sensors

MDPI

Article

Robust Target Detection and Tracking Algorithm Based on Roadside Radar and Camera

Jie Bai [1], Sen Li [1], Han Zhang [1], Libo Huang [1,]* and Ping Wang [2]

[1] Institute of Intelligent Vehicles, School of Automotive Studies, Tongji University, Shanghai 201804, China; baijie@tongji.edu.cn (J.B.); lisen@tongji.edu.cn (S.L.); hanzhang@tongji.edu.cn (H.Z.)
[2] College of Electronic Science and Technology, Tongji University, Shanghai 201804, China; pwang@tongji.edu.cn
* Correspondence: huanglibo@tongji.edu.cn; Tel.: +86-187-2135-9738

Abstract: Intelligent transportation systems (ITSs) play an increasingly important role in traffic management and traffic safety. Smart cameras are the most widely used sensors in ITSs. However, cameras suffer from a reduction in detection and positioning accuracy due to target occlusion and external environmental interference, which has become a bottleneck restricting ITS development. This work designs a stable perception system based on a millimeter-wave radar and camera to address these problems. Radar has better ranging accuracy and weather robustness, which is a better complement to camera perception. Based on an improved Gaussian mixture probability hypothesis density (GM-PHD) filter, we also propose an optimal attribute fusion algorithm for target detection and tracking. The algorithm selects the sensors' optimal measurement attributes to improve the localization accuracy while introducing an adaptive attenuation function and loss tags to ensure the continuity of the target trajectory. The verification experiments of the algorithm and the perception system demonstrate that our scheme can steadily output the classification and high-precision localization information of the target. The proposed framework could guide the design of safer and more efficient ITSs with low costs.

Keywords: target detection and tracking; sensor fusion; roadside radar and camera; intelligent transportation system

check for updates

Citation: Bai, J.; Li, S.; Zhang, H.; Huang, L.; Wang, P. Robust Target Detection and Tracking Algorithm Based on Roadside Radar and Camera. *Sensors* **2021**, *21*, 1116. https://doi.org/10.3390/s21041116

Academic Editor: Michail Antoniou
Received: 27 December 2020
Accepted: 1 February 2021
Published: 5 February 2021

1. Introduction

Road traffic safety and efficiency are the key challenges in modern transportation. According to the Global Status Report on Roads in 2018, the number of road traffic deaths is over 1.35 million per year. More than half of all deaths are among vulnerable road participants: cyclists, motorcyclists, and pedestrians [1]. Intersection collisions account for over 40% of total traffic accidents, which not only seriously threaten people's lives, but also cause severe traffic congestion [2]. Research has shown that over 60% of these collisions can be avoided if drivers receive a warning just half a second in advance [3,4]. To improve traffic problems and build smart cities, the intelligent transportation system (ITS) has been widely studied [5,6]. Especially in recent years, with the rapid development of 5G communication technology, artificial intelligence, sensor technology, and high-performance chip technology, the related technology of ITS has exploded, such as intelligent connected vehicles (ICVs) [7], vehicle-to-everything (V2X) [8,9], and edge sensing and computing [10,11]. Over-the-horizon perception for ICVs based on V2X and intelligent roadside units (RSUs) can be used for collision prevention at intersections to improve traffic safety [12,13]. Moreover, the essential task of intelligent RSUs is to build a stable and reliable perception system.

The roadside perception unit (RPU) uses cameras, lidars, and millimeter-wave (MMW) radars to detect and locate targets within the field of view. Due to the advancement of sensor technology and perception algorithms, research on roadside perception solutions can be divided into two phases. In early research, before 2011, the focus was on low-beam lidar and traditional vision-processing methods. H. Zhao et al. used multiple laser scanners located at different locations to form an observation network for intersection monitoring [14]. Wang C. at el. proposed a move-stop hypothesis tracking approach to solve the move-stop-move maneuvers by using the single-line lidar [15]. Daniel Meissner et al. utilized multiple four-layer laser scanners mounted at high parts of the infrastructure to detect and track objects inside intersections [16]. Oliver, N. M. et al. proposed a multilevel tracking approach that combined low-level image-based blob detection and high-level Kalman filtering for multi-target tracking at intersections [17]. Peyman B. developed a vision-based surveillance system for vehicle counting and tracking in an intersection by using a background-modeling technique [18]. Due to the lack of performance in target classification, radar was mainly used for monitoring the speed and distance of road vehicles during this time [19,20]. For a more stable and robust perception system, researchers have proposed some schemes to improve the target detection and tracking performance by using laser and camera fusion [21] or radar and camera fusion [22].

In the last decade, the performance of perception systems has been enormously improved with rapid developments of high-beam lidars, high-resolution radars, and deep-learning technologies. The current state-of-the-art algorithms based on convolutional neural networks (CNNs), such as yolo-V4 [23] and EfficientNet [24], can offer both high processing speed and detection accuracy. The problem of detecting small targets at a far distance also has been greatly improved. Shuai Hua et al. built a multi-vehicle tracking framework based on the Yolo network that can be used for real-time traffic applications [25]. Some scholars have also tried to use advanced image-processing methods to estimate the vehicle–pedestrian collision probability [26] or detect abnormal events [27] at intersections. The new generation of lidar has a 360-degree scanning field of view and more scanning beams, such as 32 lines, 64 lines, and 128 lines, which can provide higher detection accuracy. J. Z. et al. achieved tracking and speed estimation of vehicles at intersections using 32-line lidar with a speed estimation accuracy of 0.22 m/s [28]. Z. Z. et al. achieved large-area scenario modeling and high-resolution target tracking at intersections using 3D point clouds [29]. Some authors have also implemented real-time queue range detection [30] and collision risk analysis [31] based on roadside lidar. Similarly, the new generation of 79 GHz ultra-bandwidth radar overcomes the lack of angular resolution and is also widely used in ITS. W. L. et al. proposed a classification algorithm for pedestrians and vehicles at intersections based on point clouds of 79 GHz radar [32]. Some scholars have built safety systems for vulnerable road users [33] and traffic intersection surveillance systems [34,35] at intersections based on 79 GHz radar and V2X technology. In order to improve the robustness of RPU, some scholars have proposed methods based on radar and camera fusion for vehicle detection and width estimation in bad weather [36,37]. Christoph S. et al. proposed a two-stream CNN method for auto-calibration of a radar and camera to realize robust detection of vehicles on the highway [38]. Kaul P. at al. presented a weakly supervised multiclass semantic segmentation network to achieve semantic segmentation of multichannel radar scan inputs with the help of a camera and lidar segmentation system [39].

In the studies of and developments in ITSs, the main purpose is to improve traffic safety and efficiency. Therefore, an important prerequisite for all related research, such as traffic monitoring, behavior prediction, and collision warning, is to establish a robust target detection and tracking system. In addition to the detection accuracy, we also need to consider the cost of the whole solution for large-scale deployment [4]. In [40], the authors evaluated sensors using five criteria: range, resolution, contrast, weather, and cost. In ideal conditions, the vehicle-detection range can reach up to 100 m, and the pedestrian-detection range can reach up to 43 m using a 16-line lidar [29]. Although a higher beam lidar can increase detection accuracy and effective range, the high cost of lidar hinders its large-scale market application. According to the investigation in [40], the cost of a 16-line lidar is more than 10 times that of a 77 GHz millimeter-wave radar or a mono-camera, and the price of a 64-line lidar is between USD 40,000 and 70,000. Cameras are currently the most widely used sensor, and image-processing algorithms based on deep learning have evolved tremendously. However, the performance of the lidar and camera suffers a large degradation in bad weather scenarios [41,42] Radar has weather robustness and doppler velocity sensitivity, but its angular resolution is insufficient [43]. Cameras and radars can complement strengths and weaknesses in several aspects. Their low cost is also a hot spot of current research.

In this study, we developed a stable RPU for the detection and real-time localization of traffic participants. We hope to adopt a low-cost solution based on radar and camera fusion to realize high localization accuracy close to lidar, which can contribute to large-scale applications. There are two major problems in radar and camera perception [43]. The first is measurement loss and noise interference in complex traffic scenarios. The limited performance of the detection algorithm, target occlusion, and environmental noise can cause missed detections and false alarms. The second is the limited localization accuracy. The longitudinal range accuracy of the camera and the lateral range accuracy of the radar decrease significantly during the localization of targets at far distances. Therefore, the contribution of this paper is to propose an optimal attribute fusion algorithm, which is a detection-tracking algorithm based on the Gaussian mixture probability hypothesis density (GM-PHD) framework [44]. We introduce lost labels and attenuation functions to adaptively maintain the target life cycle and achieve continuity of the tracking trajectory.

The structure of this paper is organized as follows. Section 2 introduces the related work of target detection and data processing. Then, in Section 3, we describe the proposed optimal attribute fusion tracking algorithm. Section 4 analyzes and discusses the experimental results. Finally, Section 5 summarizes the conclusions of this paper and future works.

2. Preliminaries

In this section, we mainly introduce the detection principle and calibration process of radar and camera, as well as the pre-processing method for data fusion.

2.1. Target State Vector and Motion Model

Consider a traffic scenario at an intersection, with vehicles, pedestrians, cyclists, and other targets moving on the roadway. Let $M(k)$ denote the number of targets at time k, and the motion state of targets can be represented as the state set $X_k = \{x_{1,k}, \cdots, x_{i,k}, \cdots, x_{M(k),k}\}$. The state vector $x_{i,k}$ describes the position and velocity of target i at time k and is defined as:

$$x_{i,k} = [x\ y\ v_x\ v_y]^T,\ i \in M(k) \tag{1}$$

Let F_k be the state transfer matrix, and the movement of the target follows the motion of Equation (2):

$$x_{k+1|k} = F_k x_k + \xi_k \tag{2}$$

In the RPU, the sensor is often mounted on a light pole and has a fixed view field. All traffic participants appear and disappear independently of the sensor's field of view. Targets can be captured if they are within the sensor's range of perception. If the number of targets observed by the sensor is $N(k)$ at time k, then all observed targets can be represented by the measurement set $Z_k = \{z_{1,k}, \cdots, z_{j,k}, \cdots, z_{N(k),k}\}$. The observation vector $z_{i,k}$ is an imperfect measurement of the state of the observed target j at time k, having the same form as $x_{i,k}$. The sensor's observation model is described as:

$$z_k = H_k x_k + \varsigma_k \tag{3}$$

where H_k is the observation matrix of the linear dynamic system, and ξ_k and ς_k are system and observation white Gaussian noise with covariance $\mathcal{N}(\xi; 0, R)$ and $\mathcal{N}(\varsigma; 0, R)$, respectively. Note that the state set and observation set of the target have no correspondence and order, and $M(k)$ is not equal to $N(k)$ due to clutter interference and occlusion. Our task is to estimate the number of targets and their state from the multiple observation set.

The state and observation set of targets are considered to be random finite sets (RFS), and the number of targets in the set varies with time and has no regular ordering [45].

$$\begin{aligned} X_k &= \left\{x_{1,k}, x_{2,k}, \cdots, x_{M(k),k}\right\} \in \mathcal{F}(\mathcal{X}) \\ Z_k &= \left\{z_{1,k}, z_{2,k}, \cdots, z_{N(k),k}\right\} \in \mathcal{F}(\mathcal{Z}) \end{aligned} \tag{4}$$

where X_k and Z_k are subsets of $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Z})$, respectively, and $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Z})$ are the set of all finite subsets of state space $\mathcal{X}$ and measurement space $\mathcal{Z}$, respectively. The task of MTT is to estimate the state of targets from the sensor observations. Based on the theory of RFS, the multi-target tracking can be regarded as a filtering problem with state space $\mathcal{F}(\mathcal{X})$ and observation space $\mathcal{F}(\mathcal{Z})$. Generally, the number of elements in the state set is smaller than the observation set, i.e., $N_{(k)} \le M_{(k)}$.

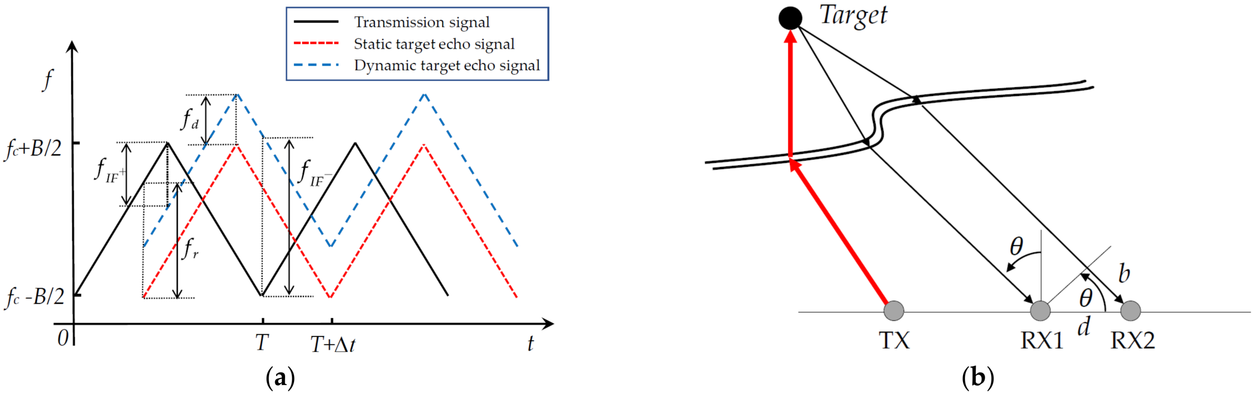
2.2. Radar Detection Model

MMW radar directionally transmits electromagnetic radio frequency signals and analyzes the echo signals of surroundings to detect targets. By measuring the time delay and phase shift of the echo signal, the distance and velocity of the target can be measured. Directional antennas or phase comparison techniques can determine the azimuth of the target [46]. As shown in Figure 1a, the echo signal produces a time delay due to the propagation of electromagnetic waves between the radar and the target, resulting in a distance frequency shift. For dynamic targets, in addition to the distance frequency shift f_d, the target movement also produces Doppler frequency shift f_r. The transmission signal and the echo signal produce two differential frequencies $f_{IF}{}^+$ and $f_{IF}{}^-$ on the rising and falling edges of the frequency, and $f_{IF}{}^+ = f_r – f_d$, $f_{IF}{}^+ = f_r + f_d$. The range R and velocity v of a target can be calculated by the following equation:

$$R = \frac{T \times c}{8B}\left(f_{IF^+} + f_{IF^-}\right) \tag{5}$$

$$v = \frac{c}{4f_c}\left(f_{IF^+} - f_{IF^-}\right) \tag{6}$$

where T and B are the period of frequency modulation and modulation bandwidth, respectively; f is the center frequency of the transmission waveform; and c is the speed of light.



(**a**) (**b**)

Figure 1. (**a**) Linear frequency modulation continuous wave (LFMCW) radar signal waveform processing; (**b**) principle of azimuth measurement of the target. TX is the transmitting antenna, and RX is the receiving antenna.

The azimuth is estimated using the phase-comparison method, as shown in Figure 1b. The target signal has a travel distance during propagation, and thus a corresponding phase difference in the echo signal. The azimuth θ of the target is calculated as shown in Equation (7):

$$\theta = \arcsin(\frac{\lambda w}{2\pi d}) \tag{7}$$

where w is the phase difference caused by the distance difference of the target echo signal, d is the distance between antennas RX1 and RX2, and λ is the wavelength. Therefore, the state of the target i can be represented as a vector $z_i^R = \left[R^R, v^R, \theta^R\right]$. After a series of data processing, radar can output sparse point-cloud information for the dynamic and static targets of the surroundings. The radar measurement set is $Z^R = \{z_1^R, \cdots, z_i^R, \cdots, z_M^R\}$, and the observation equation at time k is as follows:

$$Z_k^R = H^R X_k + W_k^R(R, \theta) \tag{8}$$

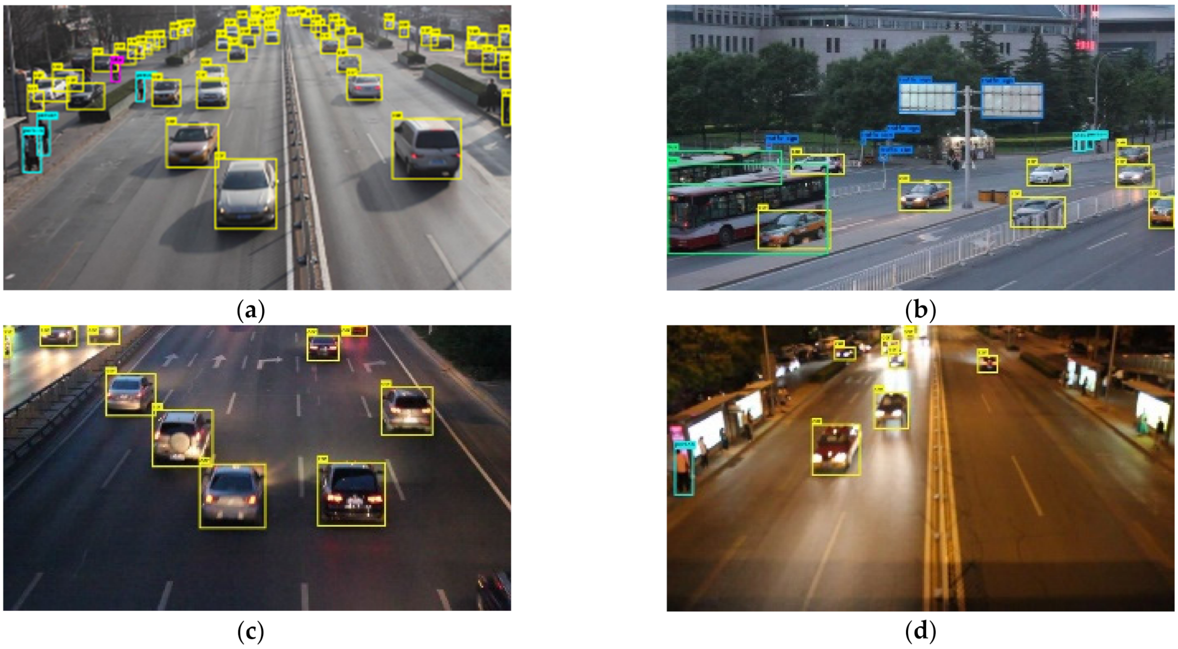
where H^R is the radar observation vector and W_k^R is observation noise, which is related to radar characteristics and environmental factors. MMW radar has high range accuracy and Doppler velocity accuracy. However, radar lacks capabilities in angle measurement and target classification, which can be compensated with the help of cameras.

2.3. Camera Detection Model

Cameras are the most widely used in intelligent vehicles and ITSs, providing rich scenario information such as target classification [47], lane lines [48], traffic signs [49], etc. Compared to automotive cameras, roadside cameras are mounted higher and have a wider observation field. The main challenges for roadside cameras are the dynamics of the background and the detection of small targets. In recent years, the rapid development of deep-learning-based image-processing algorithms has dramatically improved accuracy and real-time target-detection performance. In this paper, we retrained the YoloV4 framework, which is a one-stage target-detection algorithm, to perform image-based target detection and localization tasks [23]. Yolov4 offers a good balance of detection precision and detection speed, making it a suitable edge-computing platform for roadside units. To ensure the detection precision, we performed migration training on the roadside traffic dataset UA-DETRAC to obtain the training weights for the roadside scenario [50]. UA-DETRAC is a dataset for multi-objective detection based on real urban traffic conditions collected in China. The dataset contains a variety of traffic scenarios (urban highways, intersections, flyovers, and elevated gate crossings) under different weather conditions (sunny, rainy, and cloudy), as well as for different time periods (daytime and nighttime), covering typical Chinese urban roads. The partial test results after retraining are shown in Figure 2. The test results show that the algorithm can implement pedestrian and vehicle detection, and detect small targets over long distances in clear daylight. However, as shown in Figure 2d, some vehicles and pedestrians were not detected due to interference from the intense background light. The image detection results are the bounding box and classification information of targets. The spatial position information of the target can be obtained by converting the pixel coordinate system to the geodetic coordinate system through camera calibration, which will be described in the next section. The velocity of the target can be estimated by the position difference between adjacent frames. Similarly, the state of the target i can be also represented as a vector $z_i^C = \left[R^C, v^C, \theta^C\right]$. The camera measurement set is $Z^C = \{z_1^C, \cdots, z_i^C, \cdots, z_N^C\}$, and the observation equation at time k is expressed in a form similar to Equation (9) as:

$$Z_k^C = H^C X_k + W_k^C(R, \theta) \tag{9}$$

where H^C is the camera observation vector and W_k^C is observation noise. Compared to MMW radar, the advantage of the camera is the classification of the target and the estimation of the azimuth, but the disadvantage is the lack of accuracy in the estimation of range and velocity.



(**a**) (**b**) (**c**) (**d**)

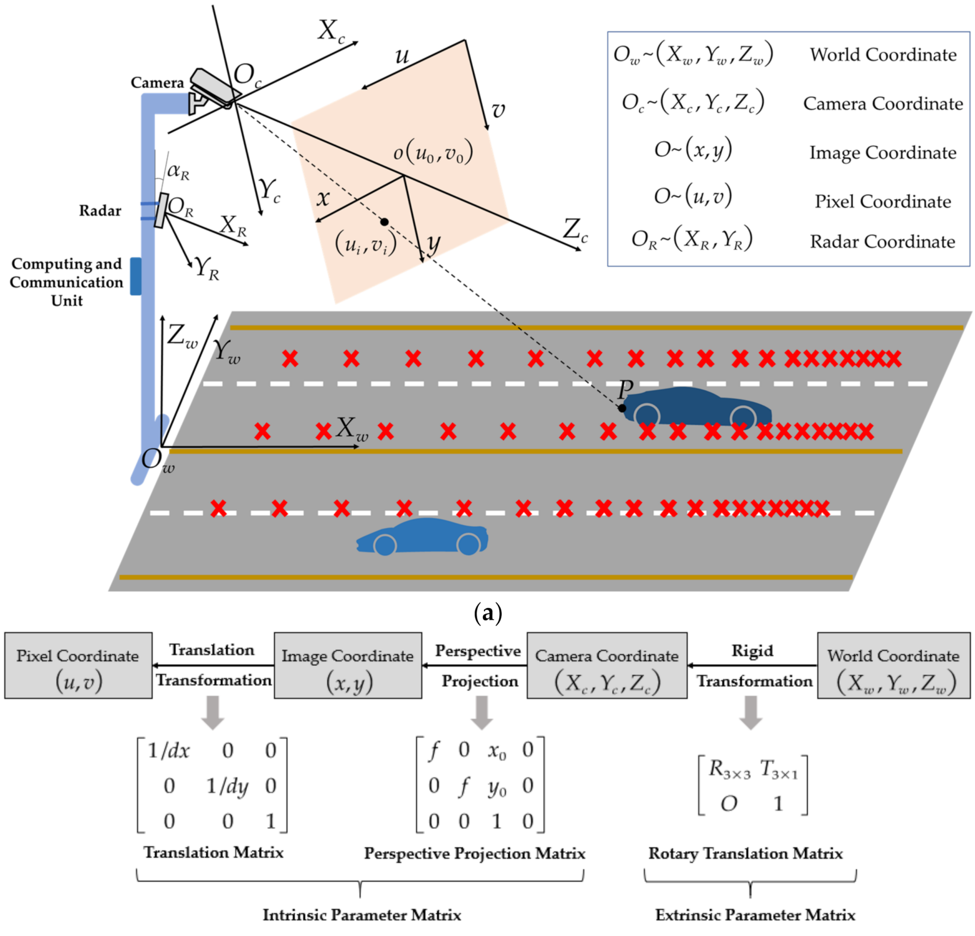
Figure 2. Partial test results for different scenarios of the UA-DETRAC dataset. (**a**) Sunny daytime; (**b**) cloudy evening; (**c**) night; (**d**) interference by strong light at night.

2.4. Sensor Calibration

In this work, we performed target detection and tracking based on the fusion of roadside radar and camera. With the joint calibration of the roadside camera and radar, as shown in Figure 3a, the radar and camera detection results can be fused with data in the same world coordinate system. According to the calibration process, the calibration parameters of the camera can be divided into an intrinsic parameter matrix and an extrinsic parameter matrix, and the mapping relationship is as follows:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_x & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{10}$$

where the first to the right of the Equation (10) is the camera’s intrinsic parameter matrix; f_x and f_y are the scale factors of the camera on the u and v axes, $f_x = f/dx, f_y = f/dy$. (u_0, v_0) is the optical center of the camera, $u_0 = x_0/d_x$, $v_0 = y_0/d_y$. Most cameras have radial and tangential distortion, which are also intrinsic parameters of the camera that need to be calibrated. The intrinsic parameters of the camera were calibrated using the classical calibration method [51]. R and T are the rotation matrix and translation matrix, respectively, which are extrinsic parameters of the camera related to the camera’s mounting position and angle [52]. If the camera’s yaw angle and roll angle relative to the world coordinate system are set to zero, the extrinsic parameter matrix can be represented by the camera’s optical center height H and pitch angle α_c. The vanishing point calibration method can be used to calibrate the extrinsic parameters of the camera [53].


Camera
O_c
X_c
Y_c
Z_c
u
v
$o(u_0,v_0)$
x
y
(u_i,v_i)
α_R
Radar
O_R
X_R
Y_R
Computing and Communication Unit
Z_w
Y_w
X_w
O_w
P
O_w~(X_w,Y_w,Z_w) World Coordinate
O_c~(X_c,Y_c,Z_c) Camera Coordinate
O~(x,y) Image Coordinate
O~(u,v) Pixel Coordinate
O_R~(X_R,Y_R) Radar Coordinate
Pixel Coordinate (u,v)
Translation Transformation
Image Coordinate (x,y)
Perspective Projection
Camera Coordinate (X_c,Y_c,Z_c)
Rigid Transformation
World Coordinate (X_w,Y_w,Z_w)
$\begin{bmatrix} 1/dx & 0 & 0 \\ 0 & 1/dy & 0 \\ 0 & 0 & 1 \end{bmatrix}$
$\begin{bmatrix} f & 0 & x_0 & 0 \\ 0 & f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
$\begin{bmatrix} R_{3\times3} & T_{3\times1} \\ O & 1 \end{bmatrix}$
Translation Matrix
Perspective Projection Matrix
Rotary Translation Matrix
Intrinsic Parameter Matrix
Extrinsic Parameter Matrix

(**a**)

(**b**)

Figure 3. (**a**) Joint calibration method of roadside cameras and millimeter-wave (MMW) radar; (**b**) coordinate mapping from the world coordinate system to the pixel coordinate systems.

The scanning range of the radar is a conical region. To guarantee a high accuracy of the radar at its observation range, only the installation height h and pitch angle α_R need to be considered. and the conversion from the radar to the world coordinate system is:

$$\begin{bmatrix} R_w \\ v_w \\ \theta_w \end{bmatrix} = \begin{bmatrix} R^R \cos\alpha_R \\ v^R \cos\alpha_R \\ \theta^R \end{bmatrix} \tag{11}$$

Since the radar and image are installed at different positions, and the measurement error increases with distance, a joint calibration is necessary for the common observation area. The sensor measurements require an additional correction to calibrate the measurement error, which can be accomplished using ground calibration points for simultaneous calibration, as shown by the red cross in Figure 3a. Corrections between calibration points can be estimated by interpolation, and the distance between the calibration points should decrease as the measurement distance increases. Although the calibration increases the workload, it is indeed important and can significantly improve the accuracy of target estimates. Ultimately, radar and image observations can be unified into a single world coordinate system for data fusion.

2.5. Data Pre-Correlation

According to the sensor detection model constructed in Sections 2.2 and 2.3, the MMW radar has higher accuracy in range and Doppler velocity. At the same time, the camera has more advantages in target classification and azimuth. Before tracking, the measurement sets of two sensors are pre-correlated by using the nearest-neighbor correlation method. Using radar-range measurements and image-angle measurements to update the target state can improve localization accuracy. The pre-correlation results of sensor measurements at time k can be divided into fusion measurement Z_k^f, uncorrelated camera measurement Z_k^C, and uncorrelated radar measurement Z_k^R, forming a new measurement set $Z_k^{New} = \left\{Z_k^f, Z_k^C,\ Z_k^R\right\}$. The new state vector of the target can be described in terms of position, velocity, and classification $x = \left[x\ \dot{x}\ y\ \dot{y}\ \xi\right]$:

$$x_k^i = \begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ \xi \end{bmatrix} = \begin{bmatrix} R_w \cos\theta^C \\ v_w \cos\theta^C \\ R_w \sin\theta^C \\ v_w \sin\theta^C \\ \xi \end{bmatrix} \tag{12}$$

where ξ represents the classification of target detected by camera. For ease of computer processing, the target's classification can be denoted by a number, e.g., "pedestrian: 1, vehicle: 2, cyclist: 3, unassociated target: 0".

3. System Overview

For real-time monitoring and security purposes, the ITS system needs to obtain continuous and accurate state information of surrounding targets in the observation area. To improve the accuracy and stability of target detection and tracking, we proposed an optimal attribute fusion algorithm based on the GM-PHD algorithm framework [44]. The GM-PHD tracking framework estimates the state and number of targets simultaneously, which is suitable for the complex traffic scenarios in which the number of targets changes over time. In our improvement strategy, the idea is to build loss tags and attenuation function to achieve the continuity of the target trajectory and use the optimal measurement to improve localization accuracy. The improved tracking algorithm is introduced in the following five processes: initialization, prediction, update, pruning, and merging.

3.1. Initialization

In the GM-PHD algorithm, Gaussian components are used to represent targets or potential targets. A Gaussian component $\{w, m, P, \varepsilon\}$ is expressed by weight w, mean state m, covariance P, and classification ε. The state distribution of Gaussian components is described by a Gaussian mixture probability density function in the observation space, and the initial probability intensity v_0 is:

$$v_0(x) = \sum_{i=1}^{J_0} w_0^i \mathcal{N}\left(x; m_0^i, P_0^i\right) \tag{13}$$

where J_0 is the number of targets or Gaussian components at initial moment; $\mathcal{N}(\cdot;\cdot,\cdot)$ denotes the distribution of Gaussian component i with weight w_0^i, mean state m_0^i, and covariance matrix P_0^i. The classification ε of the target is not changed once it has been defined. New birth targets may appear with each measurement, and the target intensity function of new targets at time K is given by:

$$\gamma_k(x) = \sum_{i=1}^{J_{\gamma,k}} w_{\gamma,k}^i \mathcal{N}\left(x; m_{\gamma,k}^i, P_{\gamma,k}^i\right) \tag{14}$$

where the covariance matrix $P^i_{\gamma,k}$ describes the spread of the birth intensity near the peak $m^i_{\gamma,k}$. The weight $w^i_{\gamma,k}$ indicates the expected number of new targets from $m^i_{\gamma,k}$, and the weight initialization function is given by:

$$w^i_{\gamma,k} = 0.25\mathcal{N}\left(x; m^1_{\gamma}, P^1_{\gamma}\right) + 0.25\mathcal{N}\left(x; m^2_{\gamma}, P^2_{\gamma}\right) + I^i_k \tag{15}$$

where m^1_{γ} and m^2_{γ} can be set as the focused point in the observation scene. If the new birth target is closer to the focal point, the higher the weighting coefficient is. I^i_k is the bias coefficient, which is determined according to the correlation results of radar and image measurement sets. If the measurement $x^i_{\gamma,k}$ is a fusion measurement, $I^i_k = w_{fu}$, otherwise $I^i_k = 0$. The measurement loss of the target is unpredictable, and each Gaussian component is assigned a loss tag to count the number of times the measurement is lost. For identified targets, measurement loss may occur due to occlusion or undetected by the sensor. So, a loss tag L_{loss} is used to count the measurement loss times of the target, which will be introduced in the update step. Finally, the initialization result of a target is $\{w, m, P, \varepsilon, L_{loss}\}$, and $L_{loss} = 0$ at initial moment.

3.2. Prediction

The posterior intensity at time $k-1$ is a Gaussian mixture form, and the predicted intensity for time k is also a Gaussian mixture form, which consists of the survival target intensity $v_{S,k|k-1}$ and the new birth target intensity $\gamma_k(x)$. In this step, ε and L_{loss} remain unchanged.

$$v_{k-1}(x) = \sum_{i=1}^{J_{k-1}} w^i_{k-1}\mathcal{N}\left(x; m^i_{k-1}, P^i_{k-1}\right) \tag{16}$$

$$v_{k|k-1}(x) = v_{s,k|k-1}(x) + \gamma_k(x) \tag{17}$$

where the $\gamma_k(x)$ is given in Equation (9), and the $v_{S,k|k-1}$ is given by:

$$v_{S,k|k-1}(x) = \sum_{i=1}^{I_{S,k|k-1}} w^i_{S,k|k-1} P_{S,k}\mathcal{N}\left(x; m^i_{S,k|k-1}, P^i_{S,k|k-1}\right) \tag{18}$$

where $P_{S,k}$ is target survival probability, which is difficult to predict directly in the actual tracking scenario. Under the condition of low-speed target or high sensor sampling, the state transfer can be approximately considered as a linear Gaussian process.

$$w^i_{k|k-1} = w^i_{k-1} \tag{19}$$

$$m^i_{k|k-1} = F_{k-1} m^i_{k-1} \tag{20}$$

$$P^i_{k|k-1} = Q_{k-1} + F_{k-1} P^i_{k-1} F^T_{k-1} \tag{21}$$

3.3. Update

The posterior density update also satisfies the Gaussian mixture distribution, consisting of a detected part and an undetected part:

$$v_k(x) = (1 - P_{D,k}) v_{k|k-1}(x) + \sum_{z \in Z_k} v_{D,k}(x; z) \tag{22}$$

where $P_{D,k}$ is the detection probability, which cannot be accurately estimated like $P_{S,k}$. $v_{D,k}$ is the posterior density of the detected part. $(1 - P_{D,k}) v_{k|k-1}(x)$ indicates that the undetected target is updated with a state prediction instead of the posterior update. A method for joint estimation of clutter distribution and detection probability was proposed in [54]. However, this approach cannot directly solve the problem of measurements loss of targets. To simplify the update process without directly estimating $P_{S,k}$ and $P_{D,k}$, the elliptical gating method was introduced in this study, inspired by ideas in [55,56]. Assume that the residual vector of Gaussian terms corresponding to the *i*-th observation value and the *j*-th prediction value is ε^{ij}, and the corresponding covariance matrix is S^j.

$$\varepsilon^{ij} = z_k^i - H_k\left(x_{k|k-1}^j\right) \tag{23}$$

$$S_k^j = H_k P_k^j H_k^T + R_k \tag{24}$$

Then the discriminant of the elliptic threshold can be expressed as:

$$\left(\varepsilon^{ij}\right)^T \left(S^j\right)^{-1} \left(\varepsilon^{ij}\right) \leqslant T_g \tag{25}$$

where T_g is the threshold. Some studies have also proposed adaptive threshold methods to improve performance [55]. According to Equation (25), the predicted Gaussian component can be divided into two parts: measurement existence and measurement loss, denoted by $\{w, m,\ P, \varepsilon, L_{loss}\}_{j=1}^{J_{match}} \in Z_{match}$ and $\{w, m,\ P, \varepsilon, L_{loss}\}_{j=1}^{J_{loss}} \in Z_{loss}$, respectively. So, the update process consists of two parts: the detection update and the missed detection update.

For the detection update, the target is detected by the sensor at the next moment, i.e., $P_{D,k} = 1$. The posterior intensity of Gaussian components is given by:

$$v_{k|k}(k) = \sum_{z_i \in Z_{\text{match}}^{\varepsilon}} \sum_{j=1}^{J_{\text{match}}} w_k^j(z_i) \mathcal{N}\left(x; m_{k|k}^j(z_i); P_{k|k}^j\right) \tag{26}$$

where the update equation for weight w, mean state m, and covariance P is as follows:

$$w_k^j(z_i) = \frac{w_{k|k-1}^j q_k^j(z_i)}{\kappa_k(z_i) + \sum_{j=1}^{I_{\text{match}}} w_{k|k-1}^j q_k^j(z_i)} \tag{27}$$

$$m_{k|k}^j(z_i) = m_{k|k-1}^j + K_k^j\left(z - H_k m_{k|k-1}^j\right) \tag{28}$$

$$q_k^j(z_i) = \mathcal{N}\left(z_i; H_k m_{k|k-1}^j, R_k + H_k P_{k|k}^j H_k^T\right) \tag{29}$$

$$P_{k|k}^j = \left[I - K_k^j H_k\right] P_{k|k-1}^j \tag{30}$$

$$K_k^j = P_{k|k-1}^j H_k^T \left[S_k^j\right]^{-1} \tag{31}$$

Equations (24)–(31) are the recursive equations of the detection part. Note that in the weight update equation (27), the detection probability $P_{D,k}$ is removed because $P_{D,k} = 1$ in this case. The classification of Gaussian components remains unchanged, and the loss tag $L_{loss} = 0$.

For the missed detection update, the target is not detected by the sensor at the next moment, i.e., $P_{D,k} = 0$, and the idea of using predicted values for status updates continues to be followed. If target *j* loses measurement of the next time at time *k*, the corresponding loss tag value is increased by 1, i.e., $L_{loss,k|k}^j = L_{loss,k|k-1}^j + 1$. However, not all Gaussian components need to be preserved, and only the targets of interest are worth maintaining. The parameter ε can be used to help select valid targets. Gaussian components with $\varepsilon = 1,\ 2,\ or\ 3$, indicating that the targets are of concern to us, must be maintained. Gaussian components with $\varepsilon = 0$ represent other noise that can be dropped. Since the weight is essential for determining the survival and extinction of Gaussian components, targets with measurement loss should be given an attenuation function:

$$w_{k|k}^j = \alpha_k^j(t) w_{k|k-1}^j \tag{32}$$

$$m^j_{k|k} = m^j_{k|k-1} \tag{33}$$

$$P^j_{k|k} = P^j_{k|k-1} \tag{34}$$

where $\alpha^j_k(t)$ is the attenuation function, and $t = L^j_{loss,k}$. In practical applications, the attenuation function can be selected according to our needs, and the Fermi–Dirac function was chosen in this study:

$$\alpha^j_k(t) = \frac{1}{1 + \exp((t - b)/a)} \tag{35}$$

where a and b are the parameters that determine the shape of the attenuation function, as shown in Figure 4. However, the targets of missed detection cannot be maintained forever, and the maximum cycle can be limited by the parameters a and b and the threshold together. For the re-identification problem of lost targets, the elliptic threshold of Equation (25) can be increased to ensure the stability of tracking.

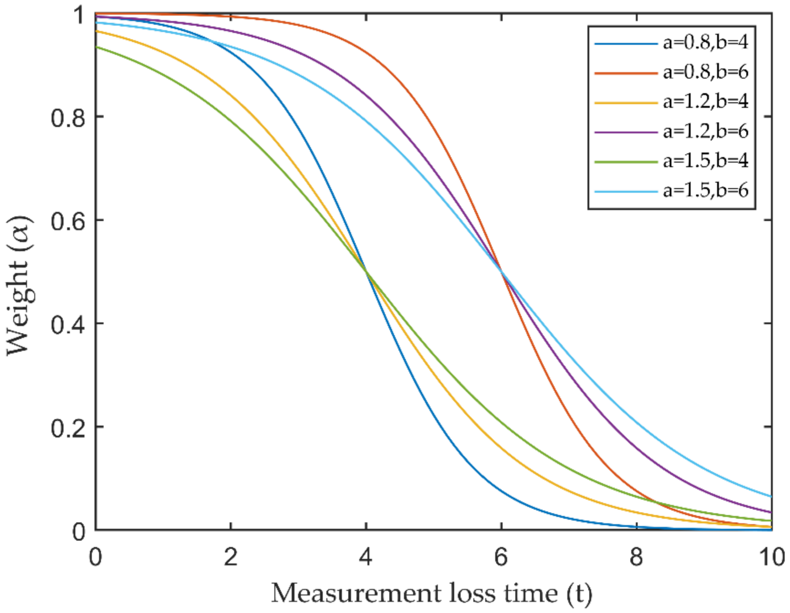


Figure 4. Visualization of weight attenuation function.

3.4. Pruning and Merging

The pruning and merging process is a key step in extracting the target and reducing the ineffective Gaussian component. The computational complexity of the heuristic pruning and merging algorithm is $\mathcal{O}\left(n_k|Z_k|^3\right)$ at each step [54]. In complex scenarios with much background noise, interference measurements consume a large amount of computational resources. Let us consider a simplified pruning process. The updated Gaussian components can be thought of as a state matrix as follows:

$$\begin{array}{cccc}
 & \textit{Upgrade} & & \begin{array}{ccccccc} \phantom{X^{(1,1)}_k} & & Z^f & & Z^C & & Z^R \end{array} \\
\begin{array}{c} J_{match} \\ \\ \\ J_{no\_match} \end{array} & \begin{bmatrix} X^1_{k|k-1} \\ X^2_{k|k-1} \\ \vdots \\ X^{J_{match}}_{k|k-1} \\ \vdots \\ X^{J_{no\_match}}_{k|k-1} \end{bmatrix} & \rightarrow & \begin{bmatrix} X^{(1,1)}_k & \cdots & X^{(1,Z^f)}_k & \cdots & X^{(1,Z^C)}_k & \cdots & X^{(1,Z^R)}_k \\ X^{(2,1)}_k & \cdots & X^{(2,Z^f)}_k & \cdots & X^{(2,Z^C)}_k & \cdots & X^{(2,Z^R)}_k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X^{(J_{match},1)}_k & \cdots & X^{(J_{match},Z^f)}_k & \cdots & X^{(J_{match},Z^C)}_k & \cdots & X^{(J_{match},Z^R)}_k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X^{(J_{no\_match},1)}_{k|k-1} & \cdots & X^{(J_{no\_match},Z^f)}_{k|k-1} & \cdots & X^{(J_{no\_match},Z^C)}_{k|k-1} & \cdots & X^{(J_{no\_match},Z^R)}_{k|k-1} \end{bmatrix}
\end{array}, \tag{36}$$

Then, the weights of each Gaussian component can be normalized by column to produce a weight matrix, which is the same as the state matrix. The state matrix and weight matrix are sparse matrices due to applicating the elliptical gating method. Assuming that each measurement has a single source, i.e., that measurement is generated by only one target. Then, each column needs to extract only one Gaussian component with the highest weight in the detection part. For the missed detection part, only Gaussian components of $\varepsilon = 1, 2,$ or 3 need to be retained. Finally, Gaussian components greater than the threshold are selected as the final estimation targets.

4. Experiment and Results

In this section, we discuss and analyze the experimental results conducted in typical traffic-intersection scenarios to evaluate the localization and perception performance of the proposed algorithm. We also demonstrated the application of the proposed algorithm in the RSU and On board Unit (OBU) platform at the system level.

4.1. Experiment Platform and Configuration

In our study, we built a movable intelligent roadside sensing and computing platform that can be used for multiple scenario testing. The roadside platform consists of the sensor unit, V2X unit, computing unit, and power unit, as shown in Figure 5.

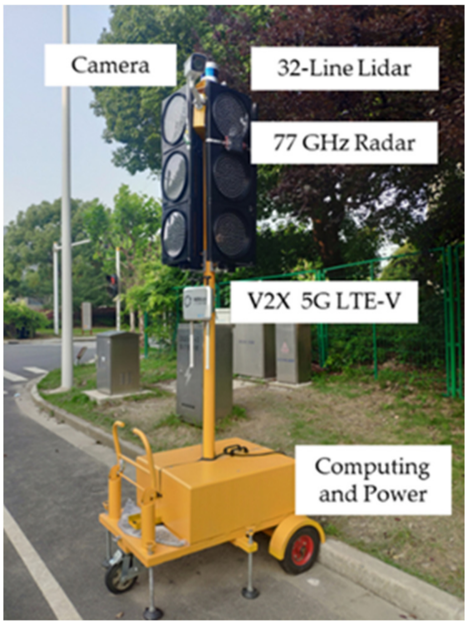


Figure 5. The movable intelligent roadside sensing and computing platform. The sensor unit consists of a radar, a camera, and a 32-line lidar; the vehicle-to-everything (V2X) unit uses the 5G LTE-V communication format; and the computing unit is an Nvidia Xavier.

During data collection, the camera resolution was 1080P (1920 × 1080) with a sampling rate of 30 Hz. Two radars were prepared for the experiment, including a high-resolution radar with 4 GHz bandwidth and a well-known 77 GHz Conti-ASR-408-21 radar. The sampling rate of the two radars was approximately 15 Hz. To verify the detection and localization accuracy of the proposed method, the detection result of high-resolution lidar was taken as the benchmark. Mechanical lidar can obtain different sampling frequencies by adjusting the rotation speed of the laser. To ensure data synchronization as synchronized as possible, the sampling frequency of the data collection system was set to 15 Hz.

The optimal sub-pattern assignment (OSPA) distance [57] is a comprehensive evaluation indicator that includes both localization and number errors, which can be taken as the evaluation criterion in this study.

$$OSPA\left(d_p^c(X,Y)\right) = \left(\frac{1}{n}\left(\min_{\pi\in\Pi_n}\sum_{i=1}^{m} d^c(x_i, y_\pi(i))^p + c^p(n-m)\right)\right)^{1/p} \quad (37)$$

where the X and Y are two sets; m and n are the dimensions of two sets; c and p are the measure factor and distance order, respectively; $c = 100$, $p = 2$.

4.2. Tracking Algorithm Performance Analysis

4.2.1. Tracking Experiment for Pedestrians

The first experiment was carried out on a wide road on campus, as shown in Figure 6. The observed objects were three pedestrians walking along the predetermined trajectory, and there was no other traffic in the test area. Pedestrians are a vulnerable group of traffic participants, and accurate detection and localization are incredibly essential to ensure pedestrian safety. The high-resolution radar was used in the first experiment. Observations were collected from the intelligent roadside platform fixed on the middle of the road. The detection results of the camera and MMW radar were converted to a ground coordinate system with the observation platform as the origin, as shown in Figure 7. The tracking results based on radar and image data fusion are shown in Figure 8.

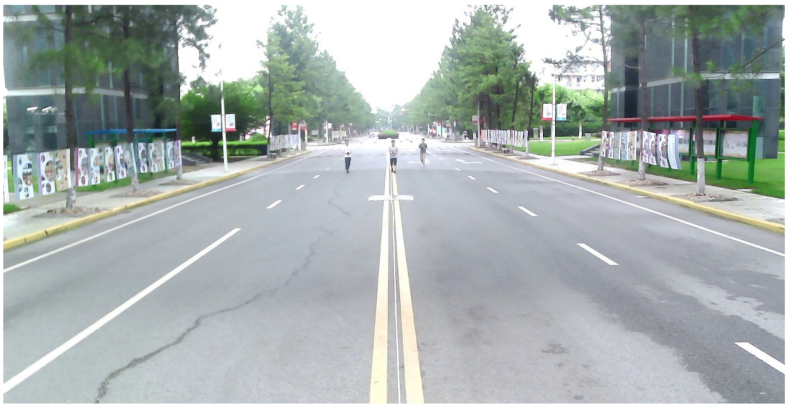


Figure 6. Pedestrian tracking experiment on a wide road.

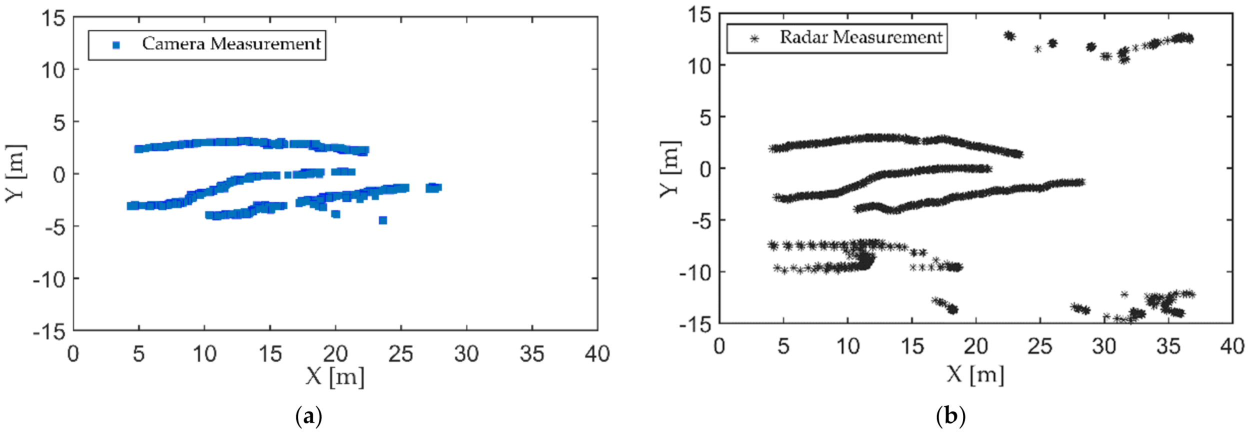


Figure 7. Detection results of targets: (**a**) camera; (**b**) radar.

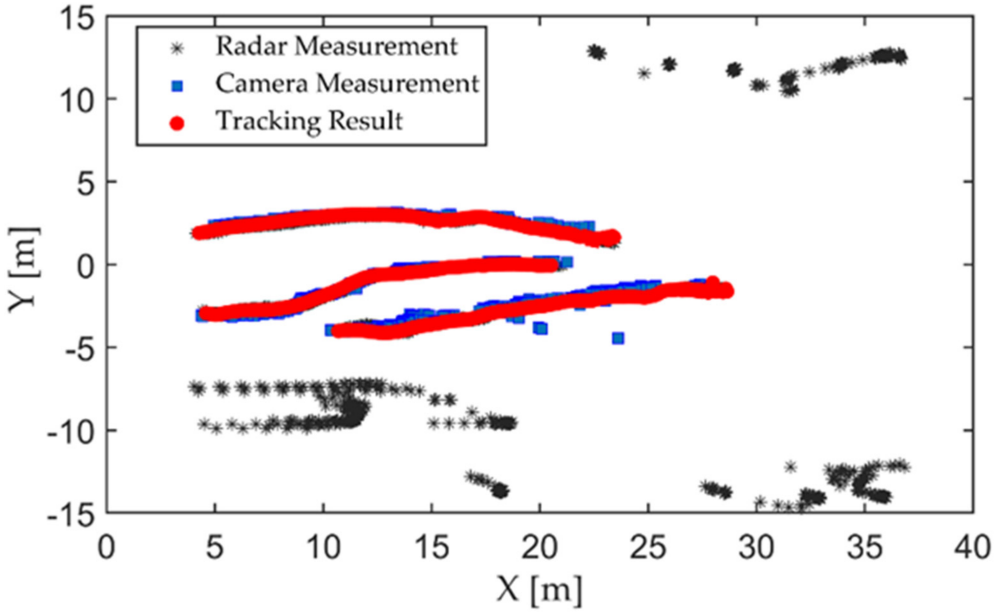


Figure 8. Tracking results of the proposed algorithm based on fusion data of the radar and camera.

As shown in Figures 7 and 8, both the camera and radar could detect pedestrians. Due to the range resolution and the localization error of the bounding box, there were some measurement errors and trajectory discontinuities in the detection results of the camera. The static targets regarded as interference measurements from the surrounding environment also appeared in radar detection results.

The improved tracking algorithm could accurately extract the actual number of targets and ensure the continuity of the target trajectory based on fusion measurements of radar and image. To illustrate the improvements of the proposed method, we compared the single sensor's tracking results and the initial GM-PHD method with our method. The single-sensor tracking algorithm adopted the tracking method in this study. All conditional assumptions of GM-PHD algorithm remained the same as [44]. The input of the GM-PHD algorithm was the fusion matrix obtained by Equation (12) without priori classification information and $P_{S,k} = 0.98$, $P_{D,k} = 0.95$. The detection results of the lidar with a higher localization accuracy were used as the benchmark. The OSPA distance is shown as Figure 9.

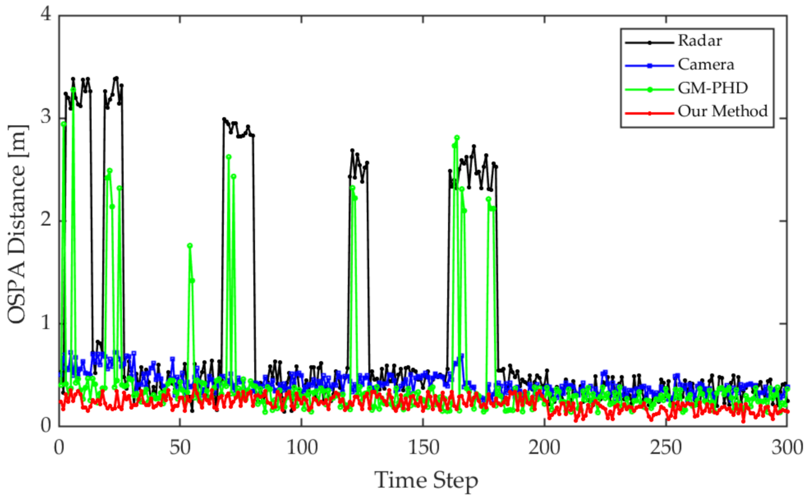


Figure 9. Optimal Sub-Pattern Assignment (OSPA) distance comparison for each tracking step.

From the detection and tracking results of the camera, the proposed algorithm can reduce the localization errors caused by missed or false camera detections. That is because the elliptic threshold and attenuation function have the effects of filtering large errors and life-cycle maintenance in the tracking algorithm. However, this approach may have negative effects on radar tracking. Due to the lack of knowledge of target class information, the radar may incorrectly extract interference measurements as targets. By using the fusion data, the GM-PHD algorithm can use radar measurements for state updating when camera measurements were lost. However, the GM-PHD algorithm also does not consider the classification information, and other interference or similar measurements may be extracted as spooky targets in the pruning and merging step. The improved algorithm can maintain stable tracking based on the prior classification labels of the measurements and the optimal detection matching. Meanwhile, using the optimal measurement properties of the radar and camera for localization, the algorithm can significantly improve the localization accuracy. In the whole process, the average OSPA distance of the proposed algorithm is about 0.14 m, which is close to the localization accuracy of the lidar.

4.2.2. Tracking Experiment for Cross Trajectory

The second test scenario was a trajectory-crossing tracking experiment for pedestrians, as shown in Figure 10. One of the pedestrians walked along the yellow line in a straight line, while the other pedestrians repeated the action of approaching and moving away, and the trajectories of the two pedestrians crossed several times. The crossing motion trajectories of pedestrians will cause occlusion, which is a tough problem in current target detection. All measurements were transformed to the ground coordinate system, and the algorithm performance comparison parameters remained the same as in the first experiment. A 77 GHz conti-ASR-408-21 radar, produced by Continental from Hanover, Germany, was used in this experiment. The detection and tracking results of targets and performance comparison are shown in the following.

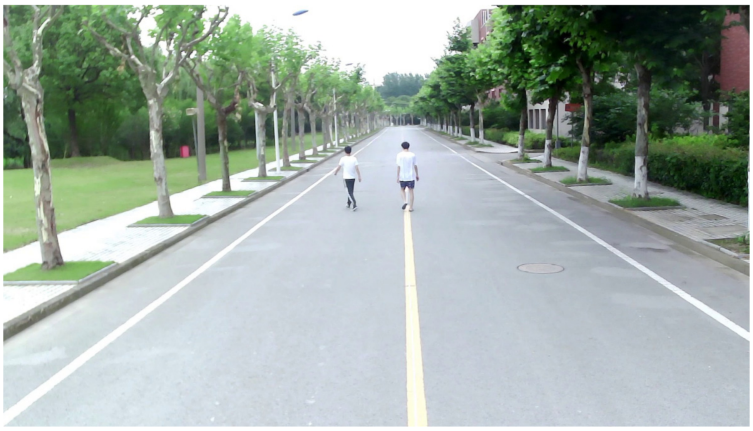


Figure 10. Pedestrian crosswalk tracking experiment.

As can be seen in Figure 11, with an increase in distance, the detection accuracy of the image decreases significantly. On one hand, the swinging arm motion of the pedestrian caused a large variation in the scale of bounding box, resulting in localization errors increasing; on the other hand, the camera could not detect the occluded pedestrian, resulting in measurement loss. Meanwhile, the lateral distance resolution of this radar was about 0.2 m, and the radar could not accurately distinguish targets when pedestrians were moving closer due to the low azimuth resolution, resulting in a loss of measurement. Within 30 m, the radar measurements suffered serious loss, while the image measurements were more stable. However, the localization errors of image increased, while the radar performed better after 30 m. It is worth noting that radar and image measurements can complement each other better. Therefore, the proposed algorithm can continuously track the target trajectories based on the radar and camera fusion data, as shown in Figure 12.

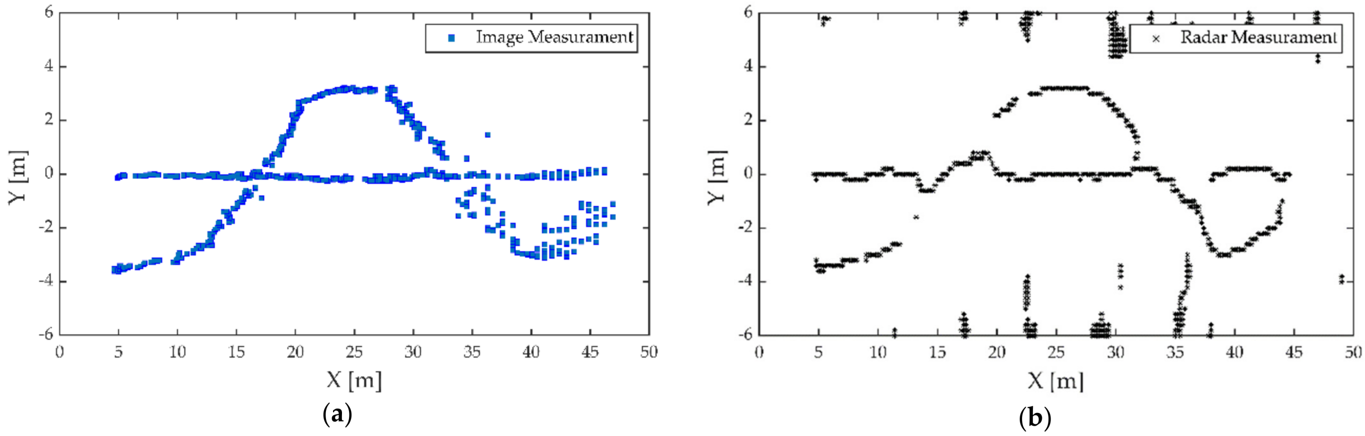


Figure 11. Detection results of targets: (**a**) camera; (**b**) radar.

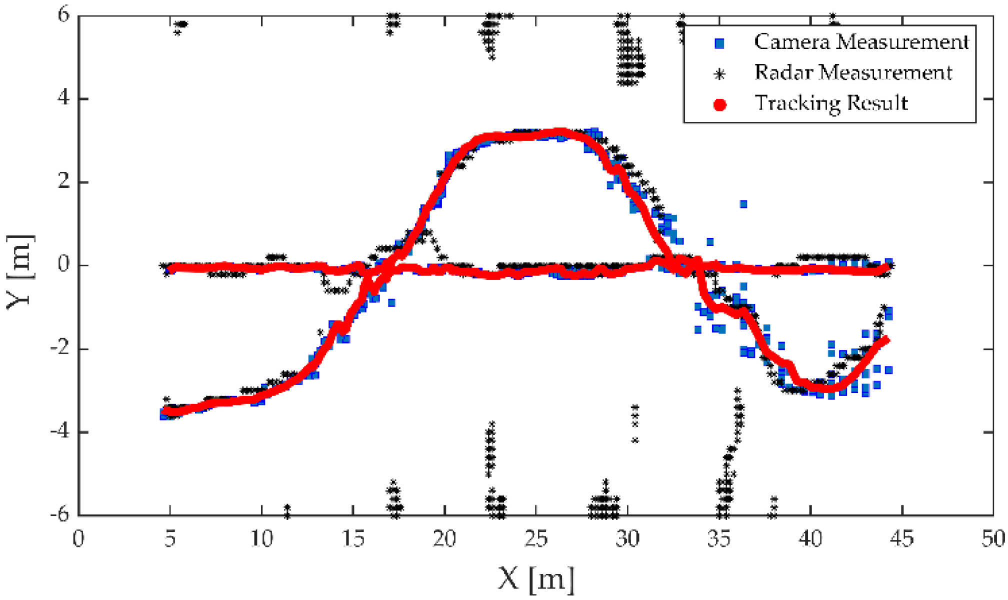


Figure 12. Tracking results for the proposed algorithm based on the radar and camera fusion data.

From the tracking results for the camera, it can be seen that the proposed algorithm could ensure the tracking stability when the measurement was lost for a short time. As seen in the tracking results for the radar, the target measurements had been lost for too long, beyond the maximum life cycle that the algorithm could have maintained. Due to the lack of classification information in the radar measurements, the trajectory maintenance for spooky targets conversely increased the tracking error.

Theoretically, when a sensor measurement is lost, the data fusion-based approach can use another sensor's measurement for target tracking. However, the performance of the GM-PHD algorithm is heavily dependent on the observation quality. As shown is Figure 13, when the radar or camera measurements are lost or have large errors, the tracking error of the GM-PHD algorithm increases in the same way. Without the guidance of the target a priori category information, the GM-PHD algorithm could also extract ghost targets while only relying on the position information. The improved algorithm can use elliptic thresholds to exclude large errors or missing measurements that occur randomly and adjust the target update weights and survival periods adaptively by the missing tags and attenuation function. By smoothing the estimation error, the improved algorithm can maintain localization accuracy and tracking trajectory continuity. The prior classification information of the camera leads the algorithm to focus on valid targets to reduce the false alarm phenomenon. The experiment results finally demonstrated that the above approaches strengthened the robustness of the perception algorithm.

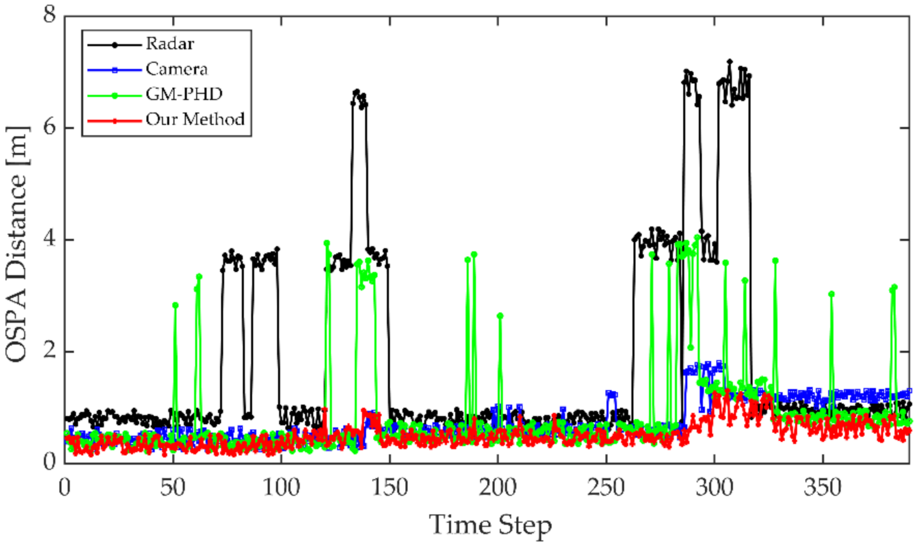


Figure 13. OSPA distance comparison for each tracking step.

4.2.3. Tracking Experiment for Vehicles

The third experiment was a vehicle-tracking experiment in the evening, as shown in Figure 14. The tested vehicle completed a lane change in a two-way lane. Due to the higher speed and stronger maneuverability of the vehicle, a farther monitoring distance was needed to give more reaction time to the intelligent networked cars or drivers. The tested vehicle was equipped with an inertial measurement unit (IMU) and a global positioning system (GPS) receiver to achieve centimeter-level localization through the real-time kinematic (RTK) technology. The GPS had a frequency of 10 Hz and a positioning accuracy of about 10 cm, the output of which was used as the benchmark for the tracking algorithm. The detection and tracking result of the vehicle is shown in Figure 15. A comparison of the tracking trajectory and GPS trajectory in a high-definition (HD) map is shown in Figure 16.

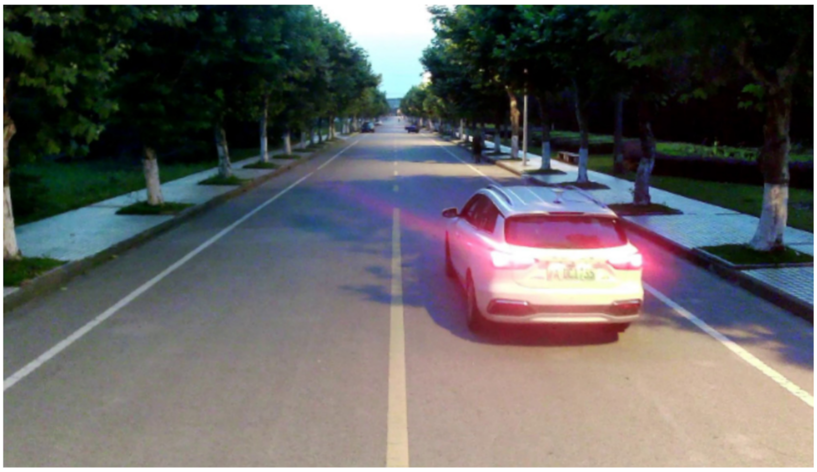


Figure 14. Vehicle tracking experiment in the evening.

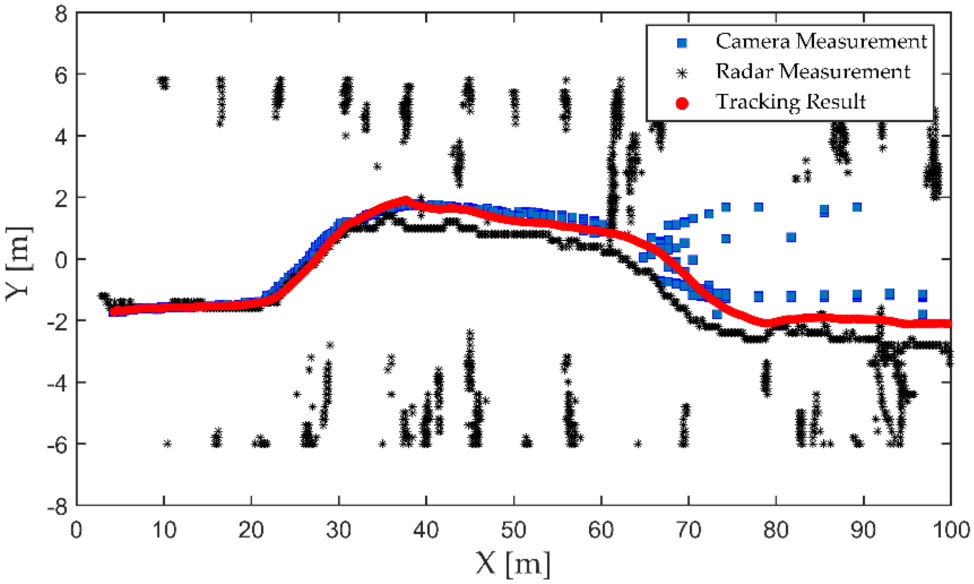


Figure 15. Vehicle detection and tracking results.

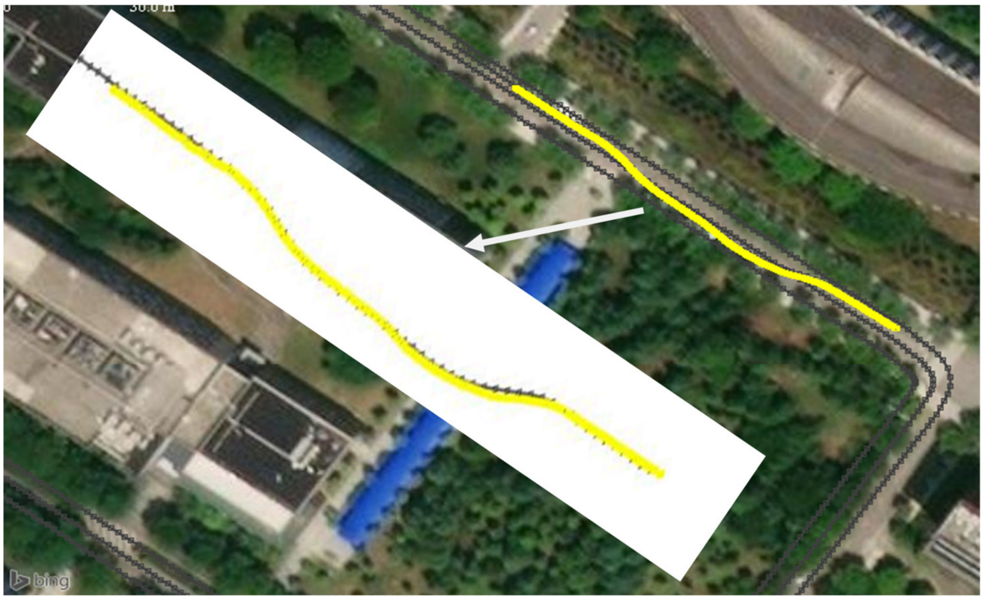


Figure 16. The tracking trajectory and global positioning system (GPS) trajectory in an high-definition (HD) map. The yellow line indicates the tracking result using the proposed algorithm, and the gray line is the measurements of the GPS system.

From the detection results of the camera and radar in Figure 15, it can be seen that the localization accuracy of both the camera and radar decreased significantly as the distance increased. The lack of longitudinal distance resolution of the camera is the main reason for its decreasing localization accuracy. Moreover, the radar's low angular resolution caused a decrease in lateral localization accuracy. By using the optimal measurement attributes, our proposed algorithm could accomplish the target state update to improve the localization accuracy. Figure 16 shows that the tracking trajectory of the vehicle and GPS measurement trajectory almost coincided. The localization error was relatively small when the vehicle was driving in a straight line. The larger localization errors mainly occurred when the vehicle changed lanes, and the maximum lateral localization error was about 0.64 m. During the lane-changing process, the attitude of the vehicle relative to the sensor also changed continuously. The bounding-box size of the vehicle in the image detection algorithm and the reflected surface area of the echo signal in the radar detection underwent unpredictable nonlinear changes, which resulted in a bias in the lateral positioning of the vehicle. The proposed algorithm could achieve lane-level detection and localization of vehicles.

To sum up, three experiments demonstrated that the proposed algorithm had a significant improvement in localization accuracy and tracking stability. The proposed algorithm could realize the accurate localization of pedestrians within 50 m, and the lane-level localization of vehicles of at least 100 m.

4.3. System Validation

The system validation is a demonstration of a vehicle-to-infrastructure (V2I) application; the system framework is shown in Figure 17. The intelligent RSU used the camera and radar to complete the monitoring of the surrounding environment. The classification and location information of targets were packaged and broadcasted by the RSU to the OBU of the surrounding ICVs. Then the ICV generated a real-time road traffic situation map based on the received information and its HD map, which helped to complete the planning and decision-making in advance. The test scenario was a complex circular intersection on campus, as shown in Figure 18. Area A contained pedestrians and buses. The ICV was driving on the right turn road of area B, with tall trees on both sides. The data acquisition synchronization rate of the camera and the radar was 15 Hz, and the data transmission rate of RSU and the data reception rate of OUB were 10 Hz. For clear observation, we only visualized the planned path and target information, as shown in Figure 19.

The domain controller decoded the target information received by the OBU and loaded it into the planner. The high-precision location information and classification information of the target were displayed on the driving map display in real time. The planner of the ICV could plan the vehicle movement in real time based on the current traffic situation to ensure safe driving. In Figure 19, the ICV slowed down in advance to prevent a collision with pedestrians. Over-the-horizon perception allowed the ICV or drivers to anticipate the traffic situation in advance, and gave more time for decision-making. Therefore, this system can be used at intersections or in accident-prone areas, which is of great significance to reduce traffic collisions and improve traffic safety.

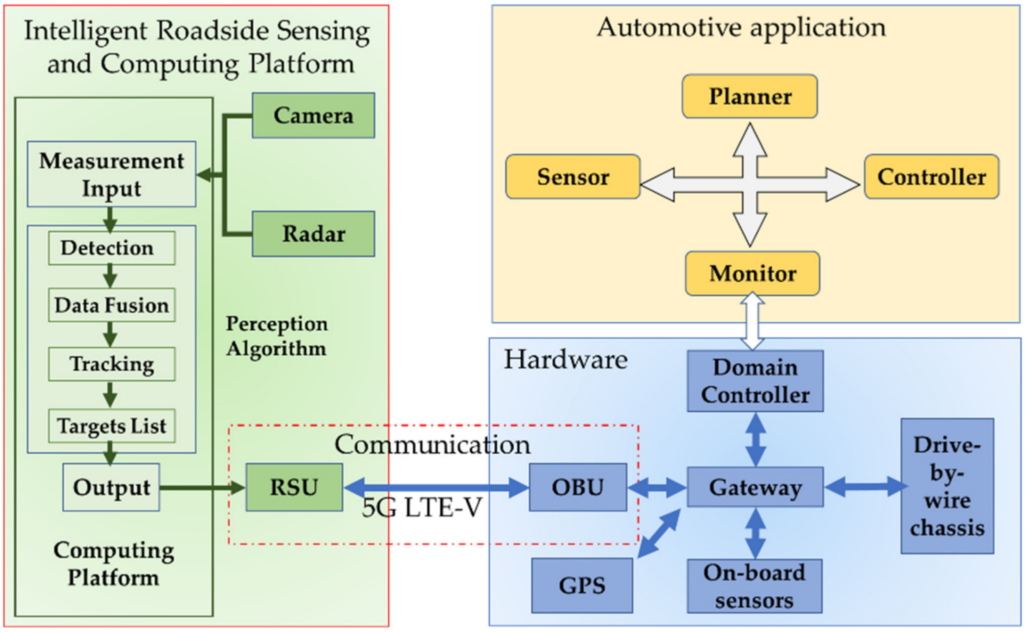


Figure 17. Main components of the system framework.

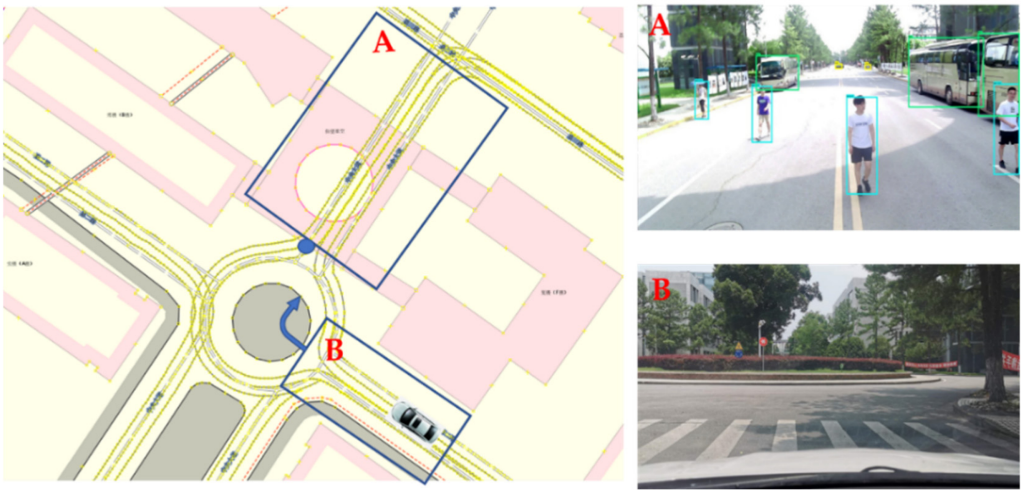


Figure 18. Test scenario for the vehicle-to-infrastructure (V2I) system. The left side is a HD map in vector format from OpenStreetMap. The blue marker is the placement of the roadside unit (RSU).

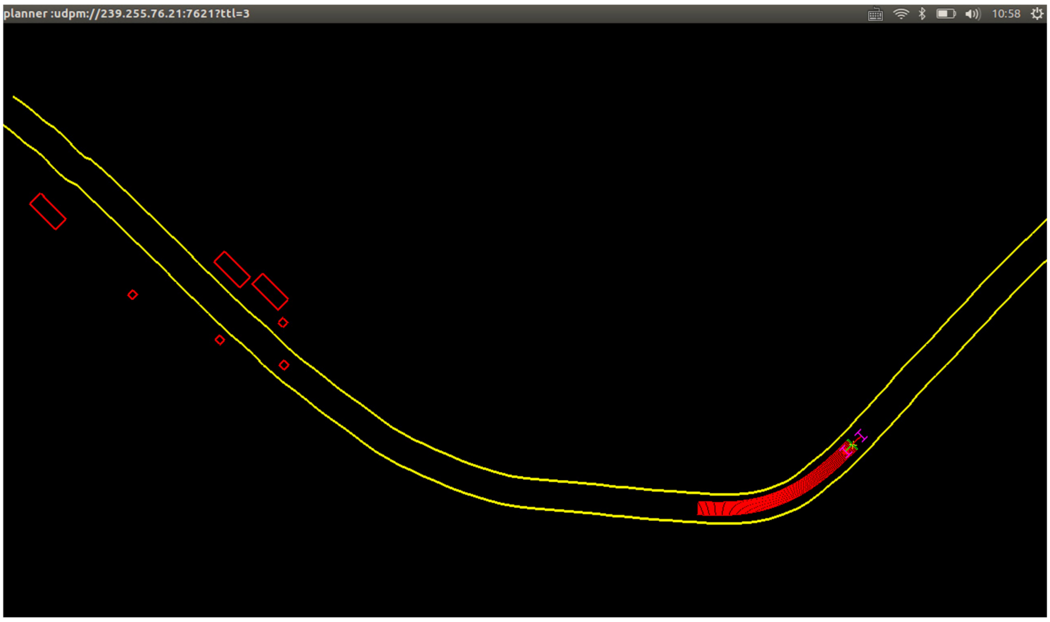


Figure 19. Visualization of the planner module. The yellow line is the driving route planned by the planner module. The red path indicates that the vehicle is slowing down or braking. Rectangles indicate vehicles. Smaller squares indicate pedestrians.

5. Conclusions

This study focused on roadside perception for ITSs. We proposed a multi-target detection and tracking algorithm based on the optimal property fusion of an MMW radar and camera. The framework could achieve classification, high accuracy localization, and trajectory tracking of targets in the observation field of view. The experiment results demonstrated that the proposed algorithm could improve the localization accuracy of targets and maintain the continuity of the trajectories. Meanwhile, this scheme realizes a high perception of confidence and stability with low-cost sensors, which is valuable for large-scale commercial applications to achieve traffic efficiency and safety. However, the whole perception process was actually a two-stage framework consisting of detection and tracking. All targets were regarded as points, ignoring the volume of targets, so that the algorithm could not track the pose-changing of targets with large volumes, such as trucks and buses. In our future research, we plan to design a one-stage end-to-end convolutional network to achieve high accuracy localization from the raw data of sensors. We will also use the 3-D bounding box to track the pose-changing and motion direction of targets.

Author Contributions: Conceptualization, J.B., S.L., and L.H.; methodology, S.L. and L.H.; software, S.L. and H.Z.; formal analysis, L.H.; investigation, S.L. and H.Z.; resources, J.B. and P.W.; data curation, J.B., S.L., and P.W.; writing—original draft preparation, S.L.; writing—review and editing, S.L. and L.H.; visualization, H.Z.; supervision, J.B. and L.H.; project administration, J.B. and L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China under Grant (2018YFB0105002) and Cross-disciplinary pre-research project of Tongji University (22120190206).

Acknowledgments: The authors acknowledge the editors and reviewers for the valuable comments and suggestions.

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

References

1. WHO. Global Status Report on Road Safety. 2018. Available online: http://www.who.int/violence_injury_prevention/road_traffic/en/ (accessed on 18 December 2020).
2. Abdelkader, G.; Elgazzar, K. Connected Vehicles: Towards Accident-Free Intersections. In Proceedings of the 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2–16 June 2020; pp. 1–2.
3. Chang, X.; Li, H.; Rong, J.; Huang, Z.; Chen, X.; Zhang, Y. Effects of on-Board Unit on Driving Behavior in Connected Vehicle Traffic Flow. *J. Adv. Transp.* **2019**, *2019*, 8591623. [CrossRef]
4. Li, H.; Zhao, G.; Qin, L.; Aizeke, H.; Yang, Y. A Survey of Safety Warnings Under Connected Vehicle Environments. *IEEE Trans. Intell. Transp. Syst.* **2020**, in press. [CrossRef]
5. Lio, M.D.; Biral, F.; Bertolazzi, E.; Galvani, M.; Tango, F. Artificial Co-Drivers as a Universal Enabling Technology for Future Intelligent Vehicles and Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 244–263. [CrossRef]
6. Guerrero-Ibáñez, J.; Zeadally, S.; Contreras-Castillo, J. Sensor Technologies for Intelligent Transportation Systems. *Sensors* **2018**, *18*, 1212. [CrossRef] [PubMed]
7. Ning, L.; Nan, C.; Ning, Z.; Xuemin, S.; Mark, J.W. Connected Vehicles: Solutions and Challenges. *IEEE Internet Things J.* **2014**, *1*, 289–299.
8. Chen, S.; Hu, J.; Shi, Y.; Peng, Y.; Zhao, L. Vehicle-to-Everything (v2x) Services Supported by LTE-Based Systems and 5G. *IEEE Commun. Mag.* **2017**, *1*, 70–76. [CrossRef]
9. Chen, S.; Hu, J.; Shi, Y.; Zhao, L.; Li, W. A Vision of C-V2X: Technologies, Field Testing and Challenges with Chinese Development. *IEEE Internet Things J.* **2020**, *7*, 3872–3881. [CrossRef]
10. Chen, C.; Liu, B.; Wan, S.; Qiao, P.; Pei, Q. An Edge Traffic Flow Detection Scheme Based on Deep Learning in an Intelligent Transportation System. *IEEE Trans. Intell. Transp. Syst.* **2020**, in press. [CrossRef]
11. Souza, A.M.D.; Oliveira, H.F.; Zhao, Z.; Braun, T.; Loureiro, A.A.F. Enhancing Sensing and Decision-Making of Automated Driving Systems with Multi-Access Edge Computing and Machine Learning. *IEEE Intell. Transp. Syst. Mag.* **2020**, in press. [CrossRef]
12. Wang, X.; Ning, Z.; Hu, X.; Ngai, E.C.-H.; 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]
13. Lefevre, S.; Petit, J.; Bajcsy, R.; Laugier, C.; Kargl, F. Impact of V2X privacy strategies on Intersection Collision Avoidance systems. In Proceedings of the 2013 IEEE Vehicular Networking Conference, Boston, MA, USA, 16–18 December 2013; pp. 71–78.
14. Zhao, H.; Cui, J.; Zha, H.; Katabira, K.; Shao, X.; Shibasaki, R. Monitoring an intersection using a network of laser scanners. In Proceedings of the 2008 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China, 12–15 October 2008; pp. 428–433.
15. Wang, C.C.; Lo, T.C.; Yang, S.W. Interacting Object Tracking in Crowded Urban Areas. In Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, 10–14 April 2007; pp. 4626–4632.
16. Meissner, D.; Dietmayer, K. Simulation and calibration of infrastructure based laser scanner networks at intersections. In Proceedings of the 2010 IEEE Intelligent Vehicles Symposium, San Diego, CA, USA, 21–24 June 2010; pp. 670–675.
17. Oliver, N.; Rosario, B.; Pentland, A. A Bayesian Computer Vision System for Modeling Human Interaction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 255–272. [CrossRef]
18. Babaei, P. Vehicles tracking and classification using traffic zones in a hybrid scheme for intersection traffic management by smart cameras. In Proceedings of the 2010 International Conference on Signal and Image Processing, Chennai, India, 15–17 December 2010; pp. 49–53.
19. Felguera-Martin, D.; Gonzalez-Partida, J.T.; Almorox-Gonzalez, P.; Burgos-Garca, M. Vehicular Traffic Surveillance and Road Lane Detection Using Radar Interferometry. *IEEE Trans. Veh. Technol.* **2012**, *61*, 959–970. [CrossRef]

20. Munoz-Ferreras, J.M.; Perez-Martinez, F.; Calvo-Gallego, J.; Asensio-Lopez, A.; Dorta-Naranjo, B.P.; Blanco-Del-Campo, A. Traffic Surveillance System Based on a High-Resolution Radar. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1624–1633. [CrossRef]
21. Zhao, H.; Cui, J.; Zha, H.; Katabira, K.; Shao, X.; Shibasaki, R. Sensing an intersection using a network of laser scanners and video cameras. *IEEE Intell. Transp. Syst. Mag.* **2009**, *1*, 31–37. [CrossRef]
22. Roy, A.; Gale, N.; Hong, L. Automated traffic surveillance using fusion of Doppler radar and video information. *Math. Comput. Model.* **2011**, *54*, 531–543. [CrossRef]
23. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. Available online: https://arxiv.org/abs/2004.10934 (accessed on 18 December 2020).
24. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceeding of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
25. Hua, S.; Anastasiu, D.C. Effective Vehicle Tracking Algorithm for Smart Traffic Networks. In Proceedings of the 2019 IEEE International Conference on Service-Oriented System Engineering, San Francisco East Bay, CA, USA, 4–9 April 2019.
26. Xun, L.; Lei, H.; Li, L.; Liang, H. A method of vehicle trajectory tracking and prediction based on traffic video. In Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications, Chengdu, China, 14–17 October 2016; pp. 449–453.
27. Zhou, Z.; Peng, Y.; Cai, Y. Vision-based approach for predicting the probability of vehicle–pedestrian collisions at intersections. *IET Intell. Transp. Syst.* **2020**, *14*, 1447–1455. [CrossRef]
28. Zhang, J.; Xiao, W.; Coifman, B.; Mills, J.P. Vehicle Tracking and Speed Estimation from Roadside Lidar. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5597–5608. [CrossRef]
29. Zhang, Z.; Zheng, J.; Xu, H.; Wang, X.; Chen, R. Automatic Background Construction and Object Detection Based on Roadside LiDAR. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 4086–4097. [CrossRef]
30. Wu, J.; Xu, H.; Zhang, Y.; Tian, Y.; Song, X. Real-Time Queue Length Detection with Roadside LiDAR Data. *Sensors* **2020**, *20*, 2342. [CrossRef]
31. Lv, B.; Sun, R.; Zhang, H.; Xu, H.; Yue, R. Automatic Vehicle-Pedestrian Conflict Identification with Trajectories of Road Users Extracted from Roadside LiDAR Sensors Using a Rule-based Method. *IEEE Access* **2019**, *7*, 161594–161606. [CrossRef]
32. Liu, W.J.; Kasahara, T.; Yasugi, M.; Nakagawa, Y. Pedestrian recognition using 79GHz radars for intersection surveillance. In Proceedings of the 2016 European Radar Conference, London, UK, 5–7 October 2016; pp. 233–236.
33. Liu, W.; Muramatsu, S.; Okubo, Y. Cooperation of V2I/P2I Communication and Roadside Radar Perception for the Safety of Vulnerable Road Users. In Proceedings of the 2018 16th International Conference on Intelligent Transportation Systems Telecommunications (ITST), Lisboa, Portugal, 15–17 October 2018; pp. 1–7.
34. Arguello, A.G.; Berges, D. Radar Classification for Traffic Intersection Surveillance based on Micro-Doppler Signatures. In Proceedings of the 2018 15th European Radar Conference (EuRAD), Madrid, Spain, 26–28 Sept. 2018; pp. 186–189.
35. Kim, Y.-D.; Son, G.-J.; Song, C.-H.; Kim, H.-K. On the Deployment and Noise Filtering of Vehicular Radar Application for Detection Enhancement in Roads and Tunnels. *Sensors* **2018**, *18*, 837.
36. Fu, Y.; Tian, D.; Duan, X.; Zhou, J.; Lang, P.; Lin, C.; You, X. A Camera–Radar Fusion Method Based on Edge Computing. In Proceedings of the 2020 IEEE International Conference on Edge Computing (EDGE), Beijing, China, 19–23 October 2020; pp. 9–14.
37. Wang, J.G.; Chen, S.J.; Zhou, L.B.; Wan, K.W.; Yau, W.Y. Vehicle Detection and Width Estimation in Rain by Fusing Radar and Vision. In Proceedings of the 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, 18–21 November 2018; pp. 1063–1068.
38. Schller, C.; Schnettler, M.; Krmmer, A.; Hinz, G.; Knoll, A. Targetless Rotational Auto-Calibration of Radar and Camera for Intelligent Transportation Systems. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 3934–3941.
39. Kaul, P.; Martini, D.D.; Gadd, M.; Newman, P. RSS-Net: Weakly-Supervised Multi-Class Semantic Segmentation with FMCW Radar. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 431–436.
40. Mohammed, A.S.; Amamou, A.; Ayevide, F.K.; Kelouwani, S.; Agbossou, K.; Zioui, N. The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review. *Sensors* **2020**, *20*, 6532. [CrossRef]
41. Hasirlioglu, S.; Riener, A. Introduction to rain and fog attenuation on automotive surround sensors. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–7.
42. Kutila, M.; Pyykonen, P.; Ritter, W.; Sawade, O.; Schaufele, B. Automotive LIDAR sensor development scenarios for harsh weather conditions. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 265–270.
43. Buller, W.; Xique, I.J.; Fard, Z.B.; Dennis, E.; Hart, B. Evaluating Complementary Strengths and Weaknesses of ADAS Sensors. In Proceedings of the IEEE 88th Vehicular Technology Conference (VTC-Fall), Chicago, IL, USA, USA, 27–30 August 2018; pp. 1–5.
44. Vo, B.N.; Ma, W.K. The Gaussian Mixture Probability Hypothesis Density Filter. *IEEE Trans. Signal Process.* **2006**, *54*, 4091–4104. [CrossRef]
45. Vo, B.T.; Vo, B.N.; Cantoni, A. Bayesian Filtering with Random Finite Set Observations. *IEEE Trans. Signal Process.* **2008**, *56*, 1313–1326. [CrossRef]

46. Patole, S.M.; Torlak, M.; Wang, D.; Ali, M. Automotive Radars: A review of signal processing techniques. *IEEE Signal Process. Mag.* **2017**, *34*, 22–35. [CrossRef]
47. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef]
48. Narote, S.P.; Bhujbal, P.N.; Narote, A.S.; Dhane, D.M.J.P.R. A review of recent advances in lane detection and departure warning system. *Pattern Recognit.* **2018**, *73*, 216–234. [CrossRef]
49. Yin, S.; Ouyang, P.; Liu, L.; Guo, Y.; Wei, S. Fast Traffic Sign Recognition with a Rotation Invariant Binary Pattern Based Feature. *Sensors* **2015**, *15*, 2161–2180. [CrossRef]
50. Wen, L.; Du, D.; Cai, Z.; Lei, Z.; Chang, M.C.; Qi, H.; Lim, J.; Yang, M.H.; Lyu, S. UA-DETRAC: A New Benchmark and Protocol for Multi-Object Detection and Tracking. *Comput. Vis. Image Underst.* **2020**, *193*, 102907. [CrossRef]
51. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]
52. Domhof, J.; Kooij, J.F.P.; Kooij, D.M. An Extrinsic Calibration Tool for Radar, Camera and Lidar. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 8107–8113.
53. Rother, C. A new approach to vanishing point detection in architectural environments. *Image Vis. Comput.* **2002**, *20*, 647–655. [CrossRef]
54. Mahler, R.P.S.; Vo, B.T.; Vo, B.N. CPHD Filtering with Unknown Clutter Rate and Detection Profile. *IEEE Trans. Signal Process.* **2011**, *59*, 3497–3513. [CrossRef]
55. Si, W.; Wang, L.; Qu, Z. Multi-Target Tracking Using an Improved Gaussian Mixture CPHD Filter. *Sensors* **2016**, *16*, 1964. [CrossRef] [PubMed]
56. Zhang, H.; Jing, Z.; Hu, S. Gaussian mixture CPHD filter with gating technique. *Signal Process.* **2009**, *89*, 1521–1530. [CrossRef]
57. Ristic, B.; Vo, B.N.; Clark, D.; Vo, B.T. A Metric for Performance Evaluation of Multi-Target Tracking Algorithms. *IEEE Trans. Signal Process.* **2011**, *59*, 3452–3457. [CrossRef]