A Novel Spatio-Temporal Synchronization Method of Roadside Asynchronous MMW Radar-Camera for Sensor Fusion

Yuchuan Du, *Member, IEEE*, Bohao Qin, Cong Zhao, Yifan Zhu, Jing Cao, and Yuxiong Ji

***Abstract*— Roadside sensors, such as camera and millimeter-wave (MMW) radar, provide traffic information beyond the visual range of intelligent vehicles in cooperative vehicle-infrastructure systems. Unlike onboard equipment, roadside sensors are affiliated with different systems and lack synchronization in both space and time. In this paper, we propose a novel spatio-temporal synchronization method of asynchronous roadside MMW radar–camera for sensor fusion, which utilizes features of the scenario to extract lane line corner points to pre-calibrate the camera. Based on the consistent time flow rate of the separate sensors, multiple virtual detection lines are set up to match the time headway of successive vehicles and conduct objective matching to track data. Finally, a synchronization optimization model is formulated and a constrained nonlinear minimization solver is applied to tune the parameters. Measure data from Donghai Bridge in Shanghai is applied to verify the feasibility and effectiveness of the method. The results determine that there are 33 frames (33*40 ms) of temporal deviation between the camera and the radar in this case. After the synchronization, the average spatial deviation is reduced from 2.47 m to 0.42 m in the X-direction and 64.06 m to 2.34 m in the Y-direction, respectively. This study provides an economical and effective way to solve the problem of spatio-temporal synchronization of roadside sensors.**

***Index Terms*— Roadside sensor fusion, camera, MMW radar, spatio-temporal synchronization, objective matching.**

I. Introduction

ROADSIDE sensors have a broad detection range and can also be equipped with powerful edge computing. 5G communication technology has led to a tremendous increase in information transmission rates [1], which makes it possible to provide roadside perception information from beyond the visual range of intelligent vehicles in real time [2]. The development of cooperative vehicle–infrastructure systems also requires high-quality perception data from roadside sensors to achieve safe and efficient driving and traffic operation [3], [4]. Intelligent highway [5] reformation is urgently needed that effectively utilizes existing roadside devices or new sensors for data fusion to improve the accuracy of roadside perception.

As the most prevalent roadside sensor, the camera relies primarily on object detection algorithms for perception. In the last decade, with the development of neural networks, state-of-art object detection algorithms (such as Yolo-V4 [6], EfficientNet [7], SSD [8], DetectoRS [9], etc.) have been widely used to detect vehicles on the road. These algorithms can assign a bounding box to each vehicle in the video and output its type and confidence level (In this paper, we also refer to camera data as video data.). For the calibrated camera, we can further calculate the relative coordinates of objects, and the detection distance can even exceed 1 km in unobstructed view conditions. Additionally, the image data of the objects can also be obtained to extract features and even recognize license plates [10].

However, the shortcomings of video-based detection are also obvious. Poor illumination conditions and local feature occlusion seriously affect the detection accuracy and trajectory continuity. Lidar is another very popular sensor in the field of autonomous driving. It has an even greater detection distance than that of a camera, and its perception information is more comprehensive. However, the price of lidar is generally very high, and it requires significant computing power and data storage capacity to process a large amount of point cloud data. Therefore, lidar has not been widely installed on roadsides. Compared with lidar, millimeter-wave radar (MMW radar or radar for short) is much cheaper, extremely sensitive to object velocity, and has been widely used in practical engineering scenarios. The structured data of the MMW radar includes the location and accurate velocity of the object in the radar coordinate system [11]. The MMW radar is almost unaffected by environmental factors and is robust for the detection of partially occluded objects. As a result, the track continuity of radar data is better. However, the MMW radar detection distance is limited, generally, not more than 300 m, and larger vehicles near the radar are prone detected as two or more objects. Both cameras and MMW radar sensors have

Manuscript received May 13, 2021; revised August 11, 2021; accepted October 1, 2021. The work of Cong Zhao was supported by the Shanghai Sailing Program under Grant 21YF1449400. This work was supported by the National Natural Science Foundation of China under Grant 52102383, and in part by the Innovation Program of Shanghai Municipal Education Commission under Grant 2021-01-07-00-07-E00092, and in part by the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0100, and in part by the Scientific Research Program of Shanghai Municipal Science and Technology Commission under Grant 19DZ1209100, and in part by the Zhejiang Province Key Research and Development Program under Grant 2021C01011. The Associate Editor for this article was Y. Hou. *(Corresponding author: Cong Zhao.)*

Yuchuan Du is with the Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China, also with the Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 201210, China, and also with the Shanghai Engineering Research Center of Urban Infrastructure Renewal, Shanghai 200032, China.

Bohao Qin, Cong Zhao, Yifan Zhu, Jing Cao, and Yuxiong Ji are with the Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China (e-mail: zhc@tongji.edu.cn).

Digital Object Identifier 10.1109/TITS.2021.3119079

advantages that make up for each other's weaknesses in detection to some extent. Hence, data fusion of these two sensors is considered an effective means to improve roadside detection accuracy.

The spatio-temporal synchronization of the two sensors is a basic but critical process in radar–camera fusion technologies, which directly affects the accuracy of subsequent data fusion [11]–[13]. The goal of this process is to match the perception data of two sensors for the same objects in spatio-temporal dimensions and minimize the system deviation. Most studies (such as Zhang calibration, four points calibration, pseudo inverse, direct linear transformation, extrinsic calibration, etc.) have focused on the calibration of the two sensors in space and assumed that the two sensors are temporally synchronized [14]–[16]. For the time dimension, other scholars mainly discussed the temporal synchronization problem caused by the inconsistent sampling frequency of the camera and radar. Fu *et al.* [17] proposed a multi-threaded temporal synchronization method that stores the data from the camera and radar in a buffer and updates in real time. A similar approach was also used by Feng *et al.* [18]. Ma *et al.* [11] used an algorithm combining Kalman filtering and Lagrangian interpolation to realize the temporal synchronization of the data from two sensors. Zhang and Cao [19] fused the data of sensors with very close timestamps ($\Delta t \leq 0.005\ s$).

Many types of sensors (i.e., cameras and MMW radars) have been installed on roadsides. The MMW radars were originally installed to measure velocity and traffic flow, and cameras were used for observing road conditions and detecting abnormal events, and radar–camera fusion was not previously considered. They are affiliated with different systems and work separately. Therefore, a camera and radar in a group may work on different timelines, which poses a challenge for calibrating the camera using radar measurement data, because the data of each at the same timestamp may deviate by several seconds [20].

In this paper, we propose a new method for spatio-temporal synchronization of roadside cameras and MMW radar. This method is intended for scenarios where the camera and the radar are out of synchronization in both space and time. This research produced the following major contributions:

1) Scenario feature-based camera pre-calibration method: by utilizing a priori information of lane lines with regular features, the corner point coordinates are estimated to generate the pre-calibration file and calibrate the camera.
2) Objective matching by multiple virtual detection lines: based on the consistent time flow rate of the camera and radar, the time headway of successive vehicles is matched via multiple virtual detection lines in the scenario.
3) Spatio-temporal synchronization optimization model: a 12-parameter spatio-temporal synchronization optimization model is developed to correct the pre-calibration file, and is solved by the constrained nonlinear minimization solver.
4) Real engineering scenario test: the proposed method is verified and evaluated with the measured data from Donghai Bridge in Shanghai. It is proved to be feasible and effective by comparing the spatio-temporal deviation before and after the synchronization.

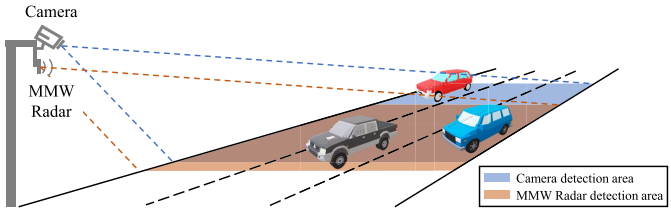

Fig. 1. Study scenario.

We present a low-cost and conveniently implemented method without field operation or additional timing devices, which only needs to simultaneously record the camera and radar data for a period of time for the synchronization.

This paper is organized as follows: we introduce the study scenario and framework in Section II. We describe the proposed spatio-temporal synchronization method in detail in Section III. Then, in Section IV, we present the results of the measured data in a case and performance analysis. Finally, we summarize the conclusions in Section V.

II. Study Scenario and Framework

This paper focuses on the scenario that is illustrated in Fig. 1. The camera and radar are installed on the roadside. They are not in the same position but their detection areas have a certain intersection, through which sensor fusion can be achieved. Additionally, they may be affiliated with different systems and have a deviation of several seconds in time. Furthermore, their data sampling frequencies are also inconsistent (such as the camera outputs the video at a frame rate of 25 fps, and the sampling frequency of the radar is 20 Hz). For the camera, each frame of the video is captured, and the object detection algorithm is performed to obtain the bounding boxes and pixel coordinates of the objects for localization. The object position data obtained by the radar are the relative coordinates of the objects in the radar coordinate system, which are also spatially unmatched with the pixel coordinate system of the camera. The lane lines in the road are clearly visible, which played a significant role in this study.

In the context of the proposed study scenario, our goal is to use the time and coordinate system of the MMW radar as the benchmark, and achieve the synchronization of the camera and radar in time and space by adjusting the camera time and transforming the pixel coordinates. Fig. 2 illustrates the framework of this study, which includes data acquisition and preprocessing, scenario feature-based pre-calibration of the camera, multi-object tracking and trajectory noise estimation, object matching by multiple virtual detection lines, and the spatio-temporal synchronization optimization model. The implementation process of each step is described in Section III.

III. Spatio-Temporal Synchronization Method

A. Data Acquisition and Preprocessing

The simultaneous acquisition of camera and the radar data over a period of time is used for preprocessing and

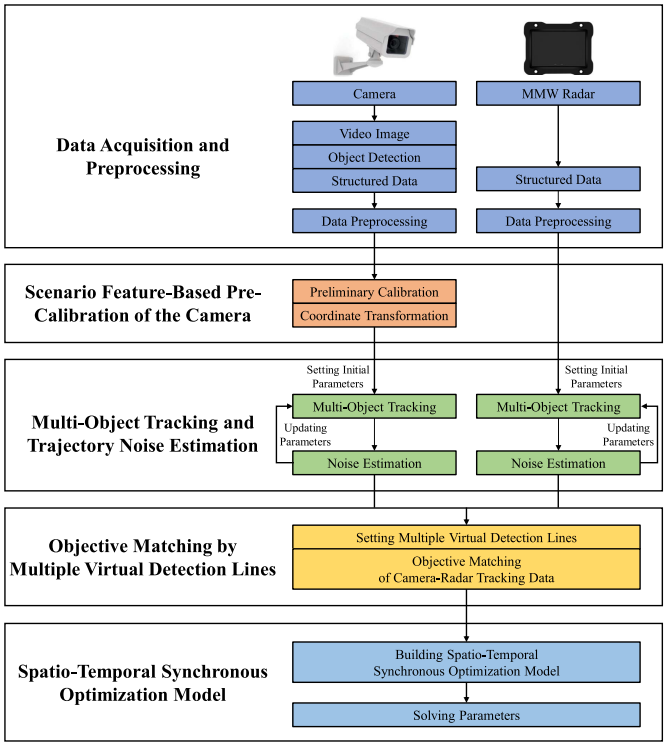

Fig. 2. The framework of roadside asynchronous MMW radar–camera spatio-temporal synchronization method.

subsequent analysis. The camera records the video at a frame rate of 25 fps, and an object detection algorithm is performed on each frame to assign the bounding boxes. We take the pixel coordinate of a specific anchor point in each bounding box to represent the object position in the image. This point could be a corner point, center point, or midpoint of an edge, etc. The selection of the anchor point can be customized for different scenarios, but it is necessary to ensure that the relative position of the point to the object does not change significantly as the object's location changes in space.

The MMW radar outputs 20 Hz detection results in its two-dimensional coordinate system, including the position, velocity, and id for each object. The radar coordinate system is located in the same plane as the road. It takes the location of the radar as the origin, the road direction as the Y-axis, and the direction perpendicular to the road as the X-axis. To maintain the consistency of the sampling frequency of the radar data with the video frames, we interpolate each vehicle trajectory to resample the radar data, as demonstrate in Fig. 3. The interpolation is performed between two radar data points, and the interval between them is 50 ms. In such a short period, the vehicle velocity changes little, and we assume that the vehicle is at a consistent velocity. Therefore, we chose the linear interpolation method. As demonstrated in the example in Fig. 3, we assume that the X coordinates of the raw radar data at $t = 50\ ms$ ($x_{t=50}$) and $t = 100\ ms$ ($x_{t=100}$) are known, and we calculate the X coordinate when according to (1) (Y coordinate and velocity interpolation method are the same). On the other hand, based on the mechanism of radar detection, it is inevitable that several additional detections or

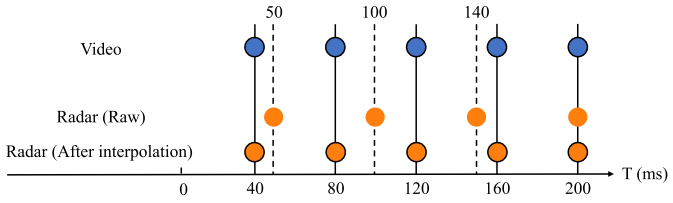

Fig. 3. Data interpolation.

false alarms are obtained. To process these data, we count the vehicle IDs and eliminate data with a very small number of IDs.

$$x_{t=80} = x_{t=50} + \frac{x_{t=100} - x_{t=50}}{100 - 50}(80 - 50) \tag{1}$$

B. Preliminary Calibration of the Camera

The purpose of the preliminary calibration of the camera is to convert the pixel coordinate system into a world coordinate system. The world coordinate system is a customized cartesian coordinate system. We can select a plane in the camera image as the plane in which the coordinate system is located and define the origin and axis direction of the coordinate system. In this paper, we select the road plane in the video image to establish the world coordinate system, which is on the same plane as the relative coordinate system of the MMW radar.

We construct vertical grids on real roads by selecting lane line corner points for the preliminary calibration of the camera. By referring to the lane line standard, we can obtain a priori information on the relative position of the lane lines in space (in China, the longitudinal length of lane lines along the lane direction is 6 m, the longitudinal distance between lane lines along the lane direction is 9 m, and the lane width is in the range of 3–4 m. [21], [22]) and estimate the approximate world coordinate of each point. At the same time, we extract the pixel coordinates of all these points from the video image. After the corresponding points of the pixel coordinates and world coordinates are obtained, the relationship between them is described by (2) [14], [23], [24]. The mapping relationship between two plane coordinates is obtained by calculating the homography matrix.

$$\begin{aligned} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \frac{1}{Z} H \begin{pmatrix} U \\ V \\ 1 \end{pmatrix} \\ &= \frac{1}{Z} \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{pmatrix} U \\ V \\ 1 \end{pmatrix} \end{aligned} \tag{2}$$

where H denotes the homography matrix, which is obtained by multiplying the intrinsic and extrinsic parameters of the camera (Equation (3)). Z denotes the scaling factor, u and v denote pixel coordinates, and U and V denote the corresponding world coordinates.

$$H = A\,(R_1\ R_2\ T) \tag{3}$$

where A denotes the intrinsic parameter of the camera, R_1 denotes the first column of the rotation matrix, R_2 denotes

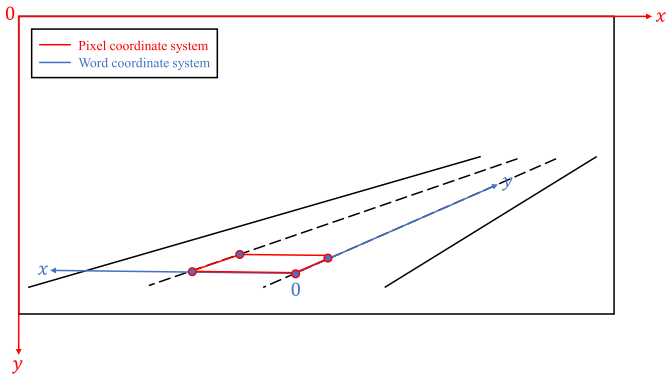

Fig. 4. Preliminary calibration of the camera.

the second column of the rotation matrix, and T denotes the translation vector.

At least four corresponding point pairs are needed to solve the homography matrix. For scenarios that involve long and straight roads, more lane corners can be selected to estimate their world coordinates, and the best homography matrix can be calculated using the least squares method. However, the world coordinates of the distant lane line corner points become difficult to estimate in scenarios involving curved roads since the curvature of the road is unknown. To ensure that the method will be robust in both straight and curved road scenarios, we specifically selected four points that are close to the camera, as illustrated in Fig. 4. We chose one point as the origin and estimated the world coordinates of the remaining three points. It should be noted that the world coordinate system established at this point still did not match the radar coordinate system, and the estimated world coordinates also contain noise. This step is only needed to obtain an initial calibration file, which will be corrected in the subsequent process.

As mentioned above, the location of each vehicle in the image can be represented by a pixel point. Then, after solving the homography matrix, the pixel coordinates corresponding to the world coordinates can be calculated according to (4) and (5). Inverting them to get (6) and (7), we can calculate the world coordinates corresponding to the pixel coordinates.

$$u = \frac{H_{11}U + H_{12}V + H_{13}}{H_{31}U + H_{32}V + H_{33}} \tag{4}$$

$$v = \frac{H_{21}U + H_{22}V + H_{23}}{H_{31}U + H_{32}V + H_{33}} \tag{5}$$

$$U = \frac{B_2C_1 - B_1C_2}{A_1B_2 - A_2B_1} \tag{6}$$

$$V = \frac{A_1C_2 - A_2C_1}{A_2B_1 - A_1B_2} \tag{7}$$

where:

$$\begin{cases} A_1 = H_{31}u - H_{11} \\ B_1 = H_{12} - H_{32}u \\ C_1 = H_{13} - H_{33}u \end{cases} \tag{8}$$

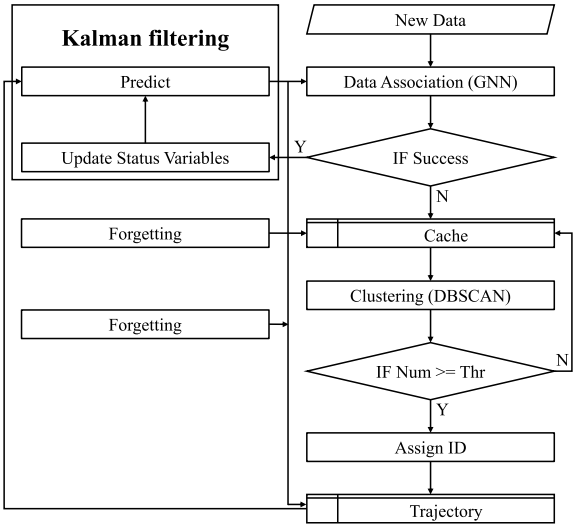

Fig. 5. The flow chart of the multi-object tracking algorithm.

$$\begin{cases} A_2 = H_{31}v - H_{21} \\ B_2 = H_{22} - H_{32}v \\ C_2 = H_{23} - H_{33}v \end{cases} \tag{9}$$

C. Multi-Object Tracking and Trajectory Noise Estimation

Kalman filtering and the global nearest neighbor (GNN) data association algorithm are performed on the objects detected by the camera and radar (since the IDs in the raw radar data are discontinuous and easy to switch, we input the data from the radar into the multi-object tracking algorithm that we write to reassign IDs) for multi-object tracking [25]. The flow chart of the multi-object tracking algorithm is illustrated in Fig. 5.

In the algorithm, the captured data at the time t is associated with the predicted values of each trajectory at time $t-1$. In the GNN association algorithm, the threshold we use is a rectangle, which is set by the mean of the state variable (location) plus or minus N times the standard deviation. The Kuhn–Munkres algorithm is used for data matching, where the distance metric is set as the euclidean distance between the objects [26], [27].

For the new successfully associated data, Kalman filtering is performed to update the state variables and predict their states at the time $t+1$ [28]. The updated data is recorded in the Trajectory Table, as shown in Fig. 5. On the other hand, the data that fails the association is put into the Cache Table, which performs DBSCAN clustering every time. When the number of data points in one category reach the threshold, they are considered a new trajectory, which is assigned a new ID ($new\ ID = max(IDs) + 1$) and moved to the Trajectory Table. The remaining data stay in the Cache Table until the next clustering. To a certain extent, the setting of the Cache Table reduced some of the abnormal trajectories caused by false alarms or additional detections.

The predicted values that fail the association are marked as missing and recorded in the Trajectory Table, along with

the number of missing counts, and the predicting process continue. Since there are no new observations to update the state variables, their covariance will gradually increase. When the missing predictions are re-associated with the new data, the missing status and the number of missing counts are reset.

A forgetting mechanism is also designed in the algorithm. When a category in the Cache Table is not cluster a sufficient amount of data after a substantial period of time, it will be removed from the Cache Table. Similarly, when the number of missing data for an object in the Trajectory Table reaches the threshold or the object exceed the sensor's detection area, the trajectory of the object will be terminated, and its prediction is also ceased.

For Kalman filtering, we set the state variables to the positions on the X and Y axes as P_x, P_y and the velocity in the X and Y directions as V_x, V_y. Additionally, we input the noise of the measurement and motion model into the filter. However, these parameters are usually unknown at first and need to be estimated according to the measured data. To iteratively calculate the estimated noise of the measurement and motion model, we use a process of setting the initial values for tracking, selecting the trajectory with better tracking, estimating the noise by fitting trajectories, updating the filter configuration, and re-tracking.

The method of curve fitting and residual calculation is used to estimate the noise of the trajectory. For each trajectory, we decompose its motion on the X and Y axes to determine the time–distance relationship. Then, we use the K-degree polynomial function to fit the curve. It is assumed that the vehicle's motion is linear with variable acceleration on each axis. The variance of the residual difference between the observed and fitted value is calculated as the measurement noise of the sensor. The motion model is illustrated in (10). The observed values of the state variables at each moment are inputted into the motion model to obtain the predicted values of the model for the next moment. The same method is applied to the predicted value of the model as the predicted noise. It should be noted that the trajectory noise consists of system noise, inherent noise and environmental noise, etc. Due to the unidirectional characteristic of the system noise, we cannot estimate the system noise of the sensor by curve fitting the observed values of the trajectory points. The main purpose of our definition of trajectory noise as the residual between the observed and fitted values of the trajectory points is to estimate the intrinsic noise of the sensor, which needs to be fed into the Kalman filter as an important input parameter.

$$\begin{bmatrix} p_{x\_t+1} \\ p_{y\_t+1} \\ v_{x\_t+1} \\ v_{y\_t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{x\_t} \\ p_{y\_t} \\ v_{x\_t} \\ v_{y\_t} \end{bmatrix} \tag{10}$$

D. Objective Matching

After the multi-object tracking of the camera and the radar data, an ID is assigned to each vehicle, and its continuous trajectory is obtained. We then match the vehicle IDs detected by the camera with the radar tracking data. In other words, we determine which vehicle in the radar tracking data correspond to a vehicle with a specific ID in the video tracking data. We also propose a method of setting multiple virtual detection lines in separate lanes to accomplish the object matching of video and radar tracking data. Using this method, we can calculate the camera and radar's coarse values of temporal deviation and Y-directional deviation in space.

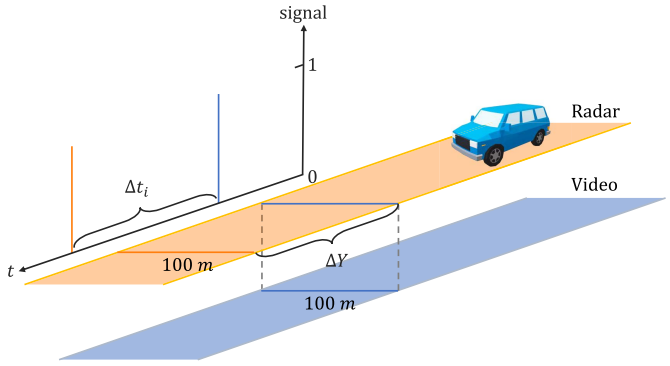

Fig. 6. Virtual detection lines.

Taking a lane as an example, we select data from the same lane in the video and radar and set a virtual detection line for both at the same value on the Y-axis. Fig. 6 demonstrates that the two virtual detection lines are not at the same location, because the camera and the radar are still spatially unified on the coordinate system. When a vehicle passes the detection line, the respective timestamps of the video and radar are recorded. There will be a time gap between the two timestamps, denoted as Δt_i, which is caused by the asynchronous coupling between the camera and radar in time and space. As illustrated in (11), Δt_i is composed of the time deviation ΔY between the video and radar, the time $\Delta Y/\bar{v}_i$ caused by the Y-direction deviation ΔY in space when the i^{th} vehicle passes by, and the error term e_i. Where e_i is assumed to follow a normal distribution with a mean of 0 ($E \sim N\left(0, \sigma^2\right)$) when there is a sufficient amount of data.

$$\Delta t_i = \Delta T + \frac{\Delta Y}{\bar{v}_i} + e_i \tag{11}$$

$$\Delta Y = Y_{radar\_act} - Y_{video\_act} \tag{12}$$

As demonstrated in Fig. 7, the timestamps of all vehicles passing the detection line are recorded. The information of each vehicle can be represented by a one-dimensional vector of length $N + 1$: $[t_{i-N} - t_i, \ldots, t_{i-1} - t_i, t_i, t_{i+1} - t_i, \ldots, t_{i+N} - t_i]$, which included the timestamp when the vehicle passes the detection line and the time headway between the vehicle and the previous and following N vehicles. When there is no vehicle in front or behind the object vehicle, the value is filled with 0. The euclidean distance (13) is used as the information distance metric between each vehicle, and the distance matrix is calculated. The Kuhn–Munkres algorithm [26], [27] is used to achieve objective matching between the video and the

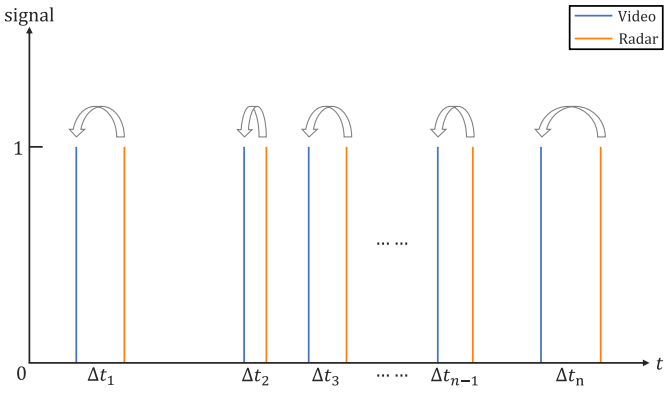

Fig. 7. The objective matching of the timestamps.

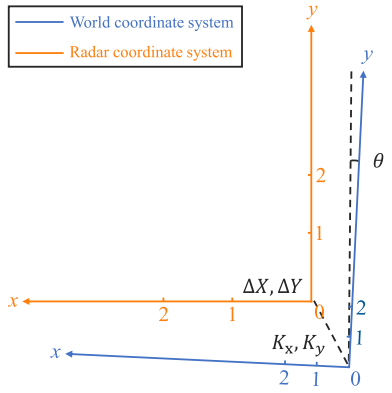

Fig. 8. The world and the radar coordinate system.

radar tracking data.

$$dis_{ij} = \sqrt{\sum_{k=-N}^{N} \left((t_{i+k} - t_i) - \left(t_{j+k} - t_j\right)\right)^2} \tag{13}$$

The objective function is designed as (14) to calculate the sum of squares of the residual differences between the Δt_i and $\Delta \bar{t}$. Where $\Delta \bar{t}$ is illustrated in (15). Fixing the position of the video detection line and constantly adjusting the position of the radar detection line minimize the value of the objective function. That is, we make ΔY close to 0. ΔT can also be approximated when the value of the objective function achieves the minimum. These two parameters can be used as the initial values of the subsequent spatio-temporal synchronization optimization model. The above calculation process for the measured data of one detection line on a single lane can be extended to other lanes or detection lines with different positions.

$$\begin{aligned} F &= \min \sum_{i=1}^{n} \left(\Delta t_i - \Delta \bar{t}\right)^2 \\ &= \min \sum_{i=1}^{n} \left[\Delta Y \left(\frac{1}{v_i} - \frac{\sum_{i=1}^{n} \frac{1}{v_i}}{n}\right) + e_i\right]^2 \end{aligned} \tag{14}$$

$$\Delta \bar{t} = \Delta T + \frac{\sum_{i=1}^{n} \frac{\Delta Y}{v_i}}{n} \tag{15}$$

E. Spatio-Temporal Synchronization Optimization Model

The preliminary calibration of the camera allows us to convert pixel coordinates into world coordinates, which are in the same plane as the radar coordinate system. However, the world coordinate system and the radar coordinate system are still not fully matched at this point. In this section, a 12-parameter $(\Delta T, \Delta X, \Delta Y, \theta, K_x, K_y, \left(e_{x\_i}, e_{y\_i}\right)_{i=1}^{3})$ spatio-temporal synchronization optimization model is built to complete the matching.

Fig. 8 demonstrates there are three deviations between the world and the radar coordinate system. Firstly, the origins of the two coordinate systems are not in the same position, and there is a certain distance between them, which can be decomposed in the radar coordinate system to obtain ΔX and ΔY. Secondly, there is a plane angle deviation between the two coordinate systems, which we set as θ. This angle is generally smaller if the Y-axis of the world coordinate system that we set during the preliminary calibration of the camera is in roughly the same direction as the Y-axis of the radar coordinate system. Finally, there are scale deviations between the two coordinate systems in the X-axis and Y-axis. Since the world coordinate points are estimated according to the prior information of relevant specifications during the camera calibration, they may not necessarily match the radar coordinate system. As a result, we set the scale deviation coefficients K_x and K_y in both directions. In the radar coordinate system, the length of 1 m in the X-direction represents $1 * K_x$ m in the world coordinate system, and the same is true on the Y-axis.

In addition, as mentioned above, we only select four corresponding points, which contain errors in the selection. Therefore, except for the origin, the errors contained in the other three points are also taken into account. Equation (16) is used to correct the world coordinates in the calibration file.

$$\begin{aligned} \begin{bmatrix} K_x & K_y \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_{w\_i} + e_{x\_i} \\ y_{w\_i} + e_{y\_i} \end{bmatrix} \\ + \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} = \begin{bmatrix} x_{wr\_i} \\ y_{wr\_i} \end{bmatrix} \end{aligned} \tag{16}$$

where K_x and K_y denote the scale factors; θ denotes the plane angle; (x_{w_i}, y_{w_i}) denote the world coordinates of the i^{th} point, and (e_{x_i}, e_{y_i}) denote the selection errors; (x_{wr_i}, y_{wr_i}) denote the coordinates of the i^{th} point modified from the world coordinate system to the radar coordinate system.

We corrected the camera time using plus ΔT : $T_{camera} = T_{camera} + \Delta T$. The same vehicle in the video and radar tracking data is selected to intercept the common detection time and calculate the euclidean distance of the vehicle coordinates for each frame. To avoid the influence of outliers on the calculation, the median of the distance during the detection time is used as the distance between the video detection coordinates and the radar detection coordinates. Finally, the average value of all vehicle distances is taken as the target of the objective function (17). The value of each parameter is constantly adjusted until the value of the objective function

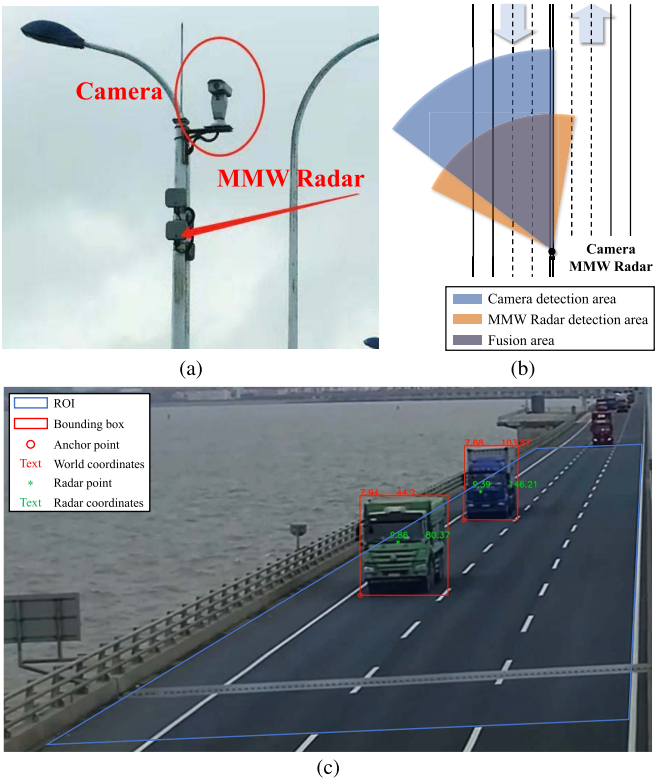

Fig. 9. The scenario of the Donghai Bridge in Shanghai, China: (a) The camera and the MMW radar; (b) The schematic diagram of the scenario; (c) A frame of the video and radar sensing.

achieves the minimum.

$$F = \min\left(mean\left(media \times \left(\sqrt{\left(x_{wr\_i} - x_{r\_i}\right)^2 + \left(y_{wr\_i} - y_{r\_i}\right)^2}\right)_{i=1}^{n}\right)_{j=1}^{k}\right) \quad (17)$$

where (x_{wr_i}, y_{wr_i}) denote the modified world coordinates of the i^{th} vehicle; (x_{r_i}, y_{r_i}) are the coordinates of the i^{th} vehicle detected by radar; n represents the number of frames in the common detection time, and k represents the total number of vehicles.

IV. Case Study

A. Scenario Description and Data Preprocessing

The Donghai Bridge (Shanghai, China) is equipped with more than 200 cameras and MMW radars to support the commercial operation of autonomous trucks for Yangshan Port. Fig. 9 demonstrates that the camera and MMW radar are installed on the same roadside pole so that they have a common detection area. The vehicles drive toward the sensors on the three-lane road from far to near. We evaluated the effectiveness and feasibility of the method using the measured data for 246 seconds.

The camera saves a video at a frame rate of 25 fps. To ensure the accuracy of the vehicle recognition, We used the detection algorithm DetectoRS [9], which has a relatively high mean average precision (mAP) on the COCO dataset, to process the video and obtain the bounding box of each vehicle. For the scenario of this study, we selected the lower left pixel coordinates of each bounding box to represent the object's location. Additionally, we set a series of constraints on the output results to eliminate abnormal or needless detection results, which include the constraints of confidence ($confidence > 0.6$), location ($0\ m < y < 300\ m$), vehicle width ($1\ m < width < 4.5\ m$), and height–width ratio of the bounding box ($0.5 < ratio < 1.5$).

The radar outputs the measured data with a frequency of 20 Hz. During the data preprocessing, we deleted the IDs that appeared less frequently ($Num < 11$). Linear interpolation is performed for the data resampling.

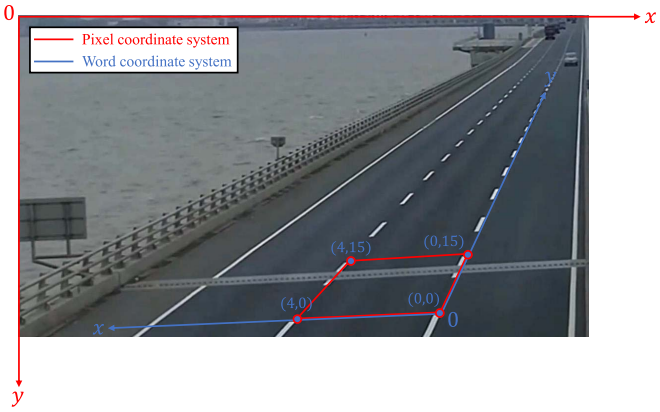

Fig. 10. Preliminary calibration of the camera.

TABLE I
The Calculated Homography Matrix

u	v	U	V	H
0	0	1420	1000	$\begin{bmatrix} 0.04 & 0.02 & -84.49 \\ 0.02 & 0.34 & -360.16 \\ -5.53E-4 & -56.51E-4 & 1 \end{bmatrix}$
4	0	936	1022	
4	15	1120	824	
0	15	1513	802	

B. Preliminary Calibration of Camera

Fig. 10 demonstrates that we captured a frame from the video and selected four lane line corners to locate the world coordinates. Here, we assumed that the length of the lane line was 6 m, the vertical lane distance was 9 m, and the lane width was 4 m. The coordinates of the other three points were estimated using the lower right corner as the origin of the world coordinate system. At the same time, the pixel coordinates of these four corners were also be obtained to calculate the homography matrix.

Table I contains the calculated homography matrix, and Fig. 11 demonstrates the conversion of the object coordinates from pixel coordinate system to world coordinate system. The object coordinates detected by the radar are also contained in Fig. 11b, which illustrates the world coordinate in the same two-dimensional plane as the radar coordinate. However, it was not completely matched because the estimated world coordinate was not consistent with the radar coordinate. Table II contains part of the structured data from the video and

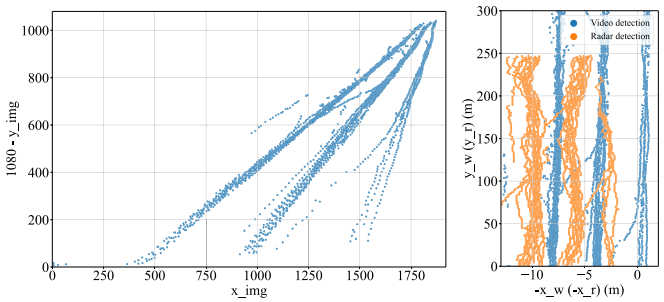

Fig. 11. Coordinate transformation: (a) Pixel coordinates of the objects; (b) World and radar coordinates of the objects.

TABLE II

THE PARTIAL STRUCTURED DATA (ONE VEHICLE) OF VIDEO AND MMW RADAR DETECTION

Frame	Video Dection		Radar Dection	
	X_word (m)	Y_word (m)	X_radar (m)	Y_radar (m)
1	3.00	189.24	5.50	229.38
2	2.99	187.81	5.50	229.38
3	2.99	187.80	5.50	229.38
4	3.03	186.12	5.46	228.90
5	3.03	186.05	5.38	227.94
6	3.08	185.02	5.38	226.82
7	3.08	185.38	5.38	225.76
8	3.04	182.99	5.38	224.79
9	3.04	182.89	5.38	223.73
10	3.07	181.23	5.38	222.98
11	3.08	181.21	5.38	222.98
12	2.98	179.58	5.38	222.33
13	2.97	179.35	5.36	221.39
14	2.94	177.27	5.28	220.24
15	2.94	177.26	5.25	219.23
16	2.99	176.14	5.25	218.32
17	2.99	176.13	5.25	217.34
18	3.03	174.99	5.25	216.33
19	3.04	175.07	5.25	215.27
20	3.01	173.42	5.25	214.36
21	3.01	173.43	5.27	213.44
22	3.05	172.17	5.36	212.47
23	3.06	172.26	5.38	212.22
24	3.04	170.81	5.38	212.04
25	3.04	170.69	5.38	210.96

radar detection of one vehicle, including the coordinates of the object in the world coordinate system and the radar coordinate system, which demonstrates that there is a deviation between the detection results of two sensors for the same object at the same time.

C. Multi-Object Tracking and Trajectory Noise Estimation

Initial tracking was performed on the video and radar data. In the tracking algorithm, the parameters were set as follows: GNN threshold: $N = 3$; clustering threshold: 5; forgetting threshold: 125 frames or times. Kalman filter initial parameters (measurement noise is the same as the model noise): The X- direction localization standard deviation: 0.5 m, and the Y-direction localization standard deviation: 5 m; the X-direction velocity standard deviation: 1 m/s, and the Y-direction velocity standard deviation: 5 m/s.

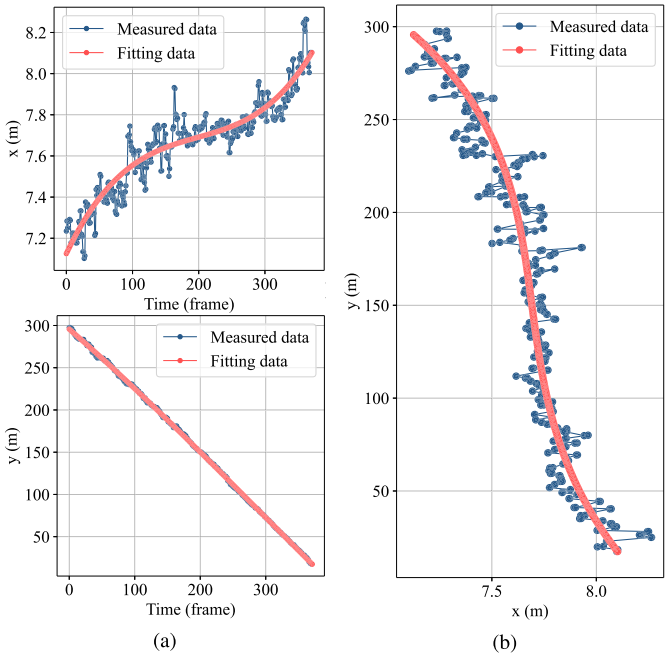

Fig. 12. Trajectory fitting: (a) Trajectory fitting in the X/Y-direction; (b) Trajectory fitting.

TABLE III

NOISE ESTIMATION RESULTS (STANDARD DEVIATION)

	Measurement		Model	
	Video	Radar	Video	Radar
Noise_location_x (m)	0.17	0.28	0.06	0.07
Noise_location_y (m)	6.43	1.39	1.80	0.53
Noise_velocity_x (m/s)	1.37	0.58	1.43	0.06
Noise_velocity_y (m/s)	38.01	1.50	36.29	0.03

After the initial tracking, the vehicle trajectories with better tracking were selected to estimate the noise. Motion decomposition was performed for each trajectory to obtain a time displacement image as illustrated in Fig. 12a, which was fitted using a K-degree polynomial function, and the K was set to 3, and Fig. 12b demonstrates the detected trajectory points and the fitting points. The noise of the trajectory was defined as the residual between the observed and fitted values of the trajectory points, and the noise of the localization and velocity in the X and Y directions was calculated for video and radar measurement, respectively (Fig. 13). The same method was also applied to the predicted value of the model as the predicted noise. Again, the estimated noise does not include the system noise of the sensor. We are more interested in the difference of the intrinsic sensor noise in the two directions than the system noise when comparing the difference of noise in the X and Y directions.

Table III contains the standard deviation of the calculated noises. As demonstrated in Table III, video measurement has lower noise than radar measurement in positioning in the X-direction, while radar measurement has lower noise in positioning in the Y-direction. For velocity measurement, the noise

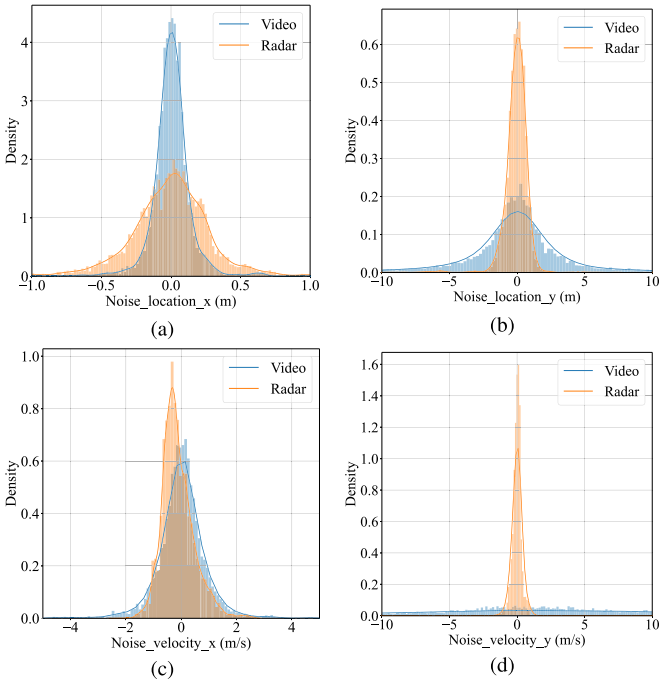

Fig. 13. Measurement noise estimation: (a) Location noise in the X-direction; (b) Location noise in the Y-direction; (c) Velocity noise in the X-direction; (d) Velocity noise in the Y-direction.

between video and radar measurement in the X-direction is relatively close, but the noise for velocity measurement in the Y-direction for the radar is much less than that of video.

We analyzed the reasons for this difference, and they can be summarized as follows: The road is narrower in the lateral direction and very long in the vertical direction. For video measurement, the lateral road has a higher calibration accuracy, and, therefore, video measurement has less noise for object localization in this direction. However, radar's measurement mechanism is more accurate in the radial direction. As a result, the measurement in the road transverse direction contains more noise. Moreover, the radar measures velocity according to the Doppler effect, and the result is much more accurate than the video velocity measurement method based on the differential between position and time.

D. Objective Matching

The virtual detection line was set at the position where the Y-direction value of the radar coordinate system and the world coordinate system are the same. However, the two virtual detection lines were actually at different positions because the two coordinate systems do not match in space. In this paper, the radar coordinate system was taken as the benchmark, so we constantly moved the position of the virtual detection line in the world coordinate system to match the detection data of the two sensors.

We set up 101 virtual detection lines on each of the three lanes (one line was set every 1 m from 10 m to 110 m). In the processing of the video and radar tracking object matching, the vehicle information was set as the timestamp when the vehicle passed the detection line and the time

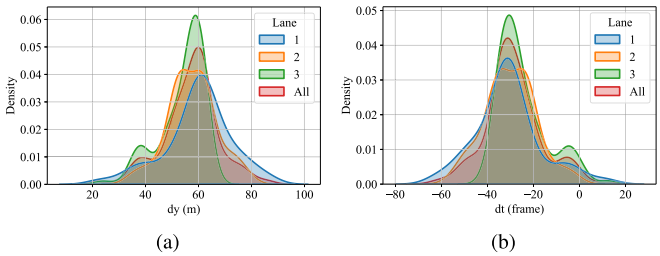

Fig. 14. Distribution of ΔY and ΔT.

TABLE IV
THE MEDIAN OF THE CALCULATED ΔY AND ΔT

Lane	ΔY **(m)**	ΔT **(frame)**
1	61.18	-32
2	57.58	-30
3	56.97	-28
All	58.38	-30

headway between the vehicle and the previous and following three vehicles (N was set as 3). We calculated the distance matrix according to (13) and used the Kuhn–Munkres algorithm [26], [27] for further matching. We constantly adjusted the position of the radar detection line to minimize the objective function (14) value (the constraint condition is set as $\Delta Y \in [0, 100]$), and we then solved ΔY and ΔT.

Fig. 14 demonstrates the statistical distribution of the calculated results for each lane and all lanes. As can be seen in Fig. 14, the distributions of ΔY and ΔT are skewed. If we take the mean value as the final result, it is susceptible to very individual outliers. The median, on the other hand, can be a good remedy for the deficiencies of the mean in skewed distributions. The median is usually used to provide relative respect for the sample's main case statistic when there are a few outliers in the sample. Its algorithm also reflects this feature in that a change in a particular value, especially on the boundary, does not necessarily change the value of that statistic. Therefore, it is more practical to use the median when the distribution is skewed. Table IV records the median of the calculated results. According to the calculation results, ΔY is about 58 m and ΔT is about -30 frames. In other words, the position at 0 m (Y-direction) in the world coordinate system corresponds to the position at 58 m (Y-direction) in the radar coordinate system, and the moment at the 0^{th} frame in the radar time corresponds to the moment at the -30^{th} frame in the video time.

Fig. 15 compares the timestamps before (Fig. 15a) and after (Fig. 15b) matching for a certain lane (the first 3,000 frames). After substituting into the calculated ΔY and ΔT, the vehicles have overlapped on the timestamps.

E. Spatio-Temporal Synchronization Optimization Model

According to (16) and (17), the spatio-temporal synchronization optimization model was built and solved using the constrained nonlinear minimization solver. All parameters

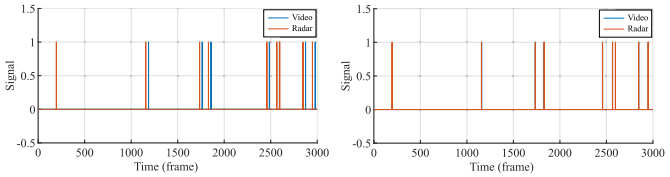

Fig. 15. Timestamps before and after synchronization: (a) Before synchronization; (b) After the synchronization.

TABLE V
PARAMETER CONSTRAINTS AND CALCULATION RESULTS OF SPATIO-TEMPORAL OPTIMIZATION MODEL

Parameter	Lower boundary	Upper boundary	Result
ΔT (frame)	-35	-25	-33
e_{1_x} (m)	-0.3	0.3	0.19
e_{1_y} (m)	-0.3	0.3	0.09
e_{2_x} (m)	-0.3	0.3	0.25
e_{2_y} (m)	-0.3	0.3	0.26
e_{3_x} (m)	-0.3	0.3	0.09
e_{3_y} (m)	-0.3	0.3	0.22
θ (°)	-1	1	-0.20
K_x	0.5	1.5	0.83
K_y	0.5	1.5	1.06
ΔX (m)	-5	5	2.50
ΔY (m)	55	65	60.53

were normalized before the solution was determined. To avoid falling into a local optimum solution due to the initial values in a single solution, we randomly generated 100 sets of initial values, solved them simultaneously, and compared their final objective function values to select the better one as the results (Table V).

As demonstrated in Fig. 16, we selected one of the trajectories with obvious features (a vehicle executed a lane change) and object matching trajectory points before the synchronization, after the temporal synchronization, and after the spatio-temporal synchronization. The dashed line in Fig. 16 connects the trajectory points in the same frame. Fig. 16a highlights that the video and radar have a significant temporal mismatch, and Fig. 16b demonstrates the object matching after the temporal synchronization. Fig. 16c is further corrected spatially.

To illustrate the effectiveness of temporal synchronization, we calculated the residuals of time-velocity curves in the X-direction of the vehicle before and after temporal synchronization. The reason why the time-velocity curve in the Y-direction was not calculated is that the main motion state of the vehicle in the Y-direction is moving at a constant velocity, without obvious change characteristics. Here, we took the trajectory of the vehicle in Fig. 17 as an example, which has significant lane change characteristics in the X-direction. We calculated the velocity of the vehicle trajectory detected by video detection and radar detection in the X-direction respectively, drew the time-velocity curve and calculated their residuals. The more accurate the temporal synchronization,

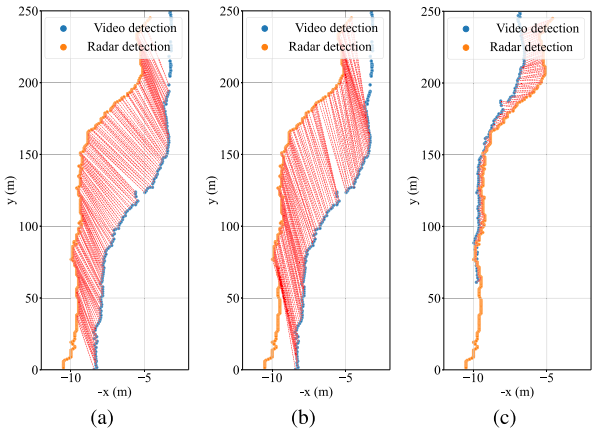

Fig. 16. The trajectory before and after the synchronization: (a) Before synchronization; (b) After temporal synchronization; (c) After spatio-temporal synchronization.

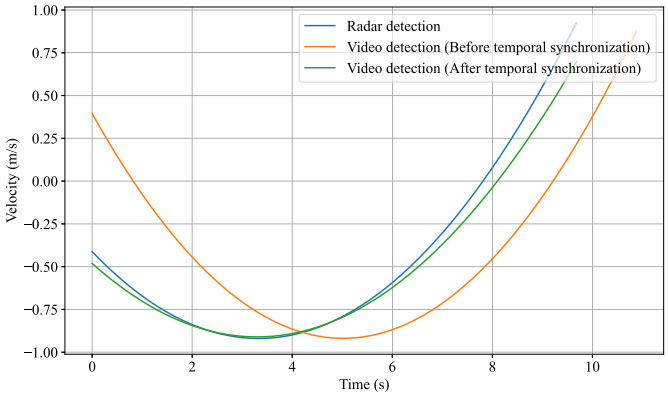

Fig. 17. The time-velocity cure before and after temporal synchronization of the trajectory in Fig. 16.

the smaller the residuals. The figure below shows the time-velocity curve of the video and radar detection trajectory in the X-direction before and after temporal synchronization. It should be noted that the velocity calculation here is based on the fitted values of this vehicle trajectory to obtain smoother velocity values. The average value of residual before temporal synchronization is 0.3836 m/s, and it decreases to 0.0541 m/s after the temporal synchronization. The residual is reduced by 85.90%, which can fully reflect the effectiveness of temporal synchronization.

Fig. 18 depicts all the vehicle trajectories detected by video and radar before and after spatio-temporal synchronization. Fig. 19 illustrates the spatial distribution of the deviation (absolute value) between the trajectories detected by video and radar before and after the spatio-temporal synchronization. Each line in Fig. 18 and Fig. 19 represents the trajectory of one vehicle. The spatial deviation between the two types of trajectories is greatly reduced after the synchronization. The average deviation in the X-direction was reduced from 2.47 m to 0.42 m, and from 64.06 m to 2.34 m in the Y-direction. For the remaining deviation, the following three errors are the main factors: 1) radar detection error; 2) video detection error; 3) the remaining deviation between the calculation results of

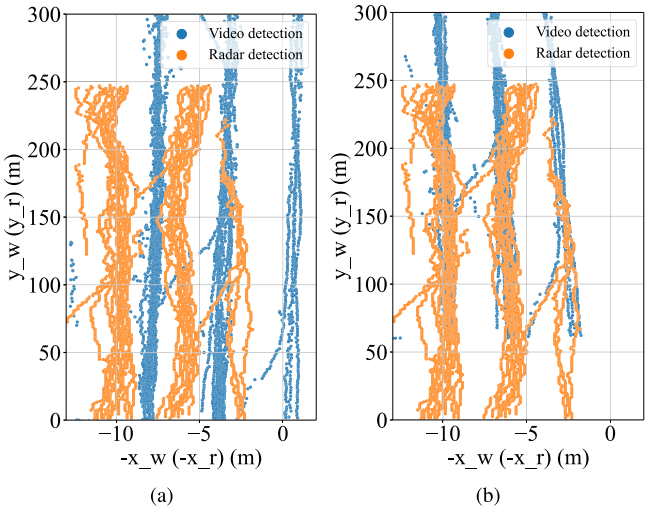

Fig. 18. World and radar coordinates of the objects before and after the synchronization: (a) Before synchronization; (b) After synchronization.

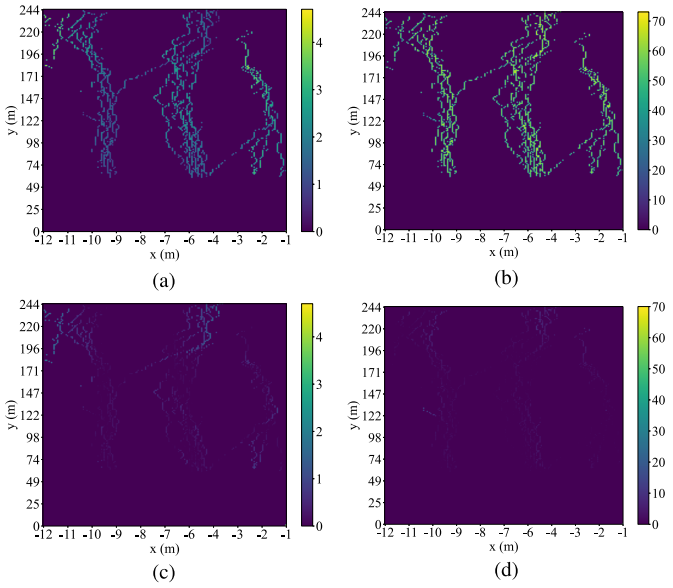

Fig. 19. Deviation in space before and after the synchronization: (a) X-direction (before synchronization); (b) Y-direction (before synchronization); (c) X-direction (after synchronization); (d) Y-direction (after synchronization).

the optimization model and the true value, that is, a systematic deviation in time and space between the radar coordinate system and the world coordinate system. The method proposed in this paper aims to minimize the deviation between the calculation results of the optimization model and the true value.

Fig. 18 and Fig. 19 demonstrate that the video trajectories within the range of 50–170 m in the Y-direction basically match those detected by radar, with a small deviation. However, the deviation between them is not large on the Y-axis beyond 170 m, while there is still a certain distance on the X-axis. As can be seen from Fig. 18, both the radar and video are skewed to the left or right relative to the video detection at a distance of more than 170 m in the Y-direction. We believe that the large deviation in the X-direction may be caused by the left and right drift of the radar for the location of the newly appeared object in the distance. As the object gradually approaches the radar, the radar detection of localization tends to be stable.

V. Conclusion

This study proposes a novel spatio-temporal synchronization method of roadside MMW radar–camera for sensor fusion. Lane line corner points are applied to pre-calibrate roadside camera in the scenario, and a multi-object tracking algorithm is utilized to obtain passing vehicles' trajectories. Then, multiple virtual detection lines are set up to match the sensing data of successive vehicles from roadside MMW radar and camera, and a 12-parameter optimization model is formulated to solve the spatio-temporal synchronization problem. Finally, it is demonstrated that the proposed method is effective and economical based on the experiments in Donghai Bridge, Shanghai, China. The experiments also found that the sensing data of roadside camera has less positioning noise in X-direction, while MMW radar is more suitable for Y-direction positioning and velocity measurement, which sheds new light on roadside sensor fusion.

References

[1] I. Rasheed, F. Hu, Y.-K. Hong, and B. Balasubramanian, "Intelligent vehicle network routing with adaptive 3D beam alignment for mmWave 5G-based V2X communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2706–2718, May 2020.

[2] Y. Zhang, G. Zhang, R. Fierro, and Y. Yang, "Force-driven traffic simulation for a future connected autonomous vehicle-enabled smart transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2221–2233, Jul. 2018.

[3] E. Arnold, M. Dianati, R. de Temple, and S. Fallah, "Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors," *IEEE Trans. Intell. Transp. Syst.*, early access, Oct. 16, 2020, doi: 10.1109/TITS.2020.3028424.

[4] C. Zhao, F. Liao, X. Li, and Y. Du, "Macroscopic modeling and dynamic control of on-street cruising-for-parking of autonomous vehicles in a multi-region urban road network," *Transp. Res. C, Emerg. Technol.*, vol. 128, Jul. 2021, Art. no. 103176.

[5] C. Liu, D. Wu, C. Zhao, Y. Du, and Y. Li, "Concept and framework of the new generation of smart highway," in *Proc. Transp. Res. Board 100th Annu. Meeting*, 2021, Paper TRBAM-21-01257.

[6] A. Bochkovskiy, C. Y. Wang, and H. Liao, "YOLOV4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*. [Online]. Available: https://arxiv.org/abs/2004.10934

[7] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, PMLR, 2019, pp. 6105–6114.

[8] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[9] S. Qiao, L.-C. Chen, and A. Yuille, "DetectoRS: Detecting objects with recursive feature pyramid and switchable atrous convolution," 2020, *arXiv:2006.02334*. [Online]. Available: http://arxiv.org/abs/2006.02334

[10] H. Li, P. Wang, and C. Shen, "Toward end-to-end car license plate detection and recognition with deep neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1126–1136, Mar. 2018.

[11] J. Ma, Z. Tian, Y. Li, and M. Cen, "Vehicle tracking method in polar coordinate system based on radar and monocular camera," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Aug. 2020, pp. 93–98.

[12] X. Wang, L. Xu, H. Sun, J. Xin, and N. Zheng, "On-road vehicle detection and tracking using MMW radar and monovision fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 2075–2084, Jul. 2016.

[13] R. O. Chavez-Garcia and O. Aycard, "Multiple sensor fusion and classification for moving object detection and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 525–534, Feb. 2016.

[14] J. Oh, K.-S. Kim, M. Park, and S. Kim, "A comparative study on camera-radar calibration methods," in *Proc. 15th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Nov. 2018, pp. 1057–1062.

[15] M. Dubská, A. Herout, R. Juránek, and J. Sochor, "Fully automatic roadside camera calibration for traffic surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1162–1171, Jun. 2015.
[16] T. N. Schoepflin and D. J. Dailey, "Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 2, pp. 90–98, Jun. 2003.
[17] Y. Fu, D. Tian, X. Duan, J. Zhou, and X. You, "A camera–radar fusion method based on edge computing," in *Proc. IEEE Int. Conf. Edge Comput. (EDGE)*, Oct. 2020, pp. 9–14.
[18] F. Liu, J. Sparbert, and C. Stiller, "IMMPDA vehicle tracking system using asynchronous sensor fusion of radar and vision," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2008, pp. 168–173.
[19] R. Zhang and S. Cao, "Extending reliability of mmwave radar tracking and detection via fusion with camera," *IEEE Access*, vol. 7, pp. 137065–137079, 2019.
[20] M. Wu and B. Coifman, "Quantifying what goes unseen in instrumented and autonomous vehicle perception sensor data–A case study," *Transp. Res. C, Emerg. Technol.*, vol. 107, pp. 105–119, Oct. 2019.
[21] R. I. of Highway Ministry of Transport, *Specification for Layout of Highway Traffic Signs and Markings*, Standard (JTG D82-2009), 2009.
[22] C. F. H. C. CO LTD, *Design Specification for Highway Alignment*, Standard (JTG D20-2017), 2017.
[23] D. Y. Kim and M. Jeon, "Data fusion of radar and image measurements for multi-object tracking via Kalman filtering," *Inf. Sci.*, vol. 278, pp. 641–652, Sep. 2014.
[24] Y. Du, C. Zhao, F. Li, and X. Yang, "An open data platform for traffic parameters measurement via multirotor unmanned aerial vehicles video," *J. Adv. Transp.*, vol. 2017, Jan. 2017, Art. no. 8324301.
[25] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.
[26] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, Mar. 1957.
[27] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistics*, vol. 52, no. 1, pp. 7–21, 2010.
[28] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.

Yuchuan Du (Member, IEEE) received the B.S. and M.S. degrees in road engineering and the Ph.D. degree in traffic engineering from Tongji University, Shanghai, China, in 1998, 2001, and 2004, respectively.

From 2003 to 2006, he was an Assistant Professor with the College of Transportation Engineering, Tongji University, where he was an Associate Professor from 2006 to 2010. He is currently a Professor with the College of Transportation Engineering, Tongji University. His research interests include innovative technology for smart transportation infrastructure and intelligent transportation systems.

Bohao Qin received the bachelor's degree in transportation engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2019. He is currently pursuing the master's degree with the College of Transportation Engineering, Tongji University.

His research interests include the traffic sensing technologies and camera-radar data fusion.

Cong Zhao received the B.S., M.S., and Ph.D. degrees in transportation engineering from Tongji University, Shanghai, China, in 2014, 2017, and 2020, respectively.

He was a Visiting Student Researcher with California PATH, University of California, Berkeley, USA, from 2018 to 2019. He is currently an Associate Research Professor with the College of Transportation Engineering, Tongji University. His research interests include intelligent transportation systems, connected and automated vehicles, infrastructure enabled automation, and machine learning.

Yifan Zhu received the bachelor's degree in transportation engineering from Tongji University, Shanghai, China, in 2019, where he is currently pursuing the master's degree with the College of Transportation Engineering.

His research interests include trajectory planning of autonomous vehicle and traffic sensing technologies.

Jing Cao received the B.S., M.S., and Ph.D. degrees in transportation engineering from Chang'an University, Xi'an, Shaanxi, China, in 2009, 2012, and 2016, respectively.

She was a Visiting Scholar with the University of Alberta from 2012 to 2014. She is currently an Associate Research Professor with the College of Transportation Engineering, Tongji University. Her research interests include intelligent transportation systems and traffic operation and control.

Yuxiong Ji received the B.S. and M.S. degrees in transport engineering from Tongji University in 2001 and 2004, respectively, and the Ph.D. degree in civil engineering from The Ohio State University, USA, in 2011.

He is currently a Professor with the College of Transportation Engineering, Tongji University. His research interests include data mining, smart transit, and traffic operation and control.