

Article

# A Unified Multiple-Target Positioning Framework for Intelligent Connected Vehicles

Zhongyang Xiao <sup>1</sup>, Diange Yang <sup>1,\*</sup>, Fuxi Wen <sup>1,2</sup> and Kun Jiang <sup>1</sup>

<sup>1</sup> State Key Laboratory of Automotive Safety and Energy, School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China; xiaozy15@mails.tsinghua.edu.cn (Z.X.); fuxi@chalmers.se (F.W.); jiangkun@mail.tsinghua.edu.cn (K.J.)

<sup>2</sup> Department of Electrical Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

\* Correspondence: ydg@mail.tsinghua.edu.cn

Received: 11 April 2019; Accepted: 24 April 2019; Published: 26 April 2019



**Abstract:** Future intelligent transport systems depend on the accurate positioning of multiple targets in the road scene, including vehicles and all other moving or static elements. The existing self-positioning capability of individual vehicles remains insufficient. Also, bottlenecks in developing on-board perception systems stymie further improvements in the precision and integrity of positioning targets. Vehicle-to-everything (V2X) communication, which is fast becoming a standard component of intelligent and connected vehicles, renders new sources of information such as dynamically updated high-definition (HD) maps accessible. In this paper, we propose a unified theoretical framework for multiple-target positioning by fusing multi-source heterogeneous information from the on-board sensors and V2X technology of vehicles. Numerical and theoretical studies are conducted to evaluate the performance of the framework proposed. With a low-cost global navigation satellite system (GNSS) coupled with an initial navigation system (INS), on-board sensors, and a normally equipped HD map, the precision of multiple-target positioning attained can meet the requirements of high-level automated vehicles. Meanwhile, the integrity of target sensing is significantly improved by the sharing of sensor information and exploitation of map data. Furthermore, our framework is more adaptable to traffic scenarios when compared with state-of-the-art techniques.

**Keywords:** vehicular localization; target positioning; high-definition map; vehicle-to-everything; intelligent and connected vehicles; intelligent transport system

## 1. Introduction

The intelligent transportation system (ITS) is one of the most indispensable components of the smart city concept that integrates sensing, control, information, and communication technologies into transportation [1]. In recent years, with the emergence of cutting-edge applications of ITS, the positioning of multiple targets, including vehicles and other elements has been playing an increasingly important role in improving safety, mobility, and efficiency [2–4]. For example, future intelligent connected vehicles (ICVs) require the positioning of their own real-time location with centimeter-level precision [5], and the awareness of all objects such as surrounding vehicles and vulnerable road users with significant integrity and confidence. In ITS, the positioning of vehicles and other targets are usually referred to as vehicular self-positioning and target localization, respectively. Although attention has been paid in these topics [6–9], there still exist many limitations that need to be eliminated.

### 1.1. Self-Positioning

Multiple self-positioning technologies are already present in the market, but none are effective under all road conditions and scenarios [10]. GNSS systems are widely employed in ITS devices, but they can support only low-precision navigation. Researchers have tried to integrate information from the base station and on-board sensors for error compensation. However, in dense urban environments where the signal is disturbed by surrounding buildings [5], even the most accurate GNSS with real-time kinematic correction and INS fusion schemes [11] cannot provide localization with adequate accuracy and stability.

Introducing new sources of information is an effective way of improving vehicular self-positioning. V2X communication, which has drawn increasing interest in recent years, renders information easily accessible to the vehicles connected [12–14]. The V2X-based (or cooperative) method aids in improving vehicular localization capability by employing the position information of other vehicles and relative measurements from their on-board sensors [15,16]. The integration of on-board sensors and V2X communication is shown to be more cost-effective than approaches based on high-quality sensors [17]. A general framework for multi-vehicular localization using pose graph optimization is proposed in [18], using vehicle-vehicle (V-V) measurements to improve the precision of vehicular localization. More recently, an implicit cooperative positioning algorithm that exploits the joint sensing of passive features is proposed in [19,20], and precludes the use of explicit V-V measurements. In addition to the use of ranging sensors, angle measurement-based cooperative localization is proposed in [21].

Maps are additional sources of information, and the locations of static elements can be used as references to improve the vehicular positioning capability [22]. In contrast to simultaneous localization and mapping (SLAM), in which a map is generated real-time [23], the map-based method assumes that maps are available in advance and aligns landmarks in the maps with on-board sensors to achieve independent positioning or as an aid to GNSS with INS system (GNSS/INS). As shown in Table 1, over the past few years, with the development of V2X, HD maps, which are characterized by high-accuracy and real-time updates, have grown to become standard and indispensable components of intelligent vehicles [4]. This also enables centimeter-level precision to be achieved in map-based localization [24].

**Table 1.** Maps for different levels of Intelligent Connected Vehicles.

| Grade                           | Title                        | Map       | Accuracy               | Typical Condition           |
|---------------------------------|------------------------------|-----------|------------------------|-----------------------------|
| DRIVER SCENARIO                 |                              |           |                        |                             |
| 1 (DA)                          | Driver Assistance            | ADAS      | Submeter               | Optional                    |
| 2 (PA)                          | Partial Autopilot            | ADAS      | Submeter               | Optional                    |
| AUTOMATIC DRIVING SYS. SCENARIO |                              |           |                        |                             |
| 3 (CA)                          | Conditional Autopilot        | ADAS + HD | Submeter<br>Centimeter | Optional                    |
| 4 (HA)                          | High-Level Automated Driving | ADAS + HD | Submeter<br>Centimeter | Essential                   |
| 5 (FA)                          | Completely Automated Driving | HD        | Centimeter             | Essential<br>(auto updated) |

The HD map-based method benefits from the high precision of the map used. For example, the digital map used in [25–27] is created from light detection and ranging (LiDAR) data and has a precision of up to 10 cm. A high-accuracy localization technique using urban environment maps for vehicles in motion is proposed in [28], and these maps are generated by integrating GNSS, LiDAR data, and on-board sensors. In addition, more features are used in the process of map alignment, which contributes to higher self-positioning precision. Traffic lights and visual lane markers are used as landmarks in [5,29], respectively. In a more recent study [26], a unified framework using more references in addition to the abovementioned features (lamp poles, traffic signs, etc.) is proposed, and self-localization with an accuracy of within 30 cm is achieved with merely a low-cost camera. While the

abovementioned studies conduct map-based localization independently, there is much research that integrates it with other on-board sensors. A lane determination system that fuses on-board sensors, GNSS, and commercially available road network maps is proposed in [30]. A proof-of-concept study using INS and maps for vehicular localization in GNSS-denied environments is conducted in [31].

### 1.2. Target Localization

Apart from vehicular self-positioning, the relative localization of targets in the surrounding environment is another fundamental technology underpinning ICVs. This task is mainly undertaken by vehicle perception modules, and the positioning result is obtained in the vehicular coordinate instead of in the world coordinate. Although the recent decades have witnessed the rapid development of on-board sensors, the current on-board sensing technology still faces the following problems [4]. First, there is a trade-off between localization accuracy and cost. For example, low-cost cameras and radars can achieve accuracies of only several centimeters, while LiDAR systems with centimeter-level ranging accuracy are expensive [32]. Second, all sensors have limited sensing ranges, and the occlusion of sensors by other vehicles and objects is a frequent occurrence [33]. Irrespective of the number of sensors equipped in the vehicle, the perception of the environment remains incomplete. Attempts to improve the perception of bicycles have already encountered bottlenecks to a certain extent.

Studies using the perceptual information of other vehicles to improve the integrity of target localization have proven to be effective. This is because other vehicles in the network may have seen a target that cannot be seen by the ego-vehicle because of occlusion or limited field of view. A vehicle-to-vehicle (V2V) communication and map merging-based cooperative perception system to extend the perception range beyond line of sight and field of view is proposed in [34]. In [35], the results of the awareness of other vehicles are integrated into the ego-vehicle's perception system as virtual sensors to achieve perception enhancement. In [36], a multi-vehicle perception framework combining image and semantic features is proposed, and experiments have proved that the problem of front-vehicle occlusion can be solved. In these studies, the problems of self-positioning and localization of other targets were considered separately, which rendered the effect of the fusion very sensitive to their relative positioning. In addition, these articles do not provide quantitative analyses of the integrity of the results perceived.

Maps also contribute to the relative positioning of targets. For example, the geometry of intersection can be directly extracted from a HD map for motion planning and control [37]. This reduces the pressure on the vehicle-mounted sensing system, but relies on vehicular self-positioning. Incorrect self-positioning greatly affects decision-making. Other works integrate the semantic and geometric prior knowledge in HD maps with the on-board sensing system to improve positioning confidence. In [38,39], a prior probability map is generated in a bird's-eye view or image plane to aid understanding of the scene. Recently, a neural network incorporating prior knowledge with on-board sensors was presented in [40]. However, these studies treat the map only as an auxiliary tool for perception to improve the recognition result without improving the integrity of perception. Moreover, these studies also rely on the accuracy of vehicular positioning.

### 1.3. Contributions

In general, recent research exploits additional sources of information to improve vehicular self-positioning and localization of other targets. However, to the best of our knowledge, V2X and HD maps are considered separately in the literature, while the positioning of the vehicles themselves and those of other targets are usually treated as different modules. In this paper, we propose a unified theoretical positioning framework for multiple targets for ICVs. The bottlenecks of vehicular self-positioning and target localization can be eliminated. Our main contributions are summarized as follows.

- A unified theoretical framework for vehicular self-positioning and relative localization of targets based on V2X is proposed, and it can integrate data from the on-board sensors in the vehicular network and HD maps with GNSS/INS measurements into a unified system.
- By cooperative positioning, accuracy of under 0.2 m can be achieved in terms of self-positioning and relative localization of targets in urban areas using low-cost GNSS/INS, on-board sensors, and widely equipped HD maps. Simultaneously, the target sensing range is extended beyond the line of sight and field of view, and this greatly improves the integrity of perception.
- Furthermore, compared with state-of-the-art techniques, the proposed framework places fewer demands on vehicular network nodes' density and the amount of vehicle-to-target measurements.

The remainder of the paper is organized as follows. In Section 2, the system model is provided. The development of the proposed joint multiple-target positioning for ICVs is detailed in Section 3. Detailed implementation aspects are introduced in Section 4. Theoretical studies are explained in Section 5. Numerical results are given in Section 6. Finally, we conclude the paper in Section 7.

## 2. Problem Formulation

Firstly, we describe the targets in a traffic scene in this section.

- Targets: All objects related to vehicle driving, including the connected vehicles themselves and the elements that constitute the environment.
- Connected vehicles: Vehicles in the vehicular network that can obtain information from other vehicles and HD maps.
- Features: Static targets that can be associated with HD maps, e.g., lamps, trees, traffic lights, and traffic signs.
- Objects: Targets, both static and moving, that do not exist in HD maps. These can be pedestrians, bicycles, and disconnected vehicles, all of which are unlabeled on the map.

Consider a vehicular network scenario with a set of  $N_v$  interconnected vehicles  $\mathcal{V} = \{1, 2, \dots, N_v\}$ , as shown in Figure 1. At time  $t$ , let  $\mathbf{x}_{i,t}^{(V)}$  be the position and orientation of connected vehicle  $i$  in the global coordinate (see Equation (1)).

$$\mathbf{x}_{i,t}^{(V)} = [\mathbf{p}_{i,t}^{(V)}, \theta_{i,t}^{(V)}] = [x_{i,t}^{(V)}, y_{i,t}^{(V)}, \theta_{i,t}^{(V)}]^T, i \in \mathcal{V} \quad (1)$$

A set of  $N_f$  static features  $\mathcal{F} = \{1, 2, \dots, N_f\}$  also exists in the scene. Their positions are stored on a HD map with noise and, although not necessary, can be captured by the on-board sensors. We use Equation (2) to describe the two-dimensional position of the  $j$ th feature.

$$\mathbf{x}_{j,t}^{(F)} = \mathbf{p}_{j,t}^{(F)} = [x_{j,t}^{(F)}, y_{j,t}^{(F)}]^T, j \in \mathcal{F} \quad (2)$$

In addition, we also consider a set of  $N_o$  objects, moving or static,  $\mathcal{O} = \{1, 2, \dots, N_o\}$ , which is described in Equation (3).

$$\mathbf{x}_{k,t}^{(O)} = \mathbf{p}_{k,t}^{(O)} = [x_{k,t}^{(O)}, y_{k,t}^{(O)}]^T, k \in \mathcal{O} \quad (3)$$

It must be noted that we do not estimate the orientations of the features and objects because the planning module usually does not require this information.

Our task is to estimate the localization and orientation of all the connected vehicles,

$$\mathbf{X}_t^{(V)} = [\mathbf{x}_{1,t}^{(V)} \dots \mathbf{x}_{N_v,t}^{(V)}], \quad (4)$$

We also attempt to localize the features and objects (see Equation (5))

$$\begin{aligned}\mathbf{X}_t^{(F)} &= [\mathbf{x}_{1,t}^{(F)} \cdots \mathbf{x}_{N_f,t}^{(F)}] \\ \mathbf{X}_t^{(O)} &= [\mathbf{x}_{1,t}^{(O)} \cdots \mathbf{x}_{N_o,t}^{(O)}]\end{aligned}\quad (5)$$

Based on Equations (4) and (5), we can obtain the relative localization of other targets by transforming their location into the vehicles' coordinate system.

In terms of the measurements, a target may be captured by a vehicle if it is within the vehicle's sensing range and without any occlusions. As shown in Figure 1, the target can be a connected vehicle, a feature, or an object, which are indicated with red, blue, and brown arrows, respectively. Its measurement model is described as Equation (6).

$$\mathbf{z}_{i,j,t}^{(\Xi)} = h^{(S)}(\mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{i,t}^{(V)}) + \mathbf{v}_{i,j,t}^{(S)} \quad (6)$$

where  $\Xi \in \{V, F, O\}$  and  $\mathbf{v}_{k,j,t}^{(S)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{k,j,t}^{(S)})$  is additive white Gaussian measurement noise with covariance  $\mathbf{R}_{k,j,t}^{(S)}$ ,  $h^{(S)}(\mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{k,t}^{(V)})$  is a function which denotes the measurement of target at position  $\mathbf{p}_{j,t}^{(\Xi)}$  from vehicle  $\mathbf{x}_{k,t}^{(V)}$ .

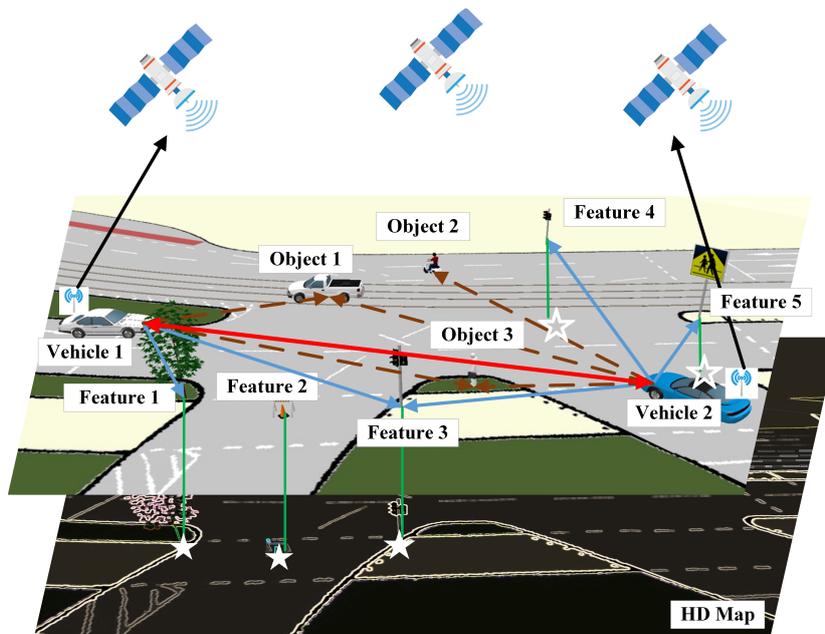


Figure 1. A demonstration of the multiple targets positioning scenario.

The connected vehicles are also equipped with GNSS/INS, which can provide measurements of their localization and orientation. The corresponding measurements are indicated with black arrows in Figure 1. Similarly, we treat the map as a virtual sensor with measurements pertaining to an associated feature as shown with green arrows. The measurement of GNSS/INS on vehicle  $i$  and that of HD map on feature  $j$  is indicated by Equations (7) and (8).

$$\mathbf{z}_{i,t}^{(G)} = \mathbf{h}^{(G)}(\mathbf{p}_{i,t}^{(V)}) + \mathbf{v}_{i,t}^{(G)} \quad (7)$$

and

$$\mathbf{z}_{j,t}^{(M)} = \mathbf{h}^{(M)}(\mathbf{p}_{j,t}^{(F)}) + \mathbf{v}_{j,t}^{(M)} \quad (8)$$

where  $\mathbf{v}_{i,t}^{(G)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{i,t}^{(G)})$  is the measurement noise of GNSS/INS, and  $\mathbf{v}_{j,t}^{(M)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{j,t}^{(M)})$  denotes the measurement noise from the map.

If we consider this problem as analogous to a distributed sensor network, the features and connected vehicles are static and mobile anchors, respectively, and their locations are constrained by the HD map and GNSS/INS. The objects are static or mobile nodes, and the on-board sensors generate constraints between the vehicle and nodes. For the vehicles, additional constraints arise from the V-V measurements.

### 3. The Unified Multiple-Target Positioning Framework

The objective of multiple-target positioning in this paper is to estimate the states  $\mathbf{X}_t = [\mathbf{X}_t^{(V)}, \mathbf{X}_t^{(F)}, \mathbf{X}_t^{(O)}]$ , from measurements,

$$\mathbf{Z}_t = [\mathbf{z}_{k,j,t}^{(S)}, \mathbf{z}_{k,t}^{(G)}, \mathbf{z}_{l,t}^{(M)}], \quad (9)$$

where  $k \in \mathcal{V}, j \in \{\mathcal{V}, \mathcal{F}, \mathcal{O}\}, l \in \mathcal{F}$ . From a probabilistic perspective, the maximum likelihood estimation of  $\mathbf{X}_t$  is given by Equation (10).

$$\mathbf{X}_t^* = \arg \max P(\mathbf{Z}_t | \mathbf{X}_t) \quad (10)$$

where  $P(\mathbf{Z}_t | \mathbf{X}_t)$  is the likelihood of the measurements  $\mathbf{Z}_t$  given the states  $\mathbf{X}_t$ . The conditional distribution of the on-board sensor measurements in Equation (6) is given by Equation (11), where  $P(\mathbf{z}_{k,j,t}^{(S)} | \mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{k,t}^{(V)})$  denotes the probability distribution of measurement  $\mathbf{z}_{k,j,t}^{(S)}$  given the states  $\mathbf{p}_{j,t}^{(\Xi)}$  and  $\mathbf{x}_{k,t}^{(V)}$ .  $\mathcal{N}(h^{(S)}(\mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{k,t}^{(V)}), \mathbf{R}_{k,j,t}^{(S)})$  denotes a normal distribution with expectation  $h^{(S)}$  and variance  $\mathbf{R}_{k,j,t}^{(S)}$ .

$$P(\mathbf{z}_{k,j,t}^{(S)} | \mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{k,t}^{(V)}) = \mathcal{N}(h^{(S)}(\mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{k,t}^{(V)}), \mathbf{R}_{k,j,t}^{(S)}) \quad (11)$$

Similarly, we get the conditional distributions of the measurement from GNSS/INS  $P(\mathbf{z}_{i,t}^{(G)} | \mathbf{p}_{i,t}^{(V)})$  and map  $P(\mathbf{z}_{j,t}^{(M)} | \mathbf{p}_{j,t}^{(F)})$  given states of vehicles  $\mathbf{p}_{i,t}^{(V)}$  or feature states  $\mathbf{p}_{j,t}^{(F)}$  (see Equations (12) and (13)). Their distributions are also normal with perception model  $h$  as expectation.

$$P(\mathbf{z}_{i,t}^{(G)} | \mathbf{p}_{i,t}^{(V)}) = \mathcal{N}(h^{(G)}(\mathbf{p}_{i,t}^{(V)}), \mathbf{R}_{i,t}^{(G)}) \quad (12)$$

and

$$P(\mathbf{z}_{j,t}^{(M)} | \mathbf{p}_{j,t}^{(F)}) = \mathcal{N}(h^{(M)}(\mathbf{p}_{j,t}^{(F)}), \mathbf{R}_{j,t}^{(M)}) \quad (13)$$

**Remark 1.** Given a set of independent and identically distributed (i.i.d.) data  $\mathcal{D} = \{\mathbf{x}_n, n = 1, 2, \dots, N\}$ , where observation  $\mathbf{x}_n \in \mathbb{R}^{D \times 1}$  is drawn from a multivariate Gaussian distribution  $\mathcal{N}(\mathbf{x}_n; \boldsymbol{\mu}_n, \mathbf{R}_n)$ . The log-likelihood of the data set can be written as Equation (14).

$$\mathcal{L}(\mathcal{D}) = -\frac{1}{2} \sum_{n=1}^N \left( \ln((2\pi)^D \det(\mathbf{R}_n)) + \mathbf{e}_n^T \mathbf{R}_n^{-1} \mathbf{e}_n \right) \quad (14)$$

where  $\mathbf{e}_n = \mathbf{x}_n - \boldsymbol{\mu}_n$ .

According to Remark 1, the maximization of  $\mathcal{L}(\mathcal{D})$  is equivalent to the minimization of  $\mathcal{J}(\mathcal{D})$  (see Equation (15)). The problem is solved with optimization as described in Section 4.

$$\mathcal{J}(\mathcal{D}) = \sum_{n=1}^N \mathbf{e}_n^T \mathbf{R}_n^{-1} \mathbf{e}_n \quad (15)$$

Considering (11)–(13) and assuming that the three types of measurements are independent, the joint probability density can be factorized as given in Equation (16).

$$\begin{aligned} P(\mathbf{Z}_t|\mathbf{X}_t) &= P(\mathbf{z}_t^{(V)}, \mathbf{z}_t^{(F)}, \mathbf{z}_t^{(O)}|\mathbf{X}_t) \\ &= \prod_{k,j} P(\mathbf{z}_{k,j,t}^{(S)}|\mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{(k,t)}^{(V)}) \\ &\quad \prod_i P(\mathbf{z}_{i,t}^{(G)}|\mathbf{p}_{i,t}^{(V)}) \prod_j P(\mathbf{z}_{j,t}^{(M)}|\mathbf{p}_{j,t}^{(F)}) \end{aligned} \quad (16)$$

The maximization of  $P(\mathbf{Z}_t|\mathbf{X}_t)$  can be reformulated as the following nonlinear least squares problem (see Equation (17)).

$$\begin{aligned} \mathbf{x}_t^* &= \arg \min \sum_k \sum_j (\mathbf{e}_{k,j,t}^{(S)})^T (\mathbf{R}_{k,j,t}^{(S)})^{-1} \mathbf{e}_{k,j,t}^{(S)} \\ &\quad + \sum_i (\mathbf{e}_{i,t}^{(G)})^T (\mathbf{R}_{i,t}^{(G)})^{-1} \mathbf{e}_{i,t}^{(G)} \\ &\quad + \sum_j (\mathbf{e}_{j,t}^{(M)})^T (\mathbf{R}_{j,t}^{(M)})^{-1} \mathbf{e}_{j,t}^{(M)} \end{aligned} \quad (17)$$

To enable insightful visualization, the nonlinear least-squares problem is interpreted in terms of inference over a factor graph [41]. This graph consists of 2 types of nodes: variable nodes, which represent the state  $\mathbf{X}_i$ , and factor nodes, which represent the constraints to on the variables. The factor nodes can be further divided into bi-directed nodes, which denote the constraints for 2 states (from the on-board sensor measurements), and directed prior nodes, which denote the constraints from the map and GNSS/INS.

As shown in Figure 2, for each measurement, we have the following factors.

- Factor between the variables  $\mathcal{V}$  and  $\Xi = \{\mathcal{V}, \mathcal{F}, \mathcal{O}\}$ , on behalf of the constraints of V-V, vehicle-feature (V-F), vehicle-object (V-O), as expressed in Equation (18).

$$\phi_{k,j,t} = P(\mathbf{z}_{k,j,t}^{(S)}|\mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{(k,t)}^{(V)}) \quad (18)$$

- Factor between the variables  $\mathcal{V}$  and GNSS/INS, on behalf of the constraints from GNSS/INS, as expressed in Equation (19).

$$\phi_{i,t} = P(\mathbf{z}_{i,t}^{(G)}|\mathbf{p}_{i,t}^{(V)}) \quad (19)$$

- Factor between the variables  $\mathcal{F}$  and the map, on behalf of the constraints from the HD map, as expressed in Equation (20).

$$\phi_{j,t} = P(\mathbf{z}_{j,t}^{(M)}|\mathbf{p}_{j,t}^{(F)}) \quad (20)$$

The joint probability in Equation (16) can then be rewritten as the product of all the factors.

$$P(\mathbf{Z}_t|\mathbf{X}_t) = \prod_{k,j,t} \phi_{k,j,t} \prod_{i,t} \phi_{i,t} \prod_{j,t} \phi_{j,t} \quad (21)$$

We can clearly see the constraints applied on each node in the factor graph.

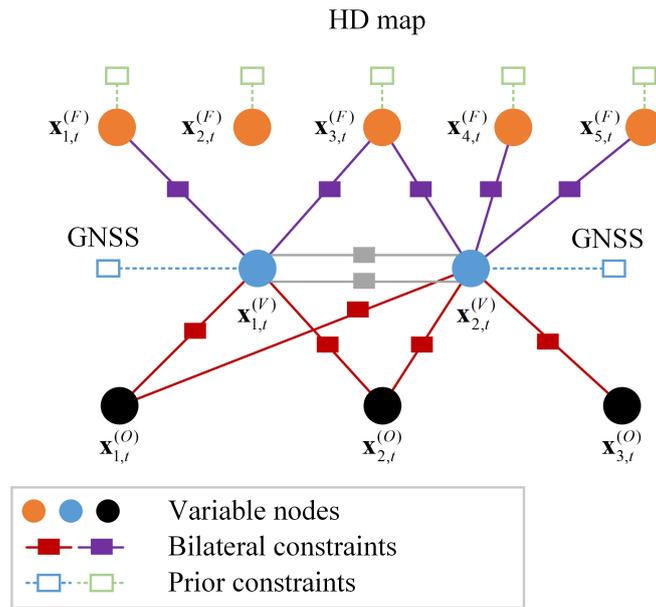


Figure 2. The proposed framework interpreted as inference on factor graphs.

#### 4. Implementation Aspects

In this section, we introduce the implementation aspects related to the hypothesis on vehicle perception, the measurement model, optimization, and data association.

##### 4.1. Perception Demands and Sensing Capability of Vehicles

We assume that due to occlusion and the limitations of perception range, the vehicle cannot completely locate the desired target. In this section, we explain the hypothesis of this work. It should be noted that our hypothesis is based on typical perceptual systems, but can be easily adapted to other forms.

As shown in Figure 3, we identify the scope of targets that need to be localized by a vehicle as “demanding space” and assume that it is a rectangle that can be quantitatively described by  $l_f$  and  $l_r$ , i.e., the distances that the vehicle requires to sense ahead of and behind itself, respectively, and  $W_d$ , the range that should be perceived laterally. We assume that the vehicle sensing range is a forward-facing cone with a radius of  $R_s$ , and the field of view is  $\theta_{FOV}$ .

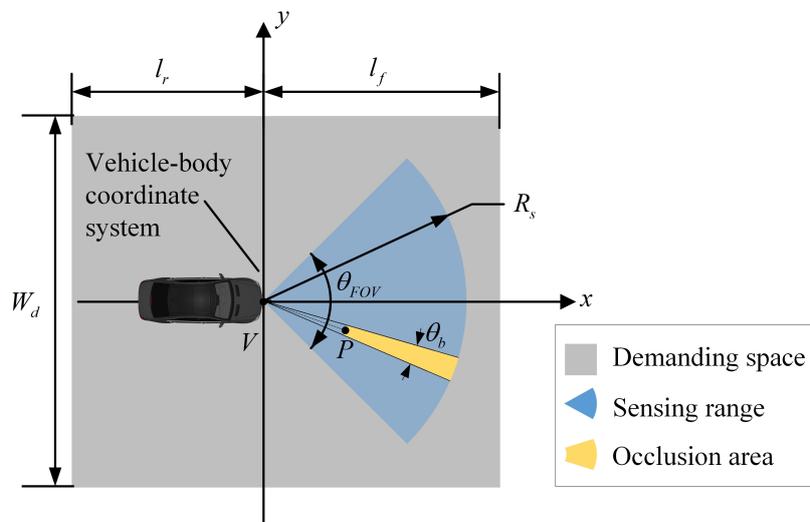


Figure 3. Sensing range and perception demand.

To consider situations of occlusion, we assume that there is an object  $P$  in the sensing range, and that the area outside  $P$  in the sector with line  $VP$  as the axis of symmetry is regarded as the occlusion area. Thus, only the blue area in Figure 3 can be perceived. The limitations in sensing range and occlusion constitute the blind spots of environment perception.

#### 4.2. Measurement Model

In this work, we assume that data from the on-board sensors are in a 2D vehicular coordinate fashion where  $h^{(S)} \sim \mathbb{R}^{2 \times 1}$ , correspond to the measurements from low-cost cameras. This can be easily adapted to other measurement types, such as polar coordinates in LiDAR measurements. The measurement model is expressed as,

$$h^{(S)}(\mathbf{p}_{j,t}^{(\Xi)}, \mathbf{x}_{i,t}^{(V)}) = \mathbf{R}^{-1}(\theta_{i,t}^{(V)}) \left[ \mathbf{p}_{j,t}^{(\Xi)} - \mathbf{p}_{i,t}^{(V)} \right], \quad (22)$$

where the rotation matrix is expressed as,

$$\mathbf{R}(\theta_{i,t}^{(V)}) = \begin{bmatrix} \cos(\theta_{i,t}^{(V)}) & -\sin(\theta_{i,t}^{(V)}) \\ \sin(\theta_{i,t}^{(V)}) & \cos(\theta_{i,t}^{(V)}) \end{bmatrix}, \quad (23)$$

and the covariance of the on-board sensors' measurements is given in Equation (24).

$$\mathbf{R}_{k,j,t}^{(S)} = \begin{bmatrix} \delta_{sensor}^2 & 0 \\ 0 & \delta_{sensor}^2 \end{bmatrix} \quad (24)$$

We assume that the GNSS/INS can provide the measurements of the coordinates and angles of the vehicles. There are many studies on modeling the measurements noise of GNSS/INS [42,43]. In this paper, we simplify the error of GNSS to Gaussian distribution, and the measurement model and uncertainty of the GNSS/INS on vehicle  $i$  are indicated as Equations (25) and (26). Our framework is also suitable for extending to other error assumptions.

$$\mathbf{h}^{(G)}(\mathbf{p}_{i,t}^{(V)}) = \mathbf{p}_{i,t}^{(V)} \quad (25)$$

and

$$\mathbf{R}_{i,t}^{(G)} = \begin{bmatrix} \delta_{GNSS,l}^2 & 0 & 0 \\ 0 & \delta_{GNSS,l}^2 & 0 \\ 0 & 0 & \delta_{GNSS,\theta}^2 \end{bmatrix} \quad (26)$$

In commercial HD maps, the coordinates of the features are provided along with noise, so we formulate the measurement mode and covariance matrix as expressed by Equations (27) and (28).

$$\mathbf{h}^{(M)}(\mathbf{p}_{j,t}^{(F)}) = \mathbf{p}_{j,t}^{(F)} \quad (27)$$

$$\mathbf{R}_{j,t}^{(M)} = \begin{bmatrix} \delta_{map}^2 & 0 \\ 0 & \delta_{map}^2 \end{bmatrix} \quad (28)$$

#### 4.3. Optimized Variable Allocation and Data Association

The optimized variables are allocated to observations within the demanding space of perception. Unlike in a traditional multi-vehicle cooperative system, barring objects and features captured by vehicles, features that are not seen by any vehicle but are within the demanding space are also included in the optimized variables, and further optimized to yield the results of perception.

Observations that are associated are merged into existing variables and form constraints in the process of optimization. There are many methods that can be applied to our framework [44,45]. In this study, we assume that the vehicle’s on-board sensors and HD maps can provide enough semantic clue to identify objects. The association algorithm itself is beyond the scope of this article.

4.4. Optimization Problem Solving

In this study, the nonlinear optimization problem in Equation (17) is solved via the Levenberg-Marquardt method [46]. We reorganize the residuals of time  $t$  into one vector, as expressed in Equation (29).

$$\mathbf{e} = \left( \mathbf{e}_{ij}^{(S)} \ \mathbf{e}_{kp}^{(S)} \ \mathbf{e}_m^{(G)} \ \mathbf{e}_q^{(M)} \right)^T \tag{29}$$

where  $\mathbf{e}_{ij}^{(S)}$  is the residual of measurement from the on-board sensor of vehicle  $i$  to vehicle  $j$ .  $\mathbf{e}_{kp}^{(S)}$  is the residual of measurement from the on-board sensor of vehicle  $k$  to feature or object  $p$ .  $\mathbf{e}_m^{(G)}$  is the residual of GNSS/INS measurement of vehicle  $m$ , and  $\mathbf{e}_q^{(M)}$  is the residual of measurement from HD map to feature  $q$ . The optimized variable at time  $t$  can then be rewritten as,

$$\mathbf{X} = \left( \mathbf{x}_i^{(V)} \ \mathbf{x}_j^{(V)} \ \mathbf{x}_k^{(V)} \ \mathbf{x}_p^{(\Xi)} \ \mathbf{x}_m^{(V)} \ \mathbf{x}_q^{(F)} \right), \tag{30}$$

where  $\Xi \in \{F, O\}$ . Let  $\mathbf{R}$  be the overall covariance matrix such that

$$\mathbf{R} = \text{diag} \left( \mathbf{R}^{(S)} \ \mathbf{R}^{(S)} \ \mathbf{R}^{(G)} \ \mathbf{R}^{(M)} \right). \tag{31}$$

The cost function can be rewritten as,

$$\mathbf{f}(\mathbf{X}) = \left( \mathbf{R}^{-\frac{1}{2}} \mathbf{e} \right)^T \left( \mathbf{R}^{-\frac{1}{2}} \mathbf{e} \right). \tag{32}$$

We can get the Jacobian matrix (see Equation (33)).

$$\mathbf{J}(\mathbf{X}) = \frac{\partial \left( \mathbf{R}^{-\frac{1}{2}} \mathbf{e} \right)}{\partial \mathbf{X}} = \mathbf{R}^{-\frac{1}{2}} \begin{bmatrix} \frac{\partial \mathbf{e}_{ij}^{(S)}}{\partial \mathbf{x}_i^{(V)}} & \frac{\partial \mathbf{e}_{ij}^{(S)}}{\partial \mathbf{x}_j^{(V)}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\partial \mathbf{e}_{kp}^{(S)}}{\partial \mathbf{x}_k^{(V)}} & \frac{\partial \mathbf{e}_{kp}^{(S)}}{\partial \mathbf{x}_p^{(\Xi)}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\partial \mathbf{e}_m^{(G)}}{\partial \mathbf{x}_m^{(V)}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\partial \mathbf{e}_q^{(M)}}{\partial \mathbf{x}_q^{(F)}} \end{bmatrix} \tag{33}$$

The initial values of the optimization iterations are given as follows. The vehicular position and attitude are calculated by the measurements of the GNSS/INS. The positions of features are determined by the map, and the initial positions of objects are calculated by converting the positions measured by the on-board sensors to the geodetic coordinate system according to the initial vehicle position and attitude. The cost function (17) can be minimized towards zero by iterations:

$$\mathbf{X}_{k+1} \leftarrow \mathbf{X}_k - \left( \mathbf{J}^T \mathbf{J} + \lambda \text{diag} \left( \mathbf{J}^T \mathbf{J} \right) \right)^{-1} \mathbf{J}^T \mathbf{f}(\mathbf{X}_{k+1}) \tag{34}$$

where  $\lambda$  is determined by the Levenberg-Marquardt method, and  $\mathbf{J}$  and  $\mathbf{f}$  are defined in Equations (32) and (33), respectively.

## 5. Theoretical Analysis on the Framework Performance

The proposed multiple-target positioning framework aims to solve a parameter estimation problem. Its performance can be evaluated either numerically or theoretically. In this section, the lower bounds on the estimation errors are determined from theoretical studies. As one of the most widely used lower bounds, the Cramér-Rao lower bound (CRLB) is chosen as the performance benchmark. The framework is performance-bound in terms of the minimum achievable variance provided by any unbiased estimators.

Assume that a deterministic signal  $s_t(\boldsymbol{\theta})$  with an unknown vector parameter  $\boldsymbol{\theta}$  is observed in white Gaussian noise as Equation (35).

$$\mathbf{z}_t = \mathbf{h}_t(\boldsymbol{\theta}) + \mathbf{v}_t \quad (35)$$

where  $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_t)$ . We wish to estimate  $\boldsymbol{\theta}$  from  $\mathbf{z}$ . The Fisher information matrix [47] is given by Equation (36).

$$[\mathbf{I}(\boldsymbol{\theta})]_{m,n} = \left[ \frac{\partial \mathbf{h}_t(\boldsymbol{\theta})}{\partial \theta_m} \right]^T \mathbf{C}_t^{-1} \left[ \frac{\partial \mathbf{h}_t(\boldsymbol{\theta})}{\partial \theta_n} \right] \quad (36)$$

Taking the inverse of  $\mathbf{I}(\boldsymbol{\theta})$ , the CRLB for the parameters is then obtained from its diagonal elements. The CRLB for  $\theta_m$  is the  $(m, m)$  entry of  $\mathbf{I}^{-1}(\boldsymbol{\theta})$ .

For the proposed framework, the following measurements are considered.

$\mathbf{z}_{i,t}^{(V)}$  vehicle  $i \in \mathcal{V}$ , measured from GNSS/INS;

$\mathbf{z}_{il,t}^{(V2V)}$  measured from vehicle  $i$  to vehicle  $l$ , where  $i \in \mathcal{V}$  and  $l \in \mathcal{V}$ ;

$\mathbf{z}_{j,t}^{(F)}$  feature  $j \in \mathcal{F}$ , measured from the HD map;

$\mathbf{z}_{ij,t}^{(V2F)}$  vehicle  $i \in \mathcal{V}$  to feature  $j \in \mathcal{F}$ , measured from the vehicle's on-board sensors; and

$\mathbf{z}_{ik,t}^{(V2O)}$  vehicle  $i$  to object  $k \in \mathcal{O}$ .

For convenience, all the measurements available are reformulated to the following compact form:

$$\mathbf{z}_t = \mathbf{h}_t(\boldsymbol{\theta}) + \mathbf{v}_t = \begin{bmatrix} z_{i,t}^G \\ z_{j,t}^M \\ z_{i,j,t}^E \end{bmatrix} + \mathbf{v}_t \quad (37)$$

where  $z_{i,t}^G$ ,  $z_{j,t}^M$  and  $z_{i,j,t}^E$  are defined in Equations (6)–(8), respectively. The unknown parameters are obtained from Equation (38).

$$\boldsymbol{\theta}^T = \underbrace{\begin{bmatrix} x_{i,t}^{(V)} & y_{i,t}^{(V)} & \theta_{i,t} \end{bmatrix}}_{i \in \mathcal{V}} \underbrace{\begin{bmatrix} x_{j,t}^{(F)} & y_{j,t}^{(F)} \end{bmatrix}}_{j \in \mathcal{F}} \underbrace{\begin{bmatrix} x_{k,t}^{(O)} & y_{k,t}^{(O)} \end{bmatrix}}_{k \in \mathcal{O}} \quad (38)$$

We observe that  $\mathbf{z}_t$  is Gaussian distributed with mean  $\mathbf{h}_t(\boldsymbol{\theta})$  and covariance matrix  $\mathbf{C}_t$ :

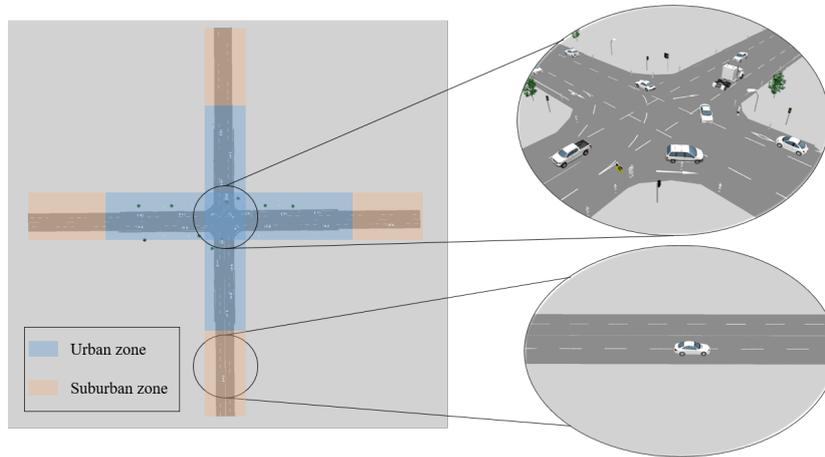
$$\mathbf{z}_t \sim \mathcal{N}(\mathbf{h}_t(\boldsymbol{\theta}), \mathbf{C}_t) \quad (39)$$

The CRLB for  $\boldsymbol{\theta}$  is obtained by substituting  $\mathbf{h}_t(\boldsymbol{\theta})$  and  $\mathbf{C}_t$  into Equation (36).

## 6. Numerical Results

In this section, we discuss the simulation experiments conducted under typical vehicular network scenarios to verify the localization and perception capacity results of the proposed algorithm. We also demonstrate its environmental adaptability in subsequent discussions on factors that influence the final performance by considering different scene configurations.

As shown in Figure 4, we build an intersection with 2 two-way two-lane roads. This scenario consists of a busy urban area and a suburban area. The trajectories of all the vehicles and objects as well as the traffic scene configuration come from VISSIM, a behavior-based traffic flow simulator [48]. Each road is 330 m long. In the middle, until approximately 200 m from the intersection, we simulate a busy urban scenario with lamps, traffic lights, and traffic signs located randomly on the roadside. Pedestrians walking around the road and across the intersection are generated. Outside of and far from the intersection, nothing is placed on the roadside, which simulates the scenario of a suburban area. In the simulation, connected and disconnected vehicles start from one end of the road simultaneously, then travel straight or turn left or right at the intersection, and exit the scene almost simultaneously. Therefore, vehicles are in the urban area in the middle section of the simulation steps, and the starting and ending segments correspond to suburban scenes.



**Figure 4.** The considered intersection simulated in VISSIM.

### 6.1. Performance in a Typical Scenario

First, we validate the effectiveness of our algorithm in a fixed scenario and compare it with the method proposed by Gloria et al. [19], as well as the theoretical bound CRLB. We set the accuracy of each measurement to that of low-cost devices. The configuration for the scenario and each measurement are as follows.

- $N_v = 6$
- $N_f = 23$  (15 lamps, 4 traffic lights, and 4 traffic signs)
- $N_o = 18$  (10 pedestrians, and 8 disconnected vehicles)
- $\delta_{sensor} = 0.25$  m, and  $\delta_{map} = 0.05$  m
- $\delta_{GNSS,l} = 2.5$  m, and  $\delta_{GNSS,\theta} = 0.1$  rad
- $\theta_{FOV} = 70$  m,  $R_s = 80$  m, and  $\theta_b = 2^\circ$
- $L_f = 100$  m,  $L_r = 30$  m, and  $W_d = 60$  m

We run the simulation 200 times, and noise is added to the measurements independently for each iteration. One localization result of the vehicles, objects, and features, and their true positions in the urban area is shown in Figure 5. As the 6 vehicles face similar scenes in every simulation step, we statistically analyze the positioning errors of all the vehicles. The root-mean-square error (RMSE) of self-positioning for all 6 vehicles at simulation time  $t$  is calculated using Equation (40).

$$\text{RMSE}_t^{(V)} = \sqrt{\frac{1}{MN} \sum_{j=1}^N \sum_{i=1}^M \|\hat{\mathbf{p}}_{i,j,t}^{(V)} - \mathbf{p}_{i,j,t}^{(V)}\|_2^2} \quad (40)$$

where  $\hat{\mathbf{p}}_{i,j,t}^{(V)}$  is the self-positioning result of vehicle  $i$  at the  $j$ th run at simulation step  $t$ , and  $\mathbf{p}_{i,j,t}^{(V)}$  is its corresponding ground truth. The self-positioning mean-square error MSE bound is calculated using Equation (41)

$$\text{CRLB}_t = \frac{1}{M} \sum_{i=1}^M (\text{CRLB}(x_{i,t}^{(V)}) + \text{CRLB}(y_{i,t}^{(V)})) \quad (41)$$

where  $\text{CRLB}(x_{i,t}^{(V)})$  and  $\text{CRLB}(y_{i,t}^{(V)})$  are the CRLBs of the  $x$  and  $y$  coordinates of vehicle  $i$  at simulation step  $t$ .

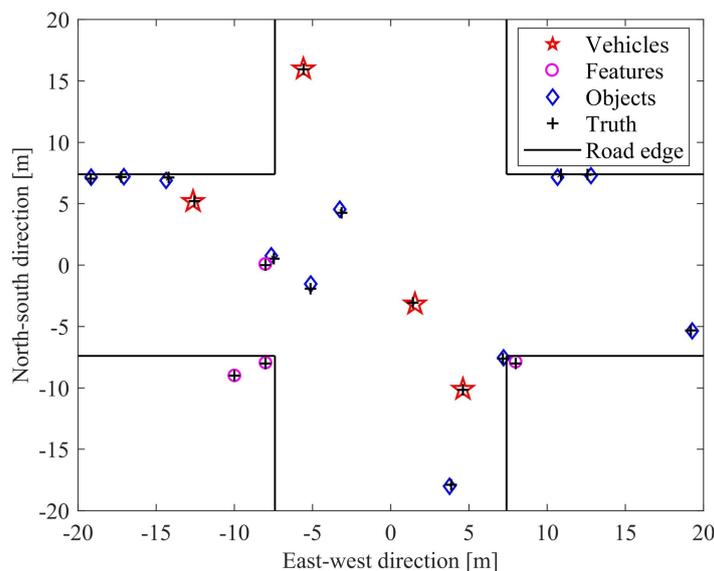


Figure 5. Multiple-target position performance in the urban area.

As shown in Figure 6, the proposed method is compared with that of Gloria et al. [19], as well as the RMSE bound  $\sqrt{\text{CRLB}_t}$  (see Equation (41)).

It is obvious that compared with the original GNSS measurements, we obtain a significantly improved positioning result by using the information from V2X and the HD map. In particular, in the urban area (simulation steps 21–55), our algorithm achieves high positioning accuracy (0.16 m), which is also lower than that of the method in [19] (1.79 m). Our positioning accuracy is close to the theoretical lower bound given by CRLB, which shows that we have effectively used all valuable information.

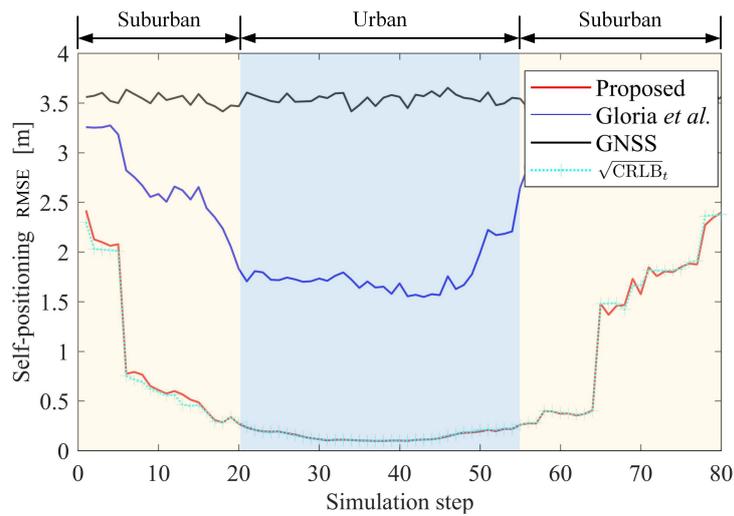


Figure 6. Self-localization error.

We also give the number of constraints at each simulation time in Figure 7. Overall, the greater the number of constraints available, the better our positioning results are. In fact, the study in [19] only uses GNSS and V-O constraints, while we have used additional constraints including V-F, V-V, and prior constraints of the HD map.

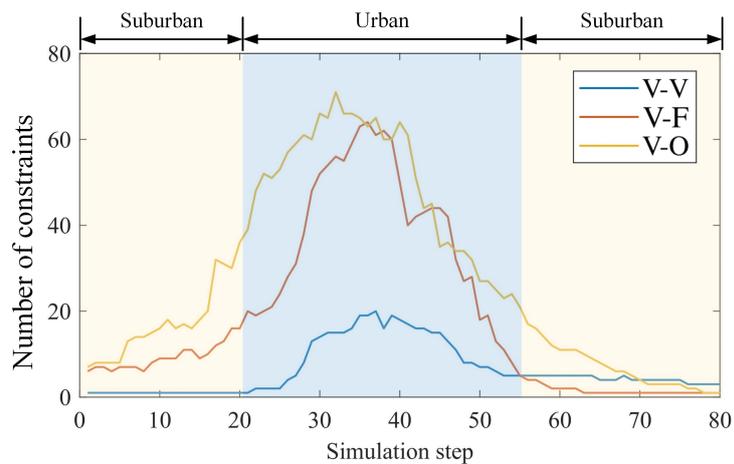


Figure 7. Amount of constraints used at every simulation step.

In the suburban area where the sensing ranges of different vehicles have little overlap, the method proposed in [19] acts ineffectively as few constraints are available. However, our positioning is still a significant improvement in these challenging areas. Environmental adaptability will be further discussed in the next section.

In terms of the positioning of other targets, we compared both the target location precision and sensing integrity. We transform the positioning results of these targets into the body coordinate system (i.e., the vehicle coordinate system shown in Figure 3 for analysis, as this analysis is consistent with the vehicle sensing system). The target positioning accuracy of a vehicle is evaluated by the RMSE of all targets within the demanding space (see Equation (42)).

$$\text{RMSE}_{i,t}^{(T)} = \sqrt{\frac{1}{ON} \sum_{k=1}^O \sum_{j=1}^N \left\| \hat{\mathbf{p}}_{k,j,t}^{(T)} - \mathbf{p}_{k,j,t}^{(T)} \right\|_2^2} \quad (42)$$

where  $\hat{\mathbf{p}}_{k,j,t}^{(T)}$  is the localization result of the  $k$ th target in the vehicle coordinate system at the  $j$ th runtime, and  $\mathbf{p}_{k,j,t}^{(T)}$  is its truth position. The RMSE of relative localization i.e.,  $\text{RMSE}_t^{(T)}$  is defined as the root mean square of the location of all vehicles. In fact, such a result is affected by both the absolute positioning and vehicular self-positioning, which makes our analysis more rigorous. The result is shown in Figure 8. In the urban area, the RMSE is 0.17 m; it is much smaller than that (0.24 m) obtained by Gloria's method as well as that (0.32 m) provided by the vehicles' on-board sensors. Higher perception accuracy is also achieved in suburban areas.

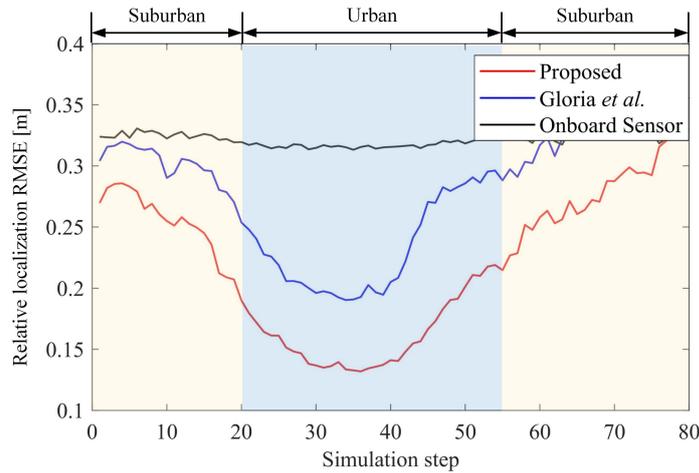


Figure 8. Relative positioning error of surrounding targets.

Figure 9 shows the improvement in sensing integrity. The blue line is the true value of the number of targets within the demanding space. Based on the raw data of the on-board sensors, only 41.81% of targets are captured in the urban area, owing to occlusion or limited sensing range, while the proposed method enables 90.42% of the targets to be captured. The improvement comes from the sharing of information between the connected vehicles, and the information provided by the real-time dynamic map.

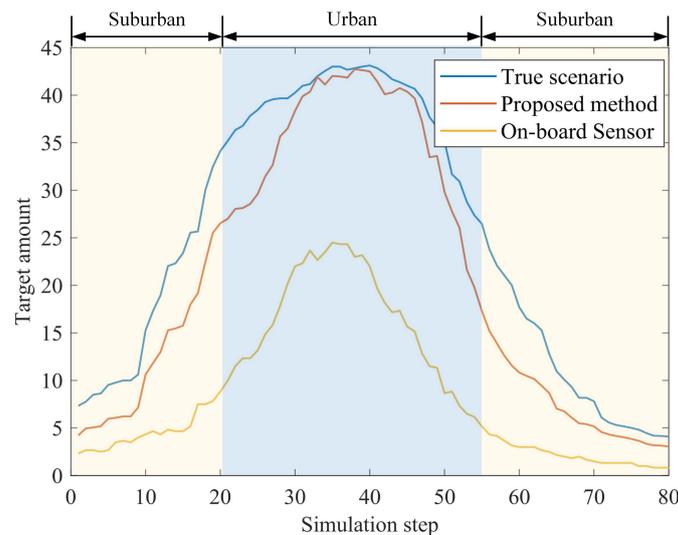


Figure 9. Perception integrity in terms of the number of targets captured.

In summary, our approach significantly improves multiple-target positioning in terms of accuracy and integrity over that achieved using the original measurements, and is also more effective than other methods.

## 6.2. Adaptability to Different Scenarios

In the following section, we analyze the impact of different elements on the results of multiple-target positioning to demonstrate the adaptability of our method to different environments. Simultaneously, we discuss the contributions of different constraint types to the results.

### 6.2.1. Number of Connected Vehicles

The vehicular positioning and relative localization of the targets in terms of the number of connected vehicles are shown in Figures 10 and 11, respectively. Except for the number of connected vehicles, the configurations of the scene are identical, with 6 features and 3 objects. The accuracy of sensing and GNSS/INS are the same as those in previous experiments.

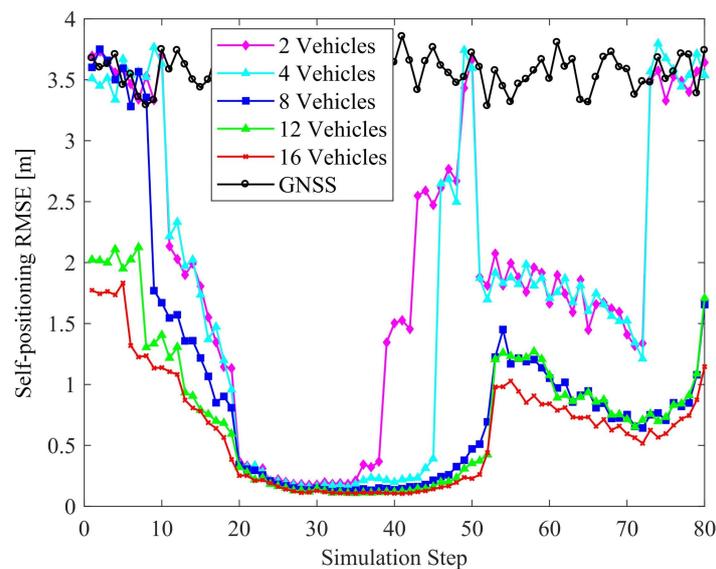


Figure 10. Effect of the number of connected vehicles on localization accuracy.

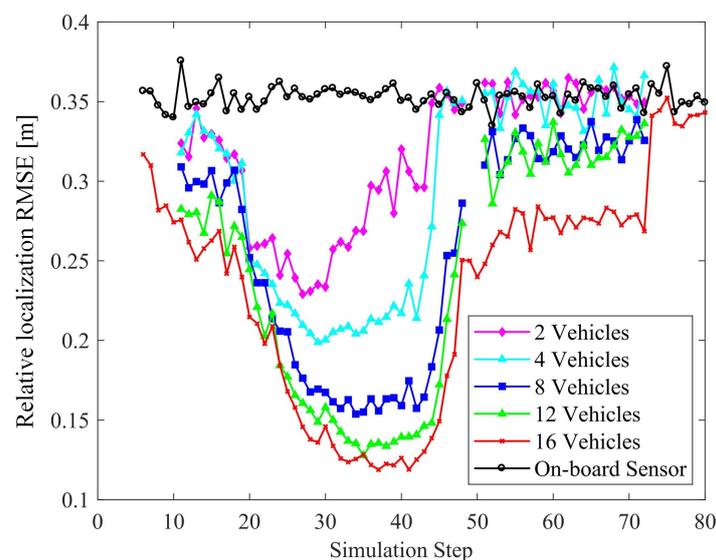


Figure 11. Effect of the number of connected vehicles on perception accuracy.

As can be seen from the figures, although there is a limited number of connected vehicles (only 2 vehicles), the positioning capacities of the vehicles and other targets are significantly improved in the urban areas. In general, the greater the number of connected vehicles, the smaller the corresponding positioning errors, as more V-V constraints are imposed between the vehicles. It is

worth noting that there is a trend that the RMSEs of self-positioning increase as the simulation step increases from 70 to 80, as the vehicles are heading the end of the roads, where features and objects are becoming increasingly sparse. In theory, the lower bound of positioning error of 12 vehicles is lower than 8 vehicles. Due to the limited simulation times, the RMSE fluctuates near the theoretical bounds. Therefore, the RMSE of 12 vehicles seems close to those of 8 vehicles. However, based on the existing results, we cannot say that there is a trend that they will exceed those of 8 vehicles.

Another interesting observation is that for a given number of connected vehicles, the positioning errors in the suburban areas are larger than those in the urban areas, but it is worth noting that the decreasing trend of RMSE with increasing the number of connected vehicles is more significant. This is because, unlike the urban area with sufficient types of constraints, the constraints in suburbs are mainly of V-V. Therefore, we argue that the number of connected vehicles is important for improving location accuracy in the suburban area, which is consistent with the results shown in Figure 10.

As for the perception accuracy, in suburban area, with the increase of the number of vehicles, there are more constraints which benefits the positioning of connected vehicles and other objects. The underlying mechanism can be explained by Equation (36). As the constraints increase, the dimension of  $h_t(\theta)$  increases, which leads to an increase in the value of the information matrix  $\mathbf{I}(\theta)$ . The CRLB for  $\theta_m$ , which can be calculated with the diagonal element of  $\mathbf{I}(\theta)^{-1}$  will decrease, which leads to the reduction of the absolute positioning error for each object. As the perception result is gained by projecting the absolute of other targets to the vehicle-body coordinate system based on the self-positioning, the reduction of the absolute positioning error finally improves the perception accuracy.

#### 6.2.2. Number of Features

The effect of the number of features are shown in Figures 12 and 13. There are four connected vehicles running on the road with four objects and different numbers of features. It is obvious that increasing the number of features improves the accuracy of vehicle positioning and relative localization of the targets. Compared to raw measurements, even with a few features, the positioning and perception errors are reduced by exploiting the vehicle-to-target constraints and HD map information. It is noteworthy that at the intersection, the positioning accuracy is high when the number of features is 5, 10, or 30. However, if there is no feature i.e., no map information is used, the positioning error is obviously higher. This reflects the contribution of the HD map to vehicle positioning.

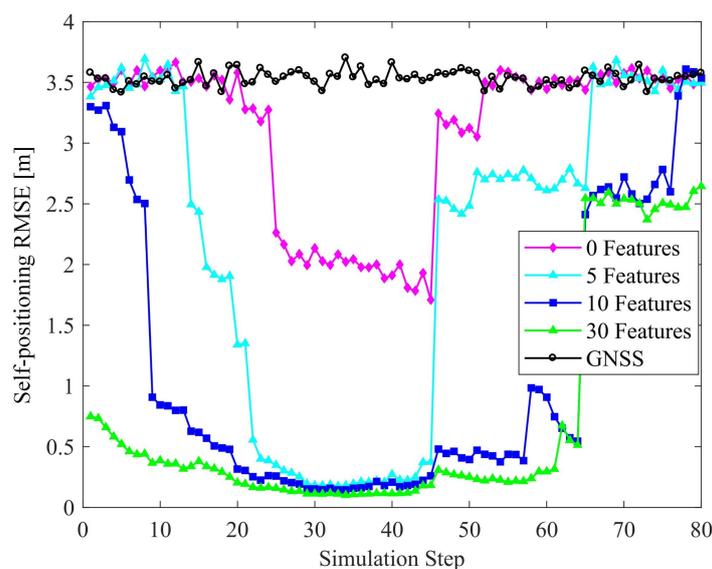
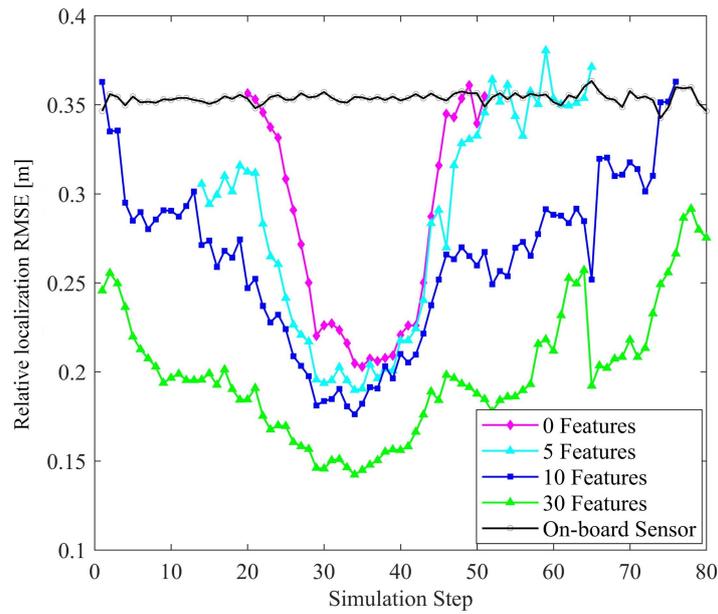


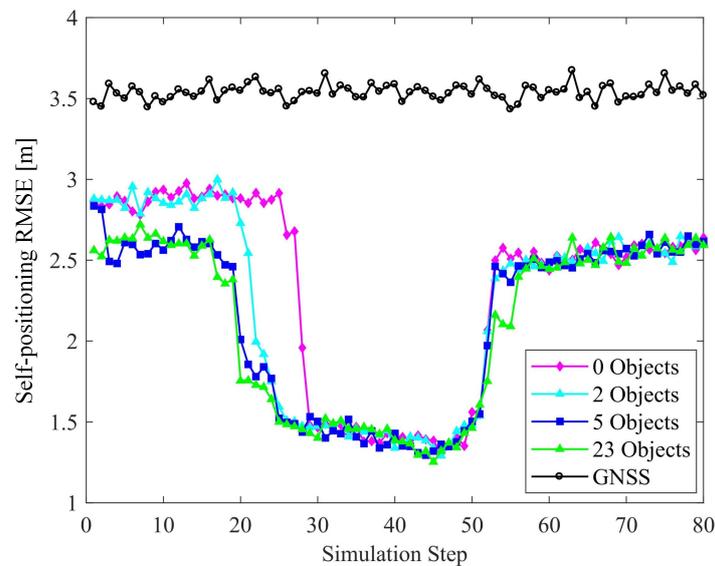
Figure 12. Effect of the number of features on localization accuracy.



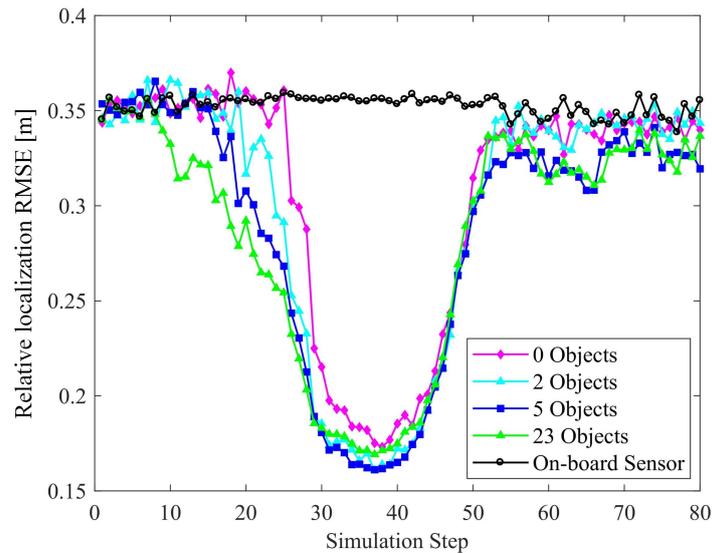
**Figure 13.** Effect of the number of features on perception accuracy.

### 6.2.3. Number of Objects

In comparing the results obtained when the number of objects varies, we set  $N_v = 8$  and  $N_f = 0$ . The corresponding results are given in Figures 14 and 15. In the suburban zones, the increase in the number of objects improves the positioning. However, it is noteworthy that such an improvement in the intersection zone is not obvious. The reason is that in the former zone, there are very few V-V constraints, and V-O constraints play the main role in improving the results. Hence, adding objects can effectively improve the positioning. However, at the intersections, the V-V constraints formed by 8 vehicles are dominant and the results approach the theoretical bounds achievable. There is no significant improvement in accuracy with the addition of objects.



**Figure 14.** Effect of the number of objects on localization accuracy.



**Figure 15.** Effect of the number object on perception accuracy.

In the case of relative localization, increasing the number of objects can reduce the overall perception accuracy, instead. As the number of objects increases, the proportion of objects among all the perceived targets participating in the perceptual precision calculation increases. The objects are less constrained relative to other targets, and the overall clarity of perception declines. Considering this and the former discussion, we argue that V-O constraints are less effective than V-V and V-F constraints in improving multiple-target positioning accuracy.

## 7. Conclusions

This study focuses on the problem of multiple-target positioning for ICVs. We propose a unified theoretical framework for positioning both vehicles and other targets, wherein sensor data from V2X and HD map data are effectively fused with GNSS/INS and on-board sensors. By jointly exploiting the vehicle-to-target constraints and HD map information, the vehicular localization accuracy can be enhanced to meet the requirements of high-level automated driving by using low-cost GNSS/INS and on-board sensors in urban areas. Meanwhile, the confidence and integrity of the results of relative localization of targets are significantly improved, realizing sensing beyond line of sight and field of view, which can improve the transportation efficiency and safety. Furthermore, the proposed framework is applicable to more challenging scenarios entailing fewer connected vehicles and sparse features and objects. In future research, we plan to remove the limiting assumption of data association employed in this study by applying association methods in the process of optimization. We will also study the formulation of communication delay of V2X in the data fusion framework.

**Author Contributions:** Conceptualization, Z.X., D.Y. and F.W.; methodology, Z.X. and F.W.; software, Z.X.; validation, K.J. and F.W.; formal analysis, D.Y.; writing—original draft preparation, Z.X.; writing—review and editing, D.Y.; visualization, F.W.

**Funding:** This work was supported in part by the International Science and Technology Cooperation Program of China (2016YFE0102200), in part by the National Key Research and Development Program of China (2018YFB0105000), in part by the National Natural Science Foundation of China (61773234 and U1864203), in part by the Project of Tsinghua University and Toyota Joint Research Center for AI Technology of Automated Vehicle (TT2018-02), and in part by the software developed in the Beijing Municipal Science and Technology Program (D171100005117002 and Z181100005918001). This work was also supported by European Commission H2020 Marie Skłodowska-Curie project under the grant agreement No. 700044 and the State Key Laboratory of Automotive Safety and Energy under Project No. KF1804.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wang, X.; Ning, Z.; Hu, X.; Ngai, E.C.; Wang, L.; Hu, B.; Kwok, R.Y.K. A city-wide real-time traffic management system: Enabling crowdsensing in social internet of vehicles. *IEEE Commun. Mag.* **2018**, *56*, 19–25. [[CrossRef](#)]
2. Usman, M.; Asghar, M.R.; Ansari, I.S.; Granelli, F.; Qaraqe, K.A. Technologies and solutions for location-based services in smart cities: Past, present, and future. *IEEE Access* **2018**, *6*, 22240–22248. [[CrossRef](#)]
3. Brummelen, J.V.; O'Brien, M.; Gruyer, D.; Najjaran, H. Autonomous vehicle perception: The technology of today and tomorrow. *Transp. Res. Part C Emerg. Technol.* **2018**, *89*, 384–406. [[CrossRef](#)]
4. Yang, D.; Jiang, K.; Zhao, D.; Yu, C.; Cao, Z.; Xie, S.; Xiao, Z.; Jiao, X.; Wang, S.; Zhang, K. Intelligent and connected vehicles: Current status and future perspectives. *Sci. China Technol. Sci.* **2018**, *61*, 1446–1471. [[CrossRef](#)]
5. Vivacqua, R.P.D.; Bertozzi, M.; Cerri, P.; Martins, F.N.; Vassallo, R.F. Self-localization based on visual lane marking maps: An accurate low-cost approach for autonomous driving. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 582–597. [[CrossRef](#)]
6. Kim, S.; Liu, W.; Ang, M.H.; Frazzoli, E.; Rus, D. The Impact of Cooperative Perception on Decision Making and Planning of Autonomous Vehicles. *IEEE Intell. Transp. Syst. Mag.* **2015**, *7*, 39–50. [[CrossRef](#)]
7. Bresson, G.; Alsayed, Z.; Yu, L.; Glaser, S. Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving. *IEEE Trans. Intell. Veh.* **2017**, *2*, 194–220. [[CrossRef](#)]
8. Kuutti, S.; Fallah, S.; Katsaros, K.; Dianati, M.; McCullough, F.; Mouzakitis, A. A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications. *IEEE Internet Things J.* **2018**, *5*, 829–846. [[CrossRef](#)]
9. Liu, J.; Liu, J. Intelligent and connected vehicles: Current situation, future directions, and challenges. *IEEE Commun. Stand. Mag.* **2018**, *2*, 59–65. [[CrossRef](#)]
10. Skog, I.; Handel, P. In-car positioning and navigation technologies—A survey. *IEEE Trans. Intell. Transp. Syst.* **2009**, *10*, 4–21. [[CrossRef](#)]
11. Jackson, J.; Davis, B.; Gebre-Egziabher, D. A performance assessment of low-cost RTK GNSS receivers. In Proceedings of the IEEE/ION Position, Location and Navigation Symposium (PLANS), Monterey, CA, USA, 23–26 April 2018; pp. 642–649.
12. Ahmed, E.; Gharavi, H. Cooperative vehicular networking: A survey. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 996–1014. [[CrossRef](#)] [[PubMed](#)]
13. Nam, S.; Lee, D.; Lee, J.; Park, S. CNVPS: Cooperative Neighboring Vehicle Positioning System Based on Vehicle-to-Vehicle Communication. *IEEE Access* **2019**, *7*, 16847–16857. [[CrossRef](#)]
14. Jeong, H.Y.; Nguyen, H.H.; Bhawiyuga, A. Spatiotemporal Local-Remote Sensor Fusion (ST-LRSF) for Cooperative Vehicle Positioning. *Sensors* **2018**, *18*, 1092. [[CrossRef](#)] [[PubMed](#)]
15. De, P.M.F. Survey on Ranging Sensors and Cooperative Techniques for Relative Positioning of Vehicles. *Sensors* **2017**, *17*, 271.
16. Severi, S.; Wymeersch, H.; Härrä, J.; Ulmschneider, M.; Denis, B.; Bartels, M. Beyond GNSS: Highly accurate localization for cooperative-intelligent transport systems. In Proceedings of the IEEE Wireless Communications and Networking Conference, Barcelona, Spain, 15–18 April 2018; pp. 1–6.
17. Hobert, L.; Festag, A.; Llatser, I.; Altomare, L.; Visintainer, F.; Kovacs, A. Enhancements of V2X communication in support of cooperative autonomous driving. *IEEE Commun. Mag.* **2015**, *53*, 64–70. [[CrossRef](#)]
18. Shen, X.; Andersen, H.; Leong, W.K.; Kong, H.X.; Ang, M.H., Jr.; Rus, D. A General Framework for Multi-vehicle Cooperative Localization Using Pose Graph. *arXiv* **2017**, arXiv:1704.01252.
19. Soatti, G.; Nicoli, M.; Garcia, N.; Denis, B.; Raulefs, R.; Wymeersch, H. Implicit Cooperative Positioning in Vehicular Networks. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 3964–3980. [[CrossRef](#)]
20. Soatti, G.; Nicoli, M.; Garcia, N.; Denis, B.; Raulefs, R.; Wymeersch, H. Enhanced vehicle positioning in cooperative ITS by joint sensing of passive features. In Proceedings of the IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6.
21. Fascista, A.; Ciccarese, G.; Coluccia, A.; Ricci, G. Angle of arrival-based cooperative positioning for smart vehicles. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 2880–2892. [[CrossRef](#)]
22. Seif, H.G.; Hu, X. Autonomous Driving in the iCity—HD Maps as a Key Challenge of the Automotive Industry. *Engineering* **2016**, *2*, 159–162. [[CrossRef](#)]

23. Durrant-Whyte, H.; Bailey, T. Simultaneous localization and mapping: Part I. *IEEE Robot. Autom. Mag.* **2006**, *13*, 99–110. [[CrossRef](#)]
24. Javanmardi, E.; Javanmardi, M.; Gu, Y.; Kamijo, S. Factors to Evaluate Capability of Map for Vehicle Localization. *IEEE Access* **2018**, *6*, 49850–49867. [[CrossRef](#)]
25. Quack, T.M.; Reiter, M.; Abel, D. Digital map generation and localization for vehicles in urban intersections using LiDAR and GNSS data. *IFAC-PapersOnLine* **2017**, *50*, 251–257. [[CrossRef](#)]
26. Xiao, Z.; Jiang, K.; Xie, S.; Wen, T.; Yu, C.; Yang, D. Monocular Vehicle Self-localization method based on Compact Semantic Map. In Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3083–3090.
27. Hsu, C.M.; Shiu, C.W. 3D LiDAR-Based Precision Vehicle Localization with Movable Region Constraints. *Sensors* **2019**, *19*, 942. [[CrossRef](#)]
28. Levinson, J.; Montemerlo, M.; Thrun, S. Map-based precision vehicle localization in urban environments. In Proceedings of the Robotics: Science and Systems III, Atlanta, GA, USA, 27–30 June 2007.
29. Wang, C.; Huang, H.; Ji, Y.; Wang, B.; Yang, M. Vehicle localization at an intersection using a traffic light map. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 1432–1441. [[CrossRef](#)]
30. Atia, M.M.; Hilal, A.R.; Stellings, C.; Hartwell, E.; Toonstra, J.; Miners, W.B.; Basir, O.A. A Low-Cost Lane-Determination System Using GNSS/IMU Fusion and HMM-Based Multistage Map Matching. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3027–3037. [[CrossRef](#)]
31. Oguz-Ekim, P.; Ali, K.; Madadi, Z.; Quitin, F.; Tay, W.P. Proof of concept study using DSRC, IMU and map fusion for vehicle localization in GNSS-denied environments. In Proceedings of the IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 841–846.
32. Campbell, S.; O'Mahony, N.; Krcpalcova, L.; Riordan, D.; Walsh, J.; Murphy, A.; Ryan, C. Sensor Technology in Autonomous Vehicles: A review. In Proceedings of the 29th Irish Signals and Systems Conference, Belfast, UK, 21–22 June 2018; pp. 1–4.
33. Kim, S.W.; Wei, L. Cooperative Autonomous Driving: A Mirror Neuron Inspired Intention Awareness and Cooperative Perception Approach. *IEEE Intell. Transp. Syst. Mag.* **2016**, *8*, 23–32. [[CrossRef](#)]
34. Kim, S.; Qin, B.; Chong, Z.J.; Shen, X.; Liu, W.; Ang, M.H.; Frazzoli, E.; Rus, D. Multivehicle Cooperative Driving Using Cooperative Perception: Design and Experimental Validation. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 663–680. [[CrossRef](#)]
35. Rauch, A.; Klanner, F.; Rasshofer, R.; Dietmayer, K. Car2X-based perception in a high-level fusion architecture for cooperative perception systems. In Proceedings of the IEEE Intelligent Vehicles Symposium, Alcalá de Henares, Spain, 3–7 June 2012; pp. 270–275.
36. Xiao, Z.; Mo, Z.; Jiang, K.; Yang, D. Multimedia Fusion at Semantic Level in Vehicle Cooperative Perception. In Proceedings of the IEEE International Conference on Multimedia Expo Workshops (ICMEW), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.
37. Paden, B.; Čáp, M.; Yong, S.Z.; Yershov, D.; Frazzoli, E. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Trans. Intell. Veh.* **2016**, *1*, 33–55. [[CrossRef](#)]
38. Yang, B.; Liang, M.; Urtasun, R. HDNET: Exploiting HD Maps for 3D Object Detection. In Proceedings of the 2nd Conference on Robot Learning, Zurich, Switzerland, 29–31 October 2018; Volume 87, pp. 146–155.
39. Kurdej, M.; Moras, J.; Cherfaoui, V.; Bonnifait, P. Map-Aided Evidential Grids for Driving Scene Understanding. *IEEE Intell. Transp. Syst. Mag.* **2015**, *7*, 30–41. [[CrossRef](#)]
40. Wang, S.; Fidler, S.; Urtasun, R. Holistic 3D scene understanding from a single geo-tagged image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3964–3972.
41. Kschischang, F.R.; Frey, B.J.; Loeliger, H.A. Factor graphs and the sum-product algorithm. *IEEE Trans. Inf. Theory* **2001**, *47*, 498–519. [[CrossRef](#)]
42. He, Z.; Hu, Y.; Wu, J.; Wang, J.; Kang, W. A comprehensive method for multipath performance analysis of GNSS navigation signals. In Proceedings of the IEEE International Conference on Signal Processing, Xi'an, China, 14–16 September 2011.
43. Hamza, B.; Nebylov, A. Robust Nonlinear Filtering Applied to Integrated Navigation System INS/GNSS under Non Gaussian Measurement noise effect. *IFAC Proc. Vol.* **2012**, *45*, 202–207. [[CrossRef](#)]

44. Rauch, A.; Maier, S.; Klanner, F.; Dietmayer, K. Inter-vehicle object association for cooperative perception systems. In Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems, The Hague, The Netherlands, 6–9 October 2013; pp. 893–898.
45. Thomaidis, G.; Tsogas, M.; Lytrivis, P.; Karaseitanidis, G.; Amditis, A. Multiple hypothesis tracking for data association in vehicular networks. *Inf. Fusion* **2013**, *14*, 374–383. [[CrossRef](#)]
46. Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [[CrossRef](#)]
47. Kay, S.M. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*; Prentice Hall: Upper Saddle River, NJ, USA, 1993.
48. Lownes, N.E.; Machemehl, R. VISSIM: A multi-parameter sensitivity analysis. In Proceedings of the Simulation Conference, Monterey, CA, USA, 3–6 December 2006; pp. 1406–1413.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).