IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 27, 2019, accepted August 16, 2019, date of publication August 28, 2019, date of current version September 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937950

Interactive Perception-Based Multiple Object Tracking via CVIS and AV

WEI SHANGGUAN[1,2,3], (Member, IEEE), YU DU[1], (Member, IEEE), AND LINGUO CHAI[1], (Member, IEEE)

[1]School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
[3]Beijing Engineering Research Center of EMC and GNSS Technology for Rail Transportation, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Wei ShangGuan (wshg@bjtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1600600, in part by the Fundamental Research Funds for the Central Universities under Grant 2019YJS023, in part by the National Natural Science Foundation of China under Grant 61773049, and in part by the Civil Aviation Administration of China Major Projects under Grant 201501.

ABSTRACT Cooperative Vehicle and Infrastructure System (CVIS) and Autonomous Vehicle (AV) are two mainstream technologies to improve urban traffic efficiency and vehicle safety in the Intelligent Transportation System (ITS). However, there remain significant obstacles that must be overcome before fully unmanned applications are ready for widespread adoption in a transportation system. To achieve fully driverless driving, the perception ability of vehicle should be accurate, fast, continuous, and wide-ranging. In this paper, an interactive perception framework is proposed, which combines the visual perception of AV and information interaction of CVIS. Based on the framework, an interactive perception-based multiple object tracking (IP-MOT) method is presented. IP-MOT can be divided into two parts. First, a Lidar-only multiple object tracking (L-MOT) method obtains the status of surroundings using the voxel cluster algorithm. Second, the preliminary tracking result is fused with the interactive information to generate the trajectories of target vehicles. Two simulation platforms are established to verify the proposed methods: CVIS simulation platform and Virtual Reality (VR) test platform. The L-MOT algorithm is tested on a public dataset and the IP-MOT algorithm is tested on our simulation platform. The results show that the IP-MOT algorithm can improve the accuracy of object tracking as well as expand the vehicle perception range via combination of CVIS and AV.

INDEX TERMS Cooperative vehicle and infrastructure system, autonomous vehicle, perception mode, multiple object tracking.

I. INTRODUCTION

As a highly complex system with a large number of different types of participants, the urban transportation system urgently needs to improve its intelligence level systematically. The coexistence of vehicles with different intelligence levels on the road is a necessary stage of the intelligent transportation system evolution from fully manual driving to fully unmanned driving. The transportation system consisting of heterogeneous traffic participants such as human drivers, pedestrians, non-motor vehicles and intelligent vehicles of different intelligence levels is called a complex mixed traffic system. The complex mixed traffic system has the characteristics of topological networking, nonlinearity, strong coupling and extensive randomness. Solving intelligent cooperative driving problems in a complex mixed traffic environment is the key task in the next generation intelligent transportation system. The combination of CVIS and AV provides the possibility to achieve wide range perception, multi-agent decision making and global optimization control.

In recent years, many advanced technologies such as computer vision, robot control and cloud computation have been applied gradually to the transportation domain and have driven forward the development of the Intelligent Transportation System (ITS). For almost all the applications of ITS, perception is the first step in the data processing and plays an irreplaceable role. The development route of ITS perception mode includes three stages: autonomous perception to interactive perception to networked perception. As shown in Fig. 1, the in-depth application of information

The associate editor coordinating the review of this article and approving it for publication was Shaohua Wan.

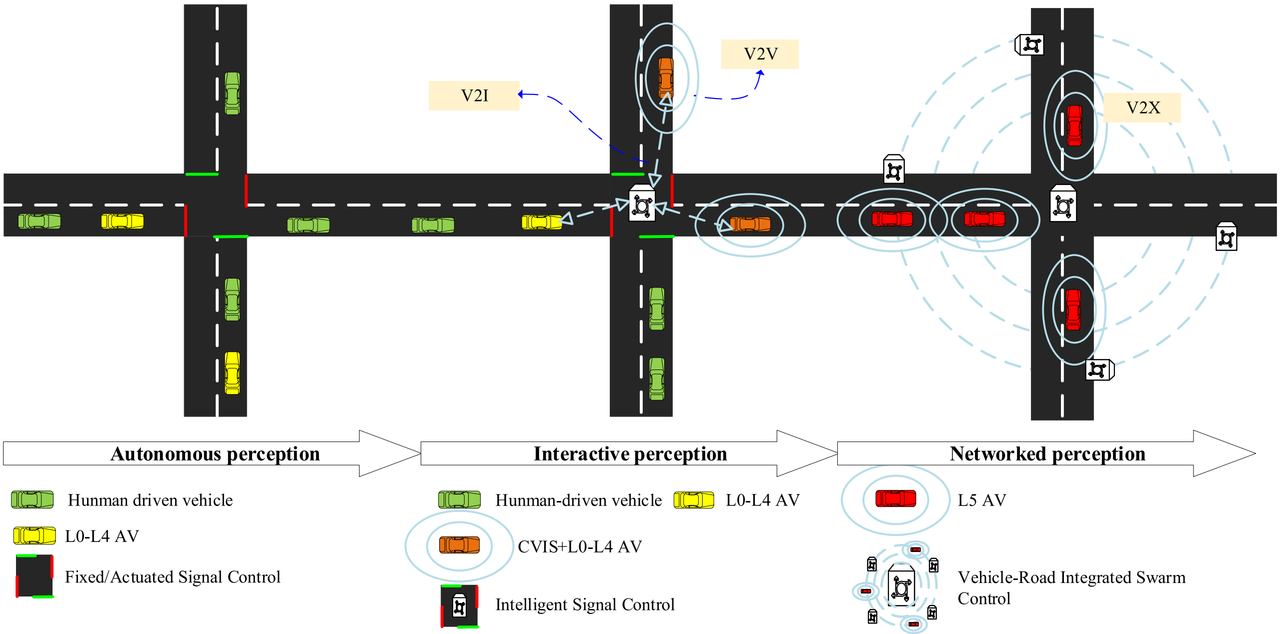

FIGURE 1. Evolution route of intelligent transportation system perception mode.

interaction technology will promote the evolution of the perception mode, make the vehicle to vehicle connection and vehicle to infrastructure connection closer. Autonomous perception, including vehicle self-perception and environment perception, uses vehicle-centric sensors as the primary means of perception. Vehicle self-perception obtains the accuracy status of an ego vehicle, such as position and attitude. Environmental perception captures the status of surroundings by vehicle-centric visual sensing technology. The limited sensing range and occlusion issues are obvious shortcomings of autonomous perception. To address these limitations, interactive perception takes audio-visual integration as the core change. In interactive perception, cooperation decision and control methods are applied via the Cooperative Vehicle and Infrastructure System (CVIS). Communication is considered as the way to expand the vehicle sensing range and improve the accuracy of sensing data, such as avoiding accidents in areas where visual sensors are blocked. With the maturity of vehicular communication technology, the reliability of Vehicle to everything (V2X) can be ensured for some time to come. In networked perception, the status of various types of vehicles, non-motor vehicles, pedestrians and traffic control system will be obtained and shared via an integrated mechanism, which connects all the transportation participants. As a result, the perception ability of ITS is gradually improving. A more detailed introduction on the definition of the stages can be found in [1].

The motivation of this paper is summarized as three points: First, multiple object real-time tracking is still a fundamental and challenging issue which is very important for vehicle obstacle avoidance and environmental situation prediction; Second, It is a development trend to combine CVIS and autonomous driving technologies to address the limitations of the vehicle-centric perception; Third, to perform fast, low-cost and flexible test verification to our methods, it is necessary to build a highly realistic simulation test environment.

Multiple object tracking (MOT) aims to capture the trajectories of surrounding objects on the road by making full use of the sensing data. The trajectory of an environmental object contains important information for vehicle safety, such as distance, velocity and acceleration. There have been lots of researches which are focusing on improving tracking accuracy of an ego vehicle using vehicle-centric perception system, however study on the use of multi-source information, including IMU/GPS/Lidar/V2V, is not enough. In this paper, the interactive perception which combines the CVIS with AV is proposed and key technologies in interactive perception are introduced in detail. Based on the interactive perception data process framework, an interactive perception based multiple object tracking (IP-MOT) method is presented and implemented. Different from the previous work, Interactive Perception Multiple Objects Tracking (IP-MOT) algorithm can continuously sense the surrounding vehicle position whether in the condition of visual occlusion or communication failure and improve tracking accuracy.

The contributions of this paper involve three aspects:

1. Interactive perception framework is proposed based on the combination of CVIS and AV.

2. A multiple object tracking algorithm is proposed based on the interactive perception, which can provide accurate trajectory information of environmental objects.

3. Two platforms are designed and implemented to provide the test and verification environment for interactive

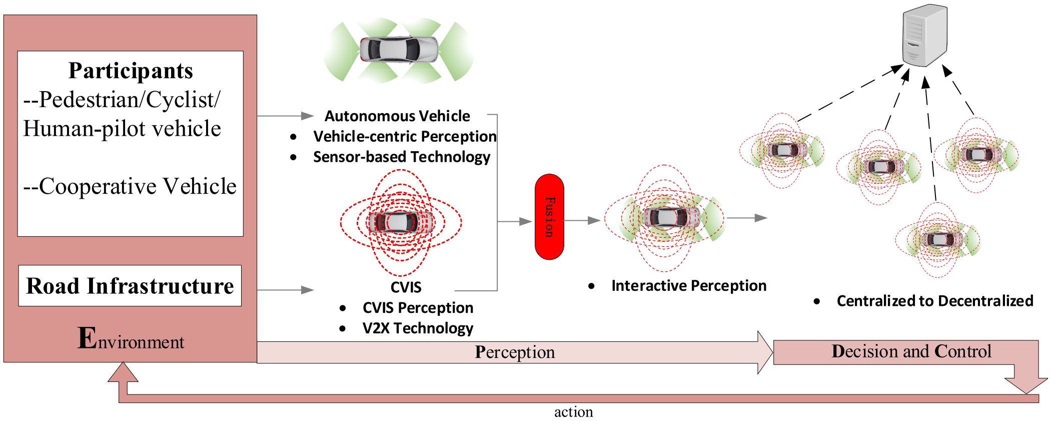

FIGURE 2. The structure diagram of Interactive Perception based Intelligent transportation system (IP-ITS).

perception applications: CVIS simulation platform and Virtual Reality (VR) test platform.

The rest of this paper is arranged as follows: In Section II, an overview of the state-of-art researches on key technologies involved in interactive perception is introduced. Section III introduces a case study about how to utilize the perception data from an autonomous vehicle and the interactive information from Vehicle to Vehicle (V2V) to achieve multiple object tracking. In Section IV, two simulation platforms for experimental verification are introduced: Cooperative Vehicle and Infrastructure System (CVIS) simulation platform and Virtual Reality (VR) test platform. In Section V, the experiments of the multiple objects tracking algorithm are implemented, analyzed and evaluated. Finally, in Section VI, the conclusion is drawn.

II. FRAMEWORK AND LITERATURE REVIEW

Fig. 2 is the schematic diagram of the Interactive Perception based Intelligent Transportation System (IP-ITS). Similar with the traditional ITS framework, it consists four main parts: environment, perception, decision and control. AV research focuses on vehicle-centric perception, ego decision-making and microscopic vehicle control, such as computer vision detection, autonomous route planning and collision avoidance. Different from the AV, in CVIS, V2V interactive information is the main way to provide the information of surroundings, and the optimization of the global traffic efficiency is considered as a critical goal under the premise of ensuring traffic safety. In IP-ITS, the interactive perception replaces the traditional vehicle-centric perception via fusion of vehicle-centric sensing data and V2X interactive data. The result of interactive perception is shared by various traffic participants for vehicle decision and traffic decision. The vehicle controller and traffic controller perform the target action given by the decision layer, and change the traffic state.

This section summarizes the key technologies involved in the interactive perception based ITS, and analyzes the state-of-art researches in these areas. Also, the relationship between AV and CVIS is discussed. And in Section III, an interactive perception-based multiple object tracking algorithm will be introduced in detail to verify that the proposed framework can improve the perception ability of vehicles.

A. VEHICLE-CENTRIC OBSTACLE DETECTION AND TRACKING

Obstacle detection and tracking are basic tasks in perception system of autonomous vehicle. Computer vision technologies are widely used to realize the perceptual functions of autonomous vehicles such as feasible area recognition, obstacle recognition and motion situation prediction [2], [3]. Convolutional Neural Network(CNN), a representative deep learning method, has replaced the traditional feature map, such as SIFT [4], HOG [5] and Haar-like [6] in most of the researches of image processing in these years. A detailed introduction on object detection using deep learning approaches can be found in [7]. However, in order to satisfy the requirement of an autonomous vehicle, it is essential to obtain the three-dimensional location of the surroundings.

In the area of three-dimensional object detection and tracking, active sensors such as Lidar and stereo camera are most common devices to detect obstacles [8]–[10]. Besides, there are lots of novel approaches using deep learning that have achieved excellent results. For example, Multi-View 3D network (MV3D) was proposed by Chen *et al.* [11]. MV3D fused Lidar and camera by projecting 3D coordinate into different 2D planes and generated multi-channel feature maps. Zan Gao et al. proposed a cognitive-inspired class-statistics matching method with triple-constraint (CSTC) for camera free 3D object retrieval [12]. AFCDL [13] fuses multiple of cameras and jointly learn the adaptive weight for each camera for human action recognition. VoxelNet is an end-to-end 3D object detection model, where the point cloud is encoded as a descriptive volumetric representation [14]. The image-based tracking algorithms provide some references for 3D tracking algorithm. Asvadi et al. projected 3D-Lidar data into a static 2.5D grid map to represent the local surrounding environment [14], [15]. A 2.5D motion grid is used to represent the moving status of surroundings, which generated the continuous static grid map. The 3D-BoundingBox (3D-BB) object models are fitted to 2.5D motion grids, followed by tracking of 3D-BB using Kalman Filter. In addition to vision-based methods, there are many studies that combine road constraints to optimize state prediction equations in the tracking process. The prior knowledge of trajectory shape constraint can be considered as a useful method to improve tracking performance. Gongjian Zhou et al. [16] proposed a trajectory shape constraint Kalman filter (TSCKF) for simultaneous filtering and smoothing.

However, due to the limitation of sensing range, it is difficult to solve the problems of weather interference and object occlusion by relying on the vehicle-centric perception system.

B. VEHICLE COMMUNICATION

A variety of heterogeneous networks are involved in V2X (Vehicle to everything) network, including mobile network, wireless personal area network/LAN and satellite network. In this heterogeneous network environment, each vehicle can be considered as a mobile node, which constitutes the Vehicular Ad hoc Networks called VANET. VANET is different from the general mobile ad hoc network due to its special application environment that leads to rapid changes of the network topology and short path life. Wireless communication devices, such as IEEE 802.11p/WAVE [17], [18] are used for VANET currently. The IEEE 802.11p and WAVE (Wireless Access for Vehicular Environment) standards form the DSRC (Dedicated Short Range Communication) for VANETs communications [19]. DSRC can support a 200 km/h vehicle speed, a wireless range between 300 and 1000 meters, and a theoretical bandwidth up to 6 to 27 Mbps [20]. The quality of the wireless channel of moving vehicle is unstable, and it is influenced by many factors, such as roadside infrastructures and road conditions. It is challenge to process large amount of real-time data for connected vehicles [21]. Xiaolong Xu *et al.* [22] applied NSGA-II (non-dominated sorting genetic algorithm II) to the Internet of connected vehicles to reduce the execution time and energy consumption and prevent privacy.

There have been many studies about the relationship among vehicle's movement, communication quality, communication security and interaction efficiency. Reference [23] proposed a swam model to describe the self-organised behaviour of the vehicle swarm in vehicular ad hoc networks and investigate the impact of vehicular communications on the mobility of multiple vehicles. S Ammoun et al. evaluated and predicted the risk of collision on a crossroad by using a standard 802.11 technology combined with a standard low-cost GPS receiver [24]. Reference [25] proposed Boneh-Boyen-Shacham, a short group signature-based reputation system, in order to improve road safety and efficiency.

C. VEHICLE POSITIONING

Global Navigation Satellite System (GNSS), GNSS/ Inertial Navigation System (INS) integration navigation system have been widely used in the vehicle navigation system. Generally, the integration navigation system performs better in accuracy and robust, which takes advantages of GNSS and INS to achieve high dynamic, real-time and high-precision position solutions, especially in dense urban areas. Loosely-coupled integration system and tightly-coupled integration system are two common integration navigation modes. Tightly-coupled integration exploits the raw information before Global Position Systems (GPS) data-computing in order to overcome the disadvantages of the loosely-coupled method that the positioning errors will accumulate quickly if the GNSS conditions are difficult [26]–[29].

For almost a decade, relative positioning attracted great attention. It can enhance the positioning accuracy by detecting the road-sign and provide accurate, reliable and continuous knowledge of the position of other traffic participants. Relative positioning methods can be roughly divided into two categories: Received Signal-based and Vision-based.

In Received Signal-based positioning, Radar/Lidar and Wireless Communications Systems can be used to measure the distance towards an object based on the time-of-flight of reflected light pulses. Received Signal Strength (RSS) is a simple method to make up for the unsuitable applications of GNSS by estimating signal transmitted by another vehicle. While RSS-based ranging has the main drawback of inaccuracy, which mostly originates from the uncertainty of the path loss exponent. Reference [30] proposed a method for dynamic estimation of the path loss exponent and distance based on the Doppler Effect and RSS. Other methods like Time of Arrival (TOA), and Time Difference of Arrival (TDOA), and direction-of-arrival (DOA) are also widely used for relative positioning.

In vision-based positioning, object detection and object tracking are served as foundations. Reference [31] proposed a IMU/Vision/Lidar integrated navigation system for providing relative navigation information in GNSS difficult environments. Reference [32] applied deep learning techniques using Camera, Lidar, Radar, and GPS data and achieved great performance of lane and vehicle detection. Yet, it is complicated to combine the detection result with the positioning knowledge.

D. CVIS AND AV

CVIS has gained great attention in recent years, because the intelligent vehicle based on vehicle-centric perception has limited sensing ability, high cost and limited synergy. CVIS attempts to establish an unmanned system based on V2X technology, in order to improve the reliability of the conflict detection, strengthen the system's coordinated operation efficiency, and transfer the cost from the user's vehicle to the city's infrastructure. Almost all the key technologies introduced above can be solved by the CVIS, such as

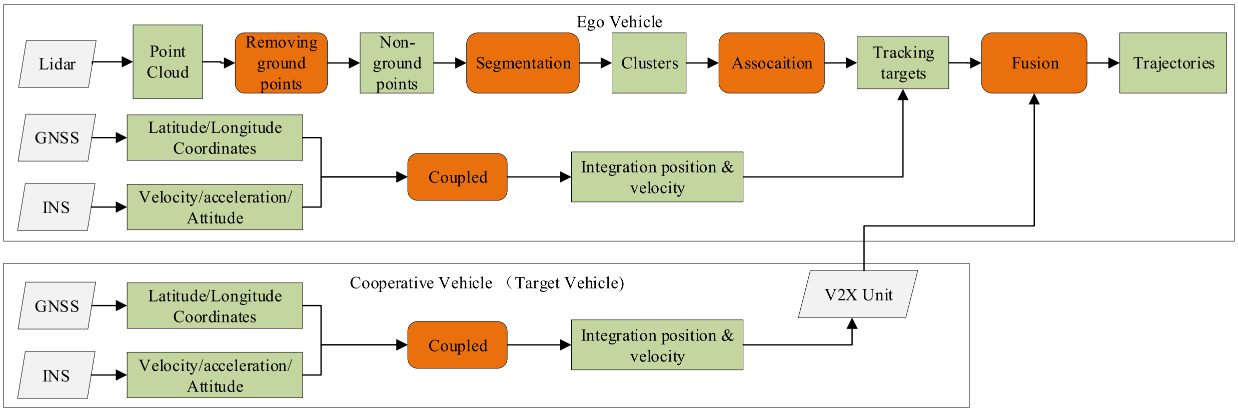
Ego Vehicle
Lidar
Point Cloud
Removing ground points
Non-ground points
Segmentation
Clusters
Assocaition
Tracking targets
Fusion
Trajectories
GNSS
Latitude/Longitude Coordinates
INS
Velocity/acceleration/ Attitude
Coupled
Integration position & velocity
Cooperative Vehicle （Target Vehicle)
V2X Unit
GNSS
Latitude/Longitude Coordinates
INS
Velocity/acceleration/ Attitude
Coupled
Integration position & velocity

FIGURE 3. **The flow chart of interactive perception-based moving object tracking algorithm.**

coordinated operation of intersections, coordinated visual collision avoidance, and vehicle speed guidance [33]–[36].

Generally, CVIS and AV are two mutually complementary and beneficial research directions. The superior vehicle-centric perception is regarded as the most important supporting for the AV, and the goal is to achieve the vehicle unmanned self-driving under all conditions. Different from AV, capturing the environmental information from roadside infrastructure or cooperative vehicles is the main way of perception in CVIS. Roadside infrastructures are used to forward and share traffic information among traffic participants and traffic managers. It seems impossible for a CVIS-based intelligent vehicle to achieve fully driverless in wild environment, but it can achieve obstacle avoidance and efficiency scheduling in a smart city.

In this paper, the interactive information provided by the CVIS and the visual knowledge captured by the AV are combined to achieve a highly reliable perception system.

E. TESTING

To ensure transport safety, various tests that fully cover the actual scenes in a controllable environment are needed before large-scale commercialization. Hence, it is necessary to build a complete and credible test system based on real-world scenarios to strictly verify the safety of the vehicle and comprehensively evaluate the collaboration capabilities of CVIS. The scale of the construction and demonstration of intelligent vehicle test sites in US, EU and China is continuously expanding. In this aspect, representative projects are Mcity in Michigan [37] and GoMentum in California [38]. Beijing, Shanghai, Shenzhen and other places in China are actively building intelligent networked vehicle test sites, involving safety, efficiency, transportation, information services, mobile services and other pilot projects.

However, the test scenarios in a closed real test site are very limited compared with the open-road environment. It has been estimated that if an autonomous vehicle drives at least 275 million miles without fatality, it could be assured that it has achieved the same level of reliability as a human-driver. Hence, various testing and calibration environments become necessary to complement the test requirements, because the time cost must to be considered. References [39]–[41] applied different control models to verify and validate the control systems of autonomous vehicles in terms of safety and performance. [42] introduced how to verify the CVIS using interactive visual simulation. Chang'an University is building a CVIS test bed which can simulate various connected vehicle scenarios with reliable heterogeneous vehicular network and support various related technologies [43].

Hardware In Loop (HIL) simulation provides a highly realistic laboratory environment to support the development, test, and verification of AV [44]–[47]. dSPACE and Carsim system are widely used in this area. How to use the HIL simulation to implement as many as possible test functions and how to employ the emerging technologies such as virtual reality or augmented reality to form a mixed test environment are the key issues for the CVIS and AV testing.

We developed a CVIS simulation platform and a VR test platform which will be introduced in the Section IV. The two platform are united to provide a verification environment for our proposed IP-MOT algorithm.

III. INTERACTIVE PERCEPTION-BASED MULTIPLE OBJECT TRACKING

A. SYSTEM OVERVIEW AND PROBLEM REPRESENTATION

The flow chart of the IP-MOT algorithm is shown in Fig.3, where gray diamonds represent sensor devices; green rectangles represent data types; and orange rounded rectangles represent data processing method. Two types of data are used in our IP-MOT approach: point cloud data captured by Lidar and interactive information about the status of cooperative vehicle received by Vehicle to Vehicle (V2V) communication unit. These two types of data are fused to enhance the perception ability of an ego vehicle.

The ego vehicle tracks the surrounding obstacles via Lidar-only multiple object tracking algorithm (L-MOT). The L-MOT algorithm basically consists of three steps as introduced in Section III-B and Section III-C. First, the ground points as well as other noise points are removed as the preprocess of the raw data. Second, the non-ground points are segmented into clusters for object representation. Third, the ego vehicle obtains the integration position and velocity by fusing the position information from the GNSS with the attitude information from the INS, and the relative position of vehicles are mapped into global coordinate system. At the same time, all the cooperative vehicles share their status to surrounding vehicles. The ego vehicle obtains the absolute position of the cooperative vehicle and matches the visually detected target vehicle with the information received from the cooperative vehicles. The wireless information interactive process is introduced in Section III-D. Finally, as introduced in Section III-E, IP-MOT fuses the result of L-MOT with the interactive information in the global coordinate system to generate the accuracy trajectories.

Both of the loosely coupled and the tightly coupled positioning algorithm have been introduced in the previous work [48]. In this section, we will introduce the multiple object tracking algorithm using Lidar and the fusion algorithm to enhance the perception accuracy using V2V.

B. FAST POINT CLOUD SEGMENTATION BASED ON VOXEL CLUSTER

1) REMOVING GROUND POINTS

Ground points typically constitute a large portion of raw point cloud data and form a large ground plane. RANSAC is used for fitting the ground plane in our method. RANSAC was proposed by Fischer and Bolles [49] in 1981 and mainly involves performing two iteratively repeated steps on a given point cloud: generating a hypothesis and verification [50]. First, a set of points are selected randomly and a hypothesis model is generated on basis of these points. Then, the rest of points are tested with the hypothesis model to verify how many points are fitted with the model. After numbers of iterations, the ground plane, which has a normal closed to z-direction and the largest number of interior points, is extracted and removed.

2) VOXEL CLUSTER ALGORITHM

The non-ground points are clustered into sub-point cloud to represent a specific object by the voxel cluster algorithm. Voxel is a cube which contains points data and it is indexed uniquely according to the spatial position. A trusted sensing area captured by Lidar with L in y axis, W in x axis and H in z axis is truncated from the raw data. Each voxel grid is binarized based on the number of the points. If a voxel contains more than Thv points, it is a positive voxel, otherwise, it is a negative voxel. Point $P(x, y, z)$ belongs to the kth voxel according to Eq.1-4, where vL, vW, vH is the length, width, and height of the voxel. For simplicity, we assume L, W, H are a multiple of vL, vW, vH, and $L' = L/vL,\ W' = W/vW,\ H' = H/vH$.

$$d_l = \lfloor x/vL \rfloor \tag{1}$$

$$d_w = \lfloor y/vW \rfloor \tag{2}$$

$$d_h = \lfloor z/vH \rfloor \tag{3}$$

$$k = d_l * W' * H' + d_w * H' + d_h \tag{4}$$

Each voxel directly connects with other 26 voxels in space. We define a binding relationship if two positive voxels are directly adjacent, and the two voxels are binding voxels for each other. Positive voxels are accessed in ascending order of index. If there is no binding voxel or all the other binding voxels are not visited yet, a new label will be created. Else, a voxel is segmented into a cluster which has minimum label among all the binding voxels.

Fig.4 is a simple case for illustration the segmentation process. In this example, the space is divided into 8*3*3 voxels and all the voxels are indexed ranging from 0 to 71. In Fig.4, only positive voxels are visible. All the positive voxels will be visited in turn and the white cubes represent those are not visited yet. As shown in Fig.4(a), new labels are created and distributed to the voxel indexed 9 and 15 respectively, because they don't have another classified binding voxel. The 9th voxel joins the cluster which is labeled 1 and colored green, and the 15th voxel joins the cluster which is labeled 2 and colored orange. For the 33th voxel, there is only one labeled binding voxel, the 9th voxel, so the label 1 is inherited to the 33th voxel. Similarly, the 37th voxel is labeled to the cluster which is labeled 2. When the 60th voxel is visited, there are three labeled binding voxels indexed 59, 53, 57 with different labels, 1 and 2, as shown in Fig.4(b). The smallest label will be choosen when conflict occurs, so the 60th voxel joins the 1th cluster and all the voxels in the 2th cluster are relabeled to cluster 1. As shown in Fig.4(c), the cluster labeled 1 and the cluster labeled 2 are merged with label 1. Finally, all the voxels are classified into one cluster as shown in Fig.4(d), and the algorithm will end until all the positive voxels have been visited.

The only parameter of the segmentation algorithm is the size of the voxel. The performance of the algorithm is influenced by the parameters. The parameters can be set according to the system resolution, which refers to the threshold distance of two object. In the open-sky road environment, the parameters are set to $vL = vW = vH = 0.2m$.

The contributions of Voxel Cluster (VC) can be summarized as follows:

1) Converting point clouds segmentation into regular voxels clustering. Raw non-ground points are visited just once to find positive voxel, after that, the cost of the calculation is greatly reduced, because only the positive voxel will be calculated.

2) VC algorithm has only one parameter, the size of voxel, representing the maximum distance between two binding voxels.

3) VC algorithm is suitable for segmentation of point cloud in outdoor condition with sparse distribution without pre-training.

C. MULTIPLE OBJECT TRACKING

A 3D shape descriptor is defined and employed for extracting the feature vector of the point cloud cluster which represents

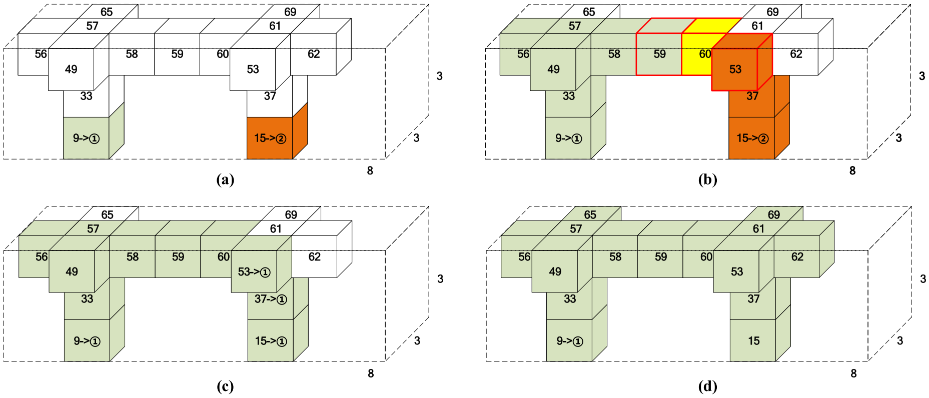

FIGURE 4. A case of getting object clusters using VC.

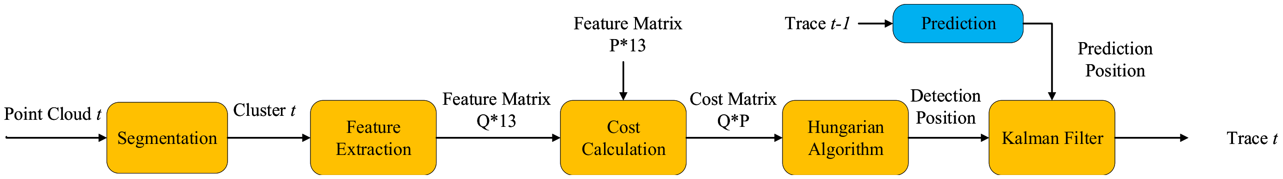

FIGURE 5. The flow chart of Lidar-based multiple object tracking.

an object. Seven shape factors are involved in the feature vector, as defined in Eq.5, where n_p is the number of the points falling into the cluster, n_v is the number of the positive voxels, $mean_p$ and var_p are the mean and variance of the points, centroid is the position of the centroid, $mean_i$, var_i are the mean and variance of the reflectance intensity. The feature of a cluster is described by using a 13 dimensional vector, where $mean_p$, centroid, and var_p are 3 dimensional vectors.

$$F\left(C\right) = \left[n_p, n_v, mean_p, mean_i, centroid, var_p, var_i\right] \quad (5)$$

We stress here that designing the best feature descriptor for 3D point clouds is not the main focus of this work. However, the simple feature vector we have chosen achieves a good tracking performance in our experiments.

The multiple object tracking issue can be regarded as an association problem which assigns the detection objects from the current frame to the historical tracks. Fig. 5 shows the flow scheme of our Lidar-only multiple object tracking (L-MOT) algorithm. First, the raw point cloud is divided into clusters for describing the targets. Then, the feature vectors of the Q targets are extracted and concatenated for generating the feature matrix. The distance of feature vectors between the last and current frame are calculated. The cost matrix is fed to the Hungarian algorithm [51] for assigning the ID to targets. Besides, the position of the targets in the last frame is predicted on the basis of constant moving speed. Finally, the Kalman Filter evaluates the trace of these target by fusing the predicted position and the detected position.

A weighted Euclidean distance is used to value the similarity between the feature vectors. The weight parameters are trained by a simple neural network model, as shown in Fig.6. For generating the training set, the feature vectors of an object in two continuous frames are collected and m-1 feature vectors of other objects are also collected as negative samples. The input samples are standardized according to the Eq.6, where $x^i_{t-1,1}$ represents value of the first feature factor of the ith object at time $t-1$, and $x^j_{t,1}$ represents the generated the value of the first feature factor of the *jth* object at time t. $d(x^j_{t,1}, x^i_{t-1,1})$ is standardized input which has been adjusted to [0,1], representing the distance between the first feature factor of the *ith* object at time $t-1$ and the *jth* object at time t.

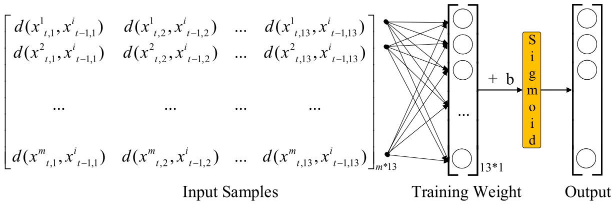
$d(x^1_{t,1}, x^i_{t-1,1})$ $d(x^1_{t,2}, x^i_{t-1,2})$ ... $d(x^1_{t,13}, x^i_{t-1,13})$
$d(x^2_{t,1}, x^i_{t-1,1})$ $d(x^2_{t,2}, x^i_{t-1,2})$ ... $d(x^2_{t,13}, x^i_{t-1,13})$
$d(x^m_{t,1}, x^i_{t-1,1})$ $d(x^m_{t,2}, x^i_{t-1,2})$ ... $d(x^m_{t,13}, x^i_{t-1,13})$
m*13
13*1
+ b
Sigmoid
Input Samples
Training Weight
Output

FIGURE 6. An artificial neural network model for training feature weights.

$$d\left(x^j_{t,1}, x^i_{t-1,1}\right) = \frac{\left\|x^j_{t,1}, x^i_{t-1,1}\right\| - \overline{x_1}}{Var\left(x_1\right)}, \quad j \in (1, m) \quad (6)$$

The output of the model is activated by the sigmoid function, which is defined as Eq.7. The label of the model is a one-hot vector, and only the ith value is set to 1. Stochastic Parallel Gradient Descent algorithm is used to train the weight parameters.

$$S(x) = \frac{1}{1+e^{-x}} \tag{7}$$

D. WIRELESS INTERACTIVE INFORMATION

The communication mode fusion mechanism proposed in the previous work [52] is employed to transfer the interactive information in this paper. We use three types of communication mode: 4G, WLAN and WAVE to support our multi-mode communication mechanism. In the communication process, a key factor which affects the accuracy of the information is communication delay.

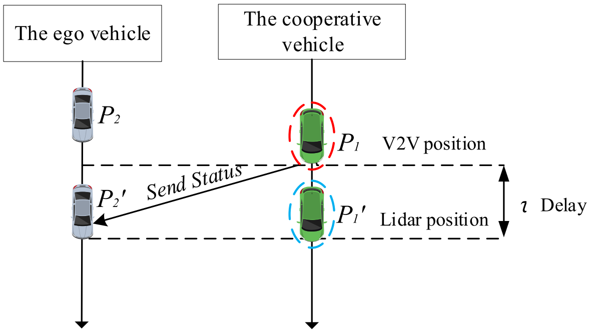

FIGURE 7. Schematic diagram of the influence of communication delay on vehicle information interaction.

As shown in Fig.7, the information obtained by the V2V from the cooperative vehicle is delayed by time τ. It is necessary to consider the delay time in the process of information fusion. Therefore, the information received by the ego vehicle is the status of the cooperative vehicle at time $t-\tau$. To obtain the delay time τ, we test the performance of our communication mechanism in simulation environment. Various traffic scenarios with different vehicle densities, velocities, and data volumes are tested, and the simulation result is shown in Table 1.

The setting of the delay parameter is based on the factor that has the greatest impact on the delay. To simplify calculations, the value of the three factors is classified into four levels respectively, and the average delay time of each level is tested as Table 1. For example, if the vehicle density is 55veh/km, vehicle velocity of the ego vehicle is 35km/h, and the data volumes is 2200bits/time, the maximum value of 0.097, 0.111, 0.104 is set to the delay time τ.

TABLE 1. Comparison of different communication mode.

(a) The delay time of different communication mode with different vehicle densities.

Dens.(veh/km) / Mode	50	150	250	400
3G	0.097	0.117	0.125	0.137
WLAN	0.105	0.107	0.109	0.115
WAVE	0.107	0.114	0.111	0.113
Fusion	0.097	0.107	0.109	0.113

(b) The delay time of different communication mode with different vehicle speeds.

Speed(km/h) / Mode	10	20	30	40
3G	0.097	0.107	0.116	0.128
WLAN	0.099	0.104	0.117	0.124
WAVE	0.102	0.106	0.111	0.117
Fusion	0.097	0.104	0.111	0.117

(c) The delay time of different communication mode with different data volumes.

Volume(bps) / Mode	1000	2000	3000	4000
3G	0.097	0.107	0.116	0.128
WLAN	0.099	0.104	0.117	0.124
WAVE	0.102	0.106	0.111	0.117
Fusion	0.097	0.104	0.111	0.117

E. INFORMATION FUSION

Similar to the Lidar based multiple object tracking algorithm, the information fusion process consists of two main steps: track association and track fusion. First, it is necessary to obtain the relationship between the Lidar sensing data and interactive information. Second, objects trajectories are generated by the fusion of two tracks. The state vector of Lidar and V2V communication are defined as Eq.8 and Eq.9, respectively. p_x and p_y are the coordinates of the target in the plane coordinate system. β^{ego} and β^{coo} are the yaw rate of the ego vehicle and the coordinated vehicle (target vehicle), respectively. v and a are the velocity and acceleration of the target vehicle. L, W, H represent the length, width and height of the target vehicle,respectively.

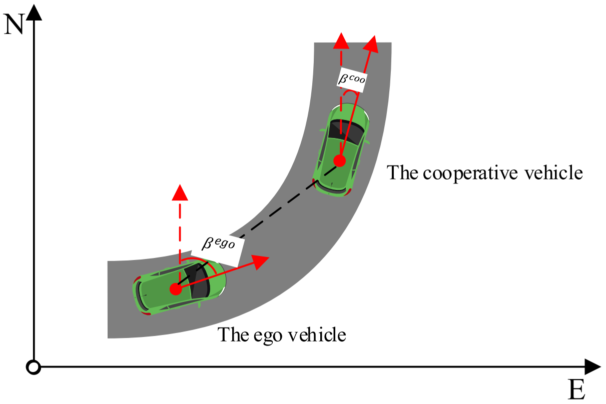

FIGURE 8. Schematic diagram of the relative position of the ego vehicle and the cooperactive vehicle.

$$\begin{aligned} x^{lidar} &= [p_x^{lidar}, p_y^{lidar}, \beta^{ego}, v_x^{lidar}, v_y^{lidar}, a_x^{lidar}, a_y^{lidar}, \\ &\quad L^{lidar}, W^{lidar}, H^{lidar}] \end{aligned} \tag{8}$$

$$\begin{aligned} x^{v2v} &= [p_x^{v2v}, p_y^{v2v}, \beta^{coo}, v_x^{v2v}, v_y^{v2v}, a_x^{v2v}, a_y^{v2v}, L^{v2v}, \\ &\quad W^{v2v}, H^{v2v}] \end{aligned} \tag{9}$$

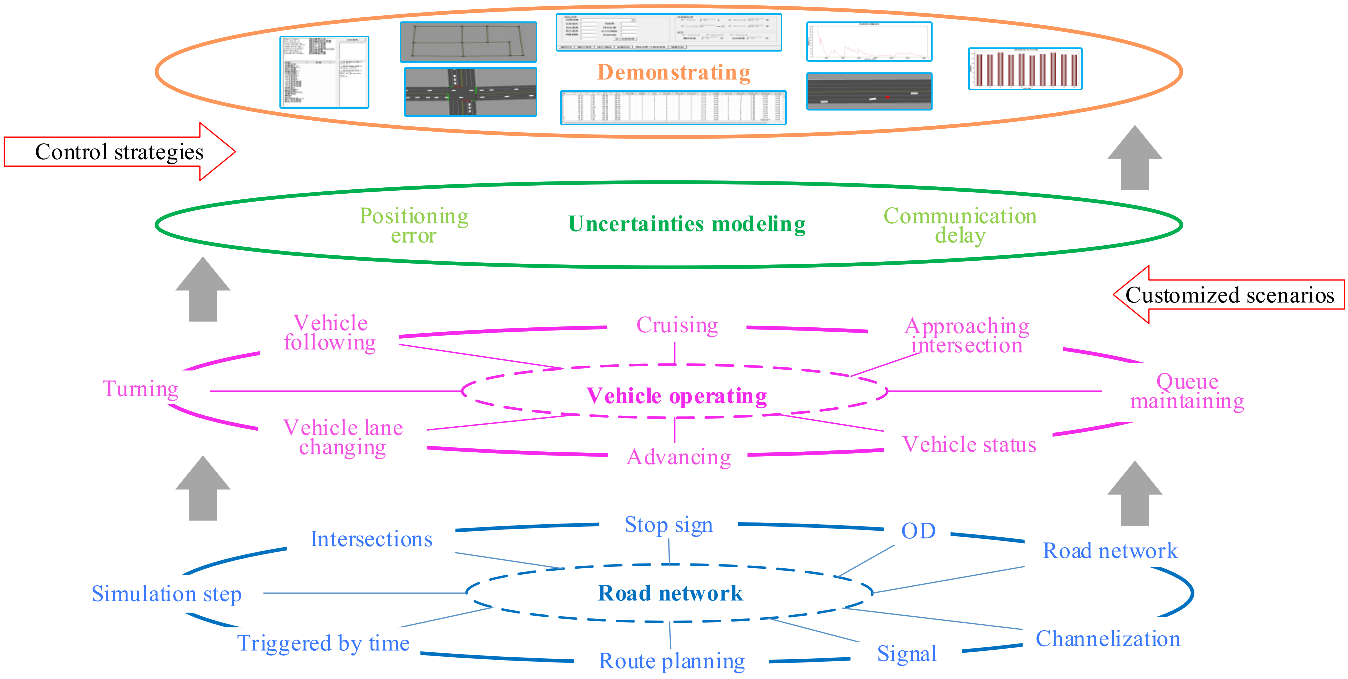

FIGURE 9. **The structure diagram of CVIS simulation platform.**

For the ego vehicle, target's position $\left[p_x^{lidar}, p_y^{lidar}\right]$ in the plane coordinate system is converted from the relative position $\left[p_x^{re-lidar}, p_y^{re-lidar}\right]$ in Lidar coordinate system as Eq.10, where $\left[p_x^{ego-lidar}, p_y^{ego-lidar}\right]$ is the global position of the ego vehicle. For the coordinated vehicle, the global position $\left[p_x^{v2v}, p_y^{v2v}\right]$ is converted from $\left[p_x^{coo-v2v}, p_y^{coo-v2v}\right]$, which is the position in the local coordinate system as Eq.11. The global positions of the vehicles are converted from the GPS plane coordinate.

$$\begin{bmatrix} p_x^{lidar} \\ p_y^{lidar} \end{bmatrix} = \begin{bmatrix} p_x^{re-lidar} \\ p_y^{re-lidar} \end{bmatrix} + \begin{bmatrix} \cos\beta^{ego} & \sin\beta^{ego} \\ -\sin\beta^{ego} & \cos\beta^{ego} \end{bmatrix} \begin{bmatrix} p_x^{ego-lidar} \\ p_y^{ego-lidar} \end{bmatrix} \tag{10}$$

$$\begin{bmatrix} p_x^{v2v} \\ p_y^{v2v} \end{bmatrix} = \begin{bmatrix} \cos\beta^{coo} & \sin\beta^{coo} \\ -\sin\beta^{coo} & \cos\beta^{coo} \end{bmatrix} \begin{bmatrix} p_x^{coo-v2v} \\ p_y^{coo-v2v} \end{bmatrix} \tag{11}$$

It is essential to determine whether two tracks represent the same target before the information fusion. In this paper, two tracks with the minimum mean square error of position, velocity, acceleration and shape size are considered to correspond to the same target. Reference [53] provided a classic method to fuse two track state vectors into a new estimated state vector based on the Kalman Filter. Reference [54] combined two estimates from the radar and V2V communication based on a Bayesian minimum mean square error (MMSE) criterion. Based on [53], [54], the measurement model of Lidar and V2V are defined as Eq.12 and Eq.13, where τ is the communication delay time.

$$z^{lidar}(k) = H^{lidar}x^{lidar}(k) + W^{lidar}(k) \tag{12}$$

$$z^{v2v}(k) = H^{v2v}x^{v2v}(k-\tau) + W^{lidar}(k-\tau) \tag{13}$$

The two state vectors are fused according to the following fusion equation:

$$\hat{x}(k) = x^{lidar}(k) + P_1 P_2^{-1}\left(x^{v2v}(k) - x^{lidar}(k)\right) \tag{14}$$

$$P_1 = P^{lidar}(k) - P^{lidar,v2v}(k) \tag{15}$$

$$P_2 = P^{lidar}(k) + P^{v2v}(k) - P^{lidar,v2v}(k) - P^{v2v,lidar}(k) \tag{16}$$

where $P^{lidar,v2v}$ is the cross-covariance matrix between x^{lidar} and x^{v2v}. P^{lidar} and P^{v2v} are error covariance matrix of Lidar and V2V, respectively.

IV. SIMULATION TEST ENVIRONMENT

Two platforms are designed and implemented to provide the test and verification environment for our proposed approaches: CVIS simulation platform and Virtual Reality (VR) test platform. CVIS simulation platform provides an integrated simulation environment which consists of 4 modules: traffic flow simulation, V2V communication simulation, three-dimensional simulation and evaluation. Virtual Reality (VR) receives the traffic data generated by the CVIS platform, and simulates the micro-movement of the controlled vehicle. The VR platform replies the sensing information of the ego vehicle, such as image data captured by the simulated camera or point cloud data captured by

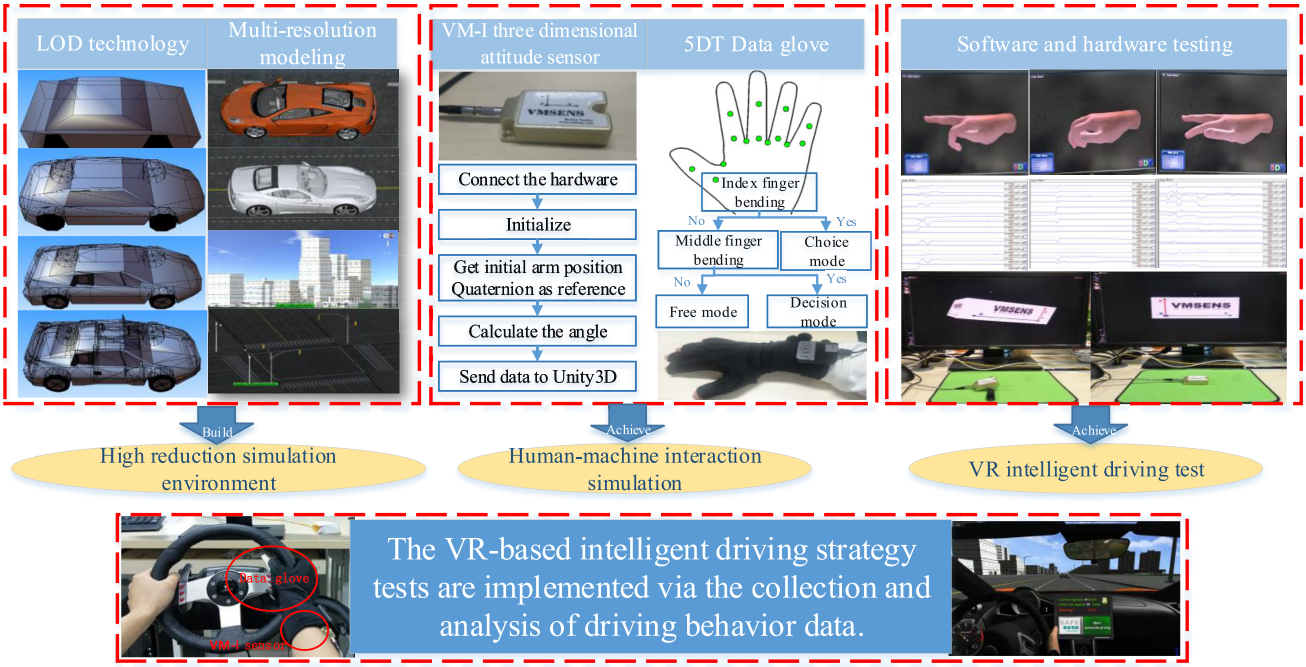
LOD technology
Multi-resolution modeling
VM-I three dimensional attitude sensor
5DT Data glove
Software and hardware testing
Connect the hardware
Initialize
Get initial arm position Quaternion as reference
Calculate the angle
Send data to Unity3D
Index finger bending
No
Yes
Middle finger bending
Choice mode
No
Yes
Free mode
Decision mode
Build
Achieve
Achieve
High reduction simulation environment
Human-machine interaction simulation
VR intelligent driving test
Data glove
VM-I sensor
The VR-based intelligent driving strategy tests are implemented via the collection and analysis of driving behavior data.

FIGURE 10. The structure diagram of VR based test platform.

simulated Lidar sensor. Besides, the VR platform can involve human drivers into the scenarios. Also, it can embed different intelligent levels of control methods for simulated vehicles.

The CVIS simulation platform generates traffic scenario and provides traffic data. The VR simulation platform executes traffic scenes with high fidelity, simulates the sensor data and verifies the proposed multiple object tracking approaches. The two simulation platforms communicate through TCP protocol.

A. COOPERATIVE VEHICLE INFRASTRUCTURE SYSTEM (CVIS) SIMULATION PLATFORM

The CVIS simulation platform applies a hierarchical simulation framework, which consists of road network layer, vehicle operating layer, uncertainties modelling layer and demonstrating layer [55], as shown in Fig.9.

The road network layer provides the basic information of the road network, such as intersection's parameters, signal control strategy and the type of road. In vehicle operating layer, vehicles generated from zone and intersections move according to the vehicle kinematic models, such as vehicle following and lane changing models. In CVIS, vehicle position error and communication delay always exist. Positioning error and communicating delay models are implemented in error modelling layer to enhance the reality of the simulation. The demonstrating layer is for visualization and evaluation.

B. VIRTUAL REALITY (VR) BASED TEST PLATFORM

The real test site has the properties of high cost, limited scenario and long construction period. To address the above issues, virtual test has attracted a lot of attention recently. We have established an interactive intelligent driving simulation platform in virtual reality environment as shown in Fig.10, which realizes a high realistic interactive intelligent driving simulation experience. The VR based test platform provides a more safety and time-saving test option for an autonomous vehicle before a real physical test bed.

The Virtual Reality (VR) based test platform consists of two parts: high realistic simulation environment and interactive human driver interface. A three-dimensional real-time simulation environment is implemented, which covers the road network, urban elements and vehicles. The Logitech G27 racing steering wheel series hardware is used to achieve high immersion intelligent driving. With the VM-I inertial sensor and 5DT data glove, information about the driver's arm that operates the vehicle is dynamically collected in real time. A communication delay model is used to simulate the information interactive process between vehicles. A series of high-fidelity driving scenarios are implemented, including of multiple driving behaviors and multiple intelligent level vehicles.

The influence of communication quality to the test result is taken into consideration. Besides, SQL database provides the services for the real-time drivers' multi-scene driving behavior data.

The application of the intelligent driving visual simulation platform in this virtual environment can be summarized as the following three aspects: First, providing an online low-cost test environment for autonomous vehicle. Second, providing intelligent driving data acquisition and analysis methods for developers. Finally, it can also provide a high immersion training platform for drivers.

V. EXPERIMENT AND ANALYSIS

The proposed interactive perception based multiple object tracking algorithm is evaluated in open dataset and our simulation platforms. We test our Lidar-based multiple object tracking algorithm using KITTI [56] dataset. However, the public dataset cannot provide the interactive information required in IP-MOT. Therefore, our simulation platforms, CVIS and VR simulation platforms, are used to evaluate our IP-MOT algorithm. The experiments were performed using a 2 core 2.9GHz processor with 4GB RAM under C++, and Point Cloud Library (PCL) is imported for visualization.

A. EVALUATION OF LIDAR-BASED OBJECT TRACKING

To verify the efficiency of proposed methods, we tested our algorithm on Object Tracking set extracted from KITTI dataset. The KITTI object tracking benchmark consists of 21 training sequences. 10 representative video sequences are extracted from the object tracking challenge set. The basic scene information can be found in Table 1.

For multiple object tracking evaluation, two indicators are used to evaluate the performance of the algorithm: multiple object tracking precision (MOTP) and multiple object tracking accuracy (MOTA) [57]. MOTP indicates the total Euclidean distance of the center of the 3D-BB between the tracking result and the ground truth of all the matching pairs over all frames, and it is averaged by the total number of pairs, as shown in Eq.17. It is an indicator that shows the center tracking ability of the tracker.

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \tag{17}$$

MOTA takes all the error elements into account, as shown in Eq.18. It consists of 3 error elements: the number of misses as Eq.19, false positives as Eq.20, and mismatches as Eq.21.

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \tag{18}$$

$$\overline{m} = \frac{\sum_t (m_t)}{\sum_t g_t} \tag{19}$$

$$\overline{fp} = \frac{\sum_t fp_t}{\sum_t g_t} \tag{20}$$

$$\overline{mme} = \frac{\sum_t mme_t}{\sum_t g_t} \tag{21}$$

The purpose of the L-MOT algorithm is to track the environmental vehicles or pedestrians during the running of the ego vehicle, and to reconstruct their absolute trajectories with the aid of the positioning system. A brief intuitive description of tracking results can be found in Fig.11, in which shows the positions and trajectories of the ego car and the tracking targets in different frames and the label of the target is colored red. Fig.12 shows the center RMSE (Root Mean Squared Error) between the groundtruth and the tracking results to see the effectiveness of the proposed L-MOT and 3D-PF algorithms. These two algorithms could successfully track the cyclist labeled 67 and the van labeled 48 at the same time in a long time. The L-MOT achieves outstanding performance that the average center error is about 0.315m (5.8 pixels) and the MOTA is about 99.3%. The red, blue, and green curves represent the trajectories of the collecting vehicle, the cyclist and the van respectively.

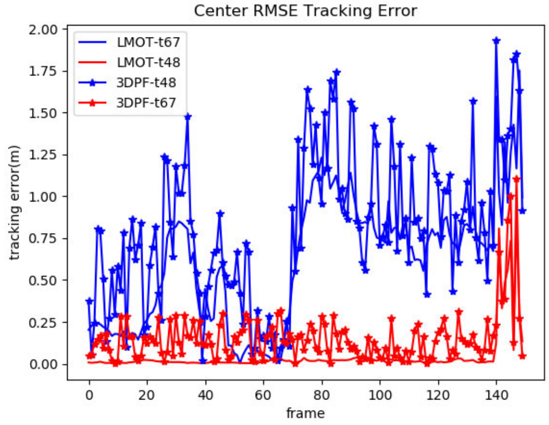

FIGURE 11. The tracking difference over all the tracking time of L-MOT and 3D-PF.

Table 2 illustrates 10 representative experimental results. No. means the index of the collected sequence in the KITTI tracking training dataset. A new multiple target tracking algorithm, Multi-LiDAR Based Multiple Object Detection and Tracking (MODT) which was proposed by Muhammad Sualeh et.al in 2019 [58], is compared with our method too. The result shows that the proposed Lidar-based multiple object tracking (L-MOT) algorithm outperforms 3D-Particle Filter (3D-PF) [59] algorithm and MODT algorithm in most of time. All of these three algorithms can achieve tracking the target in real time. The 3D-PF needs the initial position information of the targets. The L-MOT algorithm segments all the objects in the point clouds, which means no prior knowledge is needed.

The result shows that the proposed L-MOT algorithm can achieve accurate tracking of multiple targets in an open road environment. However, when there is severe occlusion, or the target is far away from the ego vehicle, the accuracy of tracking still cannot guarantee. Actually, there are many factors that affect the tracking results. Table 2 only lists some common influencing factors. One key factor is the density of the point clouds, when the target reflection point cloud is sparse, the Lidar-only based tracking method does not perform well. For example, the accuracy of the 5th scenario is much lower than the 1st and 2nd scenarios. When a van is far away from the ego vehicle, few reflect points can be received by the ego vehicle, because the surfaces of van is covered by glasses, which do not reflect laser. Besides, because the metal casing at the rear of the vehicle has more area than the front and glass windows hardly reflect the laser, therefore the target vehicle

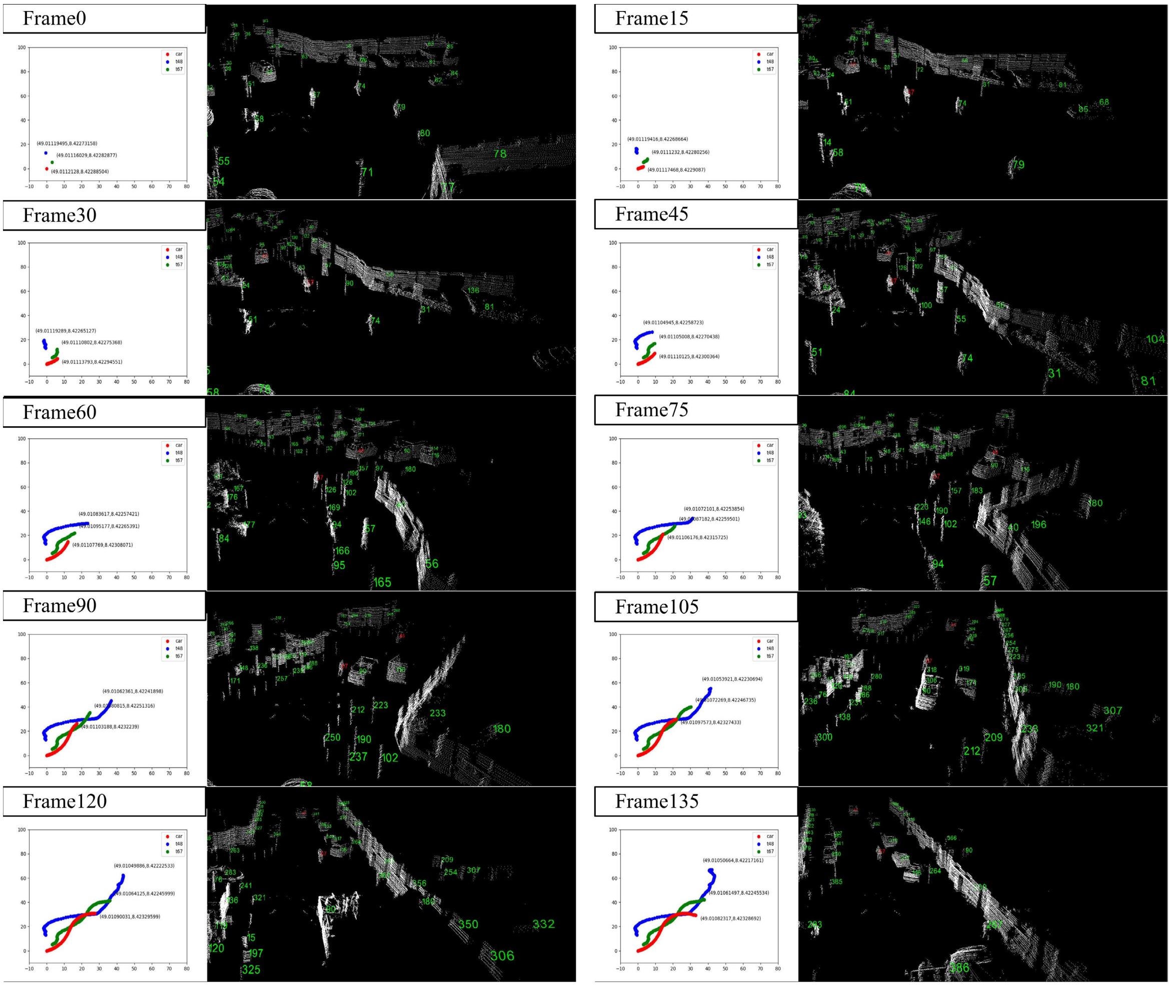

FIGURE 12. The result of multiple object tracking using L-MOT tested on KITTI.

in same direction with the ego vehicle will reflect more and denser point clouds.

B. EVALUATION OF IP-MOT

The interactive perception-based multiple object tracking algorithm is tested in simulation platform which has been introduced in Section IV. Two scenarios are designed and implemented to test our IP-MOT algorithm. In these scenarios, the ego vehicle perceives the position of surrounding vehicles through the simulated radar equipment. Cooperative vehicles in the scenario broadcast their status to nearby vehicles and communication delay model which has been introduced in Section III is employed to V2V-based information interaction. And the other non-cooperative vehicles are tracked via L-MOT algorithm.

Scenario 1 is a lane keeping scene as Fig. 13, in which the ego vehicle runs on a straight road. Scenario 2 is an intersection scene as Fig. 14, in which the ego vehicle is waiting at the stop line to cross the intersection. The parameters of simulation scenarios are shown in Table 3. The first scenario is a high-speed scenario, where all the vehicles keep moving fast. The second scenario is a low-speed scenario, where vehicles are slow down at the intersection and the ego vehicle stopped at the stop line in the first 100s, then accelerated to leave the intersection. Besides, all vehicles have a 20% lane change probability.

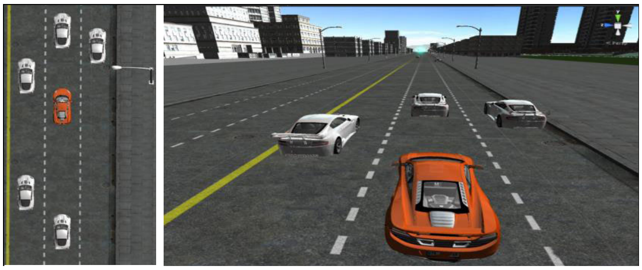

FIGURE 13. Lane keeping scenario in VR platform.

Fig. 15 and Fig. 16 show the MOTA value of the ego vehicle in lane keeping scene and intersection scene respectively using different multiple object tracking algorithm. It can be

TABLE 2. Comparison of lidar-only multiple object tracking results among L-MOT/3D-PF/MODT.

No.	Frame Count	Objects	Alg.	MOTP(m)	MOTA
1	114	11	L-MOT	0.34	96.3%
			3D-PF	0.67	82.6%
			MODT	1.10	79.2%
2	83	2	L-MOT	0.42	94.7%
			3D-PF	0.41	92.5%
			MODT	0.40	88.4%
5	160	15	L-MOT	0.38	90.6%
			3D-PF	0.52	79.1%
			MODT	0.75	88.9%
9	453	79	L-MOT	0.54	78.2%
			3D-PF	0.87	77.5%
			MODT	1.7	75.8%
13	150	2	L-MOT	0.29	98.3%
			3D-PF	0.32	89.5%
			MODT	0.3	96.1%
17	120	4	L-MOT	0.34	97.0%
			3D-PF	0.72	86.60%
			MODT	1	96.7%
18	276	11	L-MOT	0.42	88.6%
			3D-PF	0.59	86.1%
			MODT	0.82	92.6%
48	28	5	L-MOT	0.82	81.20%
			3D-PF	1.2	63.30%
			MODT	1	86.5%
51	444	37	L-MOT	1.1	83.20%
			3D-PF	1.4	74.50%
			MODT	1.3	80.6%
57	367	13	L-MOT	0.43	95.1%
			3D-PF	0.52	87.6%
			MODT	0.72	93.14%

TABLE 3. Simulation parameters.

Para. Scen.	Target number	Ego vehicle's velocity	Average velocity	Lane change rate	Tracking time
Lane keeping	5	10m/s	10.5m/s	0.2	120s
Intersection	25	2.5m/s	4.2m/s	0.2	120s

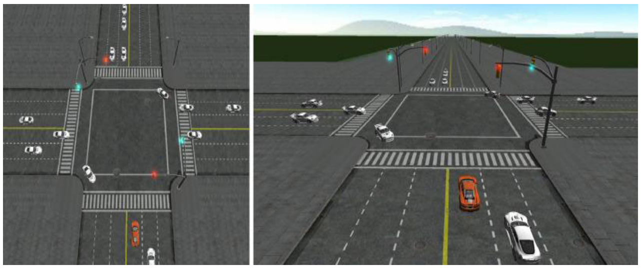

FIGURE 14. Intersection scenario in VR platform.

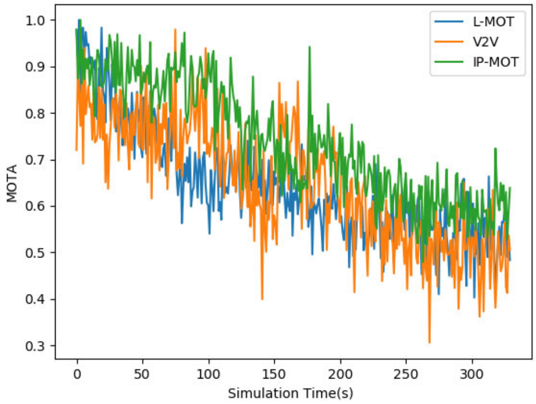

FIGURE 15. Comparison of tracking performances of IP-MOT/L-MOT/V2V using lane keeping scenario.

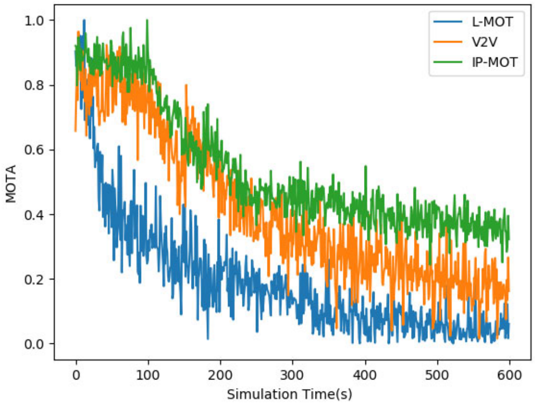

FIGURE 16. Comparison of tracking performances of IP-MOT/L-MOT/V2V using intersection scenario.

seen from the result that the proposed IP-MOT algorithm outperforms the Lidar only and V2V only approaches in both of the two scenarios.

Table 4 shows mean, variance and standard deviation of MOTA and RMSE error in these scenarios. Since the error is infinitely far if the tracking algorithm loses the target, such as V2V communication failure, we compare the central Euclidean distance in the case of successful tracking the target, as shown in the RSME column in the table.

As shown in Table 4, V2V-based method has the best accuracy, because the information from the cooperative vehicle's self-perception has a higher precision. However, if the communication is failed, targets will be lost. L-MOT and MODT are similar in RMSE. Because there is difference between the target's geometrical center and the point clouds' center, the center error of Lidar-only based methods is bigger than V2V-based methods. The results show that the proposed IP-MOT can improve the accuracy of the multiple object tracking of an ego vehicle, especially in tracking

TABLE 4. Comparison of multiple object tracking results between L-MOT and IP-MOT.

No.	Alg.	RSME	MOTA	Var.	Std.
1	L-MOT	0.41	39.35%	0.0349	0.1867
	V2V-MOT	0.27	41.62%	0.0530	0.2301
	IP-MOT	0.34	53.60%	0.0358	0.1894
	MODT	0.67	42.69%	0.0417	0.1927
2	L-MOT	0.34	74.50%	0.0151	0.1229
	V2V-MOT	0.12	64.29%	0.0178	0.1333
	IP-MOT	0.21	73.97%	0.0152	0.1237
	MODT	0.33	67.14%	0.0141	0.1224

consecutiveness, and it is more stable than the Lidar-only and V2V-only method.

VI. CONCLUSION

This paper provides a general way to combine CVIS and AV technology, interactive perception architecture, at the perceptual data level, and explains the related technologies that need to be involved. In the interactive perception, the information interaction ability is considered as the means to expand the range, improve the precision, and enhance the reliability of the perception. An interactive perception based multiple object tracking algorithm to enhance the tracking performance has been presented. The IP-MOT algorithm fuses the interactive information using CVIS and the visual sensing information from the AV. Two simulation platforms are developed and joined to provide rich test scenarios for the proposed algorithms. The results show that the IP-MOT has a better performance especially when the Lidar is limited or the V2V communication was failed.

For the future work, more applications based on the interactive perception, IP-based swarm vehicle control, and the cases about networked perception will be studied for providing more services.

REFERENCES

[1] S. Wei, D. Yu, C. L. Guo, L. Dan, and W. W. Shu, "Survey of connected automated vehicle perception mode: From autonomy to interaction," *IET Intell. Transp. Syst.*, vol. 13, no. 3, pp. 495–505, 2018.

[2] Y. Liu, X. Wang, L. Li, S. Cheng, and Z. Chen, "A novel lane change decision-making model of autonomous vehicle based on support vector machine," *IEEE Access*, vol. 7, pp. 26543–26550, 2019.

[3] C. Wu, H. Sun, H. Wang, K. Fu, G. Xu, W. Zhang, and X. Sun, "Online multi-object tracking via combining discriminative correlation filters with making decision," *IEEE Access*, vol. 6, pp. 43499–43512, 2018.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893.

[6] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, vol. 1, Sep. 2002, p. 1.

[7] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, Feb. 2018, Art. no. 7068349.

[8] V. Vineet, O. Miksik, M. Lidegaard, M. Nießner, S. Golodetz, V. A. Prisacariu, O. Kähler, D. W. Murray, S. Izadi, P. Pérez, and P. H. S. Torr, "Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 75–82.

[9] A. Broggi, S. Cattani, M. Patander, M. Sabbatelli, and P. Zani, "A full-3D voxel-based dynamic obstacle detection for urban scenario using stereo vision," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 71–76.

[10] G. Wang, J. Wu, R. He, and S. Yang, "A point cloud based robust road curb detection and tracking method," *IEEE Access*, vol. 7, pp. 24611–24625, 2019.

[11] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 1907–1915.

[12] Z. Gao, H.-Z. Xuan, H. Zhang, S. Wan, and K.-K. R. Choo, "Adaptive fusion and category-level dictionary learning model for multi-view human action recognition," *IEEE Internet Things J.*, to be published.

[13] Z. Gao, D. Y. Wang, S. H. Wan, H. Zhang, and Y. L. Wang, "Cognitive-inspired class-statistic matching with triple-constrain for camera free 3D object retrieval," *Future Gener. Comput. Syst.*, vol. 94, pp. 641–653, May 2019.

[14] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*. Jun. 2018, pp. 4490–4499.

[15] A. Asvadi, P. Girão, P. Peixoto, and U. Nunes, "3D object tracking using RGB and LIDAR data," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1255–1260.

[16] G. Zhou, K. Li, T. Kirubarajan, and L. Xu, "State estimation with trajectory shape constraints using pseudo-measurements," *IEEE Trans. Aerosp. Electron. Syst.*, to be published.

[17] A. Paier, R. Tresch, A. Alonso, D. Smely, P. Meckel, Y. Zhou, and N. Czink, "Average downstream performance of measured IEEE 802.11p infrastructure-to-vehicle links," in *Proc. IEEE Int. Conf. Commun. Workshops*, May 2010, pp. 1–5.

[18] S.-Y. Wang, C.-C. Lin, W.-J. Hong, and K.-C. Liu, "On the performances of forwarding multihop unicast traffic in WBSS-based 802.11(p)/1609 networks," *Comput. Netw.*, vol. 55, no. 11, pp. 2592–2607, Aug. 2011.

[19] S. Boussoufa-Lahlah, F. Semchedine, and L. Bouallouche-Medjkoune, "Geographic routing protocols for vehicular ad hoc NETworks (VANETs): A survey," *Veh. Commun.*, vol. 11, pp. 20–31, Jan. 2018.

[20] Y. Li, "An overview of the DSRC/WAVE technology," in *Proc. Int. Conf. Heterogeneous Netw. Quality, Rel., Secur. Robustness*. Berlin, Germany: Springer, 2010, pp. 544–558.

[21] S. Wan, Y. Zhao, T. Wang, Z. Gu, Q. H. Abbasi, and K.-K. R. Choo, "Multi-dimensional data indexing and range query processing via Voronoi diagram for Internet of Things," *Future Gener. Comput. Syst.*, vol. 91, pp. 382–391, Feb. 2019.

[22] X. Xu, Y. Xue, L. Qi, Y. Yuan, X. Zhang, T. Umer, and S. Wan, "An edge computing-enabled computation offloading method with privacy preservation for Internet of connected vehicles," *Future Gener. Comput. Syst.*, vol. 96, pp. 89–100, Jul. 2019.

[23] D. Tian, K. Zhu, J. Zhou, Y. Wang, and H. Liu, "Swarm model for cooperative multi-vehicle mobility with inter-vehicle communications," *IET Intell. Transp. Syst.*, vol. 9, no. 10, pp. 887–896, Dec. 2015.

[24] S. Ammoun, F. Nashashibi, and C. Laurgeau, "Crossroads risk assessment using GPS and inter-vehicle communications," *IET Intell. Transport Syst.*, vol. 1, no. 2, pp. 95–101, Jun. 2007.

[25] L. Chen, L. Qin, K. M. Martin, and S.-L. Ng, "Private reputation retrieval in public—A privacy-aware announcement scheme for VANETs," *IET Inf. Secur.*, vol. 11, no. 4, pp. 204–210, Jul. 2017.

[26] B. M. Aumayer, M. G. Petovello, and G. Lachapelle, "Development of a tightly coupled vision/GNSS system," in *Proc. 27th Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS)*, vol. 3, Jan. 2014, pp. 2202–2211.

[27] T. Chu, N. Guo, S. Backén, and D. Akos, "Monocular camera/IMU/GNSS integration for ground vehicle navigation in challenging GNSS environments," *Sensors*, vol. 12, no. 3, pp. 3162–3185, 2012.

[28] O. Heirich, "Bayesian Train localization with particle filter, loosely coupled GNSS, IMU, and a track map," *J. Sensors*, vol. 2016, Mar. 2016, Art. no. 2672640.

[29] E. D. Kaplan, "Understanding GPS: Principles and application," *J. Atmos. Solar-Terr. Phys.*, vol. 59, no. 5, pp. 598–599, 1996.

[30] N. Alam, A. T. Balaie, and A. G. Dempster, ''Dynamic path loss exponent and distance estimation in a vehicular network using Doppler effect and received signal strength,'' in *Proc. Veh. Technol. Conf. Fall*, Sep. 2010, pp. 1–5.

[31] S. Yun, Y. J. Lee, and S. Sung, ''IMU/Vision/Lidar integrated navigation system in GNSS denied environments,'' in *Proc. IEEE Aerosp. Conf.*, Mar. 2013, pp. 1–10.

[32] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng, ''An empirical evaluation of deep learning on highway driving,'' 2015, *arXiv:1504.01716*. [Online]. Available: https://arxiv.org/abs/1504.01716

[33] K. E. Hauer, C. Boscardin, T. B. Fulton, C. Lucey, S. Oza, and A. Teherani, ''Using a curricular vision to define entrustable professional activities for medical student assessment,'' *J. Gen. Internal Med.*, vol. 30, no. 9, pp. 1344–1348, 2015.

[34] E. Kulla, N. Jiang, E. Spaho, and N. Nishihara, ''A survey on platooning techniques in VANETs,'' in *Proc. Conf. Complex, Intell., Softw. Intensive Syst.*, 2018, pp. 650–659.

[35] J.-C. Trujillo, R. Munguia, E. Guerra, and A. Grau, ''Cooperative monocular-based SLAM for multi-UAV systems in GPS-denied environments,'' *Sensors*, vol. 18, no. 5, p. 1351, 2018.

[36] Y. Z. Zhang, Y. Cao, Y. H. Wen, L. Liang, and F. Zou, ''Optimization of information interaction protocols in cooperative vehicle-infrastructure systems,'' *Chin. J. Electron.*, vol. 27, no. 2, pp. 439–444, 2018.

[37] H. Scholl, R. Fidel, S. Liu, and K. Unsworth, ''The fully mobile city government project (mCity),'' in *Proc. Int. Conf. Digit. Government Res.*, 2006, pp. 359–360.

[38] A. Cosgun, L. Ma, J. Chiu, J. Huang, M. Demir, A. M. Anon, T. Lian, H. Tafish, and S. Al-Stouhi, ''Towards full automated drive in urban environments: A demonstration in GoMentum Station, California,'' 2017, *arXiv:1705.01187*. [Online]. Available: https://arxiv.org/abs/1705.01187

[39] M. Althoff and J. M. Dolan, ''Online verification of automated road vehicles using reachability analysis,'' *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 903–918, Aug. 2014.

[40] N. Li, D. W. Oyler, M. Zhang, Y. Yildiz, and A. R. Girard, ''Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems,'' *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 5, pp. 1782–1797, Sep. 2016.

[41] A. Carvalho, S. Lefévre, G. Schildbach, J. Kong, and F. Borrelli, ''Automated driving: The role of forecasts and uncertainty—A control perspective,'' *Eur. J. Control*, vol. 24, pp. 14–32, Jul. 2015.

[42] C. Linguo, S. Wei, W. Jian, T. Zhao, and S. Ning, ''Test sequence generating method of cooperative vehicle infrastructure system based on support index,'' in *Proc. IEEE Int. Symp. Microw.*, Oct. 2016, pp. 272–276.

[43] Z. Xu, M. Wang, F. Zhang, J. Sheng, and X. Zhao, ''PaTAVTT: A hardware-in-the-loop scaled platform for testing autonomous vehicle trajectory tracking,'' *J. Adv. Transp.*, vol. 2017, no. 6, pp. 1–11, 2017.

[44] W. Deng, Y. H. Lee, and A. Zhao, ''Hardware-in-the-loop simulation for autonomous driving,'' in *Proc. Conf. IEEE Ind. Electron. Soc.*, Nov. 2008, pp. 1742–1747.

[45] P. Jagtap, P. Raut, P. Kumar, A. Gupta, N. M. Singh, and F. Kazi, ''Control of autonomous underwater vehicle using reduced order model predictive control in three dimensional space,'' *IFAC-PapersOnLine*, vol. 49, no. 1, pp. 772–777, 2016.

[46] A. Joshi, ''Powertrain and chassis hardware-in-the-loop (HIL) simulation of autonomous vehicle platform,'' SAE Tech. Paper 2017-01-1991, 2017.

[47] S. Subramanian, T. G. Thuruthel, and A. Thondiyath, ''Real-time obstacle avoidance for an underactuated flat-fish type autonomous underwater vehicle in 3D space,'' *Int. J. Robot. Autom.*, vol. 29, no. 4, pp. 424–431, 2014.

[48] J. Wang, D. Liu, W. Jiang, and D. Lu, ''Evaluation on loosely and tightly coupled GNSS/INS vehicle navigation system,'' in *Proc. Int. Conf. Electromagn. Adv. Appl. (ICEAA)*, Sep. 2017, pp. 892–895.

[49] M. A. Fischler and R. Bolles, ''Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,'' *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[50] L. Li, F. Yang, H. Zhu, D. Li, Y. Li, and L. Tang, ''An improved RANSAC for 3D point cloud plane segmentation based on normal distribution transformation cells,'' *Remote Sens.*, vol. 9, no. 5, p. 433, 2017.

[51] Y. Liu and M. Tong, ''An application of Hungarian algorithm to the multi-target assignment,'' *Fire Control Command Control*, vol. 27, no. 4, pp. 34–37, 2002.

[52] B. Cai, C. Wang, W. ShangGuan, and W. Jian, ''Research of information interaction simulation method in cooperative vehicle infrastructure system,'' in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Oct. 2014, pp. 45–50.

[53] Y. Bar-Shalom and L. Campo, ''The effect of the common process noise on the two-sensor fused-track covariance,'' *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 6, pp. 803–805, Nov. 1986.

[54] D. Shin, B. Kim, K. Yi, A. Carvalho, and F. Borrelli, ''Human-centered risk assessment of an automated vehicle using vehicular wireless communication,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 667–681, Feb. 2018.

[55] L. Chai, B. Cai, W. ShangGuan, J. Wang, and H. Wang, ''Basic simulation environment for highly customized connected and autonomous vehicle kinematic scenarios,'' *Sensors*, vol. 17, no. 9, p. 1938, 2017.

[56] A. Geiger, P. Lenz, and R. Urtasun, ''Are we ready for autonomous driving? The KITTI vision benchmark suite,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[57] K. Bernardin and R. Stiefelhagen, ''Evaluating multiple object tracking performance: The CLEAR MOT metrics,'' *EURASIP J. Image Video Process.*, vol. 2008, pp. 1-1–1-10, Dec. 2008.

[58] M. Sualeh and G.-W. Kim, ''Dynamic multi-LiDAR based multiple object detection and tracking,'' *Sensors*, vol. 19, no. 6, p. 1474, 2019.

[59] Y. Du, W. ShangGuan, and L. Chai, ''Particle filter based object tracking of 3D sparse point clouds for autopilot,'' in *Proc. Chin. Autom. Congr. (CAC)*, Nov./Dec. 2018, pp. 1102–1107.

WEI SHANGGUAN (M'14) received the B.S., M.S., and Ph.D. degrees from Harbin Engineering University, in 2002, 2005, and 2008, respectively. From 2008 to 2011, he was a Lecturer with the School of Electronic and Information Engineering, Beijing Jiaotong University. From 2013 to 2014, he was an Academic Visitor with the University College London. He is currently a Professor and a Doctoral Tutor with Beijing Jiaotong University. His research interests include intelligent transportation system, cooperative vehicle infrastructure system of China (CVIS-C), system modeling, train control system (CTCS/ETCS/ERTMS), simulation and testing, GNSS (GPS, Galileo, Glonass and BDS)/GIS, and integrated navigation.

YU DU received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 2012, where she is currently pursuing the Ph.D. degree in transportation information engineering and control with the School of Electronic and Information Engineering. Her research interests include cooperative vehicles and infrastructure systems.

LINGUO CHAI received the B.S., M.S., and Ph.D. degrees from Beijing Jiaotong University, in 2010, 2012, and 2018, respectively, where he is currently a Lecturer with the School of Electronic and Information Engineering. From 2016 to 2017, he was a Visiting Scholar with PATH, UC Berkeley. His research interests include vehicle operational control of CVIS, CVIS modeling, and simulation, and the simulation of train control systems.