Received November 13, 2019, accepted December 23, 2019, date of publication December 26, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962554

Multi-Sensor Fusion in Automated Driving: A Survey

ZHANGJING WANG, YU WU, (Student Member, IEEE), AND QINGQING NIU
School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
Corresponding author: Yu Wu (woosual@163.com)

This work was supported in part by the National Science and Technology Project of China under Grant 61404130113.

ABSTRACT With the significant development of practicability in deep learning and the ultra-high-speed information transmission rate of 5G communication technology will overcome the barrier of data transmission on the Internet of Vehicles, automated driving is becoming a pivotal technology affecting the future industry. Sensors are the key to the perception of the outside world in the automated driving system and whose cooperation performance directly determines the safety of automated driving vehicles. In this survey, we mainly discuss the different strategies of multi-sensor fusion in automated driving in recent years. The performance of conventional sensors and the necessity of multi-sensor fusion are analyzed, including radar, LiDAR, camera, ultrasonic, GPS, IMU, and V2X. According to the differences in the latest studies, we divide the fusion strategies into four categories and point out some shortcomings. Sensor fusion is mainly applied for multi-target tracking and environment reconstruction. We discuss the method of establishing a motion model and data association in multi-target tracking. At the end of the paper, we analyzed the deficiencies in the current studies and put forward some suggestions for further improvement in the future. Through this investigation, we hope to analyze the current situation of multi-sensor fusion in the automated driving process and provide more efficient and reliable fusion strategies.

INDEX TERMS Automated driving, multi-sensor fusion strategy, multi-target tracking, environmental reconstruction, data association, intent analysis, deep learning.

I. INTRODUCTION

Multi-source and heterogeneous information fusion (MSHIF) makes integrated utilization of the information obtained by different sensors, which avoids the perceptual limitations and uncertainties of a single sensor, forms a more comprehensive perception and recognition of the environment or target, and improves the external perception ability of the system [1]. At present, MSHIF technology has been comprehensively applied in such fields as fault detection [2], [3], remote sensing [4], human health monitoring [5], [6], robot system [7], human-machine interaction [8], target recognition and tracking [9], [10], simultaneous localization and mapping (SLAM) [11] and advanced driver assistance system (ADAS) [12]. Automated driving (AD) is not a single technology, but a highly complicated system composed of many subsystems, which contains three parts:

1) perception module: It includes sensors, recognition, and decision-making;

2) client system: It includes the operating system and hardware platform;

3) cloud platform: It includes a high-precision map, model training, and stored data.

The AD system architecture proposed in [15] is shown in Fig.1, but the details are further explained along the way in this paper.

Sensors are the key to the perception of the outside world in the AD system and whose cooperation performance directly determines the safety of AD vehicles [13]. The AD vehicles primarily use seven kinds of sensors, including cameras, millimeter-wave radar (MMW-Radar), the global positioning system (GPS), inertial measurement unit (IMU), LiDAR, ultrasonic, and communication module. Various sensors have their advantages and disadvantages, so there are often different task divisions in the AD system. This paper focuses on how to realize the fusion perception by fusing multi-sensor data. The camera can acquire the optical image and accurately record the contour, texture, color distribution, and other information of the object from a certain angle. Therefore, some studies use cameras to complete target recognition and

The associate editor coordinating the review of this manuscript and approving it for publication was Datong Liu.

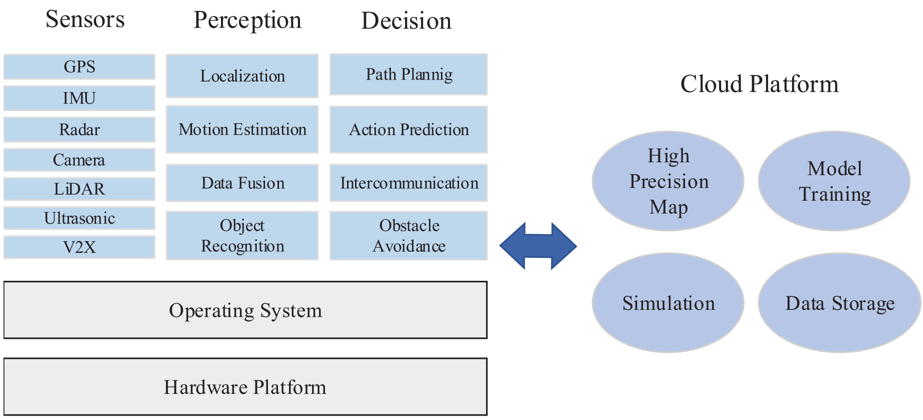


FIGURE 1. Automated driving system architecture, cited from [15].

target tracking tasks, including lane detection, pedestrian and vehicle identification, and local path planning [14]. To surmount the narrow measurable angle range of cameras, generally, AD adopts multiple cameras to form omnidirectional monitoring on the surrounding environment in the practical application [16], [17]. MMW-Radar can measure the acquisition of objects' distance through pulse compression and speed through the Doppler shift, which has extensive application for obstacle detection [18], pedestrian recognition, and vehicle recognition [19], [20]. GPS and IMU provide inertial information and global position information for AD vehicles to confirm their positions, thus enabling vehicles to update their positions in real-time in high-precision maps(HPM) [21]. The primary applications of LiDAR include positioning, obstacle detection, and environmental reconstruction. Because three-dimensional(3D) data has certain information representation advantages over two-dimensional data [22]–[24], 3D LiDAR sensors are playing an increasingly significant role in the AD system. It can maximize the restoration of traffic conditions in the authentic environment to combine the dynamic characteristics of the MMW-Radar targets, the ranging advantage of the LiDAR, and the details of the target in the optical image. Appropriate utilization of integrated information facilitates vehicles to perform diverse tasks such as intention analysis, motion planning, and autonomous driving.

However, sensor fusion alone cannot guarantee the safety of AD cars in a complex traffic environment. Society of Automotive Engineers (SAE) has divided the roles of a human driver and driving automation system by the level of driving automation in [25] including level 0 (no driving automation), level 1 (driver assistance), level 2(partial driving automation), level 3(conditional driving automation), level 4(high driving automation) and level 5(full driving automation). Reference [26] proposed an L4 framework, an enhanced Tesla Model S architecture, which can be applied for the simulation of autonomous vehicles driving on the highway. However, it remains to be seen whether there is any conflict between autonomous driving and manual driving in the future [27]. For level 4 and level 5, vehicle to everything (V2X) allows vehicles to connect to everything in the complex traffic environment and extend the scope of perception. Dedicated short-range communications (DSRC) has been testified to be competent in the vehicle-to-vehicle (V2V) applications with a tolerable latency at a data transfer rate of 27 Mbps in the scope of 1-kilometer [28]. Cellular V2X(C-V2X) bases on cellular mobile communication, including LTE-V2X and 5G-V2X, with a more extended communication range. 3G/UMTS and 4G/LTE perform better in range (up to 2km) but can not confront with time-critical scenarios [30]. 5G-V2X works in a higher frequency with ultra-high bandwidth and ultra-low time delay, it can collect and transmit more accurate environmental information in real-time, and use the cloud computing power to make decisions for the vehicle itself [31], [32]. Moreover, reference [33] discussed a flying and Ad-hoc network for communication between UAV and AD vehicles, which provides a new idea for vehicle networking.

Some studies utilized open-source data set [35], [36] or generated them from simulation software [37] to avoid the laborious collection of sensor data. The study of multi-sensor fusion requires a large amount of data, especially in the context of a large number of applications for deep learning. Therefore, the workload of data collection is enormous. In reference [38], the virtual test environment and open-source data set for AD have been analyzed and summarized in recent years. By selecting a data set that matches the research, it is possible to quickly obtain valid data during the multi-sensor fusion research process without consuming a large amount of resources and time to re-acquire. Furthermore, the targets corresponding to multiple sensors are in different coordinate systems, and the data rates of different sensors are diverse. It is necessary to map the simultaneous target information in heterogeneous information to a unified coordinate system, which the time-space alignment [39], [40]. Moreover, the presented forms of the object information have a difference with the sensors, and it is necessary to calibrate the locations of multi-sensor to acquire the final position. In the multi-sensor fusion part, the current researches have different methods, and the information fusion, fusion level, and fusion algorithm adopted by multiple sensors are different. From the fusion method, the combinations of sensors mainly include radar-camera (RC) [19], [41], camera-LiDAR (CL) [42] and radar-camera-LiDAR (RCL) [16]. Some studies integrated the vehicle location and map into the AD system, which makes the lane-level positioning possible [97]. Furthermore, the V2X sensors add nearby objects into the map in real-time, which allows AD vehicles to percept a lager scale of dynamic information [43]. Depending on the disparate forms of the fused information in MSHIF processing, the methodologies are divided into four types of information fusion, including fusion based on discernible units (FBDU) [44], [45], fusion based on complementary feature (FBCF) [46], [47], attribute-based fusion (ABF) [48], [49], and fusion based on multi-source decision making (FBMDM) [16], [50]. In general, different fusion strategies bases on different levels of abstraction of sensor data during data fusion. Before data fusion, FBDU has the lowest degree of abstraction. It usually integrates preprocessed data directly, and FBMDM has the highest and comprehensive judgment of final processing results of different sensors. Besides, in order to complete the vehicle's motion planning, it is necessary to detect obstacles and track the moving target. Due to the complexity of moving target motion, it is necessary to make the corresponding decision depending on the moving target tendency. However, a prerequisite for the realization of motion identification of moving targets is to track the target.

In this survey, we hope to summarize the specific strategies and integration goals of multi-sensor fusion in recent years. The paper first discusses the sensors and techniques used in the AD, and why and how they are used to complete AD tasks. Then, according to the specific fusion methods in different studies, the current deficiencies and areas that can be improved are analyzed and discussed. This article organized as follows: section II respective introduced the characteristics, advantages, and disadvantages of sensors. Section III summarized four fusion strategies and specific perceptual recognition methods and discussed the specific methodologies of multi-sensor fusion. Moreover, by comparing the performance of various sensors, the necessity of multi-sensor fusion is illustrated. It is necessary to establish the motion model to identify the motion intention of the target. In IV, we described the necessity to associate the data of multiple sensors to the target tracking. Section V analyzed the problems existing in the current fusion strategy and gave specific suggestions for further improvement.

II. SENSORS IN FUSION PERCEPTION SYSTEM

The type and performance of the sensors directly determine the quantity and quality of information acquired by the AD system. In addition to vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and other vehicle communications, the AD vehicle perceives the external environment by analyzing and synthesizing various sensor data. In general, the sensors adopted in the studies include radar/ultrasonic, and camera(including RGB-D, infrared camera), LiDAR, and GPS/IMU. The detection capability and reliability of different sensors are limited in different environments, and multi-sensor fusion can improve the accuracy of target detection and recognition. Table 1 summarized the advantages and disadvantages of the above sensors and the detection range, which shows that different sensors have apparent differences in operating characteristics. Meanwhile, it will improve the perception ability of the AD vehicle from all aspects to effectively guarantee the safety of the driver by fusing multi-sensor data.

In the fusion sensing process, the dominating adopted sensors are MMW-Radar, LiDAR, camera, ultrasonic, GPS/IMU, and V2X sensors. Consequently, the rest of this section will talk about the characteristics, advantages, and disadvantages of these sensors.

A. MILLIMETER WAVE RADAR

After radiating electromagnetic waves, the radar gathers the scattered wave of targets by the receiving antenna, then a series of signal processing will be performed to acquire the information of targets. At present, the mainstream frequency bands of MMW-Radars include 24GHZ, 60GHZ, and 77GHZ, and the most prevailing one is 77 GHz, while 60 GHz is a frequency band only adopted in Japan and the 24 GHz band will gradually be abolished in the future. The 79GHZ band radar has a higher resolution of range, speed, and angle, which are extensively approved and will become the mainstream frequency band of vehicle radar in the future. Compared to cameras and LiDAR, MMW-Radar has a longer wavelength, certain anti-blocking, and anti-pollution ability, which can cope with rain, snow, fog, and dark environment. The radar can not only obtain the exact distance of multiple targets, but also measure the relative velocity by the Doppler shift effect. Different types of vehicle-mounted radars waveforms are generally classified into frequency-modulated continuous-wave (FMCW) radars and pulse radars. Since the pulse radar requires to strictly isolate the transmitted signal when receiving the echo signal, while a high-power signal will be transmitted in a transitory continuous cycle, which leads high requirement of hardware and complicated structure. Therefore, most vehicle-mounted MMW-Radars adopt FMCW as the transmit waveform. FMCW radar ensures that the distance and relative speed of targets are simultaneously available, and the speed resolution and distance resolution are controllable. The distance resolution R_{res} and speed resolution V_{res} are:

$$R_{res} = \frac{c}{2B} \tag{1}$$

$$\theta_{res} = \frac{\lambda}{Nd\cos(\theta)} \tag{2}$$

where C is the speed of light, B is the bandwidth of the chirp,λ is the wavelength, and T_f is the pulse duration. Multiple receiving antennas can detect the target angle, and angle resolution is in connection with the actual angle between the target and the radiation direction. The angular resolution θ_{res}

TABLE 1. Comparison of different sensors and technologies.

Type	Advantages	Disadvantages	Max working distance
MMW-Radar	1) Long working distance 2) Available for radial velocity 3) Applicable for all-weather	1) Unapplicable for static objects 2) Generating false alarm easily	5m-200m
Camera	1) Excellent discernibility 2) Available lateral velocity 3) Available for color distribution	1) Heavy calculation burden 2) Light interference 3) Weather susceptible 4) Unavailable for radial velocity	250m (depending on the lens)
LiDAR	1) Wide field of view (FOV) 2) High range resolution 3) High angle resolution	1) Insufferable for bad weather 2) High price	200m
Ultrasonic	1) Inexpensive	1) Low resolution 2) Inapplicable for high speed	2m
DSRC	1) Applicable for high speed(up to 150 km/h) 2) Relatively mature technology 3) Low latency (0.2ms)	1) Low data rate 2) Small coverage	300-1000m
LTE-V2X	1) Long working distance 2) Relatively high data transmission rate(Up to 300Mbps)	1) High latency in long distance(> 1s) 2) Inapplicable for time-critical events	Up to 2km
5G-V2X	1) Ultra-high data transmission rate 2) Low latency(< 80ms) 3) High bandwidth 4) Applicable for high speed (up to 500km/h)	1) Immature application	100m - 300m

is:

$$V_{res} = \frac{\lambda}{2T_f} \tag{3}$$

where N is the number of antennas, d is the interval between two adjacent antennas, and θ is the real target azimuth. Nevertheless, due to radar hardware condition constraints (such as transmit power and sampling rate), the longer maximum effective working distance results in lower resolution of the parameters. In the actual application process, the AD vehicle is equipped with near-range, medium-range, and long-range radars to monitor corresponding ranges.

Most studies extract the range, angle, and speed information of pedestrians or vehicles from radar data. However, the micro-Doppler effect provides another methodology for recognizing rigid and non-rigid targets. In reference [20], the FMCW radar system was designed to provide desired recognition for both adults and children within 100 to 150 meters. In reference [21], representative target features are extracted from the received radar signal as the classification criterion of the SVM, which is applied to the classification of pedestrians, vehicles, and bicycles with an accuracy rate of up to 90%. Meanwhile, the latest IWR6843ODS radar module of TI can realize real-time target point cloud mapping, which further enriches the information content of the acquired data.

A drawback of MMW-Radar is indistinguishable for relatively static stationary targets. In addition to being disturbed by noise, AD vehicles are often suffering from false alarms produced by metal objects such as road signs or guardrails. The general processing method is to adopt the Constant False Alarm Ratio (CFAR) Detection to continuously update the decision threshold with the variance of the noise, thus obtaining a constant false alarm probability [51]. Besides, the preprocessed radar data can be generated to images by applying the generative adversarial network (GAN) [52]–[55], but the images still confronted with a problem of insufficient resolution. Moreover, with the increasing number of vehicles equipped FMCW radar, the shared frequency interference phenomenon will become a problem, and the reference [56] proposed a new radar ranging system with proper range resolution without bandwidth limitation. Compared with the same type of radar, its resolution is improved by more than an order of magnitude, which will be beneficial to the construction of high-resolution maps by radar.

B. CAMERA

The camera is one of the earliest sensors for the AD system, and which is also the primary choice for manufacturers and researchers at present. The camera is principally applied to accomplish tasks such as target recognition, environment map building, lane detection, and target tracking. In recent years, deep learning (DL) has achieved an excellent performance in

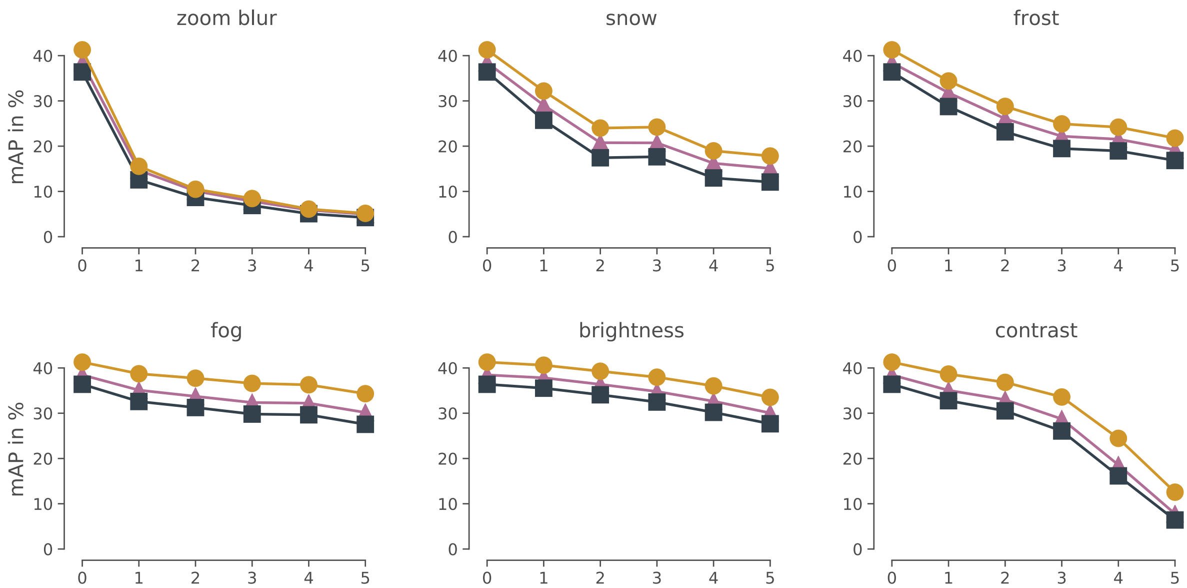


FIGURE 2. **The performance of the camera influenced by different corruptions. Source: Figure 9 in [66].**

the target recognition and tracking tasks, which can obtain powerful expression ability from massive data and replace the traditional manual features design by machine learning methods. After the AD system accurately completes the target recognition and target tracking, further decision tasks will implement further.

At present, there are two types of cameras, which is a charge-coupled device (CCD) and a complementary metal-oxide-semiconductor (CMOS). CCD has a complicated manufacturing process, higher quantization efficiency, lower noise, a high dynamic range, and high image quality in low light conditions. Compared to CCD sensors, CMOS sacrifices some performance to reduce cost. The difference between them will be wearier, and CMOS is expected to replace CCD [57].

Images captured by the camera transform the 3-D information into the 2-D one, so to obtain the position of the targets from the image, it is necessary to establish the relationship between the pixel and the physical world, which is called the camera calibration. Reference [58] reviewed the methodologies of camera calibration and divided them into optimization methods, transformation matrix method, distribution calibration method, Zhang Zhengyou calibration method, and traditional calibration method. In the actual calibration process, reference [59] proposed a flexible camera calibration method, which only requires to take pictures of a chessboard from different angles, and then establish a radial lens distortion model. The method consists of a closed-form solution and then nonlinearly solved according to the maximum likelihood criterion. Contrapose multi-sensor fusion, some camera calibration methods based on depth information are proposed in reference [60]–[64]. With computer vision applications continue to expand, it is necessary to propose novel-innovative algorithms with lower complexity and more flexible. Some studies use binocular cameras or depth cameras to obtain image data with depth information. However, in terms of the range resolution, there is still a big gap with radar or LiDAR [16], [65].

In reference [66], defaced images produced based on the existing datasets of Pascal, Coco, and Cityscapes were used to evaluate the most advanced target detection algorithms. As is shown in Fig. 2 below, the detection accuracy decreased by at least 31.1%, and the maximum decreased by 60.4% in some situations. Therefore, it is reasonable to conclude that a single camera sensor will be extremely unreliable in severe weather conditions.

The camera of AD vehicles also have poor reliability in the situation of sudden changes in light, such as exiting a tunnel. By combining the camera with GPS, HPM, and even V2X, some prior information can be introduced to adjust the camera exposure dynamically. Compared with radar, the superiority of the camera is that it can accurately capture the contour, texture, and color distribution information, which facilitates classification to the recognition of different targets under non-extreme environment conditions. However, AD vehicles have a requirement of competence to cope with all-weather environments and extreme situations. Only systems that apply MSHIF can perform 100% recognition accuracy, thus achieving the best European NCAP security level [48], [67].

The data fusion of multiple sensors is to avoid false detection of some sensors and thus generating wrong motion planning. At present, the most abundant information can obtain through the camera perception of the environment. However, it must consider that the image can become unreliable in some cases, such as the sudden change of light in entering and leaving the tunnel, or the near inability to perceive the surrounding environment at night, and the extremely vulnerable to the weather.

C. LiDAR

LiDAR calculates the interval between emissive laser pulses and scattering reflected by targets to obtain distance, which includes 2-D LiDAR and 3-D LiDAR based on the scanning structure. 2-D LiDAR is a single-layer structure device, while 3-D LiDAR is a multi-layer one. 3-D LiDAR is more prevailingly applied to AD vehicles but more expensive. With the increasing application of LiDAR and production, manufacturing costs will gradually decline, and predictably reach the situation that most automobile manufacturers can accept it. LiDAR provides practical and precise 3D perception competence in day and night. According to the presence or absence of motion units [68], LiDAR can be divided into three types: time-of-flight (TOF), triangulating LiDAR, and phase-ranging LiDAR, and the mainstream is the TOF LiDAR in AD system. In the latest research, LiDAR has been fully capable of recognizing and sensing pedestrians' multiple motion patterns and spatial states [69]. The multi-line LiDAR continuously emits a laser beam through a transmitter, and the receiver collects the target scattered light as a point cloud image, which helps in perceiving and recognizing pedestrians and vehicles.

In reference [70], a 16-line LiDAR is implemented to percept the position, velocity, and direction of pedestrians and vehicles in the streetscape. Furthermore, they used an improved density-based spatial clustering of applications with noise (DBSCAN) method for clustering LiDAR data. They then divided FOV into several sub-regions according to the target distance. Moreover, the back-propagation (BP) neural network is applied to classify and identify the targets with a 95% accuracy in 30 meters. The 64-line or 128-line LiDAR will increase the density of point cloud, which can improve the performance of background filtering, clustering, and classification, thus improving the tracking accuracy and detection range to some extent. In reference [71], a Velodyne 64-line 3-D LiDAR is applied to identify and track pedestrians with a support vector machine (SVM) method. Then a comparison of the pedestrian position and speed direction will come into play for early warning mechanisms. Notably, reference [72] provides a pedestrian identification coding method by analyzing the statistical shape of 3-D LiDAR data. Finally, the SVM and KNN algorithms are executed to target recognition. Reference [73] analyzed the influence of rainfall on LiDAR, and the results are shown in Fig. 3, Fig. 4, and Fig. 5. These pictures show when the rainfall increases, the max detectable distance, the number of points and obstacle detection range and other performance decreases rapidly. LiDAR also has the advantage of exclusivity in ranging. Reference [74] divided the road into nodes and used the dead reckoning algorithm to initialize the graph to estimate the position of other vehicles on the road. The error of their method is less than 15 cm in estimating the lateral, longitudinal, and horizontal position of the vehicle ahead, which is superior to traditional GPS methods. Reference [75] proposed a path planning method combining LiDAR and WiFi.

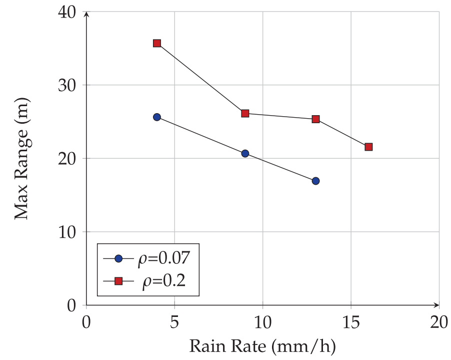


FIGURE 3. Max range influenced by rain rate [73].

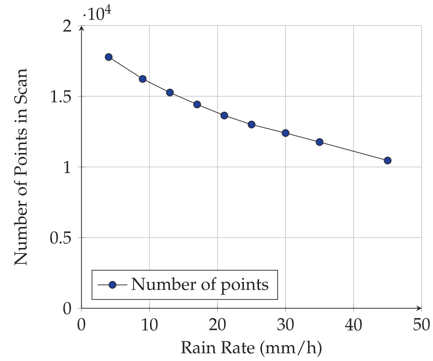


FIGURE 4. Number of points influenced by rain rate [73].

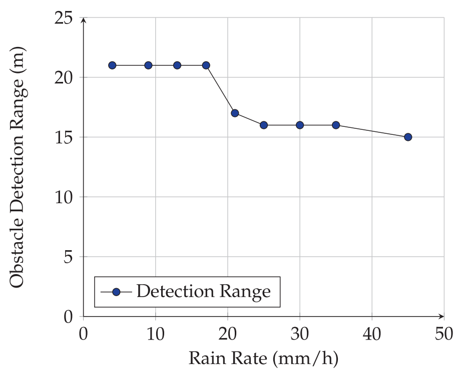


FIGURE 5. Obstacle detection range influenced by rain rate [73].

They used WiFi to determine the position of the automated platform in the resulting map of the environment, and then avoided obstacles detected by LiDAR, which enabled local navigation in low-light conditions.

Although LiDAR is superior to MMW-Radar in measurement accuracy and 3-D perception competence, whose performance is still incompetent under severe weather conditions such as fog, snow, and rain. The convergence of cameras, MMW-Radar and LiDAR data, will wipe off a portion of information redundancy, provide a reliable and efficient perception ability, but the system is too expensive.

D. GPS AND IMU

The on-board positioning equipment can solve and deal with some exclusive scenes through the cooperation of multiple sensors. The high-precision algorithm design also provides high-precision positioning for the AD car. Reference [77] believes that the combination of dual antenna and IMU can overcome the sensor biases and obtain good accuracy, but the system is too expensive. Thus, the study proposed a low-cost method estimating the lateral slip angle based on GPS and IMU. However, they believe that the camera cannot be well applied to the measurement process of the lateral slip angle due to its unreliable operation, even though the camera can provide useful angle information. At the same time, this scheme does not work well when the angular velocity changes too fast. The combination of on-board GPS and IMU can realize the positioning of its vehicle. Nevertheless, relying on GPS and IMU alone cannot achieve the requirements of lane-level positioning of AD vehicles. In reference [78], GPS and IMU are used to track the moving objects in real-time. The tracking result shows that there is still an insufferable deviation between the tracking trajectory and the actual route. For AD at L4 and L5 levels, it is evident that there are more sensors needed for data fusion. Simultaneous localization and mapping (SLAM) use camera or LiDAR data to calibrate the position through closed-loop detection to achieve accurate localization of the vehicle in the location environment. Reference [79] proposed to fuse the information of ORB-SLAM, GPS, and IMU, and improve the robustness and accuracy of autonomous vehicle positioning. This method can obtain the confidence of GPS signals through semi-supervised clustering of IMU information, which enhances the robustness and achieves better positioning even if GPS signals are lost. Besides, in reference [80], the LiDAR point cloud is fused with the GPS-IMU, and the processed data are studied through the fully convolutional neural network, to generate the safe driving route of the vehicle. Meanwhile, the possibility of a fusion of radar and camera data is discussed to further improve the sensing accuracy and sensing range of the system. In general, the fusion of more sensor data into the AD perception process will significantly enhance the vehicle perception ability and range of perception. The computational pressure brought by multiple sensors will also increase. The combination of V2V, V2I, and cloud computing will reduce the computing pressure on vehicles to process large amounts of data. Compared with DSRC, C-V2X technology has more comprehensive coverage, larger bandwidth, and which is compatible with smartphones, enabling communication between vehicles and humans.

E. VEHICULAR COMMUNICATIONS

It is challenging to deal with complex and multiple autonomous driving tasks only by relying on vehicle intelligence. The construction process of the smart city generates V2X, I2X, and P2X technologies. In recent years, the research on V2X accounts for 92.14% in connecting different road users through communication technologies [30]. V2X technology includes DSRC and Cellular-V2X (C-V2X, including LTE-V2X and 5G-V2X). Among them, DSRC is an efficient and mature communication technology, which can meet the requirements of the stability and real-time performance of the networked communication system for autonomous vehicles. However, compared with C-V2X, it has a lower data transmission rate, a smaller coverage area, and which is vulnerable to interference. Besides, the channel load of DSRC further reduced in high-speed scenarios [28]. Reference [29] considered the data sharing between AD vehicles and analyzed two situations that data can be fully shared, and data cannot be fully shared due to privacy protection. They believed that the tension between the utilitarian use of data and privacy would increase in the future.

In reference [81], DSRC is applied to communicate with surrounding vehicles, and the real-time state of vehicles determined by the vehicle dynamics model and braking system dynamics, which prevents rear-end collision with other vehicles and illustrates the reliability of DSRC communication. However, AD requires safety redundancy in the event of communication interference, and integration with other sensors is essential. In reference [82], the vehicle and pedestrian target recognition and trajectory generation completed by using the LiDAR sensor on the roadside. In this paper, the LiDAR is placed on high ground to obtain a more global traffic situation, and DSRC is applied to broadcast and receive the information in real-time. This method reduces the cost of AD vehicles and realizes the one-time processing and multi-point sharing of data. Reference [83] proposed that the LTE network can be used for real-time communication with vehicles. Each vehicle sends its position information to the adjacent base station and receives the position information of other nearby vehicles. Through the combination of trajectory prediction algorithm and vehicle motion model, the vehicle can predict the position of the surrounding vehicles and avoid the occurrence of traffic accidents. However, the vehicle communication system based on LTE cannot well adapt to the scenes of high-speed and congestion. When the vehicle speed is higher than 60 km/h or the vehicle density is higher than 1000 vehicles/km^2, the communication delay will further increase, and even real-time communication cannot realize. 5G will be well adapted to these scenarios, while the high bandwidth feature allows AD vehicles to share more sensor data or combine it with high-precision maps for safety dynamic planning [32].

F. MULTI-SENSOR FUSION AND ANALYSIS

Currently, three primary sensor-combination forms are applying for perceiving the environment in the MSHIF system, including RC, CL, and RCL. Besides, a combination of Radar-LiDAR (RL) adopted in reference [36], and MMW-Radar and the infrared camera is used in reference [89] to obtain further thermal imaging results and to sense potential organisms. We give a statistic of combination in Fig.6. The results show that the most commonly

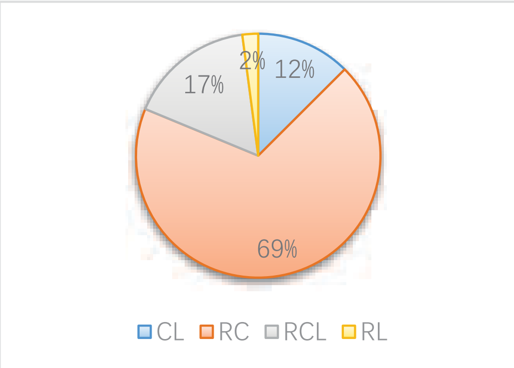


FIGURE 6. Status of multi-sensor combing form.

used sensor combination is RC because this combination can obtain excellent resolution while obtaining the distance information of the surrounding objects. Similarly, image information with depth can get from the combination of LiDAR and camera. Some studies combine LiDAR and MMW-radar with cameras to improve safety redundancy. Both of radar and camera are the full-fledged and economical technology. The performance of LiDAR is growing gradually, and the price of the high-performance product is still high. Although the camera can obtain the contour, texture, and color distribution of the target, the drawbacks are also apparent. The application of binocular and depth cameras allow image data to have depth information, but there is still a long way to go for a high accuracy. Compared with cameras and LiDAR, MMW-Radars have longer wavelengths, which can penetrate rain, snow, and fog. One fly in the ointment is that radar is more susceptible to clutter interference. LiDAR can work continuously in day and night, which provides high-resolution and long-distance 3-D data except for severe weather conditions. Therefore, the exclusive solution to satisfy various working conditions is to adopt MSHIF technology. We give another statistic of the characteristics, advantages, disadvantages, and applicable scenarios for each sensor in Fig.7, and wherein show the benefits of the MSHIF system in environment perception and target recognition. There is no doubt that MSHIF technology has achieved relatively comprehensive advantages.

However, the system is also more complex, facing challenges in the effective integration of multi-source heterogeneous data and heavy computation. One of the problems of MSHIF is that the amount of data increases and the network structure becomes more complicated to increase the recognition accuracy. References [84] proposed a fusion platform, which significantly reduced the parameters of network structure and improved the speed of network learning by Mobilenet V2. Similarly, reference [85] made the learning speed of the autonomous vehicle to the data increase by 20% based on the reinforcement learning method. Reference [86] further points out that although many studies have proposed different heterogeneous computing architectures for multi-sensor data processing, edge computing is still needed to improve the computing power in the case of limited resources.

On-board communication allows vehicles to share location information in real-time. With the application of 5G or other higher frequency communication technologies, the data transmission rate of the Internet of Vehicles will increase considerably. Although many studies are devoted to the fusion multi-sensor data like radar, camera, and LiDAR, it is necessary to combine the sensing technology and vehicle communication technology to obtain sufficient security redundancy.

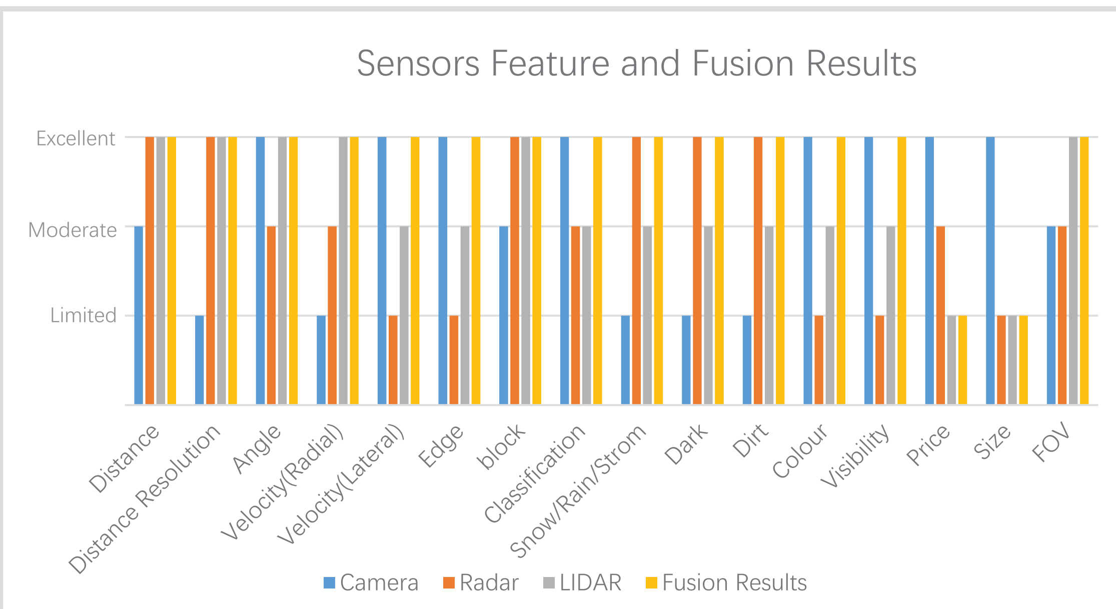


FIGURE 7. Sensor feature and fusion result.

In the next chapter, the strategies of multi-sensor fusion in the studies of recent years will be described in detail.

III. MULTI-SENSOR DATA FUSION

Different approaches to the MSHIF process represent different levels of abstraction from the original data in the fusion phase. Because different fusion strategies are adopted in different data abstraction stages, different fusion algorithms are used in multi-sensor data fusion. We classify the fusion methods used in different studies to reflect the fusion ideas adopted in these studies. These fusion methods include four categories: information fusion strategy based on discernible units, fusion strategy based on feature complementarity, fusion strategy based on target attributes of different sensors, and fusion strategy based on the decision making of different sensors. The details will be explained in the rest of this chapter.

A. FUSION STRATEGY BASED ON DISCERNIBLE UNITS

Fusion strategy based on discernible units (FSBDU), or data-level fusion, refers to the fusion process in which the data of distinguishable units of different sensors are directly fused, and then the data after fusion is further processed. FSBDU [93], [94] is abundantly adopted in multi-source image fusion for image enhancement, especially in the application of remote sensing imaging by fusing infrared images and RGB images. Because of the longer wavelength, the raw data of MMW-Radar is not conducive to imaging immediately. The spatial resolution of LiDAR is higher than that of MMW-Radar, but the horizontal resolution and vertical resolution are still far behind of optical image. At the same time, due to different sampling rates and FOV of the sensor, it is necessary to respective align them in time and space. The data processing units of multiple sensors (so-called as a frame) have different data formats and data volume sizes, so different sensor frames need to be aligned. Space alignment means that the same target detected by different sensors corresponds to a unified coordinate system in the process of FSBDU. In recent years, some studies have focused on MMW-Radar imaging [95], [96], but still not enough to distinguish multiple targets in complex scenes. Some studies using radar or LiDAR to generate raster maps and then fuse with optical images, which can also be considered as FSBDU methodology. Generally speaking, in the process of radar or LiDAR fused with the camera, FSBDU is divided into two research orientation. One is based on the results of barrier detection by radar or LiDAR and generates a raster map, which is region-based fusion. Another method is to take the optical images as the real samples and generate the radar or LIDAR images through the GAN [52]–[55].

In reference [97], the sensors, including radar, LiDAR, camera, and GPS with a map was used to create an environmental representation of the AD vehicle during driving. Multiple accumulations of observational LiDAR data are used to generate a grid map. Then each grid has a statistical barrier of observed values, and there will be a risk warning when the number is higher than a specific amount. The recognized objects will be compared with the candidate objects detected by MMW-Radar. If both of them show that the region exists target, it will be integrated into the static maps. Finally, the distance information is used to update the position error of the vehicle and construct the safe driving area. Reference [88] utilizes a deep learning method that fuses the LiDAR point cloud and camera image for road detection. The unstructured sparse point cloud is projected onto the camera image plane and then obtains a set of dense 2-D images of the encoded spatial information that applied to road separation. Moreover, a new conditional multi-generator generative adversarial network (CMGGAN) is proposed in reference [98], which can use the trained model and radar sensor data to generate the environmental image straightway, making comprehensive exploitation of all the environmental characteristics detected by the radar sensor. On this basis, the generated image and the optical image can be integrated to perform FSBDU as [97] does. Reference [99] proposed a dual static FMCW radar system, which was constructed using a simple wireless synchronization scheme and a broadband omnidirectional antenna. It adopts imaging technology compatible with the FMCW imaging system to provide high-resolution images to detect objects inside a wall, which fully demonstrates the superiority of MMW-Radar in penetration. Generative adversarial network (GAN) allows any form of data to be used directly to generate images, while it can also use the data already obtained to generate better quality data further. In reference [100], it considered that the LiDAR or MMW-Radar consumed enormous computing resources in the process of integrating with the camera. Therefore, conditional GAN was used in this paper to reconstruct vibrant semantic scene images from the LiDAR point cloud under the supervision of the image, and the effectiveness of real-time vehicle detection was verified through the KITTI data set. Similarly, a LiDAR-based feature learning framework was proposed in reference [101], which replaced the traditional feature learning framework based on geometric matching in the process of drawing construction. Besides, the mapping result of SLAM is used to realize unsupervised location identification. Reference [102] directly spliced the up-sampled point-cloud data of LiDAR with the image as the input of AlexNet, to ensure that the input of the CNN contains the image of depth information and obtain more accurate results. Their method identified 100% of pedestrians and bicycles, more than 97% of vehicles and other targets, and slightly less trucks (88.6%).

Multi-source heterogeneous pixel-level fusion in the AD process generally utilizes the resolvable unit of radar and LiDAR or generated images and then extracts the environmental characteristics and the target parameters from the fusion data for further decisions. FSBDU directly merges the data without deep information extraction [103]. Although multi-source data can be fused to the maximum extent, there is redundancy between the data, leading to low fusion efficiency.

B. FUSION STRATEGY BASED ON COMPLEMENTARY FEATURES

The fusion strategy based on complementary features (FSBCF) combines the features of multi-target extracted from corresponding sensor data and then performs classification and recognition by applying the fused multi-sensor features. Since heterogeneous sensors can capture uncorrelated dimensional features of the same target, which provides superior identification for target detection and recognition. The extracted features in AD system include target parameters extraction and data features extraction:

1) *Target parameters extraction*: it includes target information such as size, distance, direction, velocity, and acceleration of the targets extracted from the pre-processing data. Many studies extract location features of radar or LiDAR targets and assist image recognition by generating a region of interest (ROI), which directly converts the position of the radar detection target into the image to form a region.

2) *Data feature extraction*: the data feature is to extract features such as target contour, texture, time-frequency characteristics, and color distribution from the image or other processed data for classification and recognition.

In computer vision, a large number of regions of interest (ROIs) that may contain the target usually generated in an image, and these ROIs are classified through the pre-trained classification model. Moreover, the ROI with the highest confidence is the location where the target is. Determining the location of the target in this way requires enormous computation. Because of the advantages of LiDAR and MMW-Radar in detecting the target position, the calculation amount is relatively small. Therefore, many studies use radar and LiDAR to extract the distance and azimuth information of the target first and then map the position information into the image data to generate fewer ROIs. Finally, the pre-trained model is used to identify these areas of interest further and accurately classify the category of the target. After extracting ROIs, many studies applied machine learning methods for further perception tasks. Traditional machine learning methods generally require the extraction of standard features, such as Haar operator, HOG operator, and gray-level co-occurrence matrix (GLCM) to extract features from images, and then SVM [20], Adaboost [104], and other methods are applied to classify these features. Recent researches tend to use the neural network to achieve target classification and recognition, such as YOLO, CNN, and ANN. In reference [104], the application of near-infrared camera and radar enables reliable, real-time identification of pedestrians on AD vehicle platforms. At the same time, the cascading enhanced classifier is convenient for fusing the radar and camera information to the feature layer. Radar-based human motion recognition utilizes the time-frequency spectrogram of human motion, and short-time Fourier transform (STFT) is a frequently used method to analyze time-frequency characteristics [105]. By extracting the features of the same target from different sensor data, the recognition accuracy of the target will improve further, because the target features obtained by various sensors are in different dimensions.

However, the method based on deep learning has more advantages in recognition accuracy and does not need an artificial feature extraction process. Without the artificial feature extraction process, the fusion process based on complementary features needs to be embedded into the neural network architecture. Therefore, the research on feature fusion based on multi-sensor data has almost stopped with the emergence of deep learning. In recent years, FSBCF mainly uses the position characteristics of LiDAR and MMW radar for complementary fusion.

Long-term and short-term memory (LSTM) unit superposition recurrent neural network (RNN) is another new methodology to extract sequence features to automate motion classification in reference [106]. Reference [72] proposed a process of fusion sensing of MMW-Radar and camera. The radar coordinates, including distance and angle information, are transformed into corresponding image regions, and then the ROIs are classified by the deformable part model (DPM). The recognition result was inferred at a detection accuracy of 98.4%. Moreover, the candidate target coordinates are produced by LiDAR in reference [107]. Reference [108] applied a super-region model to describe targets shape in 3-D form after generating ROIs from point cloud images, and finally performed 3-D classification and target detection. Reference [109] proposed a hybrid random field model fusing camera and LiDAR feature based on the conditional random field model to segment the road in front of the vehicle. A large number of experiments have been carried out on the KITTI-ROAD benchmark dataset, and which shows that this method is superior to the existing method. It is worth mentioning that in reference [88], the data of the LiDAR and the camera are carried out by using the fully convolutional neural network (FCN). Three fusion strategies were adopted to combine LiDAR and camera data, including early fusion strategy (EFS), late fusion strategy (LFS), and cross fusion strategy (CFS), and FCN was used to generate safe driving areas on current roads. The whole FCN consists of 21 layers. EFS directly connects the LiDAR data and camera data to form the 6-dimensional tensor and then uses FCN to train the data to generate a safe area; LFS passes two different kinds of data through the first 20 layers and fuses two different outputs in the last layer; CFS introduces a weighting with another data layer at each layer and fully fused at the last layer. These fusion strategies combine the data of various sensors well and make the full fusion of the different sensors. However, due to the different amount of information obtained by sensors, it is not considered that different data should have different network depth in the fusion process.

FSBCF requires a certain degree of information extraction on the raw data and combines the uncorrelated dimensional features or parameters detected by multiple sensors. Higher dimension features have a more distinguishable ability in the target recognition, thus improving the efficiency of the fusion and breaking through the inherent defects of a single sensor [110], [111]. In recent years, the research on combining the features of multi-sensor is not carried out enough on account of straightforwardly apply the existing neural network architecture of visual pattern recognition. Most of the research is based on the method of target parameter extraction to realize the FSBCF fusion strategy.

C. FUSION STRATEGY BASED ON TARGET ATTRIBUTES

Fusion strategy based on target attributes (FSBTA) is a distributed data processing procedure, in which each sensor extracts target parameters and recognizes different target to form a target list. Multiple target lists will be fused to acquire reliable and authentic target information, avoiding false alarms, and missed inspections. In reference [16], multiple cameras, MMW-Radar groups, and LiDAR are applied to extract targets in the traffic environment and generated corresponding target lists. The generated target lists make it is possible to plan a safe driving area of AD vehicle for avoiding the potential risk of collision.

The extracted motion information of MMW-Radar provides ROIs for the image primarily in reference [48], and then the convolutional neural network (CNN) is applied to identify the target in the ROIs. Meanwhile, the target lists perceived by the MMW-Radar and camera were matched and merged, respectively. Fused information includes target type, distance, speed, angle, and angular velocity. The fusion result tolerates the missed detection of a single sensor to a certain extent and improves the robustness. Approximately in reference [112], 2-D Fast Fourier transform (2FFT), and sparse feature detection are respectively applied in a radar and vision subsystem to extract motion parameters of multi-target. Moreover, the Gaussian inverse Wishart probability hypothesis density filter (GIW-PHD) is applied to track the segmented objects. In reference [113], low-level information fusion performed by LiDAR and camera and then applied the range and angular information of LiDAR to generate ROIs in corresponding images, and they merged the target lists generated by LiDAR, MMW-Radar, and camera ultimately. Target fusion handles the complementarity between sensors. The camera provides high levels of 2-D information such as color, intensity, density, and edge information, while LiDAR provides 3D point cloud data. By obtaining as many attributes as possible, it facilitates human-computer interaction and intent recognition. The fusion processing of LiDAR and camera sensors is applied for pedestrian detection in reference [46]. Moreover, the 3-D point cloud data is adopted to detect the shape of the target further in this study to reduce the false alarm rate and cope with the target occlusion problem in camera-based pedestrian detection. In this paper, the target information of LiDAR generates ROIs for the image. At the same time, the target lists recognized by LiDAR data and images are matched to maximize the detection speed and achieve an average detection accuracy of 99.16% for pedestrian detection. Reference [114] used a stereo camera and LiDAR to detect the lane change behavior of the front vehicle. They used the neural network model based on particle swarm optimization to classify the distance, radial speed and horizontal speed of the vehicle to recognize the lane change behavior, and the final comprehensive recognition rate reached more than 88%.

When FSBTA integrates the data, the level of abstraction of the data is between FSBCF and FSBMD. This kind of fusion strategy adopts multiple sensors to perceive the targets and fuses the extracted target attributes or environmental features. This fusion strategy will improve the stability and reliability of the perception system to confront a single sensor may be subjected to false alarms or missed detection during independent detection and identification.

D. FUSION STRATEGY BASED ON MULTI-SOURCE DECISION

Fusion strategy based on the multi-source decision (FSBMD) makes a preliminary decision on the location, attributes, and categories of the target by single sensor data, and then adopt the specific fusion strategy to combine the decisions obtained by multiple sensors roundly, and appropriate methods are applied to achieve the eventual fusion result. FSBMD integration directly makes decisions for specific goals, and the accuracy of the final decision results is directly dependent on the fusion effects. FSBMD can usually be divided into decision fusion, decision making, credibility fusion, and probability fusion [115]–[118].

The methodologies of FSBMD commonly contain subjective Bayesian probabilistic reasoning method, Dempster-Shafer (D-S)-evidence-theory-based reasoning method, artificial intelligence (AI) method, and fuzzy subset theory method. Reference [16] proposed a multi-sensor platform for AD, which is applied to extract the road edges, lane signs, traffic signs, obstacles, and motion parameters of targets by data processing. Decision-making strategy based on target information in the platform can facilitate to control the motion state of AD vehicles. Reference [36] proposed a multi-modal vehicle detection system combining artificial neural network (ANN) post-fusion strategy, in which adopted optical image, 3-D LiDAR range data, and reflectance data as three different modalities with a respective detection process, and joint re-scoring and non-maximum suppression are adopted to fuse the decision of each modal. Moreover, compared with individual modal, the target detection performance of FSBMD has improved by 1.2 percentage points. Reference [113] proposed a framework for combining comprehensive fuzzy theory with the nervous system. This framework combines Kalman separation and fine processing criteria to construct an effective information combination strategy for the target tracking framework. Fuzzy sets provide new ideas for the development of data engineering, processing systems, selection, and information analysis. The adaptive neuro-fuzzy inference system (ANFIS) is one of the most influential neural system frameworks. ANFIS has strong acceptability and predictive ability and is a proficient tool for managing empirical instability in any framework [119], [120]. Moreover, an evidence fusion framework based on the

D-S evidence theory is proposed in reference [90] to cope with the vulnerability of sensor data to noise and the uncertain motion of the targets. Then a reliability function is combined with the measured value to establish a classification index to classify the target, especially pedestrian detection. Also, the confidence function in the paper redistributes the probability provided by the sensor to make reliable decisions when the uncertainty model is inaccurate. Reference [91] proposed a combination of camera, GPS, and on-board sensors for accurate positioning of vehicles. By using the extended Kalman filtering algorithm, they fused GPS information and visual odometry and improved the accuracy by 40% compared with traditional GPS positioning methods.

FSBMD synthesizes several decisions made by different sensors, and the performance of the fusion strategy determines the ultimate fusion effect. Through information fusion at this level, the final decision directly produced [121]–[123]. This method can effectively avoid the uncertainty and unreliability caused by relying on the perception results of a single sensor as the final decision. However, FSBMD does not significantly improve target detection performance from the data processing level, and data complementarity is relatively low. In some studies, this strategy often combined with others.

E. ANALYSIS OF FUSION STRATEGIES AND PERCEIVED RESULTS

Different studies apply different strategies to various tasks or scenarios, and the specific methods used in the implementation process are also different, it is difficult to compare which strategy is better. However, from the perspective of information fusion, FSBDU, and FSBCF can make the best use of the complementarity of different sensor data. Besides, some studies also adopt a combination of various strategies to improve the reliability of fusion further.

In table 2, we summarize the specific tasks accomplished by various sensors in different studies. In terms of specific sensing tasks, the current researches include two categories: moving target perception and environment perception. Among them, moving target perception includes pedestrians, vehicles, bicycles, and other obstacles. Some studies only realize the detection and recognition of this target, while others further analyze the movement trend of the target based on recognition results. By analyzing the motion trend, the target tracking loss can be avoided, because it is difficult to ensure the consistency of the detection of each frame. The next chapter discussed how to decide the existence of the object and generate the trajectory of the object. Besides, some studies combine lane line detection with obstacle detection to plan the safe driving area in front of the AD vehicle. Meanwhile, some studies have visualized the data on this basis. Of course, visualization of the information is necessary to verify that the safety zones correctly demarcated, but not essential for AD at L4 or L5 levels. Moreover, when various sensors used for target identification and tracking, due to the different spatial positions placed by different sensors, data sampling rate, and FOV, a target corresponds to several different coordinate systems in various data. It is necessary to unify these coordinate systems to obtain the target position information after fusion. Current sensor calibration mainly includes calibration of MMW-Radar and camera and calibration of LiDAR and camera.

At present, during the AD process, it mainly relies on the camera, LiDAR, and MMW-Radar to complete the target detection and recognition. The introduction of other sensors (such as communication devices, GPS/IMU) will further expand the dynamic sensing range of the vehicle. Of course, as shown in table 2, part of the research has started related research. Moreover, RCLUGIV refers to radar, camera, LiDAR, ultrasonic, GPS, IMU, and V2X, respectively. The next chapter will introduce the motion model of the target and the tracking method of multiple targets in detail.

IV. TARGET TRACKING AND DATA ASSOCIATION

In the process of driving, the driver will continuously observe the traffic situation. In addition to observing traffic lights, it is more important to analyze and predict the behavior intentions of vehicles and pedestrians. For achieving this goal, it is necessary to track multiple potential security threats in real-time and analyze the movement trend and intention of the target according to the tracking results. In this way, AD vehicles can make correct decisions in advance to avoid the occurrence of dangers. For analyzing the intention of the target, it is necessary to model the motion mode of different targets, and then judge the motion state of the target based on the motion mode, such as stationary, constant speed, accelerating the motion and turning. Moreover, in a complex traffic environment, it is difficult for different sensors to ensure the consistency of the front and rear frames of the target detection due to the existence of interference factors. For the safety decision and motion planning of the autonomous vehicle, it is necessary to ensure the accurate tracking of multiple targets even if the target lost in the current frame. Also, due to the fusion of multiple sensors, different sensors generate different tracking trajectories for multiple targets. Therefore, specific fusion strategies should be adopted in the tracking process to ensure the reliability of the final result. The following sections of this chapter will discuss in detail the motion model and the tracking strategy adopted in the multi-target tracking process.

A. MOTION MODEL

The time series of human motion has always been an arrestive research topic [132]. Pedestrian detection is a significant portion of AD, and relevant researches have carried out in recent years [133], [134]. Reference [135]–[139] proposed corresponding vehicle motion models. Specifically, to facilitate target tracking and state prediction, the constant velocity (CV) model and the constant acceleration model (CA) of the target are proposed in reference [89]. Reference [128] further proposed vehicle models including constant velocity lane keeping(CVLK) model, constant acceleration lane keeping(CALK) model, constant velocity lane changing (CVLC)

TABLE 2. Task analysis based on sensor fusion.

Senario	Specific Task	Ref	Sensor Types
Perception of Moving Objects in Traffic Environment (including pedestrians, vehicles, bicycles, etc.)	Pedestrian	[71], [72]	L
		[107]	RL
		[108], [131]	CL
		[130]	RC
	Vehicle	[18], [112], [124], [127], [129], [130]	RC
		[36], [46]	CL
		[37]	RCLUV
		[77], [78]	GI
		[79]	GL
		[81]	VG
		[82]	LV
		[107]	RL
		[128]	CLGI
	Pedestrian and Vehicle	[20]	R
		[48], [50], [76], [125]	RC
		[49]	RCLGI
		[70]	LV
		[102]	CL
	Lane Detection, Obstacle Detection, and Path planning	[12]	RCLG
		[17]	RCL
		[109], [131]	CL
		[82]	VG
		[91]	CG
		[75]	L
Reconstruction And Visualization of The Front Area	Safety Zone Division	[41]	RC
		[61]	RC
		[97]	RCLG
	Motion Analysis	[128]	CLGI
		[131]	CL
	Visualization	[71]	L
		[98]	RC
Safety Zone Construction And Collision Warning	Obstacle Avoidance	[64]	CL
		[129]	RC
		[65]	C
		[81]	VG
	Safety Zone Division	[34], [41]	RC
		[88]	LC
Multi-sensor Calibration And Data Fusion Platform	Data Fusion Platform	[16]	RCL
		[90], [92], [84]	N/A
		[84]	CLGI
	Multi-sensor Calibration	[39], [40]	RC
		[60], [63], [87]	CL

model, and constant acceleration lane changing (CALC) model.

Generally speaking, if the motion model is too complicated, the computation will increase dramatically; on the contrary, if it is too simple, the tracking performance will be affected. Different motion models include two categories: the linear motion model and the nonlinear motion model. The appropriate model can significantly improve the performance of the vehicle tracking system. Moreover, reference [140] proposed more general models including constant velocity (CV), constant acceleration (CA), constant turn rate and velocity (CTRV), constant turn rate and acceleration (CTRA), the constant curvature and acceleration (CCA) and constant steering angle and velocity (CSAV). Besides, the state transition diagram of these motion models is shown in Fig. 8, which is the angular velocity of steering, is accelerated speed and is steering angle. Tracking based on a motion model can achieve better tracking performance and accuracy. Moreover, by analyzing the motion state of the target through the motion model, the intention of the target can be further identified

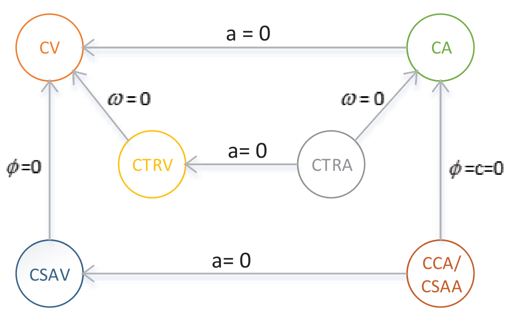


FIGURE 8. Motion model state transition diagram in [128].

[140]–[143]. Further target tracking algorithms will be discussed in the rest of this part.

B. CLASSIFICATION OF TARGET TRACKING PROBLEM

The target tracking process is to match the current measurement of the sensor with the historic track. However, when confronted with the tracking issues of multiple sensors and multiple targets, things get complicated. In different scenarios and applications, the number of targets may be different from monitored targets by multiple sensors. Multi-target tracking issues with multi-source heterogeneous data include four situations.

1) SINGLE SENSOR TRACKS SPARSE TARGETS (S2S)

S2S mainly considers multi-target tracking under sparse situations, which means multiple targets will not be interactional [144], [145]. If there is no tracking ambiguity in the multiple targets tracking, the appropriate methodologies include the nearest neighbor (NN) algorithm, corresponding improved NN algorithm, Kalman filter, and Bayesian filter-based algorithm, and both of them are relatively mature and have a wide application.

2) SINGLE SENSOR TRACKS MULTIPLE TARGETS (S2M)

Compared with multiple sparse targets, multiple targets in S2M have such situations as distance resolution blurring or overlapping occlusion, and the global data association algorithm can be adopted [146], [147].

3) MULTIPLE SENSORS TRACK SPARSE TARGETS (M2S)

Similar to S2S, the target distribution is sparse. However, regarding the same target tracked by multi-sensor, the target location between multiple sensors exists a difference in time and space [19], [113]. Besides, the data form of different sensors and processing methods vary. K-nearest neighbor (K-NN) algorithm, federated Kalman filtering (FKF), and joint probabilistic association algorithm (JPDA) and related improved algorithms are often adopted for the M2S.

4) MULTIPLE SENSORS TRACK MULTIPLE TARGETS (M2M)

For capturing the motion information of multiple targets from a single perspective, there are problems of data overlap and trajectory crossover in M2M. In the structured urban environment, the targets' behavior is often quite related to the surrounding environment, which makes the M2M problem special [89], [125], [126]. At present, the applied methods include JPDA and its improved algorithm, multi-objective hypothesis tracking (MHT), RFS theory. Specific methods are described in the next subsection.

C. DATA ASSOCIATION METHOD

As discussed above, the data association of multiple sensors includes two cases: one is the track association of a single sensor, which associates the current tracking result with the historical tracking trajectory; and another is to consider correlating the measured values of multiple sensors with the tracking trajectories of multiple targets in history. The basic ideas commonly used in current research are mainly divided into the nearest neighbor algorithm, improved algorithm based on Kalman filter, the improved algorithm based on the Bayesian filtering idea, the algorithm based on probability data association, and method based on the random finite set. These methods are discussed in detail in the following sections.

1) THE NEAREST NEIGHBOR (NN) ALGORITHM

The nearest-neighbor algorithm is simple, effective, and easy to implement, which is suitable for multi-target tracking with sparsity. When the nearest neighbor algorithm is applied to cope with the dense group targets and multi-sensor data association, the optimal solution in the statistical sense will lead to the wrong association matching. In reference [148], a data association method called RTK-GPS is proposed and compared with the NN algorithm and JPDA algorithm. The running time of NN, RT-GPS, and JPDA is 0.012193 *ms*, 0.29354 *ms*, and 2.5 *ms*, respectively. Their root means square errors were 0.9232, 0.3979, and 0.4378, respectively. Moreover, with the number of targets increases, JPDA algorithm will not be suitable for AD vehicles to track targets in real-time.In reference [149], the squared Mahalanobis distance is adopted to manage the association problem of sensor measurements and historic tracks. In reference [89], the nearest neighbor data association (NNDA) and the JPDA method combined with interactive multiple-models (IMM), and extended KF (EKF) are applied to cope with target tracking and data association, and the performance is compared and analyzed. The results showed that the root means square error of target distance (over 1000 *m*) in different directions obtained by the JPDA algorithm was about 5 *m* higher than the NNDA algorithm. Besides, the average time spent by JPDA is 5.993115 *s*, while NNDA is 1.342506 *s*. Therefore, NN has the advantage of low algorithm complexity when dealing with sparse target tracking problem while JPDA suffers from poor performance.

2) THE METHOD BASED ON KALMAN FILTER

KF is a recursive algorithm, which can estimate the current state of the target by obtaining the previously observed target state estimation and the measured value of the current state [150]. FKF can be applied to solve the problem of multiple sensors tracking multiple targets. In reference [5], a method combined with fuzzy adaptive fusion and wavelet analysis is proposed to decompose the linear process model into a series of simpler subsystems and apply multiple KF to estimate the states of these subsystems separately. Combined KF and adaptive neuro-fuzzy inference system (ANFIS) in reference [113], an effective information combination method constructed for the target tracking framework, which has better precision and performance than the traditional KF algorithm. Similarly, in reference [37], according to the weighted average of the predicted states and the estimated status update based on the current measurement, the lower weights are given to the states with higher uncertainties in the measurement update step. Besides, FKF is applied to process the measurements from radar, LiDAR, camera, and other sensors, reducing the statistical noise and other errors. In reference [148], KF is applied to measure the simple experimental scene. The problem of the basic KF is that it cannot estimate the nonlinear system accurately. KF can only accurately estimate linear systems, but it is difficult to reach the optimal estimate in nonlinear systems. Therefore, some researches adopted EKF and its improved algorithm. Due to the nonlinear motion process of vehicles, the EKF is adopted in reference [125] to linearize the nonlinear problem. The Unscented KF (UKF) is another widely applied improved KF algorithm. UKF adopts a statistical linearization technique to linearize nonlinear functions of random variables by linear regression of n points (also known as sigma points) collected in prior distribution [151]. Since UKF considers the expansion of random variables, this linearization is more accurate than Taylor series linearization in EKF.

3) THE METHOD BASED ON BAYESIAN FILTER

The Bayesian tracking method mainly tracks multi-targets with multi-modes of probability density (PD) and approximates each mode with component PD. Particle filter (PF), a modified Bayesian algorithm, can effectively cope with the nonlinear and non-Gaussian Bayesian estimation problem, which is suitable for solving the component PD estimation problem. When the Bayesian algorithm is applied for multi-target tracking to recursively estimate the state of multiple targets and determine the current target number, the sequence maximum-likelihood (ML) ratio is generally applied to verify whether the target is existed or disappeared.

According to the test results (the appearance or disappearance of targets), the mixed components representing multiple targets can be determined to add or delete, so that the multi-modal posterior PD can be updated and maintained. Reference [152] established the relationship model between target kinematics and class by adopting a Bayesian network. The Bayesian method and D-S method widely applied in recognition and sensor fusion. D-S method overcomes the problem that the Bayesian method cannot express incomplete or uncertain evidence. PF is a specific form of Bayesian filter, and the performance of PF elaborated in reference [127] for tracking vehicle targets in detail. The improved particle filter is proved to be superior to KF through practical tests. The front car tracking and prediction of the driver's intention algorithm are proposed in reference [153] combined with hybrid PF and IMM (IMM-PF). Then an improved PF algorithm is proposed to track vehicle targets to solve the problem of data correlation. The results show that the performance of the improved PF increased in an extreme situation. Reference [154] further proposed the multi-sensor track fusion algorithm, which adopted the ML fusion rule to deal with the correlation between the measurement of multiple sensors and multiple targets. The covariance matrix is applied to represent the correlation between multiple measurements updated by a fusion center decision. Moreover, compared with the naive algorithm and centralized KF, the experimental result demonstrates that the proposed Bayesian algorithm can reduce the mean square error (MSE). In reference [128], the Bayesian filtering is applied to complete multi-target tracking and introduce the road constraints into the algorithm. Due to the limitations of sensors and environment, such as short detection range, narrow field of vision, signal noise, and unexpected occupied obstacles, which lead performance constraints, the road geometry information is applied to the tracking algorithm to overcome the tracking constraints based on the vehicle sensors. With the continuous improvement of traffic facilities in the future, the structural features of roads will be more and more prominent. The introduction of structured constraints can improve algorithm performance and simplify the target motion model.

4) THE METHOD BASED ON PROBABILISTIC DATA ASSOCIATION

Probabilistic data association (PDA) is an extensively applied algorithm with many improvements. It verifies all the measures and estimates the motion state and covariance of the target, rather than merely adopting a single measure [155]. JPDA algorithm is proposed to deal with multi-target tracking, which considers the joint distribution probability of multiple marginal distributions [156].

In reference [157], an algorithm called PDA feedback PF (PDA-FPF) is proposed to solve the data association uncertainty, which shows the performance of PDA-FPF is close to the ideal JPDA. However, JPDA suffers the problem of massive computation. Furthermore, the exact nearest neighbor PDA (ENNPDA) [158], Joint integrated probabilistic data association (JIPDA) [159], and other improved algorithms are proposed to simplify the algorithm by sacrifice the precision. Reference [89] and [148] show that the JPDA algorithm has a superior correlation accuracy compared with the NN algorithm, but the computation is enormous in the meantime. JIPDA introduces the probability of the existence of the target as the trajectory quality measurement [159], which has apparent effectiveness in target automatic tracking that can cope with the tracking of adjacent targets or overlapping targets to some extent. However, it also suffers from the problem of tracking clustering. On the other hand, JIPDA still merges the trajectories of two targets in response to the problem of two targets accompanying each other for a short period. JIPDA filter combined with JPDA perfectly solves this problem [160]. In reference [125], JIPDA is adopted to manage the emergence and disappearance of multiple targets. Then the real datasets in urban traffic scenarios are applied to prove the performance of JIPDA in AD. Moreover, multiple hypothesis tracking (MHT) algorithm is a further improvement of the JPDA algorithm. Data association can be automatically started or terminated when the target appears or disappears. Its scalability is limited because it relies too much on the prior knowledge of target and noise. In reference [126], the MHT filter algorithm is applied to realize the tracking management of multiple targets. MHT filter analyzes each track and measurement value according to whether sensor data support each track. The sequential likelihood ratio test was carried out for each trajectory through two hypothesis conditions, and the scores were calculated sequentially based on the statistical decision function. If the score exceeds the threshold, the measurement is validated and updated as a track. Reference [161] combined the RFS theory and the advantages of

JPDA, presents a covariance control JIPDA filter (CCJIPDA). Meanwhile, a detailed comparison of multi-target tracking algorithms, including JIPDA, ENNJIPDA, JIPDA*, RFS, and CCJIPDA. Distinctly, the improved algorithm achieves a preferable tracking effect in the tracking process of adjacent or overlapping targets, but the relevant AD researches do not adopt it.

5) THE METHOD BASED ON RANDOM FINITE SETS

The RFS theory extends the Bayesian filtering framework of a single target directly to the multi-target tracking problem. RFS can avoid complex correlation process and estimate the number and state of targets, which is suitable for dealing with dense multi-target tracking. However, the application is difficult because the calculation process involves a complex set of integral operation. Several typical filters are proposed to change this situation in the present study. The improved algorithms include Generalized labeled multi-Bernoulli (GLMB) filters, Cardinalized PHD (CPHD), multi-target multi-Bernoulli (MeMber), and Generalized labeled multi-Bernoulli (GLMB) filters. Mahler puts forward a probability hypothesis density (PHD) method based on RFS [162], [163]. PHD filter can track a variable number of targets and estimate the number and location of targets. Data correlation is not necessary, but the correlation of the same target between frames is not possible.

Reference [127] proposed to realize multi-target tracking by combining data association with PHD filter. The first is to divide the data in the target extraction phase into clusters around each target and apply these distinctions between frames to achieve orbital continuity. The second is to apply the previous target state and motion model to estimate the target in the next frame. However, the motion model will be intractable to establish in practice when the target motion is nonlinear. Moreover, in the application of AD in reference [112], the image feature track and Gaussian mixture data based on MMW-Radar state estimation are applied to segment moving targets, and Gaussian inverse Wishart probability hypothesis density filter (GIW-PHD) is adopted to track the segmented targets. Generally, RFS has a complete theoretical basis, without complex data association, and can estimate the number and status of multiple targets at the same time. In recent years, it has achieved rapid development.

V. DISCUSSION AND CONCLUSION

A. DISCUSSION

With more and more sensors installed on AD vehicles, making full use of multiple different sensors to realize fusion perception can improve the reliability of the system, increase safety redundancy, and reduce the massive computation brought by distributed data processing. In this survey, we focus on the different information fusion strategies of different sensors in recent years. In these studies, target recognition and tracking or environment perception have realized by using the data of various sensors. The accuracy of data analysis improved by combining with the latest deep learning network, and the perceived results are also superior to a single sensor. However, through this survey, we find that there are still some shortcomings in the current multi-sensor fusion process, and we proposed the following improvements.

1) BETTER COMBINE MULTI-TARGET TRACKING RESULTS WITH IMPROVING AD SECURITY

Skilled drivers often make the right decision to control the vehicle by judging the intention of the surrounding target movement. For achieving AD at the L4 and L5 levels, it is necessary to have an accurate understanding of the intentions of the surrounding targets. However, most researches focus on the accurate classification and recognition of targets but ignore the analysis of target intention. In part A of section IV, some studies have proposed using the dynamic model of the target as the intention to identify the target, but there is a lack of multi-sensor data fusion. This process can be achieved more effortless by combining the data from multiple sensors since the speed, distance, acceleration, and other information of objects can be captured by radar or LiDAR, and the camera or LiDAR can identify the specific type and posture of targets. Also, the intention of the vehicle to turn can be identified by identifying the turn signal. However, combined with V2X, this information will be easier to capture, because each vehicle has its motion information, which can be easily broadcast through the roadside unit, avoiding double calculation, and the perception range can be greatly improved. With the popularization of 5G, mobile phones may also link to vehicle communication networks to realize information exchange. The measurement of the sensor needs to be further matched with this data to avoid errors. Therefore, the practical application of various sensor information to identify the intention of the target will make the whole process more accurate and efficient.

2) DEEP LEARNING NETWORK AND DATA FUSION MATCHING

As mentioned above, in recent years, more and more researches have focused on the application of deep learning to AD to improve the reliability of the system. However, for data fusion, the problem is that the sensor fusion process did not integrate into the network structure in most researches. The latest image recognition network has widely applied in target recognition and region segmentation, which have achieved good results. However, the end-to-end network structure is difficult to integrate heterogeneous data. Through this investigation, we found that the fusion of multiple sensors can be realized from the following two perspectives by using deep learning:

- Adopting the partial concurrent network structure to realize the fusion of multiple sensors by taking multiple sensor data as the input and concurrent processing results as the input in the middle layer of the network. Through a large number of data training, the network can learn the data characteristics of various sensors and can conduct fusion in the middle feature layer. At the same time, the depth of the network needs to take into account the data amount, sampling rate, and channel number of different sensor data.

- Another method is to use the idea of supervised learning to achieve data fusion. One sensor data supervises the data of another sensor to enhance the supervised sensor data. For example, GAN is used to monitor the radar data or the LiDAR point cloud to generate more accurate data, which is discussed in detail in FSBDU.

3) COMBINING SENSING SENSORS, LOCATING SENSORS AND COMMUNICATION TECHNOLOGIES

In section II, we have discussed and compared the detection capabilities and working distances of different sensors. Sensing sensors include radar, LiDAR, camera, and ultrasonic sensors, positioning sensors to include GPS and IMU while communication technologies cover V2X. The information obtained from the positioning device and the high-precision map is global, and the V2X also allows the vehicle to realize real-time information exchange within a range of several hundred meters. The combination of the total information will enable the autonomous vehicle to have a more comprehensive perception range. As a prior knowledge, the map should be better integrated with sensor data to provide environmental information such as lane, intersection, and tunnel. Meanwhile, the dynamic data detected by the sensor (including vehicles, obstacles, and pedestrians.) will also realize the construction of a dynamic high-precision map. Through the positioning sensor to obtain their position, combined with a high-precision map, it is easier to build a dynamic driving scene to plan a safe driving area.

B. CONCLUSION

In this survey, we focused on heterogeneous multi-sensor fusion, including radar, camera, LiDAR, ultrasonic, GPS, IMU, and V2X communication. In section II, we discuss the performance of the sensor in detail. With the improvement of sensor performance, sensor detection tasks become more diversified. In addition to cameras, radar and LiDAR can also be used to classify and identify targets, so fusion strategies have become more affluent. The performance of different sensors may be limited in some scenarios. Therefore, by analyzing these shortcomings, it is necessary to fuse the heterogeneous sensor data. In section III, we discussed heterogeneous sensor fusion strategy in the study of the last few years, and all the fusion strategy is summed up in four categories, including discernible-unit-based fusion, complementary-features-based fusion, target-attributes-based fusion, and multi-source-decision-based fusion. With the increasing application of deep learning, the relevant multi-sensor fusion strategy needs to be further improved, especially the serial network structure needs to be adjusted to adapt to the fusion of various sensor data. The purpose of multi-target tracking is to obtain the motion intention of the target, and the environment reconstruction is to generate a safe driving area. However, it is difficult to do these tasks well with sensors alone, which need to be combined with location sensors, maps, and V2X. Besides, to form the perception ability to look around, AD vehicles need the cooperation of a variety of similar sensors, which need to cooperate closely with the whole system, and future work should explain further.

REFERENCES

[1] M. L. Fung, M. Z. Q. Chen, and Y. H. Chen, "Sensor fusion: A review of methods and applications," in *Proc. 29th Chin. Control And Decis. Conf. (CCDC)*, Chongqing, China, May 2017, pp. 3853–3860, doi: 10.1109/ccdc.2017.7979175.

[2] L. Song and R. Yan, "Bearing fault diagnosis based on cluster-contraction stage-wise orthogonal-matching-pursuit," *Measurement*, vol. 140, pp. 240–253, Jul. 2019, doi: 10.1016/j.measurement.2019.03.061.

[3] V. Ankarao, V. Sowmya, and K. P. Soman, "Multi-sensor data fusion using NIHS transform and decomposition algorithms," *Multimedia Tools Appl.*, vol. 77, no. 23, pp. 30381–30402, Dec. 2018, doi: 10.1007/s11042-018-6114-2.

[4] S. Siachalou, G. Mallinis, and M. Tsakiri-Strati, "A hidden Markov models approach for crop classification: Linking crop phenology to time series of multi-sensor remote sensing data," *Remote Sens.*, vol. 7, no. 4, pp. 3633–3650, Mar. 2015, doi: 10.3390/rs70403633.

[5] G. Hernandez-Penaloza, A. Belmonte-Hernandez, M. Quintana, and F. Alvarez, "A multi-sensor fusion scheme to increase life autonomy of elderly people with cognitive problems," *IEEE Access*, vol. 6, pp. 12775–12789, 2018, doi: 10.1109/access.2017.2735809.

[6] J. Ma, P. Huang, and X. Xu, "A coordinated control strategy for rotating motion of the hub-spoke tethered space robot formation system," in *Proc. 26th Chin. Control Decis. Conf. (CCDC)*, Changsha, China, May 2014, pp. 4628–4633.

[7] X. Mao, W. Li, C. Lei, J. Jin, F. Duan, and S. Chen, "A brain-robot interaction system by fusing human and machine intelligence," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 533–542, Mar. 2019, doi: 10.1109/tnsre.2019.2897323.

[8] Y. Chen and Q. Zhao, "A novel square-root cubature information weighted consensus filter algorithm for multi–target tracking in distributed camera networks," *Sensors*, vol. 15, no. 5, pp. 10526–10546, May 2015, doi: 10.3390/s150510526.

[9] M. Nakagawa, Y. Yamada, H. Namie, T. Ebinuma, N. Kubo, T. Kawaguchi, M. Yoshida, and A. Yasuda, "Seamless navigation using various sensors: An overview of the seamless navigation campaign," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XXXIX-B4, pp. 35–38, Sep. 2012, doi: 10.5194/isprsarchives-XXXIX-B4-35-2012.

[10] H. Carvalho, A. Vale, R. Marques, R. Ventura, Y. Brouwer, and B. Gonçalves, "Remote inspection with multi-copters, radiological sensors and SLAM techniques," in *Proc. EPJ Web Conf.*, vol. 170. Uppsala, Sweden: Uppsala Univ., 2018, p. 7014, doi: 10.1051/epjconf/201817007014.

[11] S. Saponara, "Sensing and connection systems for assisted and autonomous driving and unmanned vehicles," *Sensors*, vol. 18, no. 7, p. 1999, Jun. 2018, doi: 10.3390/s18071999.

[12] X. Jia, Z. Hu, and H. Guan, "A new multi-sensor platform for adaptive driving assistance system (ADAS)," in *Proc. 9th World Congr. Intell. Control Autom.*, Jun. 2011, pp. 1224–1230, doi: 10.1109/wcica.2011.5970711.

[13] J.-P. Giacalone, L. Bourgeois, and A. Ancora, "Challenges in aggregation of heterogeneous sensors for autonomous driving systems," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Mar. 2019, pp. 1–5, doi: 10.1109/sas.2019.8706005.

[14] A. Swief and M. El-Habrouk, "A survey of automotive driving assistance systems technologies," in *Proc. Int. Conf. Artif. Intell. Data Process. (IDAP)*, Malatya, Turkey, Sep. 2018, pp. 1–12, doi: 10.1109/idap.2018.8620826.

[15] S. Liu, L. Li, J. Tang, S. Wu, and J.-L. Gaudiot, "Creating autonomous vehicle systems," *Synth. Lectures Comput. Sci.*, vol. 6, no. 1, p. i–186, Oct. 2017, doi: 10.2200/s00787ed1v01y201707csl009.

[16] J. Steinbaeck, C. Steger, G. Holweg, and N. Druml, "Design of a low-level radar and time-of-flight sensor fusion framework," in *Proc. 21st Euromicro Conf. Digit. Syst. Design (DSD)*, Aug. 2018, pp. 268–275.

[17] I. Baftiu, A. Pajaziti, and K. C. Cheok, "Multi-mode surround view for ADAS vehicles," in *Proc. IEEE Int. Symp. Robot. Intell. Sensors (IRIS)*, Dec. 2016, pp. 190–193, doi: 10.1109/iris.2016.8066089.

[18] F. A. Alencar, L. A. Rosero, C. M. Filho, F. S. Osorio, and D. F. Wolf, "Fast metric tracking by detection system: Radar blob and camera fusion," in *Proc. 12th Latin Amer. Robot. Symp. 3rd Brazilian Symp. Robot. (LARS-SBR)*, vol. 174, Uberlandia, Brazil, Oct. 2015, pp. 120–125, doi: 10.1109/lars-sbr.2015.59.

[19] A. Etinger, N. Balal, B. Litvak, M. Einat, B. Kapilevich, and Y. Pinhasi, "Non-imaging MM-wave FMCW sensor for pedestrian detection," *IEEE Sensors J.*, vol. 14, no. 4, pp. 1232–1237, Apr. 2014, doi: 10.1109/jsen.2013.2293534.

[20] S. Lee, Y.-J. Yoon, J.-E. Lee, and S.-C. Kim, "Human–vehicle classification using feature-based SVM in 77-GHz automotive FMCW radar," *IET Radar, Sonar Navigat.*, vol. 11, no. 10, pp. 1589–1596, Oct. 2017, doi: 10.1049/iet-rsn.2017.0126.

[21] M. Chen, S. Yang, X. Yi, and D. Wu, "Real-time 3D mapping using a 2D laser scanner and IMU-aided visual SLAM," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Okinawa, Japan, Jul. 2017, pp. 297–302, doi: 10.1109/rcar.2017.8311877.

[22] R. Ji, L.-Y. Duan, J. Chen, T. Huang, and W. Gao, "Mining compact bag-of-patterns for low bit rate mobile visual search," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3099–3113, Jul. 2014, doi: 10.1109/tip.2014.2324291.

[23] S. Zhao, L. Chen, H. Yao, Y. Zhang, and X. Sun, "Strategy for dynamic 3D depth data matching towards robust action retrieval," *Neurocomputing*, vol. 151, pp. 533–543, Mar. 2015, doi: 10.1016/j.neucom.2014.03.092.

[24] S. Zhao, H. Yao, Y. Zhang, Y. Wang, and S. Liu, "View-based 3D object retrieval via multi-modal graph learning," *Signal Process.*, vol. 112, pp. 110–118, Jul. 2015, doi: 10.1016/j.sigpro.2014.09.038.

[25] *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE Standard J3016_201806, 2018.

[26] K. P. Divakarla, A. Emadi, and S. Razavi, "A cognitive advanced driver assistance systems architecture for autonomous-capable electrified vehicles," *IEEE Trans. Transport. Electrific.*, vol. 5, no. 1, pp. 48–58, Mar. 2019, doi: 10.1109/tte.2018.2870819.

[27] Y. Nishimura, A. Fujita, A. Hiromori, H. Yamaguchi, T. Higashino, A. Suwa, H. Urayama, S. Takeshima, and M. Takai, "A study on behavior of autonomous vehicles cooperating with manually-driven vehicles," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. (PerCom)*, Kyoto, Japan, Mar. 2019, pp. 212–219, doi: 10.1109/percom.2019.8767390.

[28] Y. S. Song and S. K. Lee, "Analysis of periodic broadcast message for DSRC systems under high-density vehicle environments," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Jeju, South Korea, Oct. 2017, pp. 1008–1012.

[29] J. C. F. De Winter, D. Dodou, R. Happee, and Y. B. Eisma, "Will vehicle data be shared to address the how, where, and who of traffic accidents?" *Eur. J. Futures Res.*, vol. 7, no. 1, Dec. 2019, doi: 10.1186/s40309-019-0154-3.

[30] J. M. Lozano Domínguez and T. J. Mateo Sanguino, "Review on V2X, I2X, and P2X communications and their applications: A comprehensive analysis over time," *Sensors*, vol. 19, no. 12, p. 2756, Jun. 2019, doi: 10.3390/s19122756.

[31] N. Bonjorn, F. Foukalas, and P. Pop, "Enhanced 5G V2X services using sidelink device-to-device communications," in *Proc. 2018 17th Annu. Medit. Ad Hoc Netw. Workshop (Med-Hoc-Net)*, Capri, Italy, Jun. 2018, pp. 1–7.

[32] C. R. Storck and F. Duarte-Figueiredo, "5G V2X ecosystem providing entertainment on board using Mm wave communications," in *Proc. IEEE 10th Latin-Amer. Conf. Commun.*, Guadalajara, Mexico, Nov. 2018, pp. 14–16.

[33] S.-C. Choi, J.-H. Park, and J. Kim, "A networking framework for multiple-heterogeneous unmanned vehicles in FANETs," in *Proc. 11th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2019, pp. 13–15, doi: 10.1109/icufn.2019.8806105.

[34] F. Wulff, B. Schaufele, O. Sawade, D. Becker, B. Henke, and I. Radusch, "Early fusion of camera and Lidar for robust road detection based on U-Net FCN," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Changshu, China, Jun. 2018, pp. 1426–1431.

[35] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013, doi: 10.1177/0278364913491297.

[36] A. Asvadi, L. Garrote, C. Premebida, P. Peixoto, and U. J. Nunes, "Multimodal vehicle detection: Fusing 3D-LIDAR and color camera data," *Pattern Recognit. Lett.*, vol. 115, pp. 20–29, Nov. 2018, doi: 10.1016/j.patrec.2017.09.038.

[37] V. Ilic, M. Marijan, A. Mehmed, and M. Antlanger, "Development of sensor fusion based ADAS modules in virtual environments," in *Proc. Zooming Innov. Consum. Technol. Conf. (ZINC)*, Novi Sad, Serbia, May 2018, pp. 88–91.

[38] Y. Kang, H. Yin, and C. Berger, "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments," *IEEE Trans. Intell. Veh.*, vol. 4, no. 2, pp. 171–185, Jun. 2019, doi: 10.1109/tiv.2018.2886678.

[39] J. Oh, K.-S. Kim, M. Park, and S. Kim, "A comparative study on camera-radar calibration methods," in *Proc. 15th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Singapore, Nov. 2018, pp. 1057–1062.

[40] J. Kim, D. S. Han, and B. Senouci, "Radar and vision sensor fusion for object detection in autonomous vehicle surroundings," in *Proc. 10th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2018, pp. 76–78, doi: 10.1109/icufn.2018.8436959.

[41] G. Reina, A. Milella, and R. Rouveure, "Traversability analysis for off-road vehicles using stereo and radar data," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Seville, Spain, Mar. 2015, pp. 540–546, doi: 10.1109/icit.2015.7125155.

[42] T. Shimizu and K. Kobayashi, "Development of a person-searching algorithm using an omnidirectional camera and LiDAR for the Tsukuba challenge," in *Proc. 57th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, vol. 2, Nara, Japan, Sep. 2018, pp. 810–815.

[43] J. Khan, "Using ADAS sensors in implementation of novel automotive features for increased safety and guidance," in *Proc. 3rd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Noida, India, Feb. 2016, pp. 753–758, doi: 10.1109/spin.2016.7566798.

[44] X. Xiao-Zhu and X. Ya-Wei, "A comparative analysis of fusion rules for multi-sensor image fusion," in *Proc. 27th Chin. Control Decis. Conf. (CCDC)*, Qingdao, China, May 2015, pp. 3970–3973, doi: 10.1109/ccdc.2015.7162617.

[45] Z. Hang, L. Wangliang, L. Chuang, S. Linda, and W. Kaiyan, "Discussion on multi-sensor information fusion in greenhouse detection system," in *Proc. Int. Conf. Smart Grid Electr. Autom. (ICSGEA)*, Changsha, China, May 2017, pp. 64–67, doi: 10.1109/icsgea.2017.30.

[46] T.-E. Wu, C.-C. Tsai, and J.-I. Guo, "LiDAR/camera sensor fusion technology for pedestrian detection," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Kuala Lumpur, Malaysia, Dec. 2017, pp. 1675–1678.

[47] A. Cloninger, W. Czaja, and T. Doster, "Operator analysis and diffusion based embeddings for heterogeneous data fusion," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Quebec City, QC, Canada, Jul. 2014, pp. 1249–1252.

[48] Z. Zhong, S. Liu, M. Mathew, and A. Dubey, "Camera radar fusion for increased reliability in ADAS applications," *Electron. Imag.*, vol. 2018, no. 17, pp. 258-1–258-4, Jan. 2018.

[49] J. Kocic, N. Jovicic, and V. Drndarevic, "Sensors and sensor fusion in autonomous vehicles," in *Proc. 26th Telecommun. Forum (TELFOR)*, Belgrade, Serbia, Nov. 2018, pp. 1–4, doi: 10.1109/telfor.2018.8612054.

[50] M. Bouain, D. Berdjag, N. Fakhfakh, and R. B. Atitallah, "Multi-sensor fusion for obstacle detection and recognition: A belief-based approach," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Cambridge, U.K., Jul. 2018, pp. 1217–1224.

[51] J. Liu, S. Zhou, W. Liu, J. Zheng, H. Liu, and J. Li, "Tunable adaptive detection in Colocated MIMO radar," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 1080–1092, Feb. 2018, doi: 10.1109/tsp.2017.2778693.

[52] P. Ghamisi and N. Yokoya, "IMG2DSM: Height simulation from single imagery using conditional generative adversarial net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 794–798, May 2018, doi: 10.1109/lgrs.2018.2806945.

[53] F. Gao, Y. Yang, J. Wang, J. Sun, E. Yang, and H. Zhou, "A deep convolutional generative adversarial networks (DCGANs)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images," *Remote Sens.*, vol. 10, no. 6, p. 846, May 2018, doi: 10.3390/rs10060846.

[54] X. Shi, F. Zhou, S. Yang, Z. Zhang, and T. Su, "Automatic target recognition for synthetic aperture radar images based on super-resolution generative adversarial network and deep convolutional neural network," *Remote Sens.*, vol. 11, no. 2, p. 135, Jan. 2019, doi: 10.3390/rs11020135.

[55] G. Wang, M. Zhang, Y. Huang, L. Zhang, and F. Wang, "Robust two-dimensional spatial-variant map-drift algorithm for UAV SAR autofocusing," *Remote Sens.*, vol. 11, no. 3, p. 340, Feb. 2019, doi: 10.3390/rs11030340.

[56] R. Komissarov, V. Kozlov, D. Filonov, and P. Ginzburg, "Partially coherent radar unties range resolution from bandwidth limitations," *Nature Commun.*, vol. 10, no. 1, Dec. 2019, Art. no. 1423, doi: 10.1038/s41467-019-09380-x.

[57] F. Rosique, P. J. Navarro, C. Fernàndez, and A. Padilla, "A systematic review of perception system and simulators for autonomous vehicles research," *Sensors*, vol. 19, no. 3, p. 648, Feb. 2019, doi: 10.3390/s19030648.

[58] J. Sun, P. Wang, Z. Qin, and H. Qiao, "Overview of camera calibration for computer vision," in *Proc. 11th World Congr. Intell. Control Autom.*, Jun. 2014, pp. 86–92, doi: 10.1109/wcica.2014.7052692.

[59] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000, doi: 10.1109/34.888718.

[60] S. Sim, J. Sock, and K. Kwak, "Indirect correspondence-based robust extrinsic calibration of LiDAR and camera," *Sensors*, vol. 16, no. 6, p. 933, Jun. 2016, doi: 10.3390/s16060933.

[61] Y. Zhou, L. Zhang, C. Xing, P. Xie, and Y. Cao, "Target three-dimensional reconstruction from the multi-view radar image sequence," *IEEE Access*, vol. 7, pp. 36722–36735, 2019, doi: 10.1109/access.2019.2905130.

[62] V. De Silva, J. Roche, and A. Kondoz, "Robust fusion of LiDAR and wide-angle camera data for autonomous mobile robots," *Sensors*, vol. 18, no. 8, p. 2730, Aug. 2018, doi: 10.3390/s18082730.

[63] Z. Pusztai, I. Eichhardt, and L. Hajder, "Accurate calibration of multi-LiDAR-multi-camera systems," *Sensors*, vol. 18, no. 7, p. 2139, Jul. 2018, doi: 10.3390/s18072139.

[64] P. Wei, L. Cagle, T. Reza, J. Ball, and J. Gafford, "LiDAR and camera detection fusion in a real-time industrial multi-sensor collision avoidance system," *Electronics*, vol. 7, no. 6, p. 84, May 2018, doi: 10.3390/electronics7060084.

[65] Z. Han, J. Liang, and J. Li, "Design of intelligent road recognition and warning system for vehicles based on binocular vision," *IEEE Access*, vol. 6, pp. 62880–62889, 2018, doi: 10.1109/access.2018.2876702.

[66] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge, and W. Brendel, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2019, pp. 1–23.

[67] C. Berger, "Large-scale evaluation of an active safety algorithm with EuroNCAP and us NCAP scenarios in a virtual test environment–an industrial case study," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst. (ITSC)*, Las Palmas, Spain, Sep. 2015, pp. 2280–2286, doi: 10.1109/ITSC.2015.368.

[68] S. Wang, X. Dai, N. Xu, and P. Zhang, "A review of environmetal sensing technology for autonomous vehicle," *J. Changchun Univ. Sci. Technol. (Natural Sci. Ed.)*, vol. 40, pp. 1672–9870, Feb. 2017.

[69] J. Duan, K. Zheng, and J. Zhou, "multilayer Lidar's environmental sensing in autonomous vehicle," *J. Beijing Univ. Technol.*, vol. 40, pp. 1891–1898,Dec. 2014, doi: 0254-0037(2014)12-1891-08.

[70] J. Zhao, H. Xu, H. Liu, J. Wu, Y. Zheng, and D. Wu, "Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors," *Transp. Res. C, Emerg. Technol.*, vol. 100, pp. 68–87, Mar. 2019, doi: 10.1016/j.trc.2019.01.007.

[71] H. Wang, B. Wang, B. Liu, X. Meng, and G. Yang, "Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle," *Robot. Auto. Syst.*, vol. 88, pp. 71–78, Feb. 2017, doi: 10.1016/j.robot.2016.11.014.

[72] M. A. Savelonas, I. Pratikakis, T. Theoharis, G. Thanellas, F. Abad, and R. Bendahan, "Spatially sensitive statistical shape analysis for pedestrian recognition from LIDAR data," *Comput. Vis. Image Understand.*, vol. 171, pp. 1–9, Jun. 2018, doi: 10.1016/j.cviu.2018.06.001.

[73] C. Goodin, D. Carruth, M. Doude, and C. Hudson, "Predicting the influence of rain on LIDAR in ADAS," *Electronics*, vol. 8, no. 1, p. 89, Jan. 2019, doi: 10.3390/electronics8010089.

[74] M. Aldibaja, R. Yanase, T. H. Kim, A. Kuramoto, K. Yoneda, and N. Suganuma, "Accurate elevation maps based graph-slam framework for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Paris, France, Jun. 2019, pp. 1254–1261, doi: 10.1109/ivs.2019.8814007.

[75] N. Baras, G. Nantzios, D. Ziouzios, and M. Dasygenis, "Autonomous obstacle avoidance vehicle using LIDAR and an embedded system," in *Proc. 8th Int. Conf. Mod. Circuits Syst. Technol. (MOCAST)*, Thessaloniki, Greece, May 2019, pp. 1–4, doi: 10.1109/mocast.2019.8742065.

[76] S. Han, X. Wang, L. Xu, H. Sun, and N. Zheng, "Frontal object perception for Intelligent Vehicles based on radar and camera fusion," in *Proc. 35th Chin. Control Conf. (CCC)*, Chengdu, China, Jul. 2016, pp. 4003–4008.

[77] S. B. Hong, C. M. Kang, S.-H. Lee, and C. C. Chung, "Multi-rate vehicle side slip angle estimation using low-cost GPS/IMU," in *Proc. 17th Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2017, pp. 35–40, doi: 10.23919/iccas.2017.8204419.

[78] Wahyudi, M. S. Listiyana, Sudjadi, and Ngatelan, "Tracking object based on GPS and IMU sensor," in *Proc. 5th Int. Conf. Inf. Technol., Comput., Electr. Eng. (ICITACEE)*, Semarang, Indonesia, Sep. 2018, pp. 214–218, doi: 10.1109/icitacee.2018.8576928.

[79] Z. Zhang, H. Wang, and W. Chen, "A real-time visual-inertial mapping and localization method by fusing unstable GPS," in *Proc. 13th World Congr. Intell. Control Autom. (WCICA)*, Changsha, China, Jul. 2018, pp. 1397–1402.

[80] L. Caltagirone, M. Bellone, L. Svensson, and M. Wahde, "LIDAR-based driving path generation using fully convolutional neural networks," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Yokohama, Japan, Oct. 2017, pp. 1–6, doi: 10.1109/itsc.2017.8317618.

[81] M. Chen, X. Zhan, J. Tu, and M. Liu, "Vehicle-localization-based and DSRC-based autonomous vehicle rear-end collision avoidance concerning measurement uncertainties," *IEEJ Trans. Elec. Electron. Eng.*, vol. 14, no. 9, pp. 1348–1358, Sep. 2019, doi: 10.1002/tee.22936.

[82] Y. Cui, H. Xu, J. Wu, Y. Sun, and J. Zhao, "Automatic vehicle tracking with roadside LiDAR data for the connected-vehicles system," *IEEE Intell. Syst.*, vol. 34, no. 3, pp. 44–51, May 2019, doi: 10.1109/mis.2019.2918115.

[83] M. Fujinami and Y. Mizukoshi, "Study on performance of networked dead reckoning for real vehicles," in *Proc. IEEE/ACM 21st Int. Symp. Distrib. Simul. Real Time Appl. (DS-RT)*, Rome, Italy, Oct. 2017, pp. 1–8, doi: 10.1109/distra.2017.8167681.

[84] C.-M. Tsai, Y.-H. Lai, J.-W. Perng, I.-F. Tsui, and Y.-J. Chung, "Design and application of an autonomous surface vehicle with an AI-based sensing capability," in *Proc. IEEE Underwater Technol. (UT)*, Kaohsiung, Taiwan, Apr. 2019, pp. 1–4, doi: 10.1109/ut.2019.8734350.

[85] J.-J. Kim, S.-H. Cha, M. Ryu, and M. Jo, "Pre-training framework for improving learning speed of reinforcement learning based autonomous vehicles," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Jan. 2019, pp. 1–2, doi: 10.23919/elinfocom.2019.8706441.

[86] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1697–1716, Aug. 2019, doi: 10.1109/jproc.2019.2915983.

[87] J. Castorena, U. S. Kamilov, and P. T. Boufounos, "Autocalibration of lidar and optical cameras via edge alignment," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 2862–2866, doi: 10.1109/icassp.2016.7472200.

[88] L. Caltagirone, M. Bellone, L. Svensson, and M. Wahde, "LIDAR–camera fusion for road detection using fully convolutional neural networks," *Robot. Auto. Syst.*, vol. 111, pp. 125–131, Jan. 2019, doi: 10.1016/j.robot.2018.11.002.

[89] K. Ma, H. Zhang, R. Wang, and Z. Zhang, "Target tracking system for multi-sensor data fusion," in *Proc. IEEE 2nd Inf. Technol. Netw., Electron. Autom. Control Conf. (ITNEC)*, Chengdu, China, Jan. 2018, pp. 1768–1772, doi: 10.1109/ITNEC.2017.8285099.

[90] A.-M. Hellmund, S. Wirges, O. S. Tas, C. Bandera, and N. O. Salscheider, "Robot operating system: A modular software framework for automated driving," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, Nov. 2016, pp. 1564–1570, doi: 10.1109/itsc.2016.7795766.

[91] H. Kim and I. Lee, "Localization of a car based on multi-sensor fusion," *Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. XLII-1, pp. 247–250, Sep. 2018, doi: 10.5194/isprs-archives-XLII-1-247-2018.

[92] R. K. Bhadani, J. Sprinkle, and M. Bunting, "The CAT vehicle testbed: A simulator with hardware in the loop for autonomous vehicle applications," *Electron. Proc. Theor. Comput. Sci.*, vol. 269, pp. 32–47, Apr. 2018, doi: 10.4204/eptcs.269.4.

[93] Y. Zhou and M. Omar, "Pixel-level fusion for infrared and visible acquisitions," *Int. J. Optomechtron.*, vol. 3, no. 1, pp. 41–53, Feb. 2009, doi: 10.1080/15599610902717835.
[94] C. Kwan, B. Chou, L.-Y.-M. Kwan, J. Larkin, B. Ayhan, J. F. Bell, and H. Kerner, "Demosaicing enhancement using pixel-level fusion," *Signal, Image Video Process.*, vol. 12, no. 4, pp. 749–756, May 2018, doi: 10.1007/s11760-017-1216-2.
[95] F. Harrer, F. Pfeiffer, A. Loffler, T. Gisder, and E. Biebl, "Synthetic aperture radar algorithm for a global amplitude map," in *Proc. 14th Workshop Positioning, Navigat. Commun. (WPNC)*, Bremen, Germany, Oct. 2017, pp. 1–6, doi: 10.1109/wpnc.2017.8250080.
[96] R. Weston, S. Cen, P. Newman, and I. Posner, "Probably unknown: Deep inverse sensor modelling radar," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, May 2019, pp. 5446–5453.
[97] B. Kim, D. Kim, S. Park, Y. Jung, and K. Yi, "Automated complex urban driving based on enhanced environment representation with GPS/map, radar, Lidar and vision," *IFAC-PapersOnLine*, vol. 49, no. 11, pp. 190–195, 2016, doi: 10.1016/j.ifacol.2016.08.029.
[98] V. Lekic and Z. Babic, "Automotive radar and camera fusion using generative adversarial networks," *Comput. Vis. Image Understand.*, vol. 184, pp. 1–8, Jul. 2019, doi: 10.1016/j.cviu.2019.04.002.
[99] B. Yektakhah and K. Sarabandi, "All-directions through-the-wall imaging using a small number of moving omnidirectional bi-static FMCW transceivers," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2618–2627, May 2019, doi: 10.1109/tgrs.2018.2875695.
[100] Z. Ouyang, Y. Liu, C. Zhang, and J. Niu, "A cGANs-based scene reconstruction model using Lidar point cloud," in *Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl. IEEE Int. Conf. Ubiquitous Comput. Commun. (ISPA/IUCC)*, Guangzhou, China, Dec. 2017, pp. 1107–1114, doi: 10.1109/ispa/iucc.2017.00167.
[101] P. Yin, L. Xu, Z. Liu, L. Li, H. Salman, Y. He, W. Xu, H. Wang, and H. Choset, "Stabilize an unsupervised feature learning for LiDAR-based place recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Madrid, Spain, Oct. 2018, pp. 1162–1167, doi: 10.1109/iros.2018.8593562.
[102] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, "Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 4224–4231, Sep. 2018, doi: 10.1109/tii.2018.2822828.
[103] M. Li and Y. Dong, "Review on technology of pixel-level image fusion," in *Proc. 2nd Int. Conf. Meas., Inf. Control (ICMIC)*, vol. 1, Harbin, China, Aug. 2013, pp. 341–344, doi: 10.1109/mic.2013.6757979.
[104] M. Serfling, O. Loehlein, R. Schweiger, and K. Dietmayer, "Camera and imaging radar feature level sensorfusion for night vision pedestrian recognition," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2009, pp. 597–603, doi: 10.1109/ivs.2009.5164345.
[105] Y. Song, J. Hu, Y. Dai, T. Jin, and Z. Zhou, "Estimation and mitigation of time-variant RFI in low-frequency ultra-wideband radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 409–413, Mar. 2018, doi: 10.1109/lgrs.2018.2790910.
[106] M. Wang, Y. D. Zhang, and G. Cui, "Human motion recognition exploiting radar with stacked recurrent neural network," *Digit. Signal Process.*, vol. 87, pp. 125–131, Apr. 2019, doi: 10.1016/j.dsp.2019.01.013.
[107] F. Zhang, D. Clarke, and A. Knoll, "Vehicle detection based on LiDAR and camera fusion," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Qingdao, China, Oct. 2014, pp. 1620–1625, doi: 10.1109/itsc.2014.6957925.
[108] X. Han, J. Lu, Y. Tai, and C. Zhao, "A real-time LIDAR and vision based pedestrian detection system for unmanned ground vehicles," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Kuala Lumpur, Malaysia, Nov. 2015, pp. 635–639.
[109] L. Xiao, R. Wang, B. Dai, Y. Fang, D. Liu, and T. Wu, "Hybrid conditional random field based camera-LIDAR fusion for road detection," *Inf. Sci.*, vol. 432, pp. 543–558, Mar. 2018, doi: 10.1016/j.ins.2017.04.048.
[110] S. D. Thepade, R. K. Bhondave, and A. Mishra, "Comparing score level and feature level fusion in Multimodal biometric identification using iris and Palmprint traits with fractional transformed energy content," in *Proc. Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Jabalpur, India, Dec. 2015, pp. 306–311, doi: 10.1109/cicn.2015.68.
[111] D. Wu, H. Lu, C. Qiu, and X. Li, "A feature level fusion target recognition algorithm based on dynamic fuzzy integral," in *Proc. 12th Int. Conf. Signal Process. (ICSP)*, Hangzhou, China, Oct. 2014, pp. 1367–1371, doi: 10.1109/icosp.2014.7015223.
[112] F. J. Botha, C. E. Van Daalen, and J. Treurnicht, "Data fusion of radar and stereo vision for detection and tracking of moving objects," in *Proc. Pattern Recognit. Assoc. South Africa Robot. Mechatronics Int. Conf. (PRASA-RobMech)*, Stellenbosch, South Africa, Nov. 2016, pp. 1–7, doi: 10.1109/robomech.2016.7813156.
[113] D. Petković, "Adaptive neuro-fuzzy fusion of sensor data," *Infr. Phys. Technol.*, vol. 67, pp. 222–228, Nov. 2014, doi: 10.1016/j.infrared.2014.07.031.
[114] C. Wang, Q. Sun, Z. Li, H. Zhang, and K. Ruan, "Cognitive competence improvement for autonomous vehicles: A lane change identification model for distant preceding vehicles," *IEEE Access*, vol. 7, pp. 83229–83242, 2019, doi: 10.1109/access.2019.2924557.
[115] D. Ciucci, G. Wang, S. Mitra, and W. Z. Wu, "Rough sets and knowledge technology," in *Proc. 10th Int. Conf., RSKT Int. Joint Conf. Rough Sets (IJCRS)*, vol. 9436, Tianjin, China, Nov. 2015.
[116] T.-Y. Wang and J.-Y. Wu, "Further results on decision fusion in censoring sensor networks: An unknown network size," in *Proc. IEEE 12th Int. Workshop Signal Process. Adv. Wireless Commun.*, San Francisco, CA, USA, Jun. 2011, pp. 141–145, doi: 10.1109/spawc.2011.5990381.
[117] L. Sun, Y. Zhang, Z. Fu, G. Zheng, Z. He, and J. Pu, "An approach to multi-sensor decision fusion based on the improved Jousselme evidence distance," in *Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS)*. Hangzhou, China: Hangzhou Dianzi Univ, Oct. 2018, pp. 189–193.
[118] S. Yin, K. Huo, and Y. Liu, "Multi-sensor fusion recognition method based on improved D-S evidence theory," in *Proc. Int. Conf. Inf. Commun. Technol.*, Nanjing, China, 2014, pp. 1–7, doi: 10.1049/cp.2014.0595.
[119] R. Zakaria, O. Yong Sheng, K. Wern, S. Shamshirband, D. Petković, and N. T. Pavlovic, "Adaptive Neuro-fuzzy evaluation of the tapered plastic multimode fiber-based sensor performance with and without silver thin film for different concentrations of calcium hypochlorite," *IEEE Sensors J.*, vol. 14, no. 10, pp. 3579–3584, Oct. 2014, doi: 10.1109/jsen.2014.2329333.
[120] S. M. C. Noh, S. Shamshirband, D. Petković, R. Penny, and R. Zakaria, "Adaptive Neuro-fuzzy appraisal of Plasmonic studies on morphology of deposited silver thin films having different thicknesses," *Plasmonics*, vol. 9, no. 5, pp. 1189–1196, Oct. 2014, doi: 10.1007/s11468-014-9730-3.
[121] B. Lee, S. Hong, H. Lee, and E. Kim, "Gait recognition system using decision-level fusion," in *Proc. 5th IEEE Conf. Ind. Electron. Appl.*, Taichung, Taiwan, Jun. 2010, pp. 313–316, doi: 10.1109/iciea.2010.5516856.
[122] O. Yahia, R. Guida, and P. Iervolino, "Weights based decision level data fusion of landsat-8 and sentinel-L for soil moisture content estimation," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, Jul. 2018, pp. 8078–8081.
[123] A. Salentinig and P. Gamba, "Multi-scale decision level data fusion by means of spatial regularization and image weighting," in *Proc. Joint Urban Remote Sens. Event (JURSE)*, Mar. 2017, pp. 1–4, doi: 10.1109/jurse.2017.7924564.
[124] X. Wang, L. Xu, H. Sun, J. Xin, and N. Zheng, "On-road vehicle detection and tracking using MMW radar and Monovision fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 2075–2084, Jul. 2016, doi: 10.1109/tits.2016.2533542.
[125] J. Ćesić, I. Marković, I. Cvišić, and I. Petrović, "Radar and stereo vision fusion for multitarget tracking on the special Euclidean group," *Robot. Auto. Syst.*, vol. 83, pp. 338–348, Sep. 2016, doi: 10.1016/j.robot.2016.05.001.
[126] S. Graebenitz, M. Mertens, and F. Belfiori, "Sequential data fusion applied to a distributed radar, acoustic and visual sensor network," in *Proc. Sensor Data Fusion, Trends, Solutions, Appl. (SDF)*, Bonn, Germany, Oct. 2017, pp. 1–6.
[127] K.-E. Kim, C.-J. Lee, D.-S. Pae, and M.-T. Lim, "Sensor fusion for vehicle tracking with camera and radar sensor," in *Proc. 17th Int. Conf. Control, Autom. Syst. (ICCAS)*, Jeju, South Korea, Oct. 2017, pp. 1075–1077, doi: 10.23919/iccas.2017.8204375.
[128] K. Jo, M. Lee, J. Kim, and M. Sunwoo, "Tracking and behavior reasoning of moving vehicles based on roadway geometry constraints," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 460–476, Feb. 2017, doi: 10.1109/tits.2016.2605163.
[129] J.-G. Wang, S. J. Chen, L.-B. Zhou, K.-W. Wan, and W.-Y. Yau, "Vehicle detection and width estimation in rain by fusing radar and vision," in *Proc. 15th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Singapore, Nov. 2018, pp. 1063–1068.

[130] H. Jha, V. Lodhi, and D. Chakravarty, "Object detection and identification using vision and radar data fusion system for ground-based navigation," in *Proc. 6th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Noida, India, Mar. 2019, pp. 590–593.

[131] M. Dimitrievski, P. Veelaert, and W. Philips, "Behavioral pedestrian tracking using a camera and LiDAR sensors on a moving vehicle," *Sensors*, vol. 19, no. 2, p. 391, Jan. 2019, doi: 10.3390/s19020391.

[132] F. Zhou and F. De La Torre, "Generalized time warping for multi-modal alignment of human motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1282–1289, doi: 10.1109/cvpr.2012.6247812.

[133] A. Gopalakrishnan, A. Mali, D. Kifer, C. L. Giles, and A. G. Ororbia, "A neural temporal model for human motion prediction," in *Computer Science-Computer Vision And Pattern Recognition*. New York, NY, USA: Cornell Univ., 2018.

[134] C. Xu, J. He, X. Zhang, H. Cai, S. Duan, P.-H. Tseng, and C. Li, "Recurrent transformation of prior knowledge based model for human motion recognition," *Comput. Intell. Neurosci.*, vol. 2018, Jan. 2018, Art. no. 4160652, doi: 10.1155/2018/4160652.

[135] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 583–595, Jun. 2011, doi: 10.1109/tits.2011.2113340.

[136] R. Y. R. Yue, Z. L. Z. Ling, L. Z. Li Zhe, Y. W. Y. Wei, L. Y. Li Yinong, W. K. W. Ke, L. Y. Li Yusheng, and X. Z. X. Zhoubing, "The research on the vehicle collision avoidance control based on vehicle motion estimation," in *Proc. IET Int. Conf. Intell. Connected Vehicles (ICV)*, Chongqing, China, 2016, p. 1, doi: 10.1049/cp.2016.1154.

[137] W. Jeon, A. Zemouche, and R. Rajamani, "Tracking of vehicle motion on highways and urban roads using a nonlinear observer," *IEEE/ASME Trans. Mechatron.*, vol. 24, no. 2, pp. 644–655, Apr. 2019, doi: 10.1109/tmech.2019.2892700.

[138] P. G. Thomasson and C. A. Woolsey, "Vehicle motion in currents," *IEEE J. Ocean. Eng.*, vol. 38, no. 2, pp. 226–242, Apr. 2013, doi: 10.1109/joe.2013.2238054.

[139] T. Yamazato, M. Kinoshita, S. Arai, E. Souke, T. Yendo, T. Fujii, K. Kamakura, and H. Okada, "Vehicle motion and pixel illumination modeling for image sensor based visible light communication," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 9, pp. 1793–1805, Sep. 2015, doi: 10.1109/jsac.2015.2432511.

[140] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *Proc. 11th Int. Conf. Inf. Fusion*, Cologne, Germany, Jun./Jul. 2008, pp. 1–6.

[141] A. R. Chowdhury, B. Prasad, V. Vishwanathan, R. Kumar, and S. K. Panda, "Finding an operating region for a bio-inspired robotic fish underwater vehicle in the Lighthill framework," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Shenzhen, China, Dec. 2013, pp. 854–860, doi: 10.1109/robio.2013.6739569.

[142] Y. Zhang, W. Wang, R. Bonatti, D. Maturana, and S. Scherer, "Integrating kinematics and environment context into deep inverse reinforcement learning for predicting off-road vehicle trajectories," *Comput. Sci.*, vol. 16, pp. 1–12, Oct. 2018.

[143] C.-L. Hwang, B.-L. Chen, H.-T. Syu, C.-K. Wang, and M. Karkoub, "Humanoid robot's visual imitation of 3-D motion of a human subject using neural-network-based inverse kinematics," *IEEE Syst. J.*, vol. 10, no. 2, pp. 685–696, Jun. 2016, doi: 10.1109/jsyst.2014.2343236.

[144] X. Lu and D. Li, "Research on target detection and tracking system of rescue robot," in *Proc. Chin. Autom. Congr. (CAC)*, Jinan, China, Oct. 2017, pp. 5–9.

[145] B.-L. Wang, C.-T. King, and H.-K. Chu, "A semi-automatic video labeling tool for autonomous driving based on multi-object detector and tracker," in *Proc. 6th Int. Symp. Computing Netw. (CANDAR)*, Takayama, Japan, Nov. 2018, pp. 201–206.

[146] S. Wang, Q. Bao, and Z. Chen, "Refined PHD filter for multi-target tracking under low detection probability," *Sensors*, vol. 19, no. 13, p. 2842, Jun. 2019, doi: 10.3390/s19132842.

[147] L. Wang, W. Zhu, Y. Zhang, Q. Liao, and J. Tang, "Multi-target detection and adaptive waveform design for cognitive MIMO radar," *IEEE Sensors J.*, vol. 18, no. 24, pp. 9962–9970, Dec. 2018, doi: 10.1109/jsen.2018.2873103.

[148] Z. Tian, Y. Cai, S. Huang, F. Hu, Y. Li, and M. Cen, "Vehicle tracking system for intelligent and connected vehicle based on radar and V2V fusion," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Shenyang, China, Jun. 2018, pp. 6598–6603, doi: 10.1109/ccdc.2018.8408291.

[149] A. Gavriilidis, D. Muller, S. Muller-Schneiders, J. Velten, and A. Kummert, "Sensor system blockage detection for night time headlight control based on camera and radar sensor information," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst.*, Anchorage, AK, USA, Sep. 2012, pp. 78–83, doi: 10.1109/itsc.2012.6338854.

[150] B. Lagos-Álvarez, L. Padilla, J. Mateu, and G. Ferreira, "A Kalman filter method for estimation and prediction of space–time data with an autoregressive structure," *J. Stat. Planning Inference*, vol. 203, pp. 117–130, Dec. 2019, doi: 10.1016/j.jspi.2019.03.005.

[151] K. Xiong, H. Zhang, and C. Chan, "Performance evaluation of UKF-based nonlinear filtering," *Automatica*, vol. 42, no. 2, pp. 261–270, Feb. 2006, doi: 10.1016/j.automatica.2005.10.004.

[152] M. Krieg, "Joint multi-sensor kinematic and attribute tracking using Bayesian belief networks," in *Proc. 6th Int. Conf. Inf. Fusion*, 2003, pp. 1–15:15-8.

[153] S. Park, J. Pil Hwang, E. Kim, and H.-J. Kang, "Vehicle tracking using a microwave radar for situation awareness," *Control Eng. Pract.*, vol. 18, no. 4, pp. 383–395, Apr. 2010, doi: 10.1016/j.conengprac.2009.12.006.

[154] K. Lu, K. C. Chang, and R. Zhou, "The exact algorithm for multi-sensor asynchronous track-to-track fusion," in *Proc. 18th Int. Conf. Inf. Fusion*, Washington, DC, USA, 2015, pp. 886–892.

[155] E. Santos and S. Haykin, "Data association for target tracking rooted in maximum-likelihood values," *IET Radar, Sonar Navigat.*, vol. 12, no. 2, pp. 195–201, Feb. 2018, doi: 10.1049/iet-rsn.2017.0162.

[156] S. He, H. S. Shin, and A. Tsourdos, "Joint probabilistic data association filter with unknown detection probability and clutter rate," in *Proc. IEEE Int. Conf. Multisens. Fusion Integr. Intell. Syst.*, pp. 559–564, Daegu, South Korea, Nov. 2017, doi: 10.3390/s18010269.

[157] T. Yang and P. G. Mehta, "Probabilistic data association-feedback particle filter for multiple target tracking applications," *J. Dyn. Syst., Meas., Control*, vol. 140, no. 3, Mar. 2018, doi: 10.1115/1.4037781.

[158] H. Blom, "Joint probabilistic data association avoiding track coalescence," in *Proc. IEE Colloq. 'Algorithms Target Tracking'*, London, U.K., 1995, p. 1, doi: 10.1049/ic:19950667.

[159] D. Musicki and R. Evans, "Joint integrated probabilistic data association—JIPDA," in *Proc. 5th Int. Conf. Inf. Fusion*, Annapolis, MD, USA, Jun. 2003, pp. 1120–1125, doi: 10.1109/icif.2002.1020938.

[160] H. A. P. Blom and E. A. Bloem, "Probabilistic data association avoiding track coalescence," *IEEE Trans. Autom. Control*, vol. 45, no. 2, pp. 247–259, Feb. 2000, doi: 10.1109/9.839947.

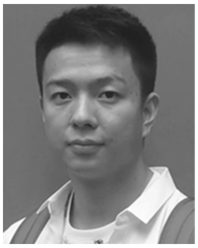
[161] Y. Zhu, J. Wang, S. Liang, and J. Wang, "Covariance control joint integrated probabilistic data association filter for multi-target tracking," *IET Radar, Sonar Navigat.*, vol. 13, no. 4, pp. 584–592, Apr. 2019, doi: 10.1049/iet-rsn.2018.5142.

[162] R. Mahler, "Multitarget bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003, doi: 10.1109/taes.2003.1261119.

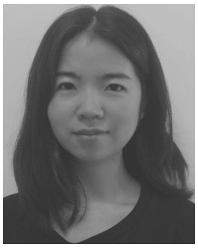
[163] L. Yulianti, B. Riyanto, and P. A. Setijadi, "Random finite sets (RFSs) approach in particle-based multi-target multisensor Bayesian filtering," in *Proc. 7th Int. Conf. Telecommun. Syst., Services, Appl. (TSSA)*, pp. 294–299, Oct. 2012, Bali, Indonesia, doi: 10.1109/tssa.2012.6366071.

ZHANGJING WANG received the B.E., M.E., and Ph.D. degrees from the University of Electronic Science and Technology of China.

Since 1999, she has been teaching and researching with the University of Electronic Science and Technology of China. Her main research interests include radar signal processing, intelligent sensing, software-based radar, heterogeneous computing, and open software architecture for embedded systems.



YU WU was born in Tongjiang, Sichuan, China, in 1995. He received the B.E. degree in communication engineering from Northeast Forestry University, in 2018. Since 2018, he has been with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, where he has been involving in radar signal processing and multisensor fusion. His research interests include signal process of intelligent radar and information fusion.



QINGQING NIU was born in Lanzhou, Gansu, China, in 1996. She received the B.S. degree from the Xi'an University of Posts and Telecommunications, in 2018. She is currently pursuing the M.S. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. Her research interests include image processing and machine learning.