of AVs.

Overall, this paper proposes a novel framework to quantify the complexity of the static environment for AVs in rural settings, providing valuable insights for the developmentand deployment of autonomous vehicles.

Discussion

What are the limitations of the previous fusion methods mentioned in the paper?

The paper mentions limitations of previous fusion methods in the context ofautonomous driving. Here are the limitations highlighted in the paper:

1. Detection-level fusion (late fusion): This approach combines sensor data from multiple sensors, but it does not fully utilize the unique characteristics provided by each sensor. It may not effectively leverage the strengths of individual modalities, such as camerasand LiDAR.

2. Point-level fusion (early fusion): This approach combines data from LiDAR point clouds with features extracted from camera images. However, camera-to-LiDAR projectionscan result in semantic loss due to sparsity, which limits the quality of fusion. This method may not capture the full semantic information from both modalities.

3. Proposal-level fusion: Notable works in this category propose initial boundingboxes using LiDAR features and refine them using camera features. While this approachimproves the fusion quality, it still has limitations in capturing comprehensive local and global relationships between sensors.

These limitations highlight the challenges faced by previous fusion methods in effectively utilizing the strengths of different sensors and capturing meaningful interactions and relationships between sensor data. The proposed method aims to address these limitationsand improve the fusion process in autonomous driving scenarios.

Are there any other fusion methods that have been proposed for autonomous drivingscenarios?

Yes, apart from the fusion methods mentioned in the paper, there have been other fusion methods proposed for autonomous driving scenarios. Here are a few additional fusionmethods:

1. Deep Sensor Fusion: This method combines data from multiple sensors, such as cameras, LiDAR, and radar, using deep learning techniques. It utilizes deep neural networks to extract features from each sensor modality and then fuses the features at different levels tomake decisions.

2. Kalman Filtering: Kalman filtering is a widely used fusion technique in autonomous driving. It combines measurements from multiple sensors, such as GPS, IMU, and odometry, using a recursive filter that estimates the state of the vehicle. Kalman filteringprovides an optimal estimate by minimizing the error covariance.

3. Bayesian Sensor Fusion: Bayesian methods, such as the Bayesian belief network orBayesian inference, are used for sensor fusion in autonomous driving. These methods model the uncertainties associated with sensor measurements and fuse the information using probabilistic reasoning.

4. Multi-Sensor Data Association: This method focuses on associating data from multiple sensors to create a coherent representation of the environment. It involves techniques such as data association algorithms, track management, and data fusion algorithmsto integrate information from different sensors.

5. Graph-based Fusion: Graph-based fusion methods model the sensor data as a graph and utilize graph-based algorithms to fuse the information. The graph representation captures the relationships and dependencies between sensor data, enabling effective fusion in complex scenarios.

These are just a few examples of fusion methods used in autonomous driving.

Researchers continue to explore and develop new fusion techniques to improve perceptionand decision-making in autonomous vehicles.

References

Abohassan, M.; El-Basyouny, K. (2024). *Leveraging LiDAR-Based Simulations to Quantifythe Complexity of the Static Environment for Autonomous Vehicles in Rural Settings*.Sensors 2024, 24, 452.

Alaba, S.Y.; Gurbuz, A.C.; Ball, J.E. (2024) *Emerging Trends in Autonomous Vehicle Perception: Multimodal Fusion for 3D Object Detection*. World Electr. Veh. J. 15, 20.

Carballo, A., Lambert, J., Monrroy, A., Wong, D., Narksri, P., Kitsukawa, Y., Takeuchi, E., Kato, S., Takeda, K., (2020). *LIBRE: The multiple 3D LiDAR dataset.* In: IntelligentVehicles Symposium. IV, IEEE, pp. 1094–1101.

Carullo, A., Parvis, M., (2001). *An ultrasonic sensor for distance measurement in automotiveapplications*. IEEE Sens. J. 1 (2), 143.

Fayyad, J., Jaradat, M. A., Gruyer, D., & Najjaran, H. (2020). *Deep learning sensor fusion for autonomous vehicle perception and localization: A review*. Sensors, 20(15), 4220.

Frenzel, L., (2021). *Ultrasonic sensors: A smart choice for shorter-range applications.*URL: https://www.electronicdesign.com/markets/automation/article/21806202/ultrasonic- sensors-a-smart-choice-for-shorter-range-applicationsultrasonic-sensors-a-smart- choice-for-shorter range-applications, [Last accessed 10 Feb 2024].

Gao, X., Roy, S., Xing, G., (2021). *MIMO-SAR: A hierarchical high-resolution imagingalgorithm for mmwave FMCW radar in autonomous driving*. IEEE Trans. Veh. Technol. 70 (8), 7322–7334.

Jiang, Haiyang & Lu, Yuanyao & Zhang, Duona & Shi, Yuntao & Wang, Jingxuan. (2024).*Deep learning-based fusion networks with high-order attention mechanism for 3D object detection in autonomous driving scenarios.* Applied Soft Computing. 152.

Kamemura, T., Takagi, H., Pal, C., Ohsumi, A., (2008). *Development of a long-range ultrasonic sensor for automotive application*. SAE Int. J. Passeng. Cars-Electron.Electr. Syst. 1 (2008-01-0910), 301–306.

Lai-Dang, Q. V., Lee, J., Park, B., & Har, D. (2023, October). *Sensor fusion by spatialencoding for autonomous driving.* In 2023 IEEE SENSORS (pp. 1-4). IEEE.

Panda, S.S.S., Panigrahi, T.; Parne, S.R.; Sabat, S.L.; Cenkeramaddi, L.R. (2021). *Recent advances and future directions of microwave photonic radars: A review*. IEEE Sens.J. 2021, 21, 21144–21158.

Patole, S.M., Torlak, M., Wang, D., Ali, M., (2017). *Automotive radars: A review of signalprocessing techniques.* IEEE Signal Process. Mag. 34 (2), 22–35.

Rexing, 2021. V1P-4K. URL https://www.rexingusa.com/products/rexing-v1p/, [Lastaccessed 05 Feb 2024].

Sae, S. (2014). International: on-road automated vehicle standards committee. *Taxonomy anddefinitions for terms related to on-road motor vehicle automated driving systems, information report*.

Teng, Siyu & Chen, Long & Li, Luxi & Li, Yuchen & Hu, Xuein & Li, Lingxi & Ai, Yunfeng. (2023). *Fusionplanner: A Multi-Task Motion Planner for Mining TrucksUsing Multisensor Fusion Method.* 10.2139/ssrn.4518295.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C., Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek, L.-E., Koelen, C., Markey, C., Rummel, C.,van Niekerk, J., Jensen, E., Alessandrini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A., Mahoney, P., (2006). *Stanley: The robot that won the DARPAGrand Challenge*. J. Field Robotics 23 (9), 661–692.

Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Varley, P., O'Dea, D., Uricár, M., Milz, S., Simon, M., Amende, K., et al., (2019). *Woodscape: A multi-task, multicamera fisheyedataset for autonomous driving*. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9308–9318.

Zhu, E., Sun, H., & Ji, M. (2024). *HawkRover: An Autonomous mmWave Vehicular Communication Testbed with Multi-sensor Fusion and Deep Learning.* arXiv preprintarXiv:2401.018