Multivehicle Multisensor Occupancy Grid Maps (MVMS-OGM) for Autonomous Driving

Xinhu Zheng, *Member, IEEE*, Yuru Li, *Student Member, IEEE*, Dongliang Duan, *Senior Member, IEEE*, Liuqing Yang, *Fellow, IEEE*, Chen Chen, *Senior Member, IEEE*, and Xiang Cheng, *Fellow, IEEE*

***Abstract*—In autonomous driving, environment perception is the fundamental task for intelligent vehicles which provides the necessary environment information for other applications. The main issues in existing environment perception can be categorized into two aspects. On the one hand, all sensors are prone to measurement errors and failures. On the other hand, in complex driving environments, vehicles may encounter a variety of blind spots caused by vehicle occlusions, overlaps, and harsh weather conditions, which will cause sensors to experience low-quality data or to miss crucial environmental information. To cope with these issues, a multivehicle and multisensor (MVMS) cooperative perception method is presented to construct the occupancy grid map (OGM) of vehicles in a global view for the environment perception of autonomous driving. Distinct from existing environment perception methods, our proposed MVMS-OGM not only provides continuous geographical information but also captures and fuses continuous information with soft occupancy probabilities, resulting in more comprehensive and raw environmental information. Simulations and real-world experiments demonstrate that the proposed approach not only expands the perception range in comparison with single-vehicle sensing but also better captures the uncertainty of sensor data by fusing the occupancy probabilities with soft information.**

***Index Terms*—Autonomous driving, cooperative sensing, multisensor data fusion, occupancy grid map (OGM) fusion, occupancy probability assignment.**

Manuscript received 30 December 2021; revised 14 April 2022; accepted 16 June 2022. Date of publication 1 July 2022; date of current version 7 November 2022. This work was supported in part the Ministry National Key Research and Development Project under Grant 2020AAA0108101; in part by the National Natural Science Foundation of China under Grant 62125101; and in part by the National Science Foundation under Grant CNS-1932139. *(Corresponding author: Xiang Cheng.)*

Xinhu Zheng is with the Intelligent Transportation Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511458, China, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR, China (e-mail: xinhuzheng@ust.hk).

Yuru Li, Chen Chen, and Xiang Cheng are with the State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics, Peking University, Beijing 100871, China (e-mail: l.yuru@pku.edu.cn; c.chen@pku.edu.cn; xiangcheng@pku.edu.cn).

Dongliang Duan is with the Department of Electrical and Computer Engineering, University of Wyoming, Laramie, WY 82071 USA (e-mail: dduan@uwyo.edu).

Liuqing Yang is with the Internet of Things Thrust and Intelligent Transportation Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511458, China, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR, China (e-mail: lqyang@ust.hk).

Digital Object Identifier 10.1109/JIOT.2022.3187827

I. Introduction

As THE core component of the next-generation transportation system and automotive technology, unmanned autonomous vehicles integrate a large variety of modules to realize different tasks, such as environment perception, intelligent decision making, autonomous driving, etc., among which, environment perception is a key step to ensure the safety of vehicle driving. The perception module uses various types of vehicular sensors, such as LiDAR, camera, and millimeter-wave radar to achieve the environment perception by detecting the surrounding environment in real time, providing effective environmental information and timely warning of potential danger. The environment perception can also allow the vehicle to make appropriate decisions based on the driving environment and provide guidance for the path planning. In typical driving scenarios, perception tasks mainly include pedestrian detection, vehicle detection, safe passing region detection, traffic sign recognition, etc.

In most existing works, the perception task is carried out by some specific sensors, such as LiDAR [1], camera [2], and millimeter-wave radar [3]. However, different sensors have different specifications in terms of the sensing range or suitable environments. Not any single type of sensor can provide reliable data in all scenarios at all times. In addition, the dynamics and complexity of the driving environment also pose many significant challenges to the perception tasks. In complicated traffic scenarios, there is a variety of blind spots that vehicles may encounter, which may be caused by vehicles' occlusions and overlaps, or harsh weather conditions, such as sandstorms, rainstorms, etc. In these cases, the vehicular sensors are prone to failure and leading to great uncertainty in the perception result.

Multisensor fusion stands for combining information from several sensors to form a more comprehensive environment or integrating information from different types of sensors to generate a more reliable description of the environment. In the environment perception, camera and LiDAR are often considered to be the two most commonly used sensors. Without loss of generality, we are considering these two types of data here in this article. The method proposed in this article can also be applicable to fuse other types of sensor data. Camera data can provide rich color information and work well on many perception tasks with the assistance of deep learning algorithms and computer vision techniques (see [2], [4]). However, the camera data is sensitive to the environment such as the lighting

condition and lacks depth information. In contrast, the point cloud data collected by LiDAR contains depth information and is more robust and accurate in ranging tasks. However, given the relatively sparse nature of its point density, it falls short in capturing the features of small objects. Therefore, combining the information from the LiDAR and camera is expected to alleviate their respective limitations and facilitate environment perception with more reliable and consistent results (see [5], [6]).

However, most existing research has only been validated in simplified environments and often with oversimplified assumptions. When they are applied to real-world driving scenarios, various issues may arise. For instance, the main reason for the first fatal Tesla car accident that occurred in Florida on 7 May 2016 [7] is because the onboard full self-driving system relying solely on the camera-based environment perception for object detection, and it is unable to detect the stationary truck in front due to the abnormal light reflection. Although the fusion of data from multiple sensors could potentially improve the perception accuracy of a single vehicle to a certain extent, it cannot completely solve the problem caused by blind spots. When the vehicle is in a poor perception perspective or encounters bad weather conditions, most of the perception area may be occluded, and the data collected by the sensor may be corrupted by significant noise. Obviously, such results would have a fatal impact.

Researchers have attempted to fuse the data from different perspectives to mitigate these issues, via multivehicle cooperative perception. In vehicular networks, vehicles can exchange sensing information via V2V and/or V2I communications to achieve cooperative perception [8]–[12]. Multivehicle cooperative perception can expand the perception range of individual vehicles, improve the data accuracy of the perceived overlapping area, compensate for the blind zones of individual vehicles, and reduce the risk of potential dangers with more comprehensive and reliable driving environment information by supplementing the perception results of multiple perspectives.

To this end, in this article, a multivehicle multisensor (MVMS) cooperative sensing environment perception approach is developed by combining the two data fusion techniques and present a more informative environment for driving scenarios through the occupancy grid map (OGM). The OGM divides the area of interest into fine grids and provides the environmental information about the target area by giving the occupancy status of all the grids. Here, we also adopt OGM as the representation of the driving environment information because it can be used as a universal representation for different sensor types and fits the multisensor fusion aspect of this work naturally. Furthermore, while most existing OGMs represent the occupancy status of grids in a discrete manner (either occupied or unoccupied), in this article, we assign continuous probability values for the occupancy status of grids to fully capture the uncertainty in the sensor data. In this way, the sensor data provided by different sensors at different vehicles could be fused according to the occupancy probabilities. We propose a rule based upon the likelihood principle to combine the occupancy probabilities and quantify the uncertainties in the fused sensing data.

In summary, our proposed MVMS-OGM, instead of using discrete information or giving binary occupancy status, will capture and fuse continuous information with soft occupancy probabilities. It provides a more comprehensive and informative environment representation for other autonomous driving tasks that require more than just binary detection results.

The remainder of this article is organized as follows. Section II further discusses the related works. Section III systematically describes the workflow of the multisensor data fusion process. Section IV presents the proposed MVMS-OGM method. Section V validates the proposed methods with simulations, public data set, and the real-world data. Section VI presents conclusion remarks and future directions for this article.

II. Related Works

A. Multisensor Data Fusion

In recent years, various multisensor data fusion methods have been adopted to improve the performance of autonomous driving perception [13]. By combining information from individual sources, the defects of single typed sensors can be well compensated, and more reliable and comprehensive perception results can be obtained. Based on the information used for the fusion algorithm, the fusion methods can be categorized into data-level fusion (see [14]), feature-level fusion (see [5], [15]), and decision-level fusion (see [16], [17]).

Data-level fusion combines the raw data from multiple sensors and provides the fusion results at the lowest data level. This method needs to process all the original information, which slows down the decision-making process and may incur extremely high computational demand and extensive data storage requirement. Feature-level fusion first extracts the features of the raw data and then combines these features into a feature map. Compared with the data-level method, the information space that this method processes is smaller and is not as computational demanding. However, the quality of the selected features directly affects the performance of this method. Decision-level fusion is to make the final decision from a set of decisions generated from multiple sensors. It has the highest level of abstraction and can be readily applied to fuse the information from groups of heterogeneous sensors.

In the field of autonomous driving, there have been numerous research works attempting to improve the perception range and accuracy via multisensor data fusion. For example, Deng *et al.* [5] proposed a real-time multisensor integration strategy for multiscale object recognition, which leads to significant improvements in the detection of objects of different sizes by introducing low-level details. Rashed *et al.* [15] designed a novel CNN architecture to fuse the RGB and LiDAR information, which operates in real time and is suitable for autonomous driving. Fu *et al.* [18] exploited a deep neural network to combine the LiDAR and RGB data and results in a denser pixel-wise depth map. Wei *et al.* [16] proposed a method to locate obstacles in the scene by comprehensively utilizing the sparse point clouds captured by a LiDAR and

the natural image from a camera. Guan *et al.* [17] combined the camera and LiDAR data at the decision level, based on bounding box fusion and improved Dempster–Shafer theory.

However, most of these works did not take into consideration of blind zones caused by occlusions and overlaps or harsh weather conditions, which are quite common in the real world. If the vehicle is in a nonideal situation, multisensor data fusion alone is insufficient to compensate for the limitation of the field of view and the influence of the surrounding objects, such as occlusion and shadowing. Therefore, in this article, we add the multivehicle perception data fusion on top of the multisensor data fusion, in order to solve the problems of vehicle blind spots.

B. Multivehicle Data Fusion

Researchers have utilized multivehicle data fusion in high-level tasks, such as decision making [19] and behavior analysis [20], but not quite much in the basic environment perception task. Rosenstatter and Englund [21] constructed a trust system to evaluate the perception information from different sources of and conduct the fusion process for multivehicle cooperative sensing. However, the limitation of this work is that the information is the location of the vehicle but not the overall environment and was only validated in the vehicle distance model (VDM) and vehicle position model (VPM).

To better demonstrate and capture the overall scene of the environment, while also providing guidance for the fusion process, a unified representation of sensing results is required to restrain different sampling rates and data formats in different sensor setups. Hence, many researchers selected the OGMs, which not only demonstrate the information in a geographical form but also quantify the uncertainty of the vehicle's location to represent the perception results. It has been widely applied and justified in previous works, including SLAM [22], object detection [23], etc. Many researchers have conducted works on expanding a single vehicle's sensing range via map concatenation. For instance, Li *et al.* [24] adopted an optimization approach based on the occupancy likelihood to achieve the merging of the OGMs. Saeedi *et al.* [25] proposed a multistep process to guide the map fusion process, including map learning, relative orientation extraction, and relative translation extraction.

However, there are still two problems to be solved in existing studies. First, most of the OGMs are constructed with discrete probability models. Specifically, they only assign the binary probability to each grid, which is not able to describe or quantitatively reflect different sensor measurements in different quality or precision. Therefore, it is necessary to study an appropriate probability approach for the OGM that will not only seize the information but sufficiently facilitate the fusion process. Some researchers introduced probability distributions to portray objects based on background subtraction [26] or likelihood-field-model [27], [28], but in terms of the general map formation, few studies have been done. Second, little has been done on the combination of different sensing information from different vehicles, most of the existing works focus on the map alignment. Furthermore, the precision of the overlapping areas where sensors are providing data from multiple vehicles is not benefited from the map merging. Few studies have been conducted for those overlapping regions. For instance, [29] adopts the largest probability value as the fusion result and another study [30] utilizes the Dempster–Shafer evidence theory to combine the probability properly. The geographical implication of occupancy probabilities is not considered in the methods mentioned above. Instead, in this article, we introduce the probability fusion into the OGM fusion process with the consideration of the physical meaning, in order to guarantee that the integrated OGM will not only expand sensing range but also capture the uncertainty quantitatively by providing the soft probability information.

C. Summary of Our Contribution

In this article, we develop an MVMS cooperative sensing approach for the environment perception, using the OGM as a unified representation of all sensor data. At the single-vehicle multisensor fusion phase, the sensing data of the LiDAR and camera are combined to exploit the advantages of these two sensors and compensate for their respective limitations. After the fusion process, the data obtained from these two sensors are represented in the form of the OGM. With the OGM, the discrete detection points provided by most existing LiDAR or camera data processing methods are now organized into a map with continuous spatial information that is crucial for driving tasks. At the multivehicle cooperative perception phase, an occupancy probability assignment algorithm and a probability fusion algorithm are proposed. The raw data of sensors are first converted into an occupancy probability density model, and then the multiview OGMs are merged at the spatial and probability levels, thereby expanding the perception range of single-vehicle sensing as well as improving the perception accuracy.

However, it is worth stressing that our MVMS-OGM is fundamentally distinct from existing environment perception in two aspects.

1) Most exciting methods process the sensor data to obtain environment representation of detection points that are discrete in space. Our MVMS-OGM will provide continuous geographical information.
2) Existing methods, regardless of whether they are OGM based or not, only provide discrete information and hard detection result or binary occupancy status. Our MVMS-OGM will capture and fuse continuous information with soft occupancy probabilities.

In other words, our proposed MVMS-OGM is novel and unique in that a more comprehensive and informative environment representation becomes possible for the first time. Not only does the MVMS fusion setup retains more information from individual sensors and perspectives but also various detection, classification, or estimation tasks that best suit the specific driving tasks become possible with such a representation that is continuous in space and soft in occupancy status. Even if the final objective is indeed hard detection decision at discrete spacial points, one could associate those hard decisions with reliability labels. All of these are crucially

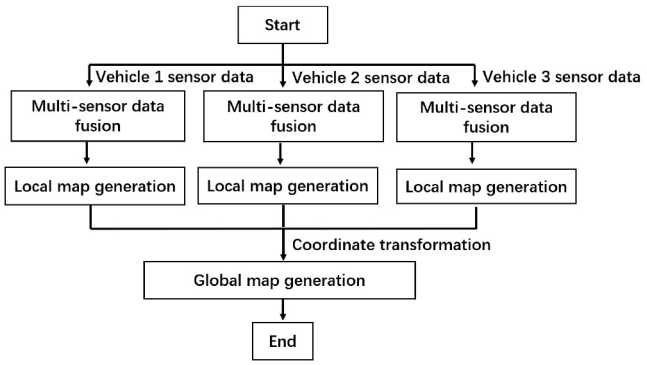


Fig. 1. MVMS cooperative perception framework.

important for fully autonomous driving, in which the required information goes far beyond the binary result of the object detection.

The overall MVMS fusion framework is shown in Fig. 1. At the single-vehicle multisensor fusion phase, the data from multiple sensors are fused to generate the local OGM for each vehicle, which contains continuous area information and continuous occupancy probability values to capture uncertainty in sensor data. At the multivehicle phase, OGMs from multiple vehicles are aligned to the same coordinate system based on the location information of the vehicles and then fused to generate a global map based upon the likelihood ratio combination principle.

The part of this work is based on our previously published paper [31], which proposed four different OGM construction methods and the map fusion algorithm. Different from that conference paper, this article proposes a more detailed multivehicle cooperative perception framework, in which the multisensor data fusion process is added at the single-vehicle multisensor fusion phase.

III. Single-Vehicle Multisensor Fusion

The single-vehicle multisensor data fusion process is presented, as shown in Fig. 2. The process starts with preprocessing the image data obtained by the camera and the point cloud data from the LiDAR to extract the object information in the image. Then, the object information is converted to the LiDAR coordinate system after depth estimation. Finally, the LiDAR point cloud data and the object information are fused and sent to the next phase for OGM construction.

A. Sensor Data Processing

LiDAR: As shown in Fig. 3, the origin O_W is the center of the LiDAR coordinate system $X_W - Y_W - Z_W$, axis Z_W points to the driving direction of the vehicle, and the axes satisfy the right-hand rule. The position of each point w in the coordinate is denoted as (x_w, y_w, z_w).

The point cloud acquired by LiDAR contains a lot of ground points, which lead to issues for subsequent object point cloud processing. Hence, the ground segmentation is usually the first step in perception for processing the 3-D point cloud. A common method is using the grid elevation map based on

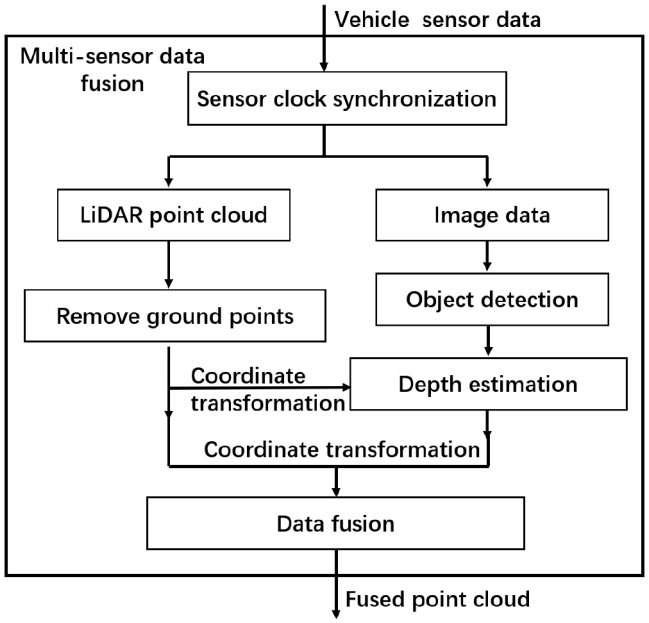


Fig. 2. Multisensor data fusion framework.

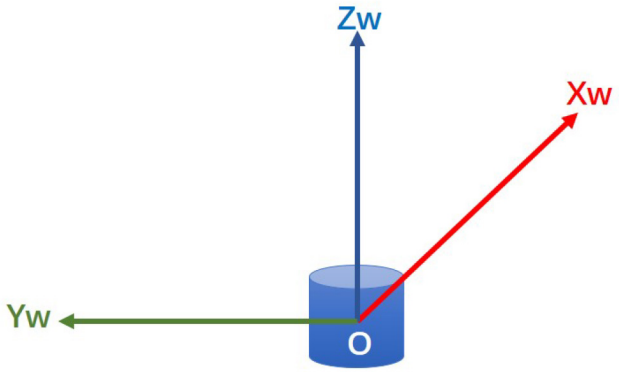


Fig. 3. LiDAR coordinate system.

cells [32]. The maximum and minimum height difference is calculated, and on the basis of the elevation information y_w, the portions are the ground areas if the difference is not greater than the threshold. After processing, the obtained 3-D point cloud will only contain the information about static objects surrounding the vehicle.

Camera: The image data obtained by the front camera on the vehicle have two different coordinates, the pixel coordinate and camera coordinate, respectively. As shown in Fig. 4, where the upper left corner image is the origin image data. $U - V$ is the pixel coordinate, the value of u is the number of columns in the image array, and the value of v is the number of rows. $X_C - Y_C - Z_C$ is the camera coordinate, with the optical center O of the camera. The axis X_C is parallel to the u axis, the axis Y_C is parallel to the v axis, and the axis Z_C is the camera optical axis, which is perpendicular to the image plane. The conversion can be expressed as

$$Zc\begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = M_1 M_2 X_w = M X_w$$

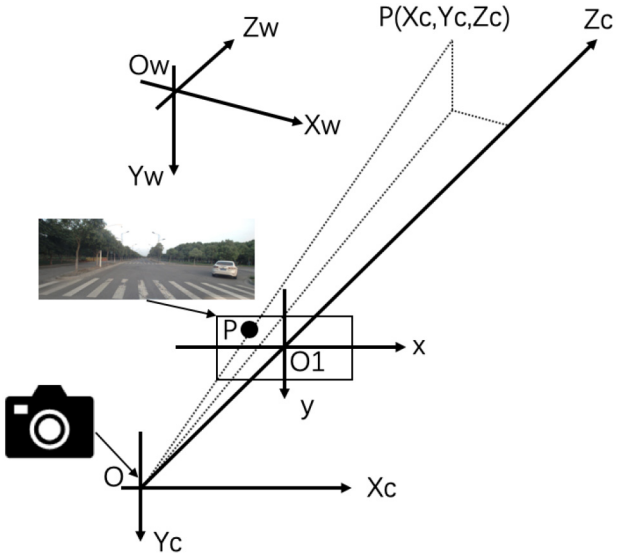


Fig. 4. Transformation relationship between pixel coordinate system and camera coordinate system.

$$f_x = f/dx, \quad f_y = f/dy \tag{1}$$

where f_x and f_y are the scale factors for axis U and axis V in pixel coordinate, respectively. (u_0, v_0) is the center of the image plane.

Spatiotemporal Consistency: Time and spatial synchronization is critical to correctly fuse data from different sensors.

Time synchronization is to align LiDAR data and camera data in time and reduce the impact of different sensors' different refresh rates and delays in the data fusion process. Based on the sensor refresh rate, we can obtain the approximate frame of multiple sensors. The GPS PPS timing source is used to synchronize the host machine and each sensor and the synchronization frames of different sensors can be obtained according to the accuracy timestamp of the host sent to the sensors.

Spatial synchronization is to align LiDAR data and camera data in space. Specifically, the objective is to map the dynamic target detected by the camera to the LiDAR coordinate system according to the conversion. Based on (1), the conversion relationship from the LiDAR coordinate $X_W - Y_W - Z_W$ system to the pixel coordinate system $U-V$ can be expressed as follows:

$$Z_c \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 M_2 X_w = M X_w \tag{2}$$

where Z_c is the distance between the object and the optical center. The camera's intrinsic matrix which contains the intrinsic parameters of the camera is denoted by M_1. The extrinsic matrix is denoted by M_2, where R is the rotation matrix and T is the translation vector. M_1 and M_2 are used to convert between camera local coordinates and global coordinates. Finally, the projection matrix is denoted by M.

Inspired by [33], assume that the Velodyne coordinate system is equivalent to the global coordinate system, because the calibration is performed when the vehicle is stationary. In (2), M_1 contains the internal parameters of the camera. The external matrix M_2 is estimated using the Caltech calibration toolbox. Then, the camera's tilt, roll, and yaw relative to the global coordinate system can be obtained based on the estimated disappearance line on the selected image.

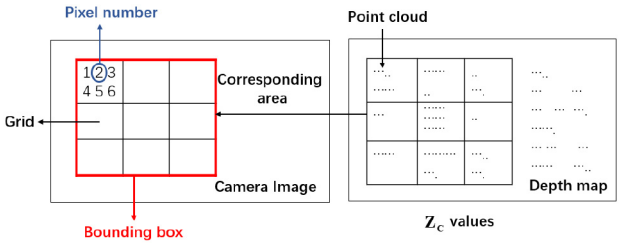


Fig. 5. Estimating the Zc value of grids in the bounding box.

B. Multisensor Fusion Algorithm

The critical step of the single-vehicle multisensor fusion process is to map the bounding box in the image to the global coordinate and fuse the camera data and the LiDAR data together. The coordinate conversion is expressed as follows:

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = Z_c \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = Z_c M_2^{-1} M_1^{-1} X_w \tag{3}$$

which is actually the inverse transform of (2). This is based on the temporal and spatial consistency of the sensor.

To calculate the coordinate via (3), Z_c is needed, which is the distance from the optical center to the object. The depth information is not presented in the camera data, the Z_c of the points in the bounding box can be obtained based on the corresponding LiDAR point cloud data.

First, the image bounding box is divided into grids as shown in Fig. 5, and then the point cloud data is mapped into the camera coordinate system as follows:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \tag{4}$$

Based on the Z_c values of the points collected by LiDAR, the Z_c value of every grid in the bounding box can be estimated by calculating the average Z_c value of the points that fall into the grid. It can be seen that the smaller the size of the grid, the more accurate the Z_c value of this grid. However, in order to avoid the case where no point falls into the small grid, the grid size cannot be too small. Therefore, the grid size should be selected based upon the resolution of the LiDAR cloud points. A good option is to make it to be slightly larger than the minimum distance between LiDAR cloud points.

Based on the positions of the grids in the bounding box and the corresponding Z_c values, the RGB values of these grids can be mapped into the world coordinate system using (3). Then, the dynamic target information becomes (x, y, z, R, G, B).

IV. MVMS-OGM Construction

The data of multiple sensors are now combined with spatial and temporal consistency. While the density of point clouds can be improved by fusing LiDAR and camera data, the LiDAR point cloud is discretely distributed in nature and there is no guarantee that every grid occupied by the vehicle has some data points. In other words, the fused result will only contain isolated points in space rather than continuous areas. In addition, regardless of whether it is LiDAR data or image data, there will also be a certain level of uncertainty in the data that need to be quantified during the fusion process. Hence, an assignment algorithm of the occupancy probabilities based on kernel density estimation is proposed, in order to quantify the uncertainty in sensor measurements. To this end, discrete sensing data are converted into a continuous PDF by mapping those data into continuous areas on the grid map. Finally, the multivehicle maps are fused.

A. Introduction of OGM

There are typically three methods to integrate the LiDAR and the camera data. They are adding depth information to images, adding RGB information to point clouds, or converting the image and point cloud data into other unified data representation formats, such as OGMs, Graphs, and Trees. The first method loses the depth information when converting the data to the 2-D image plane. The fusion efficiency of the second method is affected by the mismatch between high-resolution images and low-resolution point clouds [34]. In autonomous driving, the depth information and the fusion efficiency are both very important. Hence, in this article, the third method is adopted and the image data and point clouds data information are converted into an OGM before the fusion process. The advantages of the OGM are: 1) it gives a unified representation of data from different types of sensors and extracts the features of them in a joint manner and 2) the OGM provides the sensing information in the form of continuous geographical areas instead of isolated detection points, which makes the fusion greatly simplified and provides crucial information for driving. In an OGM, the entire area of interest would be divided into many grids uniformly, and the state of each grid s_i in the map is either occupied $s_i = 1$ or unoccupied $s_i = 0$. For grid i, $p_{s_i=1}$ and $p_{s_i=0}$ indicate the occupied probability and unoccupied probability, respectively, and the sum of these two values equals to 1 as follows:

$$p_{s_i=0} + p_{s_i=1} = 1\ . \tag{5}$$

B. Kernel Density Estimation

The fixed bandwidth kernel density estimator [35] is defined as

$$\widehat{p_h(x)} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(\frac{x - x_i}{h}\right) \tag{6}$$

where $(x_1, x_2, \ldots, x_n)$ are independent and identically distributed samples collected from PDF $p(x)$ and K is the kernel function. The kernel function is usually a single-peak symmetric function. The positive smoothing parameter denoted by h is also regarded as the bandwidth. The scaling kernel K_h is expressed as

$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right). \tag{7}$$

The selection of the bandwidth h can be troublesome. If the bandwidth is too large, the estimation result will be too smooth, masking the data structure of the ground truth, resulting in a large deviation from the true probability density. If the bandwidth is too small, though the deviation between the estimation result and the ground truth will be reduced, the variance of the estimated value will be large and the result will be sharp. The choice of h is a tradeoff between deviation and variance. However, the choice of the kernel function is not that important in the KDE, because the effect of kernel function on the estimation error is a constant shift [36]. Here, in this article, the Gaussian kernel is selected.

C. Occupancy Probability Assignment

Assume that there are N vehicles cooperating in the traffic scene, and in the ego vehicle's map, there are M sensor data points $z_t^1 = (x_t^1, y_t^1), z_t^2 = (x_t^2, y_t^2), \ldots, z_t^M = (x_t^M, y_t^M)$ in the target object t's occupied area. Denote the output of probability assignment as $p(m(x, y)|z_t^1, z_t^2, \ldots, z_t^M)$, where $m(x, y)$ means the grid at the (x, y) is occupied. The output indicates the occupancy probability based on the measurements $z_t^1, z_t^2, \ldots, z_t^M$. No sensing information of the environment is available at initialization, and the initialization for all grids is set to 0.5.

The kernel function is formulated as

$$K_d(z) = \frac{1}{2\pi}e^{-1/2z\cdot z^T}. \tag{8}$$

The kernel estimator is formulated as

$$p_{\mathrm{KDE}}(z) = \frac{1}{M}\sum_{i=1}^{M} K_H\left(z, z_t^i, \boldsymbol{H}\right) \tag{9}$$

where the bandwidth is denoted by $\boldsymbol{H}$, a 2-D matrix. $K_H(z, z_t^i, \boldsymbol{H})$ is the scaled kernel function, given by

$$K_H(z, z_t^i, \boldsymbol{H}) = \frac{1}{|\boldsymbol{H}|}K_d\left((z - z_t^i)\boldsymbol{H}^{-1}\right). \tag{10}$$

Then, the h-bandwidth matrix is determined following the generalized rule of thumb [37]. Assume the target vehicle's occupancy probability distribution follows a Gaussian distribution. Then, denote h-bandwidth matrix as

$$\boldsymbol{H} = M^{-1/6}\hat{\boldsymbol{\Sigma}}^{1/2}. \tag{11}$$

Substituting $\boldsymbol{H}$ and K_H into (9), the occupancy probability of the target object can be obtained, given by

$$p(m(z)|z_t^1, z_t^2, \ldots, z_t^M) = p_{\mathrm{KDE}}(z). \tag{12}$$

D. Probability Mapping to OGM

Note that the vehicle is continuously sensing the environment. Hence, the OGM should be updated continuously when new measurements are taken. The main goal of this mapping process is to update the occupancy probabilities based on the

Algorithm 1 Probability Mapping to OGM

Require: $p_t(m(x_i, y_i)|z_t^1, z_t^2, \ldots, z_t^M)$; $p_{t-1}(x_i, y_i)$;
Ensure: $p_t(x_i, y_i)$;
1: if $p_{t-1}(x_i, y_i) = 1$ or $p_t(m(x_i, y_i)|z_t^1, z_t^2, \ldots, z_t^M) == 1$ then $p_t(x_i, y_i) = 1$;
2: elseif $p_t(m(x_i, y_i)|z_t^1, z_t^2, \ldots, z_t^M) < 0.5$ then $p_t(x_i, y_i) = 0.5$;
3: elseif $p_{t-1}(x_i, y_i) > 0.5$ then
$$l_p = \log \frac{p_{t-1}(x_i, y_i)}{1 - p_{t-1}(x_i, y_i)} + \log \frac{p_t(m(x_i, y_i)|z_t^1, \ldots, z_t^M)}{1 - p_t(m(x_i, y_i)|z_t^1, \ldots, z_t^M)}$$
$$p_t(x_i, y_i) = \frac{e^{l_p}}{1 + e^{l_p}};$$
4: else $p_t(x_i, y_i) = p_t(m(x_i, y_i)|z_t^1, z_t^2, \ldots, z_t^M)$;
5: **return** $p_t(x_i, y_i)$.

most recent measurements obtained via occupancy probability assignment, to construct the accurate area occupied by the objects in the map.

First, normalize the occupancy probability distribution obtained previously, denoted by $p_t(m(x, y)|z_t^1, z_t^2, \ldots, z_t^M)$. For a grid located at (x_i, y_i) in the map, denote $p_t(x_i, y_i)$ and $p_{t-1}(x_i, y_i)$ as its current occupancy probability at t and the previous moment probability at $t - 1$, respectively.

The pseudocode of the probability mapping is given by Algorithm 1. If any input probability is 1, this means that the corresponding grid is occupied and hence its occupancy probability is assigned to 1, as shown in step 1. Next, if the probability returned is less than 0.5, it indicates that it is included in the occupied area of the object, and its state is assigned as "uncertain." The probability distribution of the unoccupied area will be discussed in Section IV-E. In step 3, if the input probability is greater than 0.5, it indicates that there is an object that may be near the vehicle, resulting in overlapping of some grids. Here, the log likelihood is introduced to combine the probabilities which will be further explained in detail in Section IV-F. In step 4, if the current input does not satisfy any cases above, the probability is directly mapped onto the map based on the current measurement value. Finally, the algorithm returns the updated occupancy probability value at current moment time t.

E. Unoccupied Area Probability Assignment

The surrounding environment of the ego vehicle is divided according to the occupancy probability distribution within the maximum detection distance D of the sensor. For simplicity, assume that there is only one object in the detection range, as shown in Fig. 6, the blue part is the detected object. The reset portion image, the darker the color, the greater the occupancy probability. And the uncertain state is represented by shadows. For multiple objects, the unoccupied area can be determined by the same beam model except that all beams returning objects need to be involved.

First, for the grids that are out of D, which means that objects cannot be detected by the sensor, they are treated as under the uncertain state with the probability of 0.5 at

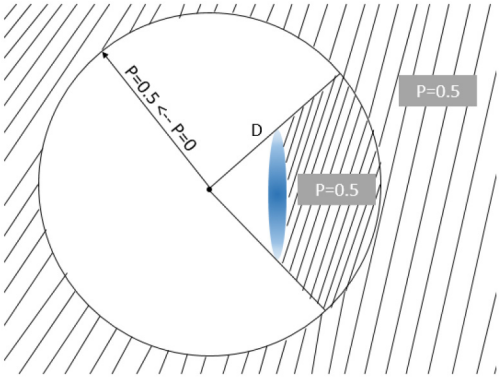


Fig. 6. Beam model of the sensor.

first. Then, if the returned distance is the maximum range of the sensor, which means that within the detection range in this direction, there is no object. Hence, all grids in this direction are now assigned with the unoccupied state. And for the area occluded by the objects that cannot be detected, the state is also set as the uncertain state with the probability of 0.5. Conversely, along the path from the sensor to the occupied part, there is no occlusion, then the grids in this area are assigned with the unoccupied state.

For the grids with the unoccupied state, the occupancy probability values are always below 0.5, and increasing linearly according to the distance from the sensor. The maximum value of the unoccupied grids is 0.5 in the boundary of the detectable area, because the state of the maximum range of the sensor cannot be determined.

F. Multivehicle Map Fusion

Multiple vehicles share their OGM with a fusion center, and it will return a global OGM to each vehicle by map fusion that integrating the local OGMs. The integration process consists of two steps: 1) map alignment and 2) probability fusion. Many existing studies for the fusion of different maps can provide good alignment results. To this end, we assume that the fusion center knows the position of each participated vehicle in a global coordinate. Then, the local OGMs can be easily mapped into the global coordinate via coordinate transformation. After the coordinate conversion, probabilities of the overlapped region are combined based on the log-likelihood ratio.

For the overlapped regions, assuming the detecting process of the occupancy status of a grid is independent, and then the log-likelihood ratio is the best way to combine the perception information from different vehicles without losing. In addition, the log-likelihood ratio can avoid unnecessary truncation errors because it is in the range of $[-\infty, \infty]$.

Following the previous assumptions, there are N vehicles cooperating at time t, denoted as $V_1, \ldots, V_N$. Denote the corresponding maps under the global coordinate are expressed as $M_{V_1,t}, \ldots, M_{V_N,t}$ after the coordinate transformation. Then, the log-likelihood ratio corresponding to vehicle V_i for a

overlapped grid r in the global OGM can be expressed as

$$L_{r,t}^{V_i} = \log \frac{p_t((x_r, y_r)|M_{V_i,t})}{1 - p_t((x_r, y_r)|M_{V_i,t})} \tag{13}$$

where (x_r, y_r) is the coordinates of the rth grid; and $p_t((x_r, y_r)|M_{V_i,t})$ is the occupancy probability of grid r in map $M_{V_i,t}$.

To fuse the OGMs, the occupancy probability of N vehicles is a summation of the log-likelihood ratio of each vehicle by converting the summation of the log-likelihood ratio into a probability value. The probability is formulated as

$$L_{r,t}^{f} = \sum_{i=1}^{N} L_{r,t}^{V_i} = \sum_{i=1}^{N} \log \frac{p_t((x_r, y_r)|M_{V_i,t})}{1 - p_t((x_r, y_r)|M_{V_i,t})}$$

$$p_{r,t}^{f} = \frac{e^{L_{r,t}^{f}}}{1 + e^{L_{r,t}^{f}}} \tag{14}$$

where $p_{r,t}^{f}$ is the occupancy probability of the grid r after fusion.

V. Simulations and Experiments

In this section, the proposed MVMS-OGM method will be validated with the simulated data, an open data set, and the real-world data, respectively. For the multivehicle cooperative perception scenario, it requires multiple vehicles equipped with sensors and V2V equipment, which is difficult to realize, therefore, the multivehicle cooperative perception algorithm proposed in Section IV will be first validated on the MATLAB simulation platform, with a detailed description presented in Section V-B. Then in Section V-C, the multisensor and multivehicle data fusion algorithm is demonstrated on the challenging open data set KITTI [38], which contains the real sensing data collected by LiDAR and camera. Finally, some real-world experiments are conducted via an autonomous driving platform to validate the algorithms in Section V-D. The sensor data are collected from a multivehicle cooperation scene where each vehicle is equipped with LiDAR and camera.

A. Experimental Setup

The simulation uses the MATLAB [39] autopilot toolbox visionDetectionGenerator and lidarPointCloudGenerator preset settings, the maximum detection distance is limited to 500. The two vehicles KITTI validation is based on the KITTI data set, 2011_09_26. Three frames of image and LiDAR data are chosen, 207, 235, and 269. The sensor setup is provided on the official KITTI website [40]. The data acquisition platform for the KITTI data set was assembled with two gray-scale cameras, two color cameras, a Velodyne 64-line 3-D LiDAR, four optical lenses, and a GPS navigation system. Real-world experiments were conducted via an unmanned vehicle platform. The raw data set was obtained on 19 November 2018. The data acquisition platform for the real-world experiment was equipped with one color camera with a 1392×512 resolution, with a total number of three LiDAR, the left and right ones are 16-line LiDAR, Velodyne VLP-16, and the middle one is 32-line LiDAR, Velodyne VLP-32C. The following sections present the results from each experiment.

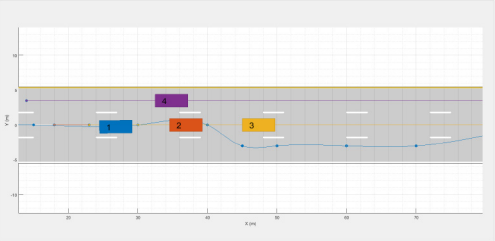


Fig. 7. Overtaking scene aerial view.

B. Simulations

The implementation of simulation is conducted in the MATLAB, which consists of four phases: 1) scene construction; 2) data acquisition; 3) map formation; and 4) map fusion. The scene construction and data acquisition modules are based on the Automated Driving Toolbox in MATLAB.

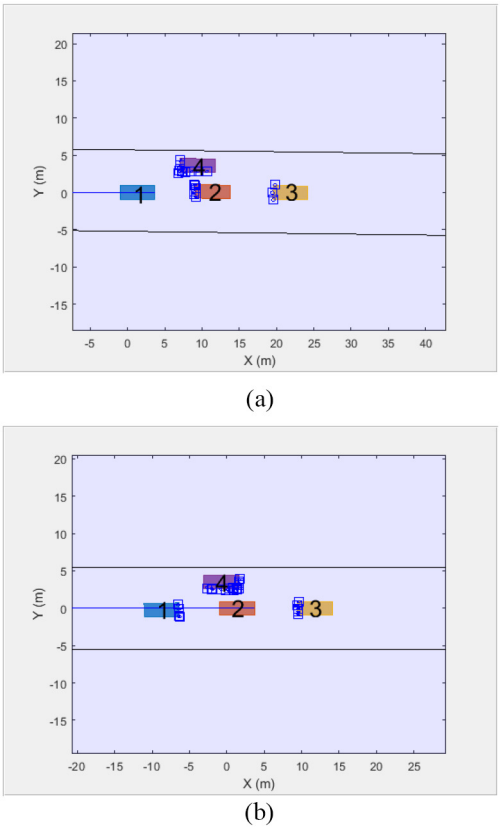
For scene construction, a total duration of 4.45 s typical scenario is considered. There are four vehicles on a three-lane road. Three vehicles numbered 1–3 are driving in the middle lane and vehicle 4 is in the left lane. Vehicle 1 attempts to overtake vehicle 2 at $t = 1.95$ s. The aerial view of the scene at $t = 1.95$ s is shown in Fig. 7. However, vehicle 2 blocks part of the view of vehicle 1, making vehicle 1 unable to observe the traffic situation beyond vehicle 2, making it difficult to make an overtaking decision.

In the data acquisition module, only vehicles 1 and 2 are equipped with the same configuration of the sensor, and the sensor scanning radius is set to 360°. The aerial views in vehicle 1's and vehicle 2's ego coordinates at $t = 1.95$ s are shown in Fig. 8, respectively, where the blue boxes represent the points detected by the vehicle's sensor at the corresponding moment. The figures are shown in aerial views while vehicles may have different heights which are not shown in aerial views. For this specific scenario, vehicle 3 has detection points due to the difference in height, and it only has three detection points which is much less than regular detection results.

Local OGM: For map formation, the construction of the local OGM of each vehicle is using the data from its own sensors. Based on the occupancy probability assignment algorithm, the local OGMs of vehicles 1 and 3 are shown in Fig. 9, respectively.

In Fig. 9, the blue area represents the occupied area, and the white part represents the unoccupied area. The occupancy probability of the gray area is 0.5, the indefinite state. The position of the vehicle itself (the origin of the map) is circled on the map as an ellipse. In order to evaluate the OGM, the yellow dotted lines are provided in the figure to identify the true contour and position of vehicles at the corresponding time.

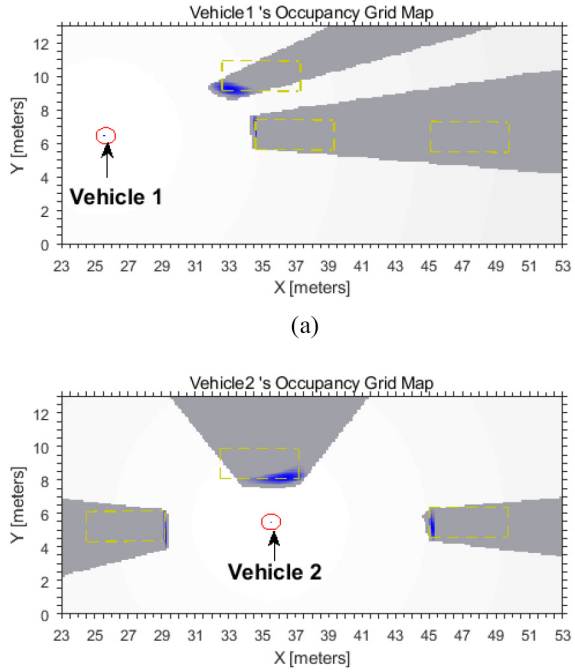
Fig. 9(a) shows the local OGM formed by vehicle 1. In the unoccupied area, the occupancy probability is increasing with the distance from vehicle 1's sensor, as shown by the fact that the color gradually darkens with the increasing distance. In the occupied area, the discretely distributed detection points are continuously distributed in the map based on the KDE. Due to the occlusion of the preceding vehicle and the poor detection position, vehicle 1 cannot detect vehicle 3 and only partial contours of vehicle 2 and vehicle 3 can be identified.



(a)

(b)

Fig. 8. Aerial views under four vehicles' ego coordinates. (a) vehicle 1. (b) vehicle 2.



(a)

(b)

Fig. 9. Local OGMs of (a) vehicle 1 and (b) vehicle 2.

Compared with vehicle 1, vehicle 2 is in a better detection position, which can detect three vehicles around, although the contour information is still incomplete compared with the yellow dotted line, as shown in Fig. 9(b).

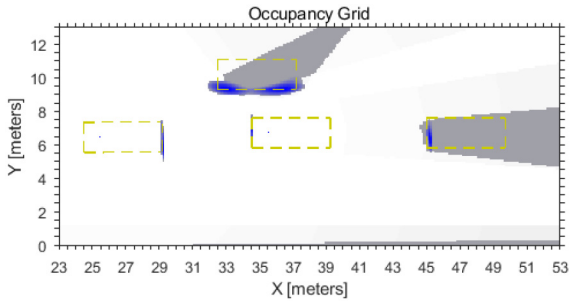


Fig. 10. Integrated global OGM based on kernel density estimation.

Grid Map Fusion: The map fusion phase combines the two local OGMs of vehicles 1 and 2 to generate the global OGM, including the occupied area and the unoccupied area.

Fig. 10 shows the fusion result of the two local OGMs. It can be observed that the information of the global map is more comprehensive, with more detail and complete boundary information of vehicles in the environment. Furthermore, the accuracy of the detection results is improved in the overlapped areas. For instance, as shown in Fig. 9, the two vehicles both have fewer detection points for vehicle 4, resulting in only a small portion of the occupied area being detected. After fusion, the boundary of vehicle 4 is more accurate and concrete. Meanwhile, the occupancy area of vehicle 3 is reflected in the map, which cannot be detected from vehicle 1's perspective. Some of the occluded and uncertain areas in the single-vehicle maps are also determined after the fusion.

C. Validation With KITTI Data Set

To the best of our knowledge, most of the publicly available perception data sets of outdoor scenes are collected by a single vehicle, and no public data set contains the perception data from multiple vehicles under the same scenario. The KITTI data set [38] contains the sensing data from two high-resolution color cameras, two gray-scale video cameras, a Velodyne laser scanner, and the real-time positioning information provided by a GPS localization system. To overcome this problem and utilize the KITTI data set, we treat several consecutive frames of data collected by a single vehicle as the perception data of the environment from different perspectives, which is similar to the approached adopted by [41] to further verify the proposed MVMSOGM method. With this approach, the data frames will contain information about the same environment, but they will be obtained from different views since the vehicle is at different locations when taking them. This makes these frames a good approximation to the case of multivehicle sensing from different views at the same time instant.

As shown in Fig. 11, a T-shaped intersection scene is selected. The ego vehicle is going to pass through the intersection and then turn left. Its camera can only see vehicle 1 at the intersection in its vision of range, as shown in Fig. 12(a), while its LiDAR can additionally see vehicle 10 on its front right as shown in Fig. 13(a). The local OGM from

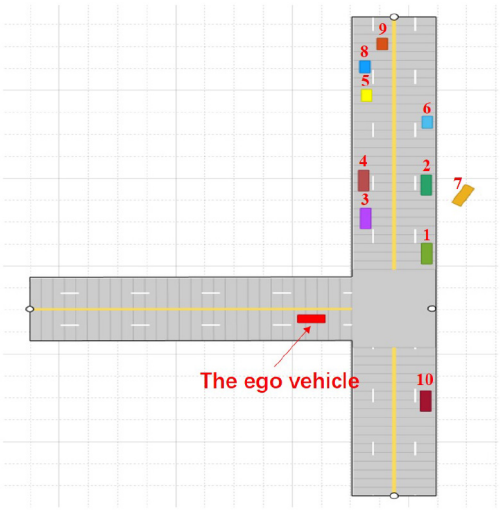
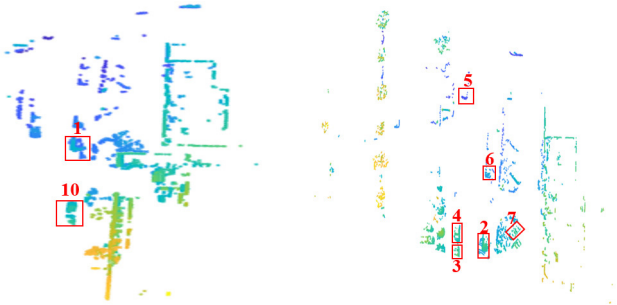


Fig. 11. Bird's-eye view of the T-shaped intersection scene.

(a) (b)

Fig. 12. Object detection results of the images collected by the cameras of the (a) ego vehicle and (b) vehicle 1.

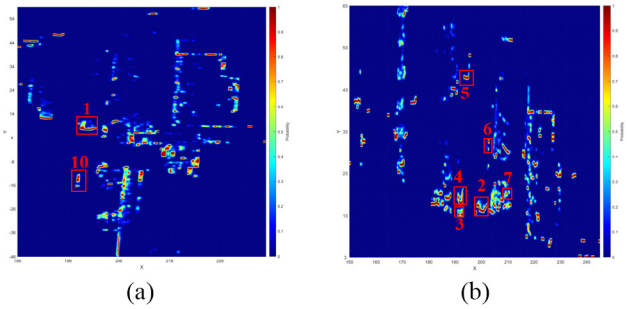


(a) (b)

Fig. 13. Point cloud maps detected by the LiDAR of the (a) ego vehicle and (b) vehicle 1.

the perspective of the ego vehicle is shown in Fig. 14(a), which is constructed based on the fusion of its camera and LiDAR data. It can be seen that the fusion of multisensor data can compensate for the short sensing range of the camera and the sparseness of the point cloud from the LiDAR. The outlines of vehicle 1 and vehicle 10 can be displayed after constructing the OGM.

However, due to the occlusion of buildings in the environment, from the perspective of the ego vehicle, it is impossible to perceive the traffic conditions of the lane that it is going to join, whereas vehicle 1 can obtain the information because it is already in this lane. The local OGM from the perspective of vehicle 1 is obtained based on the same method as shown in Fig. 14(b). It can be seen that from the perspective of vehicle 1, the traffic conditions in this lane can be seen quite well. Therefore, the local OGMs from the perspective of the



(a) (b)

Fig. 14. Local OGMs generated after fusing the camera and LiDAR data of the (a) ego vehicle and (b) vehicle 1.

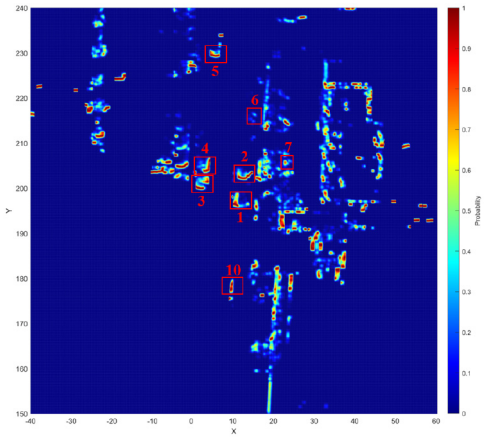
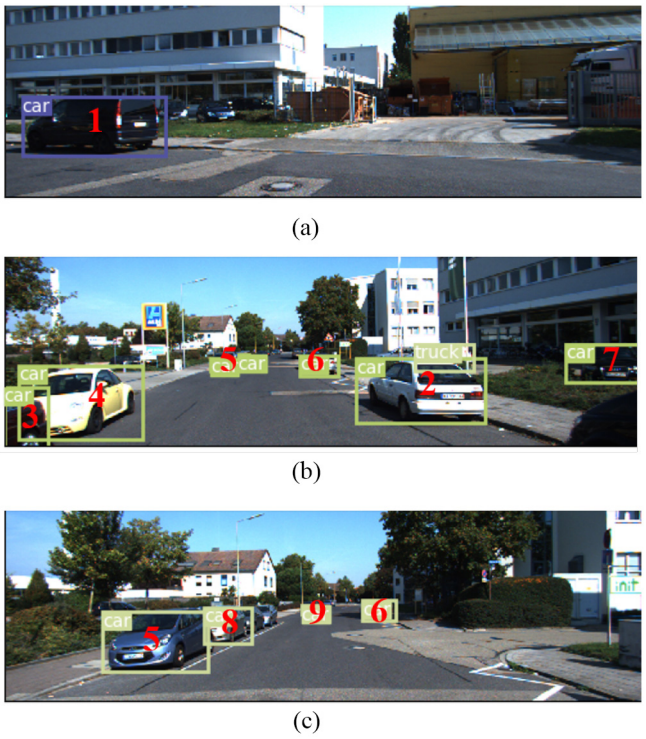


Fig. 15. Fusion result of the OGMs generated by the perception data of the ego vehicle and vehicle 1.

ego vehicle and vehicle 1 are fused and the result is shown in Fig. 15. It can be observed that, via the cooperative perception with surrounding vehicles, the ego vehicle can sense beyond its own vision range and obtain more comprehensive perception data for its turn. As a result, subsequent safer decisions or timely warnings can be made if needed.

Next, the perception information of vehicle 2, which is beyond vehicle 1, is fused. The vision fields of these three vehicles are shown in Fig. 16 and the integrated global perception map is shown in Fig. 17.

In order to observe the changes in the occupied area during the fusion process in detail, vehicle 6 is taken as the observation object. It can be seen from Fig. 16 that only vehicles 1 and 2 can observe vehicle 6. However, due to the long distance, the perception information about vehicle 6 is limited. Zoomed-in version of the area occupied by vehicle 6 in the OGMs of vehicles 1 and 2 is provided in Fig. 18(a) and (b), respectively. It can be seen that only one side or two sides have a higher occupancy probability and the probability in the middle area is small with great uncertainty. The corresponding area in the merged map is shown in Fig. 18(c). In the merged map, the contour of vehicle 6 can be clearly seen, which shows that by fusing the occupancy probabilities of these two maps, the uncertainty of the single vehicle's perception data can be further reduced, providing more accurate perception results, and more helpful to the subsequent semantic analysis and scene understanding process.



(a) (b) (c)

Fig. 16. Object detection results of the images collected by the cameras of the (a) ego vehicle, (b) vehicle 1, and (c) vehicle 2.

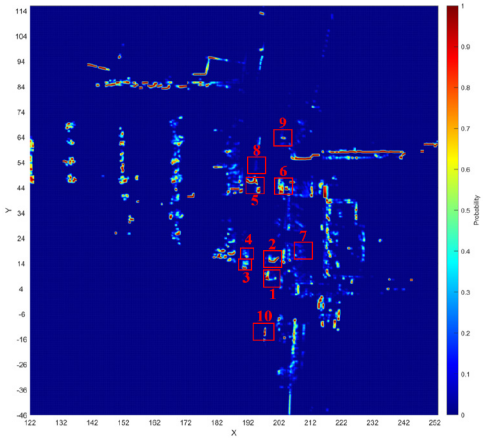
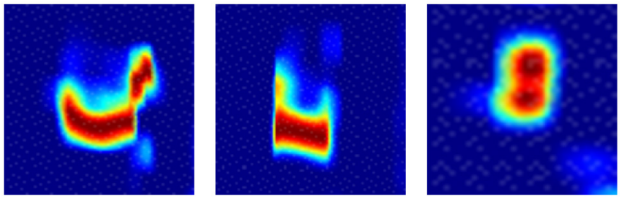


Fig. 17. Fusion result of the OGMs generated by the perception data of the ego vehicle, vehicle 1, and vehicle 2.



(a) (b) (c)

Fig. 18. Area occupied by vehicle 6 in the three OGMs. (a) Area in vehicle 1's map. (b) Area in vehicle 2's map. (c) Area in the fusion map.

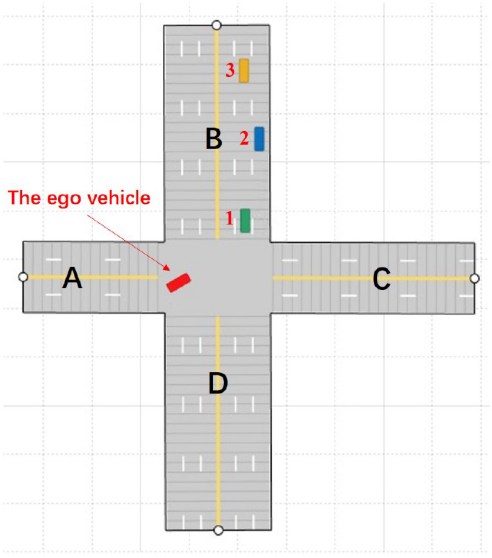


Fig. 19. Bird's-eye view of the crossroad in the real-world experiment.

(a) (b)

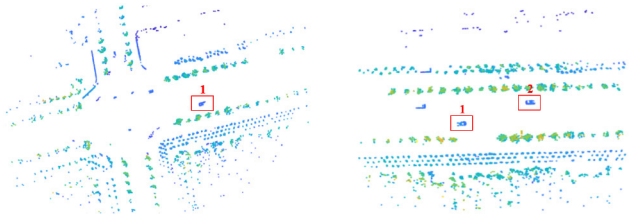
Fig. 20. Object detection results of the images collected by the cameras of the (a) ego vehicle and (b) vehicle 1 in the real-world experiment.

D. Real-World Experiments

To evaluate the performance of the proposed algorithm in the real world, the multivehicle perception data provided by the unmanned vehicle platform are used, which was collected by two autonomous vehicles equipped with LiDAR and cameras.

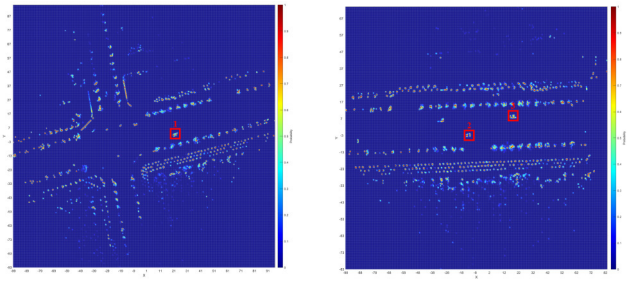
The experimental scene is shown in Fig. 19, which is an intersection. The ego vehicle is in the central area of the intersection and is about to turn into lane B. The vehicle closest to the ego vehicle in lane B is vehicle 1. Figs. 20–22, respectively, show the camera data, LiDAR data, and the generated OGMs from the perspective of the ego vehicle and vehicle 1. It can be seen that the ego vehicle can only see vehicle 1 due to the limitation of its own vision range and cannot obtain the traffic status of lane B.

By cooperating with vehicle 1, the ego vehicle can receive the perception data of vehicle 1, thereby obtaining the information of lane B. The fusion result of the ego vehicle's map is displayed in Fig. 23 and vehicle 1's map in Fig. 22. It can be seen that by fusing the perception data of these two vehicles, the ego vehicle can perceive vehicle 2 and vehicle 3 in the lane to be merged before turning and obtain richer perceptual information, which helps to predict the incoming risks in advance and improve the safety during the turning process at the intersection.



(a) (b)

Fig. 21. Point cloud maps detected by the LiDAR of the (a) ego vehicle and (b) vehicle 1 in the real-world experiment.



(a) (b)

Fig. 22. Local OGMs generated after fusing the camera and LiDAR data of the (a) ego vehicle and (b) vehicle 1 in the real-world experiment.

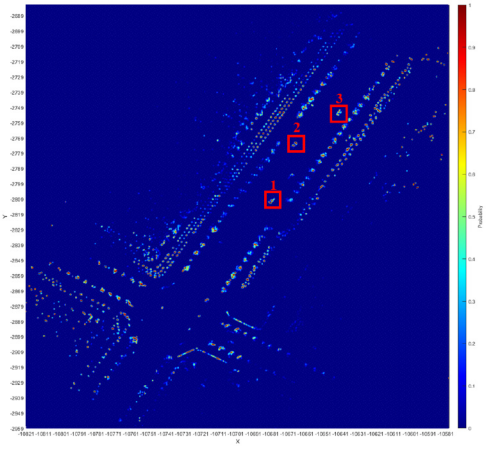


Fig. 23. Fusion result of the OGMs generated by the perception data of the ego vehicle and vehicle 1 in the real-world experiment.

VI. Conclusion

In this article, we proposed an MVMS cooperative perception framework for autonomous driving by the construction and fusion of OGMs. The OGM is adopted as a unified representation of the sensing information to facilitate the fusion of sensing information collected by different sensors at different vehicles. At the multisensor level, our proposed fusing algorithm ensures the spatiotemporal consistency in different kinds of sensor data; at the multivehicle level, our proposed fusion scheme conducts the fusion with considerations of the quantified uncertainty in sensor data by likelihood ratio principle. Our proposed MVMS-OGM not only provides continuous geographical information but also captures and fuses continuous soft, analog occupancy probabilities, resulting in a more comprehensive and raw environmental information. MATLAB simulations, public KITTI data, and real-world experiments were used to validate and demonstrate our proposed algorithm. In the future, we would consider more factors in sensor data uncertainty such as the weather conditions in the OGM construction and also the analysis of OGMs for various driving tasks, such as driving decision making, multivehicle interaction, and so on.

Acknowledgment

The authors would like to thank Prof. Georgios B. Giannakis (University of Minnesota, Twin Cities) for his valuable discussions and suggestions

References

[1] B. Li, T. Zhang, and T. Xia, "Vehicle detection from 3D LiDAR using fully convolutional network," Aug. 2016, *arXiv:1608.07916*.

[2] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transport. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[3] G. Reina, J. Underwood, G. Brooker, and H. Durrant-Whyte, "Radar-based perception for autonomous outdoor vehicles," *J. Field Robot.*, vol. 28, no. 6, pp. 894–913, May 2011.

[4] B. Li, W. Ouyang, L. Sheng, X. Zeng, and X. Wang, "GS3D: An efficient 3D object detection framework for autonomous driving," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2019, pp. 1019–1028.

[5] Q. Deng, X. Li, P. Ni, H. Li, and Z. Zheng, "Enet-CRF-LiDAR: LiDAR and camera fusion for multi-scale object recognition," *IEEE Access*, vol. 7, pp. 174335–174344, 2019.

[6] G. A. Kumar, J. Lee, J. Hwang, J. Park, S. H. Youn, and S. Kwon, "LiDAR and camera fusion approach for object distance estimation in self-driving vehicles," *Symmetry*, vol. 12, no. 2, p. 324, Feb. 2020.

[7] "A Tragic Loss." Tesla. Jun. 2016. [Online]. Available: https://www.tesla.com/blog/tragic-loss

[8] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[9] X. Cheng, C. Chen, W. Zhang, and Y. Yang, "5G-enabled cooperative intelligent vehicular (5GenCIV) framework: When benz meets marconi," *IEEE Intell. Syst.*, vol. 32, no. 3, pp. 53–59, May 2017.

[10] X. Cheng, R. Zhang, and L. Yang, "Wireless toward the era of intelligent vehicles," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 188–202, Feb. 2019.

[11] D. T. Kanapram *et al.*, "Collective awareness for abnormality detection in connected autonomous vehicles," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3774–3789, May 2020.

[12] X. Cheng, D. Duan, L. Yang, and N. Zheng, "Societal intelligence for safer and smarter transportation," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9109–9121, Jun. 2021.

[13] H. Cho, Y.-W. Seo, B. V. Kumar, and R. R. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, Hong Kong, China, May 2014, pp. 1836–1843.

[14] M. Kang, S. Hur, W. Jeong, and Y. Park, "Map building based on sensor fusion for autonomous vehicle," in *Proc. 11th Int. Conf. Inf. Technol. New Gener.*, 2014, pp. 490–495.

[15] H. Rashed, M. Ramzy, V. Vaquero, A. El Sallab, G. Sistu, and S. Yogamani, "FuseMODNet: Real-time camera and LiDAR based moving object detection for robust low-light autonomous driving," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 2393–2402.

[16] Y. Wei, J. Yang, C. Gong, S. Chen, and J. Qian, "Obstacle detection by fusing point clouds and monocular image," *Neural Process Lett.*, vol. 49, no. 3, pp. 1007–1019, Jun. 2019.

[17] L. Guan, Y. Chen, G. Wang, and X. Lei, "Real-time vehicle detection framework based on the fusion of LiDAR and camera," *Electronics*, vol. 9, no. 3, p. 451, Mar. 2020.

[18] C. Fu, C. Mertz, and J. M. Dolan, "LiDAR and monocular camera fusion: On-road depth completion for autonomous driving," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Oct. 2019, pp. 273–278.

[19] Y. Rahmati and A. Talebpour, "Towards a collaborative connected, automated driving environment: A game theory based decision framework for unprotected left turn maneuvers," in *Proc. IEEE Intell. Veh. Symp.*, Redondo Beach, CA, USA, Jun. 2017, pp. 1316–1321.

[20] M. A. S. Kamal, S. Taguchi, and T. Yoshimura, "Efficient driving on multilane roads under a connected vehicle environment," *IEEE Trans. Intell. Transport. Syst.*, vol. 17, no. 9, pp. 2541–2551, Sep. 2016.

[21] T. Rosenstatter and C. Englund, "Modelling the level of trust in a cooperative automated vehicle control system," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1237–1247, Apr. 2018.

[22] S. Kohlbrecher, O. Von Stryk, J. Meyer, and U. Klingauf, "A flexible and scalable SLAM system with full 3D motion estimation," in *Proc. IEEE Intell. Symp. Safety, Security Rescue Robot.*, Kyoto, Japan, Nov. 2011, pp. 155–160.

[23] R. Dubé, M. Hahn, M. Schutz, J. Dickmann, and D. Gingras, "Detection of parked vehicles from a radar based occupancy grid," in *Proc. IEEE Int. Veh. Symp.*, Dearborn, MI, USA, Jun. 2014, pp. 1415–1420.

[24] H. Li, M. Tsukada, F. Nashashibi, and M. Parent, "Multivehicle cooperative local mapping: A methodology based on occupancy grid map merging," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2089–2100, Oct. 2014.

[25] S. Saeedi, L. Paull, M. Trentini, and H. Li, "Neural network-based multiple robot simultaneous localization and mapping," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2376–2387, Dec. 2011.

[26] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Washington, DC, USA, Jun. 2004, p. 2.

[27] T. Chen, R. Wang, B. Dai, D. Liu, and J. Song, "Likelihood-field-model-based dynamic vehicle detection and tracking for self-driving," *IEEE Trans. Intell. Transport. Syst.*, vol. 17, no. 11, pp. 3142–3158, Nov. 2016.

[28] M. Schütz, N. Appenrodt, J. Dickmann, and K. Dietmayer, "Multiple extended objects tracking with object-local occupancy grid maps," in *Proc. IEEE Int. Conf. Inf. Fusion*, Salamanca, Spain, Jul. 2014, pp. 1–7.

[29] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA, USA: MIT Press, 2005.

[30] M. Aeberhard, S. Paul, N. Kaempchen, and T. Bertram, "Object existence probability fusion using Dempster-Shafer theory in a high-level sensor data fusion architecture," in *Proc. IEEE Intell. Veh. Symp.*, Baden-Baden, Germany, Jun. 2011, pp. 770–775.

[31] Y. Li, D. Duan, C. Chen, X. Cheng, and L. Yang, "Occupancy grid map formation and fusion in cooperative autonomous vehicle sensing," in *Proc. IEEE Int. Conf. Commun. Syst.*, Chengdu, China, Dec. 2018, pp. 204–209.

[32] C. Tongtong, D. Bin, L. Daxue, Z. Bo, and L. Qixu, "3D LiDAR-based ground segmentation," in *Proc. IEEE Asian Conf. Pattern Recognit.*, Beijing, China, Nov. 2011, pp. 446–450.

[33] G. Zhao, X. Xiao, J. Yuan, and G. W. Ng, "Fusion of 3D-LiDAR and camera data for scene parsing," *J. Vis. Commun. Image Represent.*, vol. 25, no. 1, pp. 165–183, Jan. 2014.

[34] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, and D. Cao, "Deep learning for image and point cloud fusion in autonomous driving: A review," 2020, *arXiv:2004.05224*.

[35] G. R. Terrell and D. W. Scott, "Variable kernel density estimation," *Ann. Stat.*, vol. 20, no. 3, pp. 1236–1265, Sep. 1992.

[36] Y.-C. Chen, "A tutorial on kernel density estimation and recent advances," *Biostat. Epidemiol.*, vol. 1, no. 1, pp. 161–187, Apr. 2017.

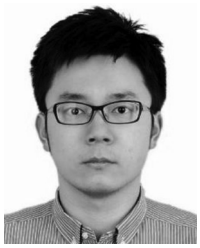
[37] W. K. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and Semiparametric Models*, vol. 1. Berlin, Germany: Springer, 2004.

[38] 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, Aug. 2013.

[39] *MATLAB, Version 9.7.0.1190202 (R2019b)*. Natick, MA, USA: MathWorks, Inc., 2018.

[40] A. Geiger *et al.* "The KITTI Vision Benchmark Suite." 2021. [Online]. Available: http://www.cvlibs.net/datasets/kitti/setup.php

[41] J. Guo *et al.*, "CoFF: Cooperative spatial feature fusion for 3-D object detection on autonomous vehicles," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11078–11087, Jul. 2021.



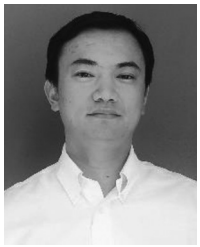
Xinhu Zheng (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2022.

He is currently an Assistant Professor with The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. His current research interests are data mining in power system and intelligent transportation system by exploiting different modality of data, leveraging optimization and machine learning techniques.

Yuru Li (Student Member, IEEE) received the M.S. degree in communication and information system from Peking University, Beijing, China, in 2021.

She is currently an Algorithm Researcher with Tencent, Guangzhou, China. Her research interests include multivehicle cooperative perception, neural network sensitivity, and reliability.

Ms. Li has received the Best Paper Awards at ICCS'18.



Dongliang Duan (Senior Member, IEEE) received the B.S. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2006, the M.S. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 2009, and the Ph.D. degree in electrical engineering from Colorado State University, Fort Collins, CO, USA, in 2012.

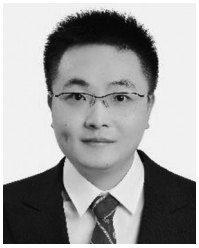
Since graduation, he joined the Department of Electrical and Computer Engineering, University of Wyoming, Laramie, WY, USA, where he is currently an Associate Professor. His current research interests are signal processing and data analytics for power and intelligent transportation systems.

Dr. Duan is currently a Subject Editor for *IET Communications* and a Guest Editor for the Special Issue on the IoT for Power Grid on IEEE INTERNET OF THINGS JOURNAL.

Liuqing Yang (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2004.

She has been a Faculty Member with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA; Colorado State University, Fort Collins, CO, USA; and the University of Minnesota, and is currently a Chair Professor with The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. Her general interests are in communications, sensing, and connected intelligence—subjects on which she has published more than 380 journal and conference papers, four book chapters, and six books.

Dr. Yang was the recipient of the ONR Young Investigator Program Award in 2007, the NSF Faculty Early Career Development (CAREER) Award in 2009, and the Best Paper Award at IEEE ICUWB'06, ICCC'13, ITSC'14, GLOBECOM'14, ICC'16, WCSP'16, GLOBECOM'18, ICCS'18, and ICC'19. She is the Editor-in-Chief for *IET Communications*, an Executive Editorial Committee Member for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and a Senior Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING. She has also served as the Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE INTELLIGENT SYSTEMS, and *PHYCOM: Physical Communication*, and as the program chair, track/symposium chair, or TPC chair for many conferences.



Chen Chen (Senior Member, IEEE) received the Ph.D. degree from Peking University, Beijing, China, in 2009.

He is currently an Associate Professor with Peking University. Since 2010, he has authored or coauthored five books and over 120 journal and conference papers, including one "ESI hot paper" (top 0.1%) and four "ESI highly cited papers" (top 1%). His research interests include wireless communications and networking, signal processing, and vehicular and satellite communication systems.

Dr. Chen was a recipient of two Outstanding Paper Awards from the Chinese Government of Beijing in 2013 and 2018, respectively, and three Best Paper Awards of IEEE conferences (ICNC'17, ICCS'18, and Globecom'18). He has served as the symposium co-chair, session chair, and a member of the Technical Program Committee for several international conferences. He is currently an Associate Editor of the *IET Communications*. He has been the principal investigator of over 20 funded research projects and has led the development of a key subsystem in a 12th–13th Five Year Plan Program of China, which has won a First Prize of State Science and Technology Award.

Xiang Cheng (Fellow, IEEE) received the Ph.D. degree jointly from Heriot-Watt University, Edinburgh, U.K., and the University of Edinburgh, Edinburgh, in 2009.

He is currently a Boya Distinguished Professor with Peking University, Beijing, China. His general research interests are in areas of channel modeling, wireless communications, and data analytics, subject on which he has published more than 280 journal and conference papers, nine books, and holds 17 patents.

Prof. Cheng is a Distinguished Young Investigator of China Frontiers of Engineering, a recipient of the IEEE Asia–Pacific Outstanding Young Researcher Award in 2015, a Distinguished Lecturer of IEEE Vehicular Technology Society, and a Highly Cited Chinese Researcher in 2020. He was a co-recipient of the 2016 IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Best Paper Award: Leonard G. Abraham Prize, and *IET Communications* Best Paper Award: Premium Award. He has also received the Best Paper Awards at IEEE ITST'12, ICCC'13, ITSC'14, ICC'16, ICNC'17, GLOBECOM'18, ICCS'18, and ICC'19. He has served as the symposium lead chair, co-chair, and member of the Technical Program Committee for several international conferences. He is currently a Subject Editor of *IET Communications* and an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE WIRELESS COMMUNICATIONS LETTERS, and the *Journal of Communications and Information Networks*. In 2021, he was selected into two world scientist lists, including World's Top 2% Scientists released by Stanford University and Top Computer Science Scientists released by Guide2Research.