A Roadside Camera-Radar Sensing Fusion System for Intelligent Transportation

Lefei Wang[1], Zhaoyu Zhang[2], Xin Di[3], Jun Tian[4]
Fujitsu Research and Development Center Co., Ltd, China
{[1]wanglefei, [2]zhangzhaoyu, [3]dixhappy, [4]tianjun}@cn.fujitsu.com

***Abstract* —Cooperative Vehicle Infrastructure System (CVIS) is one of the most valued technologies in intelligent transportation system. By use of roadside sensing, it provides extended coverage and more dimensions traffic information compared with independent perception by vehicle sensors. Sensing fusion of camera and radar can overcome the shortcomings of both sensors so that it is the trend of roadside sensing in CVIS. In this paper, we propose a novel roadside sensing fusion system. It filters out background objects from radar detection to avoid wrong calibration and fusion with camera detection, and makes calibration of camera and radar automatically to reduce time cost of system implementation. Experiment results show that it can be fast implemented to automatically acquire accurate roadside sensing information.**

***Keywords* —intelligent transportation system, CVIS, sensing fusion.**

I. Introduction

Intelligent Transportation System (ITS) is a comprehensive transportation management and service system, which combines various advanced technologies, such as sensors, communication, information systems, controllers, and advanced mathematical methods with the conventional world of transportation infrastructure [1].

Cooperative Vehicle Infrastructure System (CVIS) is a subsystem of ITS, which acquires vehicle and road information by use of roadside sensing and wireless communication technologies [2]. Compared with autonomous vehicle that rely solely on their own perception systems [3], CVIS with roadside sensing provides additional information with extended coverage and more dimensions [4]. Camera-based perception is widely used in roadside sensing, but it is prone to being influenced by light, rain, fog and other environmental factors [5]. On the contrary, radar-based perception is not sensitive to these environmental factors, while due to lower imaging resolution than camera, the objects detected by radar is hard to be classified. Consequently, to overcome the shortcomings of both sensors by one another sensor's complementary advantage, sensing fusion of camera and radar is more suitable for roadside sensing in CVIS compared with camera-only sensing [6].

In roadside sensing fusion system [7], two issues need to be considered. Firstly, besides vehicles and pedestrians, some strong reflectors on road with slight movement, such as metal street signs, can be detected by radar as well. In CVIS, they are static objects rather than dynamic objects which are necessary to be detected in roadside sensing. In camera detection, these objects are not detected because they are not trained as target identified objects. These background objects detected by radar can lead to wrong sensing fusion since the number of objects detected by radar and camera cannot be matched. Secondly, before sensing fusion, camera and radar detect objects independently and simultaneously, and the detection results of two sensors must be calibrated in the same coordinate system to make fusion. For calibration, a sample set including several matched objects pairs from camera and radar detection results must be prepared. The sample set is usually made by manual so that it is very time-cost and leads to low efficiency and weak robustness to roadside fusion sensing system implementation [8]. To solve the issues, a novel roadside sensing fusion system is presented in this paper. In this system, we present a radar processing algorithm to filter out the background objects while still maintain dynamic objects, and another algorithm we called auto-calibration to make the sample set for calibration automatically.

This paper is organized as follows. A novel roadside sensing fusion system is proposed In Section II. In Section III, two proposed modules of the system, background objects filtering and auto-calibration, are presented. In Section IV, series of experiments evaluate the proposed roadside sensing fusion system. Finally, conclusion is drawn in Section V.

II. Page Limit and Page Layout

Structure of the proposed roadside camera-radar sensing fusion system is shown in Fig. 1. In the system, camera and radar detect objects independently and simultaneously. Then they are calibrated and fused with each other.

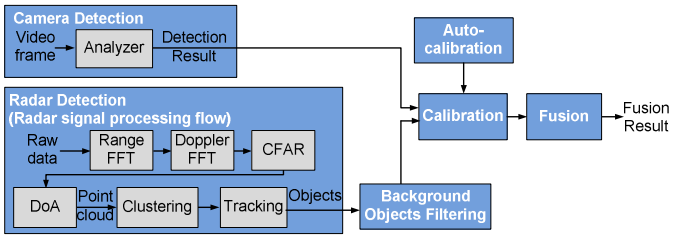


Fig. 1. Structure of proposed roadside camera-radar sensing fusion system.

In the camera detection, video frame is input into an analyser periodically. The analyser is a classifier trained by AI

2020 17th European Radar Conference (EuRAD) | 978-2-87487-061-3/21/$31.00 ©2021IEEE | DOI: 10.23919/EuRAD48048.2021.00079

algorithm based on amount of training data. After analysing, detection result is output as video frame as well. The result mainly includes bounding box, object type and property, etc.

In the radar detection, a radar signal processing flow is implemented. The raw data is obtained from various front end channels, and input into the 1D (range) FFT sub-module. Then, 2D (velocity) FFT processing is performed to give a (range, velocity) matrix, which also includes the CFAR detection in Doppler direction. After that, direction of arrival (azimuth) estimation is to map the X-Y location of object, and the output is point cloud. Based on the point cloud, clustering algorithm, such as dBscan algorithm, is used for obtaining the mean location of object. Then, tracking algorithm, such as Extended Kalman Filte (EKF), is implemented and outputs detected objects which are formed as radar frame periodically. A frame may include several objects described with X-Y location and velocity. The radar detection results may include some strong reflectors with slight movement, e.g., metal street signs. These objects are not needed in calibration and fusion, and filtered out by the background objects filtering module.

The camera detection result and the radar filtering result are input into the calibration module. This module is to find a relationship between radar coordinate system and video coordinate system. The relationship is actually a projection function. Before calibration, auto-calibration is implemented to automatically find a large amount of radar and video object pairs as the sample set for calibration. With the calibration results, each radar detected object is fused with one of camera detected objects. As a result, position and velocity of the object detected by radar can be matched to its corresponding bounding box marked by camera detection.

III. Two Proposals of Roadside Sensing Fusion System

A. Proposal of Background Objects Filtering

Motion feature of objects can be analysed according to continuous radar detection results. We propose some principles based on motion feature:

- Velocity: if velocity magnitude of an object below a threshold, the object should be filtered out.
- Trace: if an object is filtered previously, it should be filtered in the following frames.
- Tracking: if an object moves towards to the roadside radar or moves away from the roadside radar continuously, this object cannot be filtered.

According to these principles, the flow of background objects filtering is described as in Fig. 2. The input is the radar detection results with radar frames. The first step is to determine whether two objects in two frames are the same object, which depends on a distance threshold (th_d). Assume that the maximum velocity of a car is v km/h (only consider the velocity magnitude), frame period is Δt ms, and n is frame number, then th_d can be calculated as $th_d = n \cdot v \cdot \Delta t/3600$, which indicates the displacement of an object during $n \cdot \Delta$t ms is not greater than th_d meters. If the displacement is greater than th_d, there must be two different objects. The Euclidean distance between the objects in the $(i\text{-}n)_{\text{th}}$ frame and the i_{th} frame is denoted as $d_{k,j}^{i-n,i}$, where k and j indicates the k_{th} object in the $(i\text{-}n)_{\text{th}}$ frame and the j_{th} object in the i_{th} frame, respectively, and n is the frame numbers between the two frames. Select the minimum $d_{k,j}^{i-n,i}$ of an object in the i_{th} frame, which is denoted as $d_{m_j}^{\ i}$, then select the minimum $d_{m_j}^{\ i}$ of all objects in the i_{th} frame. If $\min\{d_{m_j}^{\ i}\} < th_d$, the object in the $(i\text{-}n)_{\text{th}}$ frame and the object in the i_{th} frame, which are corresponding to $\min\{d_{m_j}^{\ i}\}$, are the same object in the frames. Otherwise, the object in i_{th} frame is a new object. Repeat to find the same objects in the two frames with excluding the compared objects until the objects in the i_{th} frame have been compared. Consequently, all objects in two frames can be distinguished, and the same objects can be aligned. For example, the j_{th} object in i_{th} frame and the j_{th} object in $(i\text{-}n)_{\text{th}}$ frame is the same object.

For labelling the objects, a velocity magnitude threshold (th_v) can be set by experience, such as 1m/s, which indicates the velocity of background object is smaller than the threshold. Denote the velocity of an object in the i_{th} frame as v^i, and the label of the object is denoted as l^i. For each object in every frame, set the initial label as zero. If an object is new detected, set the label of the object as -1; if the object is a background object, set the label as 1. If the velocity of the object is small than th_v in previous frames, it must be a background object. So if $v^i < th_v$, and $l^{i-1}=1$, set $l^i=1$. Background objects are filtered with a sliding window, and the window size is denoted as w. In the first w frames, if $v^i < th_v$, filter the object, and update the object label as 1. From the $(w+1)_{\text{th}}$ frames, for the each object labelled as zero, if $v^i < th_v$, search the object in previous frames until find the frame where the label of the object is not zero. The number of the searched frames is denoted as f_N. Consider the previous f_N frames from current frame if $f_N < w$, and only consider the w frames if $f_N \geqslant w$. Whether an object moving towards to or moving away the roadside radar can be detected by radar, which is reflected by the velocity direction. If an object moves towards to (or away from) the roadside radar, and the distance between the object and roadside radar in successive frames does not decrease (or increase) continuously in f_N or w frames, the object should be filtered. Then, update the label as 1 for all filtered objects.

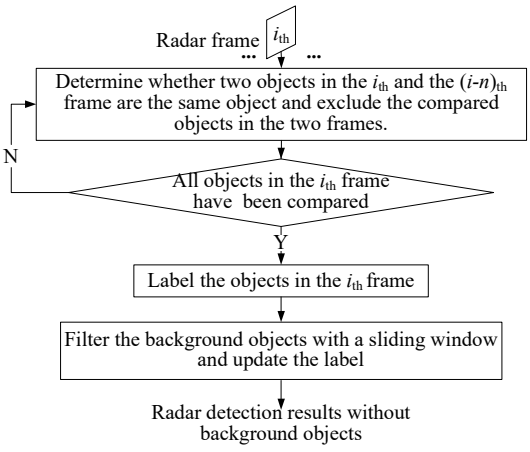


Fig. 2. Method of background objects filtering

B. Proposal of Auto-calibration

The aim of calibration is to find a relationship between radar coordinates and video coordinates, so that a radar detected object can be projected and spatially aligned with its corresponding object in camera coordinate system. This projection can be described as $(u, v) = f(x, y)$, where (u, v) is the pixel coordinates of the object in video coordinate system, and (x, y) is the position of the object in radar coordinate system. The projection function f can be estimated via lots of radar and video object pairs. To manually find the object pair is very time-cost. Therefore, in our proposed system, we use an automatically calibration method to find the object pair.

The approach of auto-calibration is shown in Fig.3. The input is sensing data of objects in each time-synchronized radar and video frame. At the beginning of the approach, it sets trajectory id and trajectory direction for radar and video detected objects, respectively. In radar frame process, we find the nearest object in the $(i\text{-}n)_{\text{th}}$ frame for each object in the i_{th} frame. Assuming that $d_{k,j}$ is the distance between the object k and j, where j is one of objects in the i_{th} frame, and k is its nearest object in the $(i\text{-}n)_{\text{th}}$ frame, then if $d_{k,j}$ is smaller than pre-set threshold d, assign the trajectory id k to the object j. Otherwise, a new trajectory id is assigned to the object j. The process of video frame is same as the radar process. The trajectory direction is defined and updated in every time window *tw*. Radar trajectory direction is determined by mean velocity direction in *tw*. Radar can detect whether the object moves toward or away from the radar. In each frame, -1 is used to indicate that the object is moving away from the radar, and +1 to indicate that the object is moving to the radar. If average value of the velocity indicator in each frame during *tw* is positive, then the radar trajectory direction is set as 'toward radar'. Otherwise, the radar trajectory direction is set as 'away from radar'. Video trajectory direction is determined by object's position in the last and first frame during *tw*. Coordinate pair $(u_{\text{first}}, v_{\text{first}})$ and $(u_{\text{last}}, v_{\text{last}})$ stands for the object's pixel coordinates in the first and the last frame during *tw*. If $v_{\text{last}} - v_{\text{first}} > 0$, the video trajectory direction is set as 'toward camera'; otherwise, it is set as 'away from camera'.

Whether the trajectory is new or not can be known after set trajectory id of each object. If it is new, put it into an unmatched trajectory set. If it is a trajectory in former frames and unmatched, put it into the unmatched trajectory set; if it is a matched trajectory, put it into matched trajectory set. In both unmatched radar (*urt*) and video trajectory set (*uvt*), the sorting of trajectories is from left to right or from far to near, based on position of each trajectory in its own coordinates. If the number of *urt* and *uvt* with the same trajectory direction equals, trajectories are matched by the sorted sequence, e.g., the most left trajectory in radar system matches the most left one in video system; the second left trajectory in radar system matches the second left one in video system.

After sorting trajectories in both *urt* and *uvt*, all the new matched and former matched trajectories are put into matched trajectories set. The objects in each matched trajectory are paired, e.g., the r_{th} radar trajectory and the m_{th} video trajectory are matched, (x_r^i, y_r^i) is the position of the detected object in the r_{th} radar trajectory of i_{th} frame, and the (u_m^i, v_m^i) is pixel coordinate of the object in the m_{th} video trajectory of the i_{th} frame. The radar object and video object are set as paired data $[(x_r^i, y_r^i), (u_m^i, v_m^i)]$. The paired data are used to estimate the projection function f by mathematical method, such as least square fitting method or interpolation method.

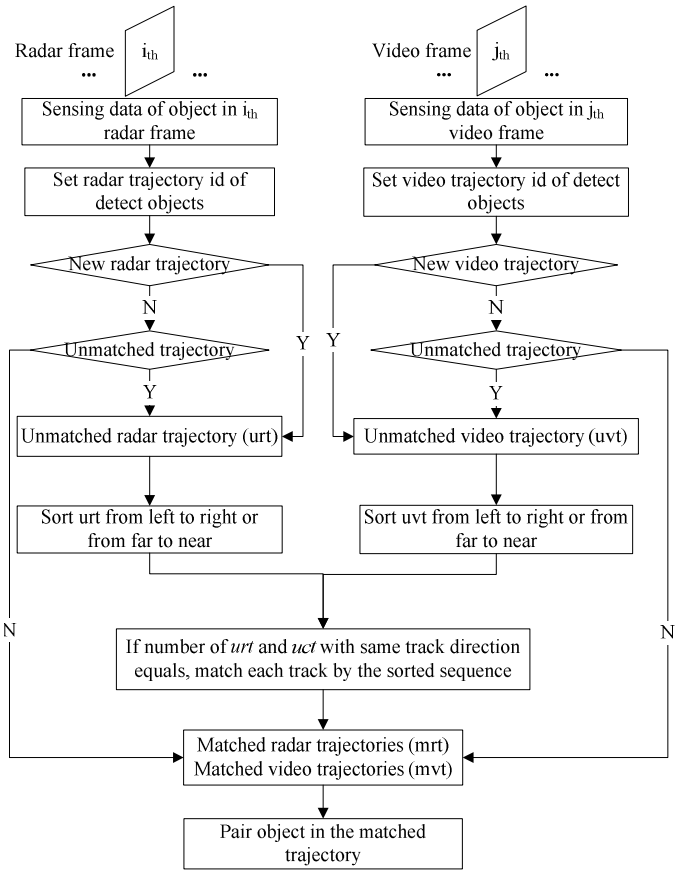


Fig. 3. Approach of auto-calibration

IV. Experiments and Results

A. Parameters of Fusion Sensors: Radar and Camera

In the experiments of roadside sensing fusion system, camera is a Hikvision product, and the fusion video frame rate is 10 fps. Radar is a TI AWR1642 Evaluation Module, which is an mmWave radar using Frequency-modulated continuous-wave (FMCW), and works on 76 to 81 GHz frequency band. The radar works in multi-mode. One is named as short range radar mode, in which the maximum range is 80 m, the range resolution is 36.6 cm, the maximum velocity is 90 km/h, and the velocity resolution is 0.52 m/s. The other one is named as ultra-short-range radar mode, in which the maximum range is 20 m, the range resolution is 4.3 cm, the maximum velocity is 36 km/h, and the velocity resolution is 0.32 m/s.

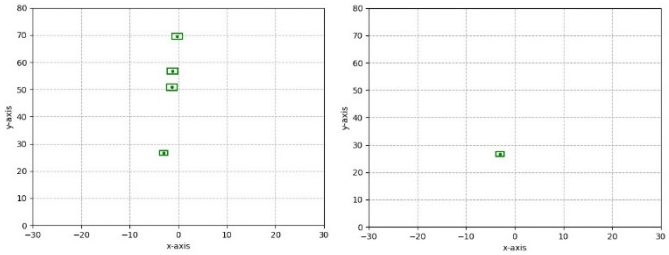
B. Experiment of Background Object Filtering

The experiment is implemented at a two-way road shown in Fig.4. (a), and the radar and camera are installed at a tripod deployed at roadside. The sensors face to the roads directly.

While radar captures sensing data, camera is also recording video frames for verification. The proposed background objects filtering algorithm is used to process the detected objects by radar continuously. We illustrate one moment as an example, which is shown in Fig. 5. From its corresponding video frame shown in Fig.4. (b), there is only one dynamic object at that moment, a vehicle, on the road in two directions. However, in Fig.5. (a), four objects marked as green rectangles are detected by radar without filtering. The result of background objects filtering is shown in Fig.5. (b). It shows that all background objects are filtered out and the vehicle is reserved after filtering.

(a) (b)

Fig. 4. Experiment of background object filtering (a) L-shaped intersection with two-way road observed by camera and radar; (b) a video frame captured by camera.



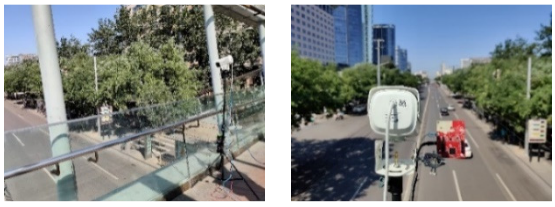
(a) (b)

Fig. 5. Objects filtering result: (a) before filtering; (b) after filtering

C. *Experiment of Background Object Filtering*

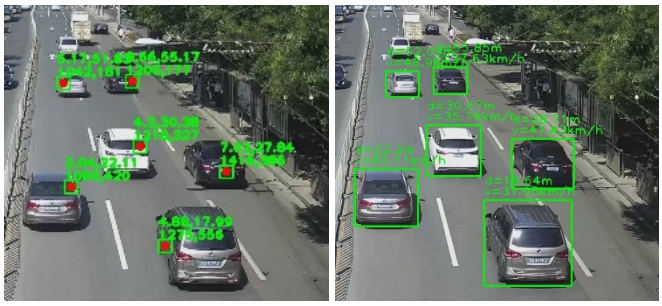
The camera and the radar are installed at a tripod on an overpass shown in Fig.6. In this scenario, both radar and camera can detect three lanes where vehicles move away from sensors.

In the experiment, we record 19483 video frames in total and capture radar sensing data simultaneously. Among them, we use 5000 frames and corresponding radar sensing data to make sample set for calibration. Totally, there are 296 pairs of radar and camera data which are extracted based on auto-calibration. These pairs of data are used for calibration and fusion, and interpolation fitting method is used for calibration. The calibration based on auto-calibration result is shown in Fig.7. (a). The red points are the radar detection results which are calibrated to the coordinate system of camera. The sensing fusion results are shown in Fig.7. (b). Camera detection outputs the bounding box for each car. With the calibration results, each radar detected object is fused with one of camera detected objects. As a result, its position and velocity detected by the radar is labelled above its corresponding bounding box marked by camera detection. These experiment results show that the sample set made by auto-calibration is good enough to make accurate sensor calibration and then sensing fusion.



(a) (b)

Fig. 6. Experiment scenario for auto-calibration and fusion: (a) sensors are setup at an overpass; (b) roads observed by camera and radar.



(a) (b)

Fig. 7. Auto-calibration and fusion: (a) calibration result based on auto-calibration; (b) fusion result.

V. Conclusion

A roadside camera-radar sensing fusion system for intelligent transportation with the method of background objects filtering and the auto-calibration approach is presented in this paper. Experiments results show that the proposed system can be fast implemented to automatically acquire accurate roadside sensing fusion information.

References

[1] Y. Lin, P. Wang, and M. Ma, “Intelligent Transportation System (ITS): Concept, Challenge and Opportunity,” in *Proc. IEEE 3rd bigdatasecurity*, 2017, pp. 167-172.

[2] Y. Zhang, Y. Cao, Y. Wen, L. Liang, and F. Zou, "Optimization of Information Interaction Protocols in Cooperative Vehicle-Infrastructure Systems," *Chinese Journal of Electronics*, vol. 27, pp. 439-444, Mar. 2018.

[3] L. Wu, L. Zhang, H. Li, and H. Ding, "Vehicle sensing data acquisition and analysis," in *Proc. DATA'18*, 2018.

[4] G. T. S. Ho, Y. P. Tsang, C. H. Wu, W. H. Wong and K. L. Choy, “A Computer Vision-Based Roadside Occupation Surveillance System for Intelligent Transport in Smart Cities,” *Sensors*, vol. 19. pp. 1-26, Apr. 2019.

[5] M. S. Shirazi and B. T. Morris, “Looking at Intersections: A Survey of Intersection Monitoring,” *IEEE Transaction intelligent Transportation Systems*, vol. 18, pp. 4-24, Jan. 2017.

[6] Q. Jiang, L. Zhang, D. Meng, "Target Detection Algorithm Based on MMW Radar and Camera Fusion," In *Proc. IEEE ITSC'19*, 2019.

[7] A. Krämmer, C. Schöller, D. Gulati, and A. Knoll, "Providentia - A Large Scale Sensing System for the Assistance of Autonomous Vehicles", in *Proc. RSS'19*, 2019.

[8] C. Schöller, M. Schnettler, A. Krämmer, G. Hinz, M. Bakovic, M. Güzet, and A.s Knoll, “Targetless Rotational Auto-Calibration of Radar and Camera for Intelligent Transportation Systems,” in *Proc. IEEE ITSC'19*, 2019.