Received 19 May 2022
Revised 24 June 2022
Accepted 4 July 2022

Research on multi-target tracking method based on multi-sensor fusion

Bolin Gao
Tsinghua University, Beijing, China

Kaiyuan Zheng
Hebei University of Technology, Tianjin, China and Tsinghua University, Beijing, China

Fan Zhang, Ruiqi Su and Junying Zhang
Xiangyang Daan Automobile Testing Center Co., Ltd, Wuhan, China, and

Yimin Wu
Hebei University of Technology, Tianjin, China

Abstract

Purpose – Intelligent and connected vehicle technology is in the ascendant. High-level autonomous driving places more stringent requirements on the accuracy and reliability of environmental perception. Existing research works on multitarget tracking based on multisensor fusion mostly focuses on the vehicle perspective, but limited by the principal defects of the vehicle sensor platform, it is difficult to comprehensively and accurately describe the surrounding environment information.

Design/methodology/approach – In this paper, a multitarget tracking method based on roadside multisensor fusion is proposed, including a multisensor fusion method based on measurement noise adaptive Kalman filtering, a global nearest neighbor data association method based on adaptive tracking gate, and a Track life cycle management method based on M/N logic rules.

Findings – Compared with fixed-size tracking gates, the adaptive tracking gates proposed in this paper can comprehensively improve the data association performance in the multitarget tracking process. Compared with single sensor measurement, the proposed method improves the position estimation accuracy by 13.5% and the velocity estimation accuracy by 22.2%. Compared with the control method, the proposed method improves the position estimation accuracy by 23.8% and the velocity estimation accuracy by 8.9%.

Originality/value – A multisensor fusion method with adaptive Kalman filtering of measurement noise is proposed to realize the adaptive adjustment of measurement noise. A global nearest neighbor data association method based on adaptive tracking gate is proposed to realize the adaptive adjustment of the tracking gate.

Keywords Roadside perception, Multisensor fusion, Multitarget tracking, Data association, Kalman filter

Paper type Research paper

Smart and Resilient Transportation
Vol. 4 No. 2, 2022
pp. 46-65
Emerald Publishing Limited
e-ISSN: 2632-0495
p-ISSN: 2632-0487
DOI 10.1108/SRT-05-2022-0010

1. Introduction

In recent years, with the new round of scientific and technological revolutions such as artificial intelligence, big data and cloud computing, the field of autonomous driving has attracted widespread attention around the world. Among them, accurate and comprehensive environmental perception technology is the premise of decisions and control of autonomous vehicles (Li *et al.*, 2017). Single-vehicle autonomous driving relying on in-vehicle autonomous intelligent systems is the main research interests at present, but there are still many inherent problems, such as limited in-vehicle perception field of view, poor measurement stability, relatively high cost, and it is difficult to support high-level autonomous driving. The roadside sensor is usually located at a higher position from the ground, and the perception field is wide, and the spatial position and relative motion relationship of the traffic participants within the sensing range is clearer. In addition, the installation platform of the roadside sensor is relatively fixed, and it is not easy to shake violently, and its perception accuracy and perception stability are higher (Zhang *et al.*, 2019). Due to different working mechanisms, single sensor such as camera, radar and lidar cannot be fully qualified for the environmental perception task of autonomous driving. To make full use of the characteristics of different sensors and achieve complementary advantages between sensor characteristics, multisensor information fusion is required (Zhao and Wang, 2013). In the roadside perception system, it is difficult to meet the needs of high-level autonomous driving only by detecting the targets within the perception range. It is necessary to track the targets within the range, continuously obtain the motion state information of the target and transmit the continuous multitarget motion trajectory to the upper-level decision-making and control module. Therefore, multitarget tracking based on roadside multisensor fusion can accurately and stably provide continuous state information of dynamic and static traffic participants within the perception range, which is of great significance for the realization of high-level autonomous driving.

The traditional Bayesian estimation theory lay an important foundation for the development of multisensor information fusion technology, which mainly includes the Kalman filter (KF; Xiu and Guo, 2013) evolved from the Bayesian Filter and its related variants for nonlinear systems: extended KF (Deng *et al.*, 2013), unscented KF (UKF; Wan and Merwe, 2000) and Partical Filter (Yibing *et al.*, 2021), the most complete inheritance of Bayesian theory. In addition, some logical reasoning theories and artificial intelligence methods are also widely used in the field of multisensor information fusion, such as D-S evidence theory (Zhang *et al.*, 2019), Random Finite Set (Wu *et al.*, 2016), deep learning, genetic algorithm. Caltagirone *et al.* (2018) used deep learning to propose a road detection algorithm based on the fusion of lidar point cloud and camera image. First, the unstructured sparse point cloud was projected to the camera image plane, and then it was sampled to obtain the encoding space A dense 2D image of information enables road segmentation. Lekic and Babic (2019) proposed a multisensor information fusion algorithm based on the original point cloud of radar and camera image. The proposed method is a completely unsupervised machine learning algorithm, which converts the original point cloud of millimeter-wave radar into a camera-like environment image and fuses it with the camera image. Experiments show that the information generated by the proposed algorithm is more accurate and effective than a single sensor. Jo *et al.* (2012) proposed a fusion localization algorithm based on the interacting multiple model (IMM) filter using low-cost GPS and on-board sensors, which effectively improved the localization accuracy.

Multiple object tracking (MOT; Bar-Shalom, 2001) is one of the essential key technologies in autonomous driving environment perception, and it is an intersecting theory involving multiple disciplines and fields. It usually implements functions based on cameras, radars

and lidars. The multitarget tracking technology determines the target number and status of traffic participants within the sensing range by receiving the measurement information of the sensor and forms the corresponding target track to provide a basis for subsequent decision-making and planning. Mobus (Mobus and Kolbe, 2004) proposed a multitarget fusion tracking method based on infrared sensors and radar. This research uses the IMM method to switch the motion model of the target, and each sensor tracks multiple targets within the sensing range individually. In the data association link, the research uses PDA to fuse the tracking results of the two sensors. To solve the disadvantage that the number of targets needs to be manually modified in the PDA method, Otto *et al.* (2012) proposed a multisensor fusion tracking method based on joint integrated probability data association (JIPDA), which can automatically add or delete tracking targets to achieve multitarget tracking of pedestrians. Aiming at the problem that the sensor perception characteristics are difficult to describe in the process of multitarget tracking, Josip *et al.* (2016) proposed a multitarget tracking method based on JIPDA by using radar and stereo vision camera and established sensor uncertainty model in polar coordinate system. To achieve high-precision and low-cost multitarget tracking in clutter environment, Eltrass and Khalil (2018) proposed a multitarget tracking method based on vehicle radar. This study uses JPDA to association the measurement of multiple targets from a single sensor. After the success, the UKF and constant turn rate and acceleration models are used to estimate the target state, and a linear regression algorithm is introduced to improve the accuracy of the state estimation. In addition, the M/N test is used to manage the life cycle of the multitarget trajectory, and the number and status of the targets in the tracking process are clearly defined. In the experimental verification link, compared with three same type of multitarget tracking methods, this method has the highest tracking accuracy. However, this research only relies on radar for multitarget tracking, which is difficult to support high-level autonomous driving. Besides traditional methods, MOT can also be achieved by deep learning. Sadeghian *et al.* (2017) proposed a hierarchical recurrent neural network structure that integrates the motion, appearance and interaction features of each tracked target. The processor extracts multiframe appearance and context features.

This paper proposes a multitarget tracking method based on roadside multisensor fusion based on the application background of roadside fusion perception in the intelligent connected environment, and the target-level data of roadside cameras and radars as information sources. To carry out the research on the multisensor fusion tracking method from the roadside perspective, the error characteristics of the camera and the radar from the roadside perspective are first calculated to provide the basis for parameter setting for the subsequent research on the multitarget tracking method. Second, in view of the problem that the fixed-size tracking gate in the data association link is difficult to adapt to the dynamically changing sensor measurement error, a data association method based on the adaptive tracking gate is proposed to achieve the many-to-many data association between the measurement values of different sensors and the tracking target. Third, for the uncertainty of the motion state of the target vehicle and the dynamic perception characteristics of the sensor in the state estimation process, a multisensor fusion method based on measurement noise adaptive KF is proposed. Finally, the real vehicle test verifies the effectiveness of the proposed multitarget tracking method based on multisensor fusion.

2. Space–time synchronization and performance evaluation of roadside fusion perception system

To study the multitarget method based on roadside multisensor fusion, this section builds a roadside perception platform and analyzes the characteristics of the camera and radar of the roadside perception platform. First, the time synchronization of various heterogeneous

sensors is carried out. Second the local coordinate system of the sensor and the geodetic coordinate system are unified to realize the time–space synchronization of the roadside sensing system. Third, the perception characteristics of the sensors are statistically analyzed to provide a theoretical basis for the multisensor fusion tracking method.

2.1 Deployment of roadside sensors

The deployment of roadside multisensors is the basis for building a roadside fusion perception system. At present, traffic sensors are generally installed in urban road networks, such as various traffic radars and security cameras, which are mainly used for traffic flow density monitoring and illegal capture. At the same time, there are also various poles such as street light poles and signal poles, among which the poles that meet the requirements can be directly used to install roadside sensors. In addition to considering the existing infrastructure, roadside sensor deployment should also consider the geometric factors of the actual road to ensure full coverage of the sensing area as much as possible.

Based on this, three roadside sensors are deployed at an intersection, including lidar, camera and radar. The lidar uses Velodyne HDL-64E lidar. The camera uses a 2-megapixel professional camera with a resolution of 1920 × 1080 for industrial use. The radar uses a 77 GHz traffic detection radar. Among them, the 64-line lidar is used as the ground truth value of the roadside fusion perception system to analyze the sensor characteristics of roadside cameras and radars and to verify the method proposed in this paper. The deployment of each sensor is shown in Figure 1: the coverage of the camera and the radar is approximately an isosceles trapezoid, and the lidar is approximately a circular ring, covering the entire T-shaped intersection.

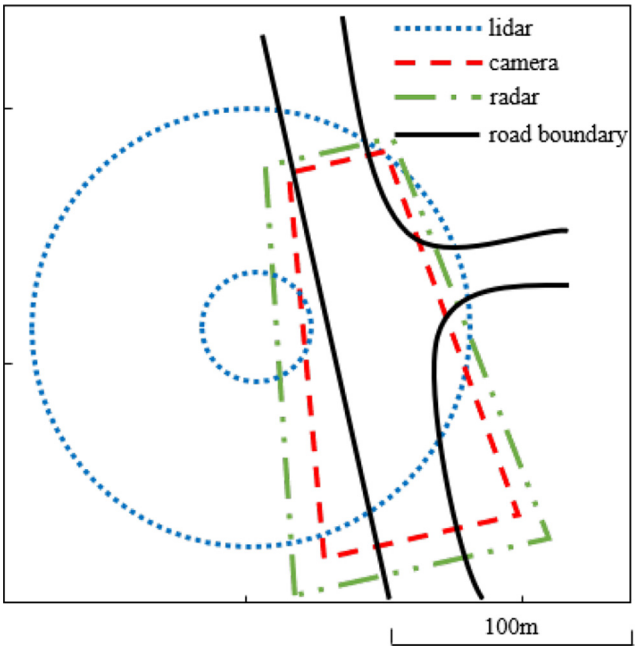

Figure 1.
Schematic of roadside sensor deployment

The specific installation positions and orientations of each sensor are shown in Table 1. Table 1 lists the installation positions of cameras, radars and lidars under the Universal Transverse Mercator grid system (UTM) coordinate system and the Y+ axis orientation of each sensor's local coordinate system.

2.2 Time synchronization

Due to different working modes, software and hardware levels and other factors, the sampling frequency and sampling start time of each sensor are different. Time synchronization needs to be performed before information fusion. The synchronization methods are mainly divided into two types: hard synchronization and soft synchronization. Hard synchronization refers to sending physical signals directly through hardware triggers to trigger sensors to collect information. Soft synchronization refers to providing the same time source to multiple sensors, and stamping a timestamp on the recorded data, and using a mathematical method to synchronize the time according to the timestamp.

In this paper, because the sensors are not all deployed on the same rod, it is difficult to unify the time through hard synchronization, so this paper adopts the soft synchronization method, that is, the timestamps of the three sensors are used to synchronize the time of the three sensors. To facilitate the follow-up sensor performance statistics and experimental verification, the lidar perception time is used as the reference time for soft synchronization, and the target state is predicted for the latest camera and radar perception data before the lidar sensing time, and the equivalent measurement at reference time is obtained. The frequency of the lidar and camera deployed on the roadside perception platform in this study is 10 Hz, and the frequency of the radar is about 14.5 Hz. During the movement of the maneuvering target, the magnitude and direction of its speed are changing all the time, but in a short time (for example, no more than one sampling period), it can be treated as a state of uniform linear motion, so this paper adopts the constant velocity (CV) model to perform state prediction on perception data from cameras and radars.

2.3 Spatial synchronization

In a fusion perception system, multiple sensors are generally deployed, and the measurement of each sensor are based on its own local coordinate system. Before information fusion, the data of each sensor needs to be spatially synchronized. Compared with vehicle perception, roadside perception focuses more on the absolute positioning of the target, so it is necessary to convert the data of each sensor into a unified geodetic coordinate system. In this paper, the sensors on the roadside perception platform have completed the corresponding detection and tracking in their respective local coordinate systems. Now the output target-level data needs to be transformed into the UTM coordinate system. Therefore, the method of linear transformation is used to achieve the above purpose.

2.4 Performance evaluation of roadside fusion perception system

Roadside perception focuses on absolute positioning. After spatial synchronization, the data of each sensor was converted to the UTM coordinate system. However, the sensor's

Table 1.
Sensor installation position and orientation

	Absolute horizontal position/m	Absolute portrait position/m	Height/m	Orientation/(°)
Camera	326288.857	3462327.119	12.4	153.15
Radar	326288.857	3462327.119	12.4	153.15
Lidar	326282.340	3462271.990	12.4	70.06

perception error characteristics are generally based on experimental statistics based on its own local coordinate system. To obtain its perception error characteristics in the geodetic coordinate system, target-level data of roadside camera, radar and lidar are collected within a certain period of time. Using the perception results of the lidar as the true value of the target state, the error statistics of the roadside camera and the radar are performed to obtain the characteristics of the perception error, which provides a basis for the setting of relevant parameters in the subsequent multitarget tracking.

2.4.1 Camera error characteristic statistics. To obtain the perception error characteristics of the roadside camera, in an environment with good lighting conditions and good weather, the roadside perception platform that has been built is used to collect real road data for a certain period of time for statistical analysis, and the data is shown in Table 2.

In this paper, the target-level perception results of lidar are used as the true state value of road traffic participants, and the target-level perception results of roadside cameras at different distances are compared with the target-level perception results of roadside lidar to obtain the mean absolute error, and then determine the ranging performance of roadside cameras. As shown in Figure 2, the trapezoidal perception area of the roadside camera is divided into several sub-areas according to its Y+ axis direction, and the target sensing results of the roadside camera in each sub-area are statistically analyzed.

The sensor accuracy is represented by the absolute deviation between the roadside camera perception results and the roadside lidar perception results. The perception data is divided into 7 groups at equal intervals of 10 meters, and the average absolute position accuracy of all target data in each group is counted. Figure 3 shows the results of the mean

Table 2. Collection of road data

Collection time/s	Target No.	Target type	Target behavior
148.479	68	car	straight/turn

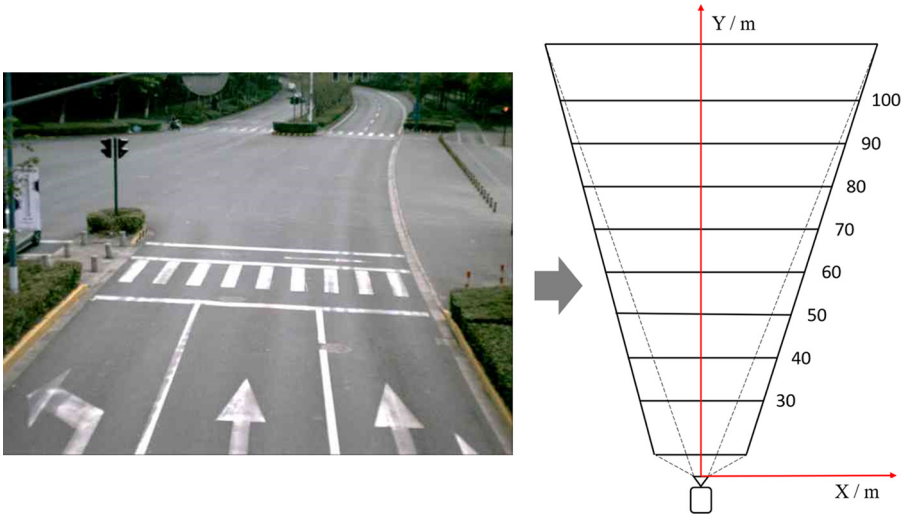

Figure 2. Perceptual region division of roadside camera

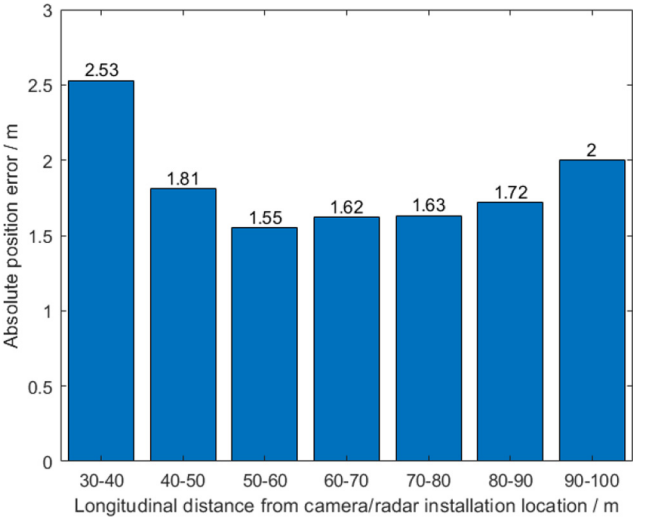

Figure 3.
Statistical results of absolute position error of roadside camera

absolute position error of the roadside camera at different longitudinal distance gradients from its installation location.

It can be seen from Figure 3 that the ranging accuracy of the roadside camera in the range of 50–80 meters is better than that of 30–50 meters and 80–100 meters. The farther away in an image, the lower the proportion of pixels occupied by the target, and the greater the spatial distance represented by each pixel. Therefore, the ranging accuracy of the roadside camera within the range of 50∼100 meters continues to decrease with the increase of the longitudinal distance between the target and the camera.

2.4.2 Radar error characteristic statistics. To obtain the error characteristics of the roadside radar, the real road data collected in Table 2 is statistically analyzed, and the target-level perception result of the roadside lidar is used as the true state value of the road traffic participants. The target-level perception results of the roadside radar are compared with the target-level perception results of the roadside lidar to obtain the mean absolute error, which is used to determine the ranging performance of the roadside radar. Referring to the statistical method of the error characteristics of the roadside camera in Section 2.4.1, the trapezoidal sensing area of the roadside radar is divided into several sub-areas according to the direction of its Y+ axis, as shown in Figure 4. Statistically analyze the target perception results of roadside radar in each sub-area.

The sensor accuracy is represented by the absolute deviation between the roadside radar perception results and the roadside lidar perception results. The perception data are divided into 7 groups at equal intervals of 10 meters, and the average absolute position accuracy of all target data in each group is counted. Figure 5 shows the average absolute position error results under different longitudinal distance gradients from the roadside radar to its installation position. It can be seen from Figure 5 that the average ranging accuracy of the roadside radar is less sensitive to distance changes than roadside camera.

In conclusion, the ranging accuracy of camera and radar in the roadside fusion perception system can provide a basis for the setting of related parameters in subsequent multitarget tracking.

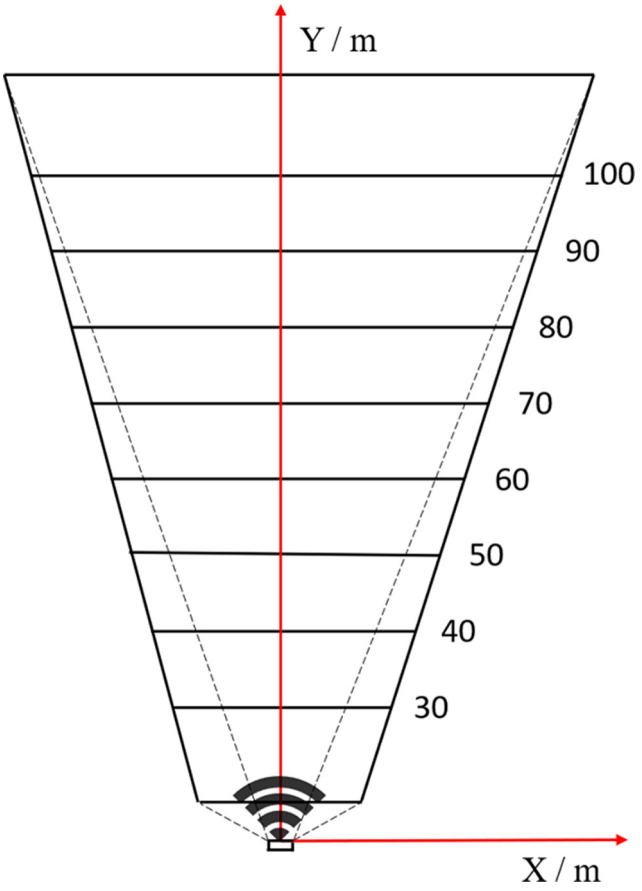

Figure 4.
Perceptual region division of roadside radar

3. Multitarget tracking algorithm based on multisensor fusion

3.1 Multitarget tracking framework based on multisensor fusion

The overall framework of the method is shown in Figure 6, which mainly includes four parts, namely, the data input module, data association module, target state estimation module and track life cycle management module. The input of the algorithm is the target-level data obtained by the roadside camera and the radar through their respective detection and tracking algorithms. The information is used including timestamp, target type, spatial location and velocity. The output is the trajectory sequence of the tracking target, which is used to provide the upper layer for decision and control.

3.2 Target state estimation algorithm based on adaptive Kalman filter

The aim of multitarget tracking is to optimally estimate the state of the moving target at each moment, and a model-based method is generally used to describe its system equation. In the real environment, the motion state of the vehicle changes all the time. However, after the time synchronization, the data frequency of the roadside camera and radar is unified to 10 Hz. Within 0.1 s, the moving target can be regarded as a state of uniform linear motion,

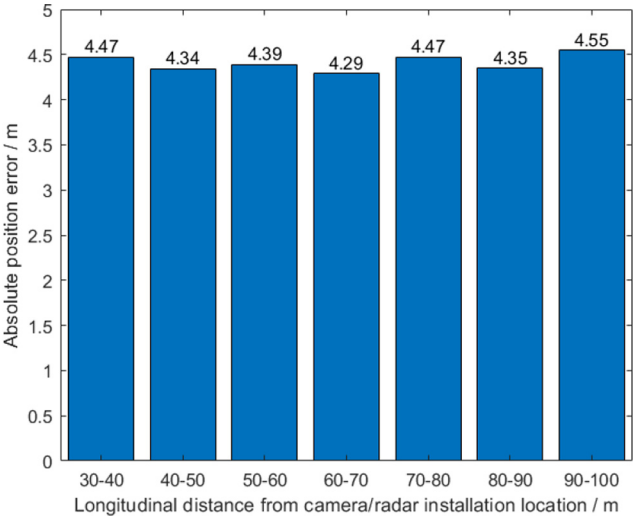

Figure 5.
Statistical results of absolute position error of roadside radar

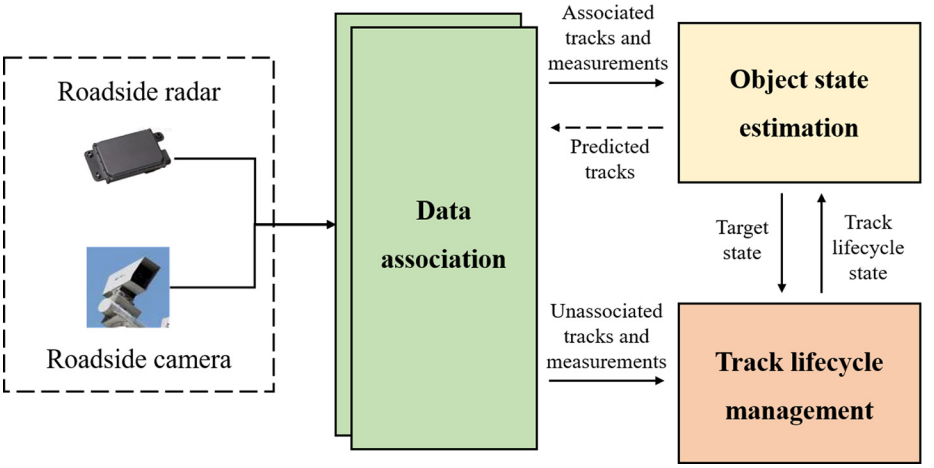

Figure 6.
Schematic of multitarget tracking framework

assuming that the magnitude and direction of its velocity remain unchanged. Let the state vector of the target be $x = [x, y, v_x, v_y]^T$, where x and y represent the absolute lateral distance and absolute longitudinal distance of the target vehicle in the UTM coordinate system, respectively, and v_x and v_y represent the absolute lateral speed and absolute longitudinal speed of the target vehicle in the UTM coordinate system, respectively.

Considering the uncertainty of the motion state of the target vehicle and the dynamic perception characteristics of the sensor, the measurement noise changes dynamically. In the traditional KF algorithm, the measurement noise is a fixed value, which easily makes the

state estimation result difficult to adapt to the dynamic characteristics of the sensor. To ensure the stability of the state estimation, the correction coefficient is proposed to adaptively adjust the measurement noise. The AKF fusion algorithm process is as follows:

- Using the CV model, the system state transition equation is established. If T is the time step, the transition formula of each state vector component is as follows:

$$x_{k+1} = x_k + v_{x_k} T$$

$$y_{k+1} = y_k + v_{y_k} T$$

$$v_{x_{k+1}} = v_{x_k}$$

$$v_{y_{k+1}} = v_{y_k}$$

According to the above state transition equation, the solution state transition matrix $\boldsymbol{F}_k$ is as follows:

$$\boldsymbol{F}_k = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Time update stage. The target state xk and the state covariance matrix Pk are predicted by the state transition matrix Fk calculated in the first step.
- Determine the observation matrix. The multisensor information is fused by the means of measurement fusion, as shown in Figure 7, measurement fusion is to combine the measurement of all sensors in the form of vectors, and then filter.

The measurement of the target vehicle is obtained by roadside camera and radar. The measurements from each sensor are combined as a vector and then filtered. Let the measurement vector be $\boldsymbol{z} = [\boldsymbol{z}_c, \boldsymbol{z}_r]^{\mathrm{T}}$, where $\boldsymbol{z}_c = [x_c, y_c, v_{cx}, v_{cy}]$ is the measurement vector of

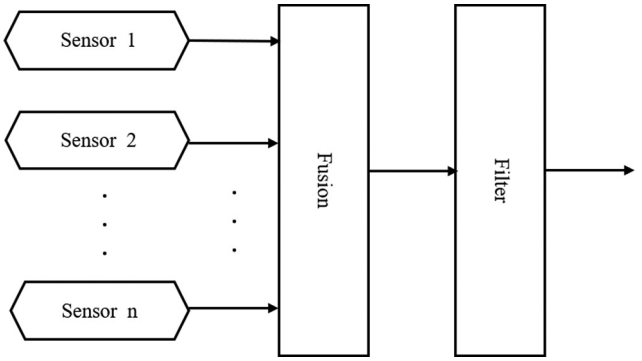

Figure 7.
Schematic of measurement fusion

the roadside camera, where x_c, y_c, respectively, represent the absolute lateral distance and absolute longitudinal distance of the target vehicle in the UTM coordinate system provided by the roadside camera, v_{cx}, v_{cy}, respectively, represent the absolute lateral speed and absolute longitudinal speed of the target vehicle in the UTM coordinate system provided by the roadside camera. Similarly, $\boldsymbol{z_r}$ is the measurement vector of the roadside radar, and the corner marks c and r represent the camera and the radar, respectively. According to the state vector x and the measurement vector z, the observation matrix $\boldsymbol{H}$ is determined as follows:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(4) Adaptive adjustment of measurement noise. Considering the uncertainty of the motion state of the target vehicle and the dynamic perception characteristics of the sensor, the measurement noise changes dynamically. In the traditional KF algorithm, the measurement noise is a fixed value, which easily makes the state estimation result difficult to adapt to the dynamic characteristics of the sensor. In this paper, a correction coefficientwas defined to adjust the measurement noise adaptively to ensure the stability of the state estimation. First, fit the measurement of the sensor for a period of time, and obtain the fitting terms of each measurement component of the sensor:

$$\hat{\mathbf{z}}_{ik} = \mathrm{p}(\mathbf{z}_{ik}, k)$$

Where $\hat{\boldsymbol{z}}_{ik}$ is the measurement sequence of the i-th component of the measurement vector $\boldsymbol{z}$ in time k, $\boldsymbol{z}_{ik}$ is the fitting term of the measurement sequence of the i-th component of the measurement vector $\boldsymbol{z}$ in time k, and $p(\mathrm{s}^2)$ is fitting function, which can be polynomial, spline, trigonometric, etc.

The residual of the measurement sequence of the i-th component of the measurement vector z at time k is as follows:

$$\Delta \mathbf{z}_{ik} = \mathbf{z}_{ik} - \hat{\mathbf{z}}_{ik}$$

By detecting the stability of the sensor measurement, the correction coefficient u is generated. The correction coefficient of the i-th measurement component at time k is as follows:

$$u_{ki} = \begin{cases} \dfrac{|z_{ki} - \hat{z}_{ki}|}{\delta_i}, & |z_{ki} - \hat{z}_{ki}| > \delta_i \\ 1, & |z_{ki} - \hat{z}_{ki}| \leq \delta_i \end{cases}$$

Where z_{ki} is the measurement of the i-th measurement component at time k, $\hat{z}_{ki}$ is the fitting value of the i-th measurement component at time k, δ_i is the threshold of the i-th measurement component, which reflects the volatility of the measurement, which can be a constant or time-varying function. In this paper, the volatility of the measurement of the

sensor is constrained to a fixed value, so the constants δ_i is used and δ_i are used to determine the percentile of the absolute value. The absolute value $\Delta \boldsymbol{z}_{ik}$ is arranged from small to large, and its appropriate percentile is selected as δ_i to constrain the fluctuation of the sensor measurement within a small range. If the residual $z_{ki} - \hat{z}_{ki}$ of the i-th measurement at time k is less than or equal to δ_i, indicating that the measurement of the sensor is stable, then the correction coefficient u_{ki} of the i-th measurement component is set to be equal to 1; if the residual $z_{ki} - \hat{z}_{ki}$ of the measurement value at time k is greater than δ_i, indicates that the measurement of the sensor is unstable, then set u_{ki} to a value greater than 1 to amplify the noise variance of the measurement component.

The measurement noise correction matrix is constructed as:

$$\boldsymbol{N}_k = \mathrm{diag}(u_{k1}, u_{k2}, \ldots, u_{k8})$$

Using $\boldsymbol{N}_k$ to adaptively adjust the initialized measurement noise matrix $\boldsymbol{R}_{ok}$, the new measurement noise matrix is obtained as:

$$\boldsymbol{R}_k = \boldsymbol{N}_k \boldsymbol{R}_{0k}$$

3.3 Data association algorithm based on adaptive tracking gate

Considering the dynamic perception characteristics of the sensor, the use of a fixed-size tracking gate may cause the correct measurement to fail to fall into the tracking gate or cause too many irrelevant measurements in the tracking gate, resulting in data association errors. Therefore, this paper dynamically adjusted the size of the tracking gate based on the position error perception characteristics of the sensor. The setting of the tracking gate size follows two principles:

(1) As far as possible, ensure that the measurement of the sensor falls into the door to reduce leakage correlation.

(2) As far as possible, ensure that irrelevant measurement values fall outside the door to reduce false associations.

Therefore, based on the above principles, the absolute position errors of the camera and the radar at different longitudinal distances from the sensor installation position are obtained by using the target-level perception results of the camera, radar and lidar in the previous section:

$$error_{region}^{type} = D_{region}^{type} - D_{region}^{lidar}$$

where $error_{region}^{type}$ represents the absolute position error of different sensors at different longitudinal distances, type represents the sensor type, including cameras and radars, and region represents the perception area at different longitudinal distances. D_{region}^{type} represents the absolute position of different sensors at different longitudinal distances, and D_{region}^{lidar} represents the absolute position of the lidar at different longitudinal distances.

The 3δ principle is used to eliminate outliers in the position errors of the camera and the radar, and the maximum value in the statistical results is used as the tracking gate at different longitudinal distances between the camera and the radar:

$$Gate_{region}^{type} = max\left(error_{region}^{type} < u_{region}^{type} + 3\delta_{region}^{type}\right)$$

Calculate the Euclidean distance d_{M} between the target track and the sensor measurement using the position similarity between the two:

$$d_{\mathrm{M}} = \sqrt{\left(\boldsymbol{z}_k^{x,y} - \hat{\boldsymbol{z}}_k^{x,y}\right)^2}$$

where $\boldsymbol{z}_k^{x,y}$ and $\hat{\boldsymbol{z}}_k^{x,y}$ represent the absolute position components of the sensor measurement and the target track prediction at the k-th time, respectively.

According to the calculated Euclidean distance d_{M}, determine whether the sensor measurement is within the tracking gate of the target track. If $d_{\mathrm{M}} \leq Gate_{region}^{type}$, it means that the sensor measurement is within the tracking gate. If $d_{\mathrm{M}} > Gate_{region}^{type}$ it means that the sensor measurement is outside the tracking gate, and then determine the size of each matrix element in the cost matrix. If the sensor measurement is inside the tracking gate, the size of the element in the cost matrix is d_{M}. If the sensor measurement is outside the tracking gate, the size of the element in the cost matrix is $Gate_{region}^{type}$. Finally, the Munkres algorithm is used to solve the cost matrix, and the optimal matching combination is obtained, and the matching result is used as the input of the state estimation module to update the state of each target track.

3.4 Life cycle management of target trajectory

The life cycle management of the track is used to realize the continuous tracking of the target, which mainly includes two parts: one is the definition of the target track state, and the other is the rule boundary of the track state switching. To manage the target track more effectively, a track life cycle management method based on M/N logic rules is adopted. The life state of the target track is divided into three types, including temporary, mature and dead. The state is judged and updated by the logical rules formulated by humans. First, use the position information of the sensor measurement and the predicted value of the target track to associate the two data. Second, for the target track that has not been successfully associated, determine whether to delete the track according to the set logic rules, for the sensor measurement that have not been successfully associated, define it as a temporary track and determine whether the temporary track is a new target or a sensor noise measurement according to the subsequent data association results; for the successfully associated target track, the state estimation module uses the matching sensor measurement to update the motion state of the target track. Figure 8 shows a schematic of the framework for target track life cycle management.

3.4.1 Temporary track. The generation of target tracks is an important link in track life cycle management. The data association module inputs the sensor measurement that are not associated successfully to the track life cycle management module to determine whether it is a new target. In this paper, the sensor measurement that has not been successfully associated is defined as a temporary track. Only when the temporary track has new sensor measurement associated with it in the subsequent frames, the temporary track is regarded as a new target.

3.4.2 Mature track. In the data association module, for the target track that has been successfully associated with the sensor measurement for many times, the track life cycle management module determines the status of the track maturity. In this paper, two of the three consecutive frames are successfully associated with the target track. Defined as a mature track, it is retained in the tracking track sequence, and data association and status updates are continuously performed on it.

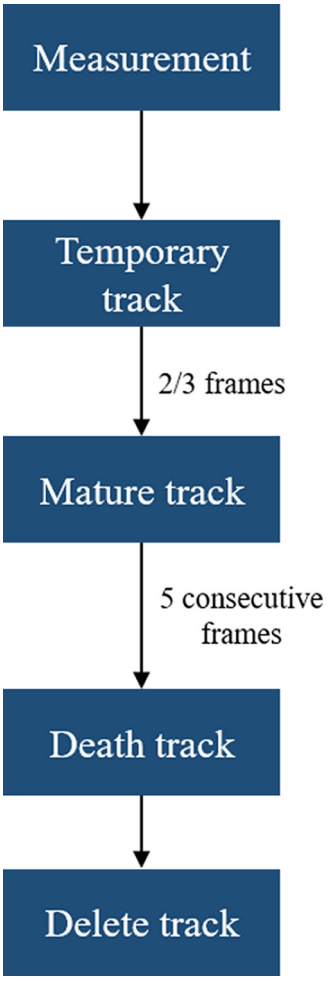

Figure 8.
Distribution of equivalent force in traffic safety field

3.4.3 Dead track. After the data association module completes the data association between the target track and the sensor measurement values, there will be some target tracks that are not associated with any sensor measurement. There may be two situations: one is that the target leaves the sensor's perception field of view, in which case the target track needs to be deleted, the other is that the target is in the sensor's perception field of view, but due to target occlusion, under-segmentation, etc., the sensor does not accurate identification; this situation requires the retention of the target track. Based on this, the target track without any sensor measurement associated with it for five consecutive frames is defined as a demise track, and the track is deleted.

4. Experiment

To verify the performance of the proposed multitarget tracking method based on roadside multisensor fusion. This section uses the roadside perception platform built in Section 2 to

collect target-level perception results of roadside camera, radar and lidar for a certain period. The perception result includes the target type, position, speed, etc., where the perception results of the 64-line lidar are used as the true value. And three indicators are used to evaluate the method proposed: the number of missed associations, the number of false associations and the mean absolute error.

4.1 Verification results and analysis

In this section, the above processed target-level data is used as the input of the proposed multitarget tracking algorithm based on roadside multisensor fusion, and the algorithm is run to obtain the tracking results. To better verify the performance of the proposed multitarget tracking algorithm, the performance verification is divided into two aspects: association performance and estimation performance. In terms of association performance, the improvement of data association performance of the proposed adaptive tracking gate compared to the fixed-size tracking gate is compared and analyzed. In terms of estimation performance, the improvement of estimated performance of the proposed multitarget tracking method to a single sensor and the comparison method is compared and analyzed. Moreover, to illustrate the rationality of the parameter settings in life cycle management, several demonstration specific tracking segments are used to analyze the impact of different parameter settings on the tracking effect.

To verify the performance of the proposed multitarget tracking algorithm under actual working conditions, three roadside sensors are deployed at the T-junction, including lidar, camera and radar. The data of Velodyne HDL-64E lidar is used as reference true value. A 2-megapixel industrial edition camera is used with a resolution of 1920×1080. The radar uses a 77 GHz traffic radar. All three heterogeneous sensors can directly detect traffic participant and output the target-level perception results, among whose the target-level perception results of the camera and radar are used as the fusion perception algorithm input. With good illumination and weather conditions, a length of about 60 s real road data section is selected for demonstration. The specific data is shown in Table 3. The data segment not only contains the processed target-level perception results of each sensor but also collects its image sequence while collecting the camera's target-level data for the performance verification of the auxiliary algorithm.

4.1.1 Association performance comparison. To verify the association performance of the multitarget tracking algorithm, this section takes the above data fragments as input, runs the proposed algorithm to obtain the tracking results and retains the predicted value of each frame of the track during the tracking process as the verification data for the association performance. As shown in Table 4, it is the size of the tracking gate under different longitudinal distance gradients from the camera and the radar to their installation positions. Next, centered on the predicted value of each target track in each frame, the maximum and minimum fixed tracking gate of the camera and the radar are respectively constructed, and the data association results are obtained and compared with the proposed adaptive tracking gate.

As shown in Tables 5 and 6, the data association results of the camera and radar using a fixed-size tracking gate and an adaptive-size tracking gate are, respectively, where the number of comprehensive errors is the sum of the number of false association errors and the number of missed association errors. It can be seen from Table 5 that, for the camera,

Table 3.
Collection of road data

Collection time/s	Target No.	Target type	Target behavior
58.631	28	car	straight/turn

compared with the maximum fixed tracking gate, the tracking gate of the adaptive size has fewer false associations and more missed associations. This is mainly because the larger the tracking gate is, the more likely false associations will occur, and the less likely missed associations will occur. Compared with the minimum fixed tracking gate, the adaptive size tracking gate has more false associations and fewer missed associations. This is mainly because the smaller the tracking gate is, the less likely false associations will occur and the more likely missed associations will occur. From a comprehensive point of view, compared with the fixed-sized tracking gate, the adaptive-sized tracking gate has the least number of comprehensive errors and the highest comprehensive performance of data association. It can be seen from Table 6 that for the radar, the data association results are similar to those of the camera. Compared with the maximum fixed tracking gate, the tracking gate with adaptive size has fewer false associations and more missed associations. For the smallest fixed tracking gate, the number of false associations is more and the number of missed associations is less. Compared with the fixed-size tracking gate, the number of comprehensive errors is the least, and the comprehensive performance of data association is the highest. However, compared with cameras, radars are more prone to false associations and missed associations, mainly because radars have lower positional accuracy and are prone to noise measurements and false detection measurements. To sum up, in the data association link, the adaptive size of the tracking gate can be used to comprehensively improve the association performance in the multitarget tracking process.

4.1.2 Estimated performance comparison. This section compares and analyzes the improvement effect of the proposed multitarget tracking method compared with the single sensor measurement and the comparison method baseline, in which baseline adopts the traditional Kalman filtering method with fixed measurement noise in the state estimation module.

As shown in Figure 9, the mean absolute errors of the absolute position and velocity of the method in this paper, the measurement of a single sensor, and the baseline method at different longitudinal distances, respectively. It can be seen from the Figure 9 that the

Table 4. Size of the tracking gate under different longitudinal distance gradients

Area/m	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Camera gate/m	5.40	3.95	3.87	4.21	4.90	6.01	6.29
Radar gate/m	7.92	7.83	7.98	8.16	8.14	8.17	8.87

Table 5. Data association results of camera under different tracking gate

Method	No. of false associations	No. of missed associations	Total no. of errors
Maximum tracking gate – 6.29	21	10	31
Minimum tracking gate – 3.87	12	27	39
Adaptive tracking gate	15	11	26

Table 6. Size of the tracking gate under different longitudinal distance gradients

Method	No. of false associations	No. of missed associations	Total no. of errors
Maximum tracking gate – 8.74	56	22	78
Minimum tracking gate – 7.83	51	32	83
Adaptive tracking gate	52	24	76

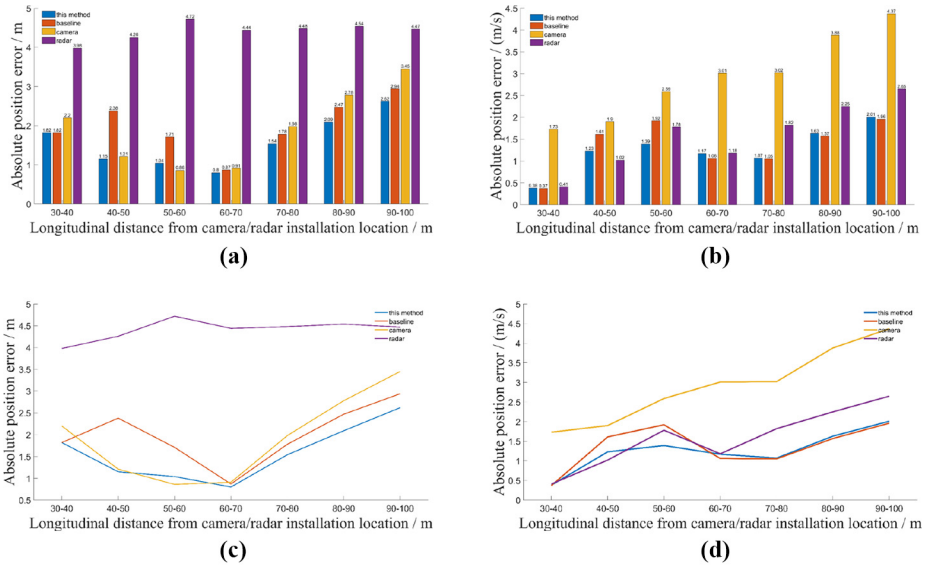

Figure 9.
Comparison of absolute position estimates and measurements

Notes: (a) Comparison of absolute position estimates and measurements; (b) comparison of absolute velocity estimates and measurements; (c) comparison of absolute position estimates and measurements; (d) comparison of absolute velocity estimates and measurements

ranging accuracy of the camera is higher than that of the radar, but it is greatly affected by the distance. The speed measurement accuracy of the radar is higher than that of the camera. Compared with the single-sensor measurement, the multitarget tracking method proposed in this paper has improved accuracy after fusion. In addition, compared with the baseline method, the proposed method can improve the estimation accuracy of absolute position and velocity and reduce noise interference.

The total mean absolute error of the estimated value of the multitarget tracking algorithm and the measurement of a single sensor relative to the reference value of the lidar is statistically calculated. Table 7 lists the total mean absolute error for absolute position (D) and velocity (v) of the measurement of each sensor, this method and the baseline method.

According to Figure 9, the multitarget tracking method proposed in this paper improves the estimation accuracy of absolute position and velocity. After calculation, compared with the measurement of a single sensor, the position estimation accuracy is increased by 13.5%, and the speed accuracy is increased by 22.2%. Compared with the baseline method, the accuracy of position estimation is increased by 23.8%, and the accuracy of speed estimation is increased by 8.9%.

5. Discussion

In terms of process of setting the relevant parameters in the life cycle management link, it hopes that the temporary track can be switched to the mature tracking as quickly as possible at the beginning of a new target tracking process. While, if the tracking is deleted, it is better to obtain enough evidence to show that the mature track has faded out of the sensor's view field.

To demonstrate the rationality of the parameter settings at the start of the tracking, we compare one of two consecutive frames that are successfully associated with the sensor measurements. A track segment with 1/2 as the starting condition of the track is selected as an example. As shown in Figure 10, the black dotted line is the road boundary. The triangles are camera and radar deployed on the roadside. Solid circles and asterisks represent tracked multitarget trajectories, where asterisks represent false targets and the rest are real targets. When 1/2 is used as the threshold for the start of the track, because the threshold is set too small, a false target for one frame is generated, as shown by the asterisked target in Figure 10. Therefore, at the beginning of the track, 1/2 is more likely to generate false targets than 2/3.

In the other side, to demonstrate the rationality of parameter settings when the track is deleted, we compare with three consecutive frames that are not associated with any sensor measurement. A track segment that using 3/3 as the track deletion condition is selected as an example. The tracking result is shown in Figure 11. ID3 is a vehicle turning right from southeast to east, and ID1 and ID2 are two tracking IDs of the same target. ID3 briefly blocked the targets represented by ID1 and ID2 when turning. When the track was deleted, the set threshold was too small, resulting in two tracking IDs generated for the targets that were briefly occluded. Therefore, when the track is deleted, compared with 5/5, using 3/3 is more likely to affect the continuity of tracking.

Sensor or algorithm	D/m	v/m/s
Sensor		
Camera	1.41	2.89
Radar	4.51	1.58
Algorithm		
Baseline	1.60	1.35
This method	1.22	1.23

Table 7. MAE index of sensor measurements, the baseline and this method

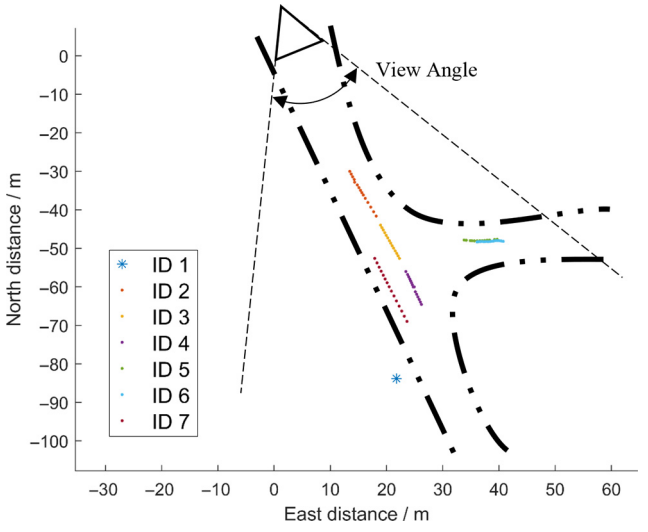

Figure 10. Track segment with track start threshold of 1/2

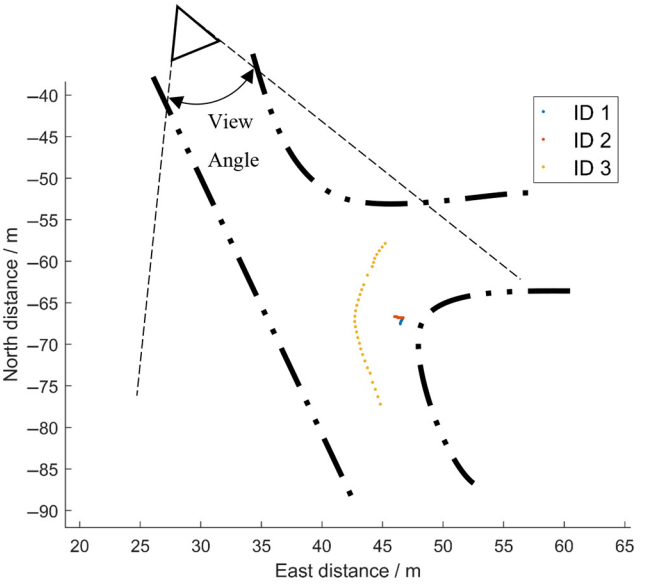

Figure 11.
Track segment with track deletion threshold of 3/3

6. Conclusion

To meet the environmental perception requirements of high-level autonomous driving and improve the driving performance of intelligent connected vehicles, a multitarget tracking method based on roadside multisensor fusion is proposed. The main work of the paper is as follows.

First, a roadside perception platform is built, and the roadside camera, radar and lidar in the platform are synchronized in time and space. At the same time, using the real road data after synchronization, the perception error characteristics of each sensor are evaluated and analyzed.

Second, to solve the problem that the fixed-size tracking gate is difficult to adapt to the dynamically changing sensor measurement error, a data association method based on the adaptive tracking gate is proposed. The position error characteristic of the heterogeneous sensor determines the size of the tracking gate under different longitudinal distances and realizes the adaptive adjustment of the tracking gate.

Third, in view of the uncertainty of the motion state of the target vehicle and the dynamic perception characteristics of the sensor in the real environment, a multisensor fusion method based on measurement noise adaptive KF is proposed to realize the target-level data fusion of heterogeneous sensors. The fusion method detects the stability of the sensor measurement, generates the correction coefficient of the measurement noise, adaptively adjusts the measurement noise and reduces the influence of the fixed noise value on the system estimation.

The experimental results of the real city intersection senario show that the multitarget tracking method based on multisensor fusion proposed in this paper improves the position estimation accuracy by 13.5% and the speed estimation accuracy by 22.2%. Compared with the baseline method, the method proposed in this paper improves the position estimation accuracy by 23.8% and the velocity estimation accuracy by 8.9%.

References

Bar-Shalom, X. (2001), *Estimation with Applications to Tracking and Navigation*, Wiley.

Caltagirone, L., Bellone, M., Svensson, L., *et al.* (2018), "LIDAR-Camera fusion for road detection using fully convolutional neural networks", *Robotics and Autonomous Systems*, Vol. 111, p. 111.

Deng, F., Chen, J. and Chen, C. (2013), "Adaptive unscented Kalman filter for parameter and state estimation of nonlinear high-speed objects", *Journal of Systems Engineering and Electronics*, Vol. 24 No. 4, pp. 655-665.

Eltrass, A. and Khalil, M. (2018), "Automotive radar system for multiple-vehicle detection and tracking in urban environments", *IET Intelligent Transport Systems*, Vol. 12 No. 8, pp. 783-792.

Jo, K., Chu, K. and Sunwoo, M. (2012), "Interacting multiple model Filter-based sensor fusion of GPS with in-vehicle sensors for real-time vehicle positioning", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13 No. 1, pp. 329-343.

Josip, M., et al. (2016), "Radar and stereo vision fusion for multitarget tracking on the special Euclidean group", *Robotics and Autonomous Systems*, Vol. 83 No. C, pp. 338-348.

Lekic, V. and Babic, Z. (2019), "Automotive radar and camera fusion using generative adversarial networks", *Computer Vision and Image Understanding*, Vol. 184, pp. 1-8.

Li, K., Dai, Y., Li, S., *et al.* (2017), "State-of-the-art and technical trends of intelligent and connected vehicles", *J Automotive Safety Energy*, Vol. 8 No. 1, pp. 1-14. (in Chinese)

Mobus, R. and Kolbe, U. (2004), "Multi-target multi-object tracking, sensor fusion of radar and infrared", *Intelligent Vehicles Symposium*, IEEE.

Otto, C., Gerber, W., Leon, F.P., *et al.* (2012),. "A joint integrated probabilistic data association filter for pedestrian tracking across blind regions using monocular camera and radar", *Intelligent Vehicles Symposium*, IEEE.

Sadeghian, A., Alahi, A. and Savarese, S. (2017), "Tracking the untrackable: learning to track multiple cues with long-term Dependencies", *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE.

Wan, E.A. and Merwe, R. (2000), "The unscented Kalman filter for nonlinear estimation", *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000, AS-SPCC*, The IEEE.

Wu, W., Jiang, J., Feng, X., Liu, C. and Qin, X. (2016), "A survey of multi-target tracking algorithms based on random finite sets", *Electronics Optics and Control*, Vol. 23 No. 3, pp. 1-6.

Xiu, C.B. and Guo, F.H. (2013), "Wind speed prediction by chaotic operator network based on Kalman filter", *Science China: Technical Science in English*, Vol. 5, p. 8.

Yibing, Z., Liu, C., Zheng, Z., Guo, L., Ma, Z. and Han, Z. (2021), "Intelligent vehicle localization method based on multi-sensor information fusion", *Chinese Journal of Automotive Engineering*, Vol. 11 No. 1, pp. 1-10.

Zhang, C., Bao, H., Xiantong, D., *et al.* (2019), "The application and future of roadside perception in intelligent networked vehicles", *Artif Intell*, Vol. 1, pp. 58-66. (in Chinese)

Zhang L., Zhang S., Tao Y., *et al.* (2019), "Sensory task assignment based on Dempster-Shafer theory and multi-attribute fusion in mobile sensor networks", *IEEE Access*.

Zhao, J. and Wang, W. (2013), "Development of electric vehicle autonomous driving system based on sensor fusion technology", *Manufacturing Automation*, Vol. 35 No. 9, pp. 43-46.

Corresponding author
Yimin Wu can be contacted at: wuyimin2000@126.com