Vehicle queuing at complex intersection scenarios via mmWave radar-camera fusion

Haiyang Zhu [1, a], Yuejuan Wang [1, b], Leilei Wang [1, c]

[1] Shanghai Limradar Electronic Technology Co., Ltd, Shanghai, 200090 China.

[a] zhuhaiyang@limradar.com, [b] wangyuejuan@limradar.com, [c] wangleilei@limradar.com

Abstract. The queuing effect of vehicles perceived by road-side sensing equipment is very important which will directly affect the control quality of the signal. We proposed a vehicle queuing algorithm via mmWave radar-camera fusion to optimize queuing in complex intersection scenarios. Three types of stopped state was introduced into the tracker, which are FirstStop, VisionStop and LostStop. Shaded area caused by the vehicles was calculated and the lost target in the shaded area would be predicted along the lane. The shaking phenomenon of the stopped vehicle would be eliminated, the queue vacancies would be alleviated, and the timely location update would be realized when the stopped vehicle restarts. Through extensive evaluations of real-world datasets, the proposed method demonstrates remarkable improvements in the accuracy of queue length, which is more than 90%, and the ID switch of the stopped target is reduced by about 3 times/vehicle.

Keywords: traffic monitoring; mmWave radar-camera fusion; queuing algorithm; shaded area.

1. Introduction

Vehicle-Road-Cloud Integration System (VRCIS) can not only promote the development of autonomous driving to be completely unmanned by making up for the lack of individual vehicle intelligence, but also build efficient and safe transportation systems when integrating with smart cities[1]. Compared with the vehicle-end perception, longer range monitoring and more continuous trajectory tracking are required in road-end perception. Therefore, the combination of mmWave radar and camera becomes the optimal choice for performance, sensing range and cost. However, the split solution consisting of mmWave radar, short-focus camera, long-focus camera and MEC (Multi-access Edge Computing) has so many problems, such as high price, poor real-time, difficult deployment, and high maintenance cost, which is difficult to meet the need of large-scale deployment. Radar-camera integrated device integrating the above devices gradually replaces the split solution because of it’s small size, full functions, good real-time, and low cost.

The road-end perception needs to solve the accuracy, stability and reliability of multi-object detection and tracking in all-day, all-weather, all-scene and large-range. Radar-Camera fusion can be summarized as three types of fusion method: date-level fusion[2,3], feature-level fusion[4,5] and decision-level fusion[6,7]. Compared with data-level fusion and feature-level fusion, decision-level fusion has the following advantages:

(1) Less affected by spatial and temporal alignment. Alignment errors can be corrected at the target level.

(2) The system design is relatively simple. The multi-target tracking framework based on Kalman filter is commonly used.

(3) Less dependence on upstream modules. The post-fusion module only needs to get the output of the detection target from the upstream sensing module, and does not care about the specific structure of the sensing module. If the sensor or perception algorithm is replaced later, it will have little impact on the fusion module. Due to the rapid development of image detection technology, its network model is constantly iterated, and the use of decision-level fusion can use the latest technology in a longer term.

Therefore, most of the landing projects and products adopt the decision-level fusion. In complex traffic scenarios, such as the queuing scene at the entrance of the intersection, the targets are often lost or false targets are generated, because the detection values are unstable, and the multi-target tracking algorithm can not meet the tracking of all moving targets. Therefore, a vehicle queuing algorithm via decision-level fusion is proposed to solve the problem of ID switch and queuing accuracy.

2. Problem formulation

Radar-camera integrated device is composed of a mmWave radar, a short-focus camera and a long-focus camera, which is installed on the traffic light pole at the intersection to monitor the vehicle in entrance, as shown in Fig.1. The blue beam is the coverage of the mmWave radar, the red area is the coverage of the short-focus camera, and the yellow area is the coverage of the long-focus camera.

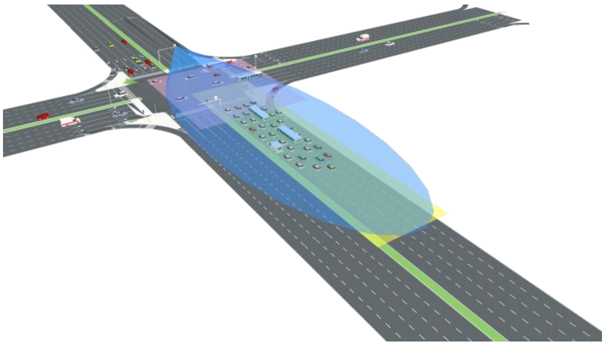

Fig. 1 The coverage of mmWave radar, short-focus camera and long-focus camera

When decision-level fusion is applied, the radar tracking data and the camera tracking data are projected into the same coordinate system, which can be the local Cartesian (LC) coordinate system of radar or the World Geodetic System (WGS-84) [8,9,10]. Radar-camera fusion can not only enrich the structured data of the trajectory, such as license plate, vehicle type, body color, etc. , but also make up for the instability of detection of a single sensor, such as small targets far away are unstably detected by camera, stationary and cross targets are unstably detected by radar, so that the trajectory will not be interrupted, as shown in Fig.2(a).

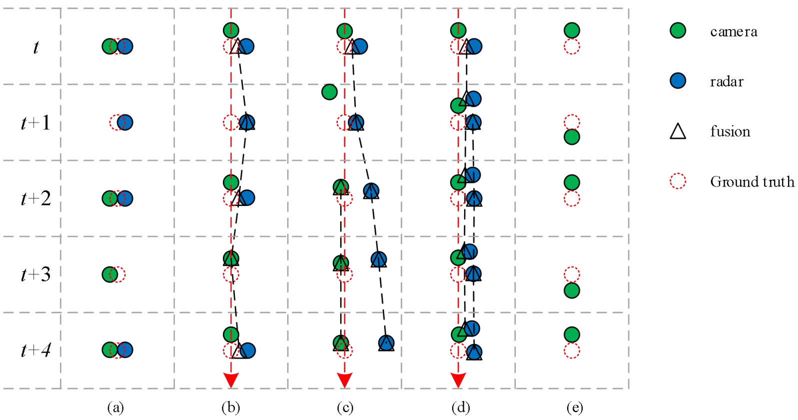

Fig. 2 Data feature of radar-camera decision-level fusion

Fig.2(b) shows the impact of data loss on the trajectory. It may be that the lost frame of original data, lost frame caused by temporal alignment, or one of sensors can not detect this target, resulting side-to-side shake of the fusion data. Fig.2(c) shows the influence of radar date uncertainly on trajectory, once the radar data deviates too much, the fusion data will split, resulting in the trajectory splitting. Fig.2(d) shows the impact of radar data splitting on trajectory. Especially for large vehicles such as buses and trucks, radar data often has multiple scattering points. Once the camera data is merged with the radar data from the vehicle body, the trajectory will be inaccurate and will easily split. Fig.2(e) shows the up-and-down shake in the detection of the stopped target by camera. Since radar has no detection ability to the stopped target, the tracking can only rely on the camera data. As a result, it is difficult to filter its speed to 0 km/h by traditional tracker, which causes the speed of target to bounce around 0 km/h repeatedly, and also causes the trajectory to shake. The above is still a problem in the case that at least one sensor can detect the target. In the scene of the intersection entrance, there are a lot of shaded problems, resulting in that target disappears in the two sensors, which brings a huge challenge to queuing algorithm.

3. Method

The LC coordinate system of radar is used as the base BEV (Bird's Eye View) coordinate system. Inverse Projective Mapping (IPM) is employed to project the bottom center of the bounding box from camera to the radar BEV coordinate system, as shown in the green and red dots in Fig. 3.

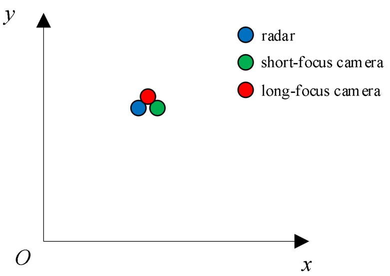

Fig. 3 BEV coordinate system

The fusion data formed after the clustering of the detection data of various sensors is input as the measurement value of the multi-object tracker, and the tracking trajectory is obtained by Kalman filter algorithm. When the target is not shaded, the radar data disappears near or after the stop, and the multi-object tracker can only update the state according to the camera data. As mentioned in the previous section, the up-and-down shake of the camera detection will cause the shake of the stop target. In addition, if the target is shaded, the radar and camera data of the target will all disappear, resulting in the loss of the target and a vacancy in the queue.

Therefore, we introduce the trajectory stop state to deal with the above problems, and the tracking state can be divided into TrackState=(New, Tracked, Lost, Stopped, Removed), and the update of tracking state is shown in Fig.4.

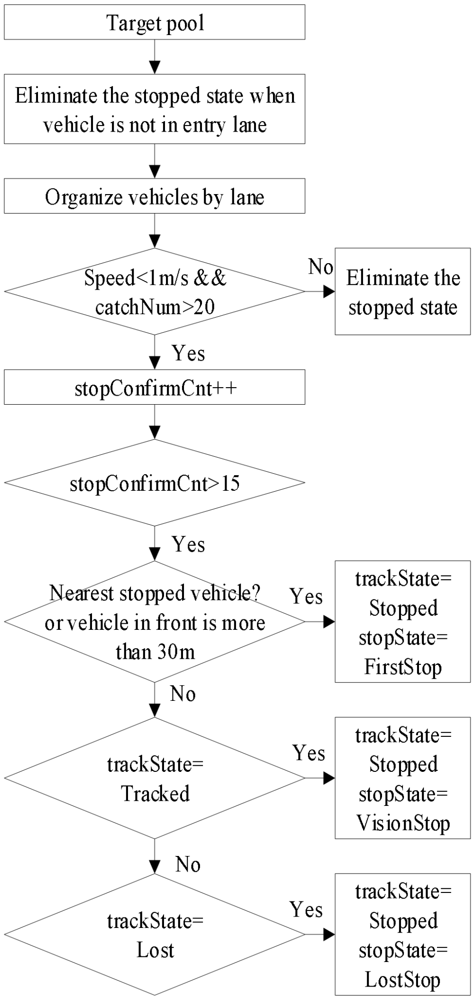

Fig. 4 The update of tracking state

The target is considered to be stopped if it is stably tracking in the entrance and the speed is less than 1m/s in a continuous period. According to the position of the vehicle in the lane and the tracking state, it is labeled as FirstStop, VisionStop and LostStop respectively. The tracking state will be updated every period. For the first stopped target, it can be stably detected because there is no occlusion relationship. Therefore, for FirstStop targets, the location can be updated by the tracker, while for VisionStop and LostStop targets, the location cannot be updated by the tracker due to the serious occlusion in the queue, which cannot guarantee stable detection or inaccurate detection values.

For the lost track, the multi-target tracker can not always predict, and it needs to be deleted after a certain time. However, in the environment of the entrance of the intersection, there are a lot of shaded situation between vehicles, resulting in the loss of the target. Fig. 5 shows the shaded area of the vehicle. Once a target enters the red area in the figure, neither radar nor the camera can detect it. Therefore, the lost target entering the shaded area need to be specially processed. The lost number in the tracker is no longer added, and keep predicting along the lane until it occupies the vehicle in front and becomes LostStop.

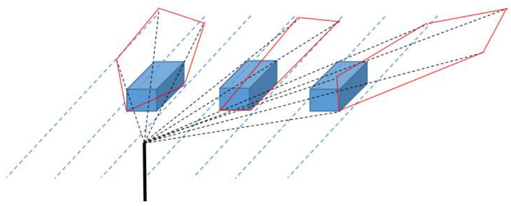

Fig. 5 Shaded area of the vehicle

4. Experimental results

We evaluated the performance of the proposed queuing algorithm through radar-camera integrated device which has a mmWave radar, a short-focus camera, a long-focus camera and two groups of fill light. The radar-camera integrated device is shown in Fig. 6.

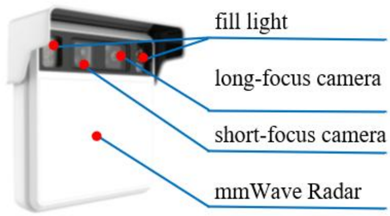

Fig. 6 Radar-camera integrated device

The data acquisition of the intersection was tracked and queued, as shown in Fig.7. The black box in the figure means the target is Tracked; Smaller black boxes indicate pedestrians or non-motor vehicles; The yellow box indicates that the target is Lost and in the shaded area; The gray box area is the shaded area, and only the shaded area formed by the large vehicle is drawn in the simulation. Red, magenta, and green boxes indicate that the target is in the Stopped state and represent FirstStop, VisionStop, and LostStop, respectively.

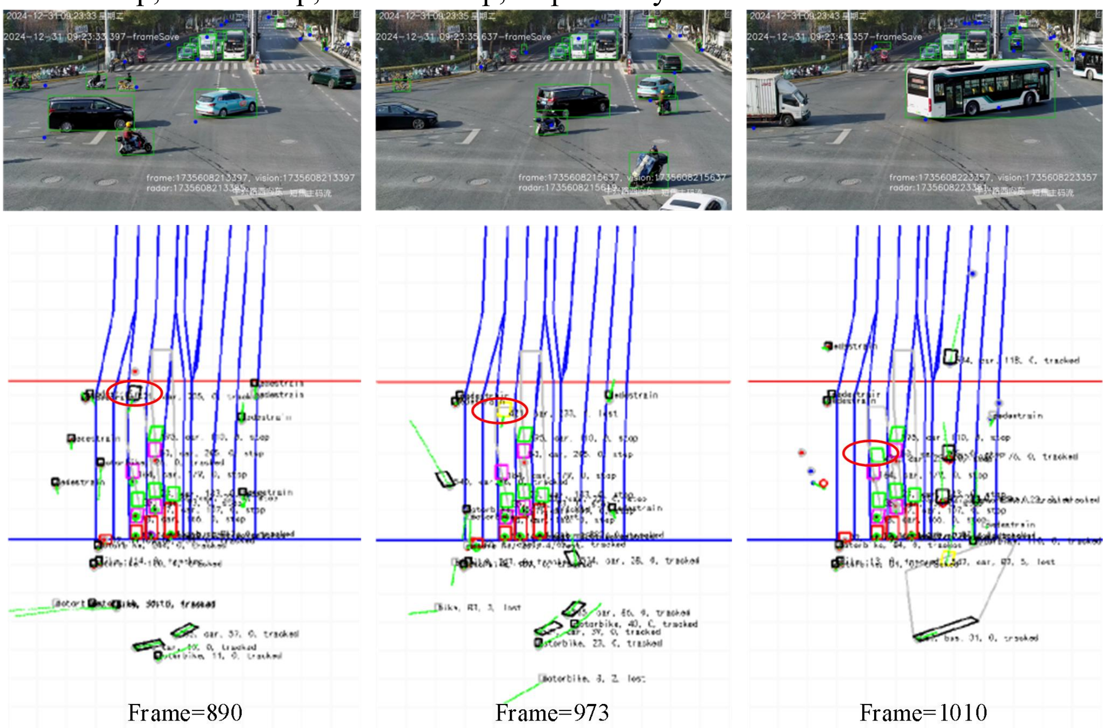

Fig. 7 Queuing algorithm simulation

The target marked by the red ellipse in Figure 7, from the normal Tracked at 890 frame to the Lost at 973 frame into the shaded area, and then to the Stopped at 1010 frame, completely shows the algorithm's process of the lost target in the shaded area. Although the target is completely invisible in radar and camera, it can be queued normally after processing by the queuing algorithm. As can be seen from the third column in Fig. 7, the overall queuing performance is completely consistent with the actual scene.

Compare the performance with and without the queuing algorithm, as shown in Fig. 8. At the 1010 frame, there is no obvious missing target in the queuing performance. At the 1010 frame, the processing with the queuing algorithm assigns an ID of 576 to the new target, while the processing without the queuing algorithm assigns an ID of 614 to the new target. Since the queuing algorithm is mainly aimed at the target processing in the entrance, the 38 additional ID switches are generated by the queuing target as shown in Fig. 8, and the number of queuing targets is 13, then the ID switch of the processing with queuing algorithm is about 3 times/vehicle less than that of the processing without queuing algorithm.

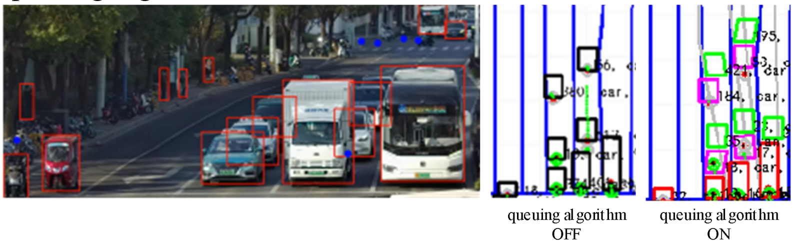

Fig. 8 Simulation with and without queuing algorithm

The queuing algorithm is deployed in the equipment that runs normally at the intersection, and the queuing performance is shown in Fig.9. The accuracy of the queue length is more than 90%, and there is no vacancy in the queue.

Fig. 9 Actual queuing performance

5. Conclusion

This paper focuses on the queuing optimization problem of mmWave radar-camera fusion in the entrance area of the intersection. Based on the analysis of the data feature of radar and camera and the impact of the fusion measurement on the tracker, a queuing algorithm is proposed. By introducing the Stopped state into the tracker, the stopped vehicle is subdivided into FirstStop, VisionStop and LostStop, and different update strategies are adopted according to the different stopped states, which can not only eliminate the shake of the stopped target, but also update the position in time when the vehicle restarts. At the same time, the ID switch of queuing target is greatly reduced. In view of the large number of shaded problems at the intersection, the lost targets in the shaded area are predicted along the lane to reduce the problem of queue vacancies. Future work will focus on smoothness, accuracy, reliability of the tracker itself with the cluttered radar and camera detection data.

References

[1] Q. Xu, K. Li, J. Wang, Q. Yuan, Y. Yang, and W. Chu, “The status, challenges, and trends: An interpretation of technology roadmap of intelligent and connected vehicles in China (2020),” J. Intell. Connected Vehicles, vol. 5, no. 1, pp. 1–7, Feb. 2022.

[2] U. Kadow, G. Schneider, and A. Vukotich, “Radar-vision based vehicle recognition with evolutionary optimized and boosted features,” in Proc. IEEE Intell. Vehicles Symp., Istanbul, Turkey, Jun. 2007, pp. 749–754.

[3] R. Nabati and H. Qi, “RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles,” May 15, 2019. doi: 10.1109/ICIP.2019.8803392.

[4] F. Nobis, M. Geisslinger, M. Weber, J. Betz, and M. Lienkamp, “A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection,” May 15, 2020, arXiv: arXiv:2005.07431. Accessed: Oct. 29, 2024. [Online]. Available: http://arxiv.org/abs/2005.07431.

[5] S. Chang et al., “Spatial Attention Fusion for Obstacle Detection Using MmWave Radar and Vision Sensor,” Sensors, vol. 20, no. 4, p. 956, Feb. 2020, doi: 10.3390/s20040956.

[6] A. Sengupta, L. Cheng, and S. Cao, “Robust Multiobject Tracking Using Mmwave Radar-Camera Sensor Fusion,” IEEE Sens. Lett., vol. 6, no. 10, pp. 1–4, Oct. 2022, doi: 10.1109/LSENS.2022.3213529.

[7] L. Cheng, A. Sengupta, and S. Cao, “Deep Learning-Based Robust Multi-Object Tracking via Fusion of mmWave Radar and Camera Sensors,” IEEE Trans. Intell. Transp. Syst., pp. 1–16, 2024, doi: 10.1109/TITS.2024.3421339.

[8] Y. Li and W. Zhang, “Traffic flow digital twin generation for highway scenario based on radar-camera paired fusion,” Sci. Rep., vol. 13, no. 1, p. 642, Jan. 2023, doi: 10.1038/s41598-023-27696-z.

[9] Z. Tao, Y. Li, P. Wang, and L. Ji, “Traffic Incident Detection Based on mmWave Radar and Improvement Using Fusion with Camera,” J. Adv. Transp., vol. 2022, pp. 1–15, Apr. 2022, doi: 10.1155/2022/2286147.

[10] C. Jin, B. Zhu, J. Deng, Z. Hu, and X. Wen, “Automatic calibration and association for roadside radar and camera based on fluctuating traffic volume,” Meas. Sci. Technol., vol. 35, no. 5, p. 055118, May 2024, doi: 10.1088/1361-6501/ad29e6.