Intersection Object Detection Algorithm Based on Camera and Radar Data Fusion

Xiaoguang Chen[1,a], Feng Yan[2,b], Tienan Pan[2,c], Suining Wu[1,d], Bin Li[1,e] and Zhixin Wang[1,f]

[1]*China Railway Signal & Communication Research & Design Institute Group Co.,Ltd. Beijing,China*

[2]*School of Instrumentation and Optoelectronic Engineering Beihang University Beijing, China*

a. chenxiaoguang@crscd.com.cn, b. yanfengbuaa@163.com, c. ptn1209@qq.com, d. wusuining@crscd.com.cn, e. libin3345@crscd.com.cn, f. wzx@crscd.com.cn

Keywords: Object detection, data fusion, camera, radar.

Abstract: Intersections are the weakest link in tram operations, where collisions between trams and cars often occur. This paper proposes an intersection object detection algorithm based on camera and radar data fusion using the method of Intersection over Minimum. For camera, this paper proposes a method of constructing the Total Object List for inter-frame target matching, which realizes the measurement of position and velocity. For radar, a combined filtering method is proposed in this paper to improve the detection rate. After experimental tests, it can be seen that the intersection object detection algorithm proposed in this paper makes up for the shortcomings of a single sensor and shows a higher overall performance.

1. Introduction

With the advancement of urbanization in China, the operating mileage of trams is rapidly increasing. However, intersections are the weakest link in tram operations. As of 2018, the operating mileage of trams has reached 327.1 kilometers, increased by 91.13 kilometers in 2018, which means that the operating mileage for two consecutive years has increased by more than 30%. Traffic junctions are areas where trams and motor vehicles have joint ownership, and are the key sections of the accident. In 2018, in Wuhan Optics Valley, there were four collisions between trams and cars. Most of the accidents were caused by illegal intrusion of cars and insufficient braking distance of trams. Therefore, it is of great significance to develop an intersection object detection system to prevent obstacles from invading the intersection in front of the train.

At present, at home and abroad, Honeywell has developed a Radar scan system based on scanning radar. IHI developed a 3-D laser radar system based on laser radar. Crossing obstacle detection technology based on image recognition was developed by Japanese company. The detection method based on the ground induction loop has limitations such as damage to the road surface and single detection type[1, 2]. The advantages of the infrared-based detection method are low cost and strong penetration in foggy weather, but its anti-interference ability is weak and it is greatly affected by

weather conditions[3]. In summary, the millimeter-wave radar has strong anti-interference ability for rain and snow weather, but it cannot perform identification analysis[4-7]. As a more universal and widely popularized technology, video recognition can accurately classify objects[8, 9]. Therefore, this paper will use the image recognition and millimeter wave radar fusion scheme for research.

2. Algorithm

An intersection object detection algorithm based on camera and radar data fusion using the method of 'Intersection over Minimum' (IoM) is proposed in this paper. The system contains a camera and a millimeter-wave radar, and its structure is illustrated as Figure 1. The camera is used to obtain 2D data and the radar is to obtain the 3D data of object velocity and distance to the system. The fusion of 2D and 3D data can realize high accuracy and high dimension detection of invaders.

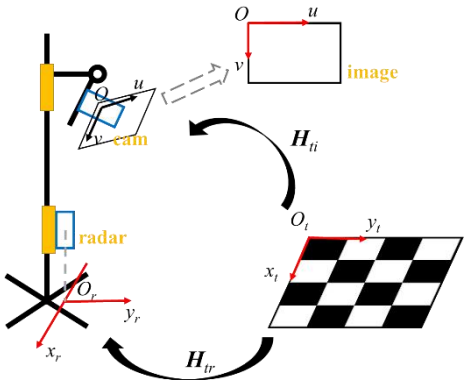

Figure 1: The structure of the system.

2.1. The Hardware and the System Calibration

As shown in Figure 1, the reference frames involved in this paper includes Radar plane Reference Frame (RRF) that is also the World Reference Frame (WRF) $O_r - x_r y_r$, the Target plane Reference Frame (TRF) $O_t - x_t y_t$ and the camera Image plane Reference Frame (IRF) $O - uv$. The RRF is set at the ground vertically corresponding to the radar, and it is used as the world reference frame. As the objects detected by the camera and the radar need to fusion, the system structure parameters describing the relationship between IRF and RRF need to calibrate.

The calibration is achieved by a planar chessboard target. First, the coordinate axis of TRF and the RRF are set parallel artificially, and the translation vector of the two reference frames is measured using meter ruler. After that, the homography matrix $\boldsymbol{H}_{tr}$ between TRF and RRF is obtained:

$$\boldsymbol{P}_r = \boldsymbol{H}_{tr}\boldsymbol{P}_t \tag{1}$$

where the $\boldsymbol{P}_t$ is the feature point coordinate in the TRF and the $\boldsymbol{P}_r$ in the RRF.

Then, the image of target is captured using camera, and at least 4 feature points are extracted from the image. According to[10], the homography matrix $\boldsymbol{H}_{ti}$ between TRF and IRF can be obtained. $\boldsymbol{H}_{ti}$ satisfies the following formula:

$$\boldsymbol{p} = \boldsymbol{H}_{ti}\boldsymbol{P}_t \tag{2}$$

The p is the feature point coordinate in the IRF.

Therefore, the target can be used as the intermediary to calculate homography matrix $\boldsymbol{H}_{ri}$ between RRF and IRF:

$$\boldsymbol{H}_{ri} = \boldsymbol{H}_{ti}\boldsymbol{H}_{tr}^{-1} \tag{3}$$

And

$$\boldsymbol{p} = \boldsymbol{H}_{ri}\boldsymbol{P}_{r} \tag{4}$$

So far, the calibration of the system structure parameters has been completed.

2.2. Camera Data Reception

In this paper, the camera is used to capture the images in the intersection scene in real time, and the YOLOv3 neural network model is used to detect the images collected by the camera frame by frame. YOLOv3 is a One-Stage object detection network with high real-time performance, which turns the object detection problem into a regression problem and solves it[11]. The backbone of the YOLOv3 is Darknet-53, which uses a full convolutional structure. The size of the tensor in the forward propagation process is changed by changing the stride of the convolution kernel. It also uses the layer-hopping connection method like ResNet to ensure that the network structure can still converge even though it's very deep. YOLOv3 uses a multi-scale prediction method similar to FPN, using 13 × 13, 26 × 26, and 52 × 52 feature maps. Multi-scale prediction can make YOLOv3 have higher accuracy when detecting objects of different scales.

The images collected by the camera are detected frame by frame to obtain the category information and position coordinate information of each object in each frame. Objects in two adjacent frames are matched to find the correspondence between the same object. This article adopts the method of constructing the Total Object List (TOL), and uses the sum of Euclidean distances as a measurement to match the objects of two adjacent frames. The TOL is a list that stores object information in the scene, including the ID, category, number of missed detections, position coordinates, and detection time of each object. The position coordinates are calculated from image coordinates using the homography matrix $\boldsymbol{H}_{ri}$. The TOL stores the objects information of multiple frames. For the image coordinate [u, v] of an object, its corresponding world coordinate [X, Y] is calculated using Eq. (5).

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \boldsymbol{H}_{ri} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{5}$$

$\boldsymbol{H}_{ri}$ is the homography matrix from the radar coordinate system to the image coordinate system.

For the objects detected in the current image, the world coordinates are arranged to get their total permutations. Take one of the permutations, which represents the correspondence between the world coordinates of each object in the current image and those in the TOL. Calculate their Euclidean distances, and add the results. Each permutation corresponds to a sum of Euclidean distances, and the permutation corresponding to the minimum value is the best matching result of the objects in the current frame and those in the TOL. If the number of objects in the current frame is not equal to the

number of objects in the TOL, you need to fill them up and record the Euclidean distance at the corresponding position as 0 when calculating the sum of Euclidean distances. After finding the best matching result, the current frame information of each object is added to the TOL.

If the number of objects in the current frame is not equal to the number of objects in the TOL, you need to fill them up. When calculating the sum of Euclidean distances, record the Euclidean distance of the corresponding position as 0, and add 1 to the number of missed detections. After finding the best matching result, the current frame information of each object is added to the TOL, and objects that have been missed too much are deleted.

With the real-time shooting of the camera, the TOL is constantly updated. When the number of the information of an object in the TOL reaches a certain number, the speed of the object can be calculated. The speed is calculated using the method of successive minus. Assume that an object has N (N> 4 and N is even number) frames of the world coordinate information in the TOL. Calculate the distances between the world coordinates of the first frame and N/2+1th frame, the second frame and N/2+2th frame ... and the N/2th frame and the Nth frame. Divide the distances by the corresponding time intervals, and take the average of each result, which is the speed of the object.

So far, the position and speed of each object detected by the camera are obtained.

2.3. Radar Data Reception

The millimeter wave radar sends out pauses at a frequency and receives echoes to detect the objects. The Doppler principle and FMCW technique are used to complete the object velocity measurement. However, the original information received by the radar has serious misdetections and leak detections. This phenomenon is particularly serious in indoor spaces affected by wall reflections. To this end, this paper proposes a combined filtering method, which includes two methods: Parameter Matching Filtering (FMP) and Interframe Mutual information Filtering (IMF).

PMF refers to the completion of filtering by determining whether the object position and velocity returned by two adjacent frames of the radar match. Based on a large number of experiments, it is observed that there are two conditions occurring frequently under the influence of reflection or other conditions that (1) the ultra-low speed objects are detected and (2) the velocity and displacement do not match. Aimed at condition (1), the object whose velocity is lower than a certain threshold will be filtered out. and for condition (2), the object will be filtered out whose difference of the detected displacement and the calculated theoretical displacement of two adjacent frames exceed the threshold.

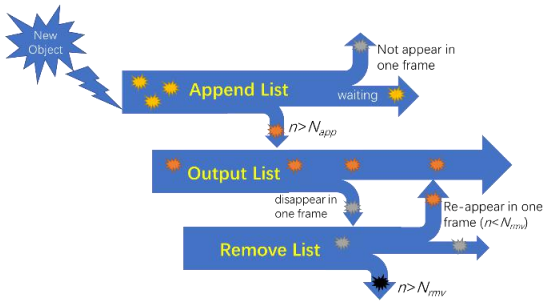

Figure 2: The basic flow of IMF.

IMF means that if an object appears for multiple consecutive frames, it is considered to be a real target for stable detection. In addition, if the object disappears for multiple consecutive frames, it is considered to be completely disappeared instead of being instantaneously missed, and then removed. Adding the time dimension on the basis of the space dimension, and using the mutual information between multiple frames to achieve the object stable detection, can greatly reduce the rate of missed

detection and false detection. The basic flow is shown in Figure 2, when one object appears for the first time, it will be added to the Append List (AL), and only when the objects in the AL appears for N_{app} frames consecutively, they can be add to the Output List (OL). However, if the object disappears for some on frame during in the AL, it will be removed and recount when it appears again. When one object in the OL disappears in some one frame, it will be added to the Remove List (RL) at first, and for the objects in the RL, only when they disappear for N_{rmv} frames, they are regarded as disappearance really, and then will be removed. If one object in the RL appears again, it will return to the OL, and the disappearance will be treated as missed detection.

2.4. Algorithm of Camera and Radar Fusion

The fusion of camera data and radar data requires the synchronization in space and time. The synchronization of space requires camera and radar have the common field of view and the structure parameters of them should be calibrated, which are used to convert radar detection result from radar reference frame to camera image reference frame. The calibration can be realized by the method described in the Section 2.1. The synchronization of time requires the acquisition of camera data and radar data should be synchronized, which can be realized by multi-thread process. The processing of real-time camera data acquisition and fusion can be implemented in the main thread, and the processing of radar data acquisition and filtering in the separate thread.

The match and data fusion of camera and radar fusion is achieved by the method of 'Intersection over Minimum' (IoM). The output of camera data processing contains the position in the image, the class as well as the distance and the velocity calculated by the homography matrices, while the output of radar contains the precise distance and velocity of the objects. In the fusion, the objects of the sensors should be matched first, and for the ones detected by the camera and the radar at the same time, their distance and velocity information come from radar and others from camera.

The match of the objects obeys the principle of the IoM. The IoM is defined as the ratio of the intersection of two rectangle areas and the minimum square between them. The definition is stated as Eq. (6), and the illustration is as Figure 3.

$$IoM = \frac{S_A \cap S_B}{\min(S_A, S_B)} \tag{6}$$

The IoM refers to the intersection over the minimum of two rectangular areas. S_A and S_B are the areas of two rectangles, respectively. This calculation method can measure the degree of coincidence of two rectangles, which determine whether the two rectangles represent the same area. At the same time, compared with the method of IoU, the IoM can alleviate the calculating failure problems caused by the large difference between the areas of the two rectangles.

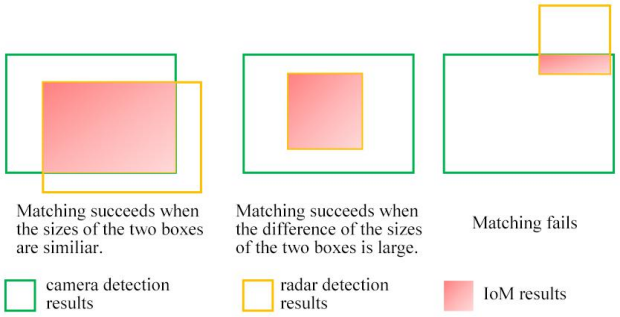

Figure 3: The illustration of IoM.

Among them, scale is the scaling factor of the bounding box. The specific bounding box size is $scale[w,h]$. And y is the y coordinate of the object detected by the radar in radar coordinate system.

After expanding the radar object position into an area, according to the IoM calculation formula, go through all radar objects and all camera objects in the current frame to determine whether there are object pairs which satisfy $IoM > T_{fusion}$. T_{fusion} refers to the threshold value for determining whether the object is the same object according to the IoM. If the same pair of objects are determined consecutively, the radar and camera are considered to have detected the same object together, and the object size and category information of the camera are fused with the distance and speed information detected by the radar to obtain the complete information of the object. Otherwise, it is considered to be a false detection by a single device and will not be fused. Finally, all fusion objects are used as the final fusion output result of the current frame.

(a)

(b)

(c)

Figure 4: (a) A part of radar detection results. (b) A part of camera detection results. (c) A part of fusion results.

3. Experiment Results

Experiments of the proposed detection system are conducted indoors. In the experiment, the landmarks are arranged on the ground. When object moves to the landmarks, data collection is carried out. Namely, the original output of radar and camera, as well as the fusion output. We collected 2400 groups of experimental data. Figure 5 shows the proposed detection system.

Figure 5: The proposed detection system.

Figure 4 is a part of radar detection results, camera detection results, and fusion results. It can be seen that when the radar works alone, the detection effect on distant objects and dense objects is not good, causing missed detection. The detection effect of the camera is better than that of the radar, and the positions of the bounding boxes are more accurate. The fusion result absorbs the advantages of camera detection and makes up for the shortcomings of radar in detection rate.

Figure 6 is a bird's-eye view of a part of experimental results. The area is 20m × 10m. The red dot represents the real position of the object, the blue dot represents the position detected by the radar, and the yellow dot represents the position detected by the camera. It can be seen that compared with the camera, the position detected by the radar is closer to the real position, and the position error is smaller. The result of the fusion uses the object position obtained by radar detection, which absorbs the advantages of radar detection and makes up for the lack of position measurement accuracy of the camera.

In Table 1, we compare the proposed method with camera-based method, radar-based method and fusion-based method.

Table 1: Comparison of the three methods.

Sensors	Performance		
	Detection Rate	Miss Rate	Mistake Rate
Radar	77.11%	22.89%	3.04%
Camera	98.93%	1.07%	2.87%
Fusion	99.59%	0.41%	5.91%

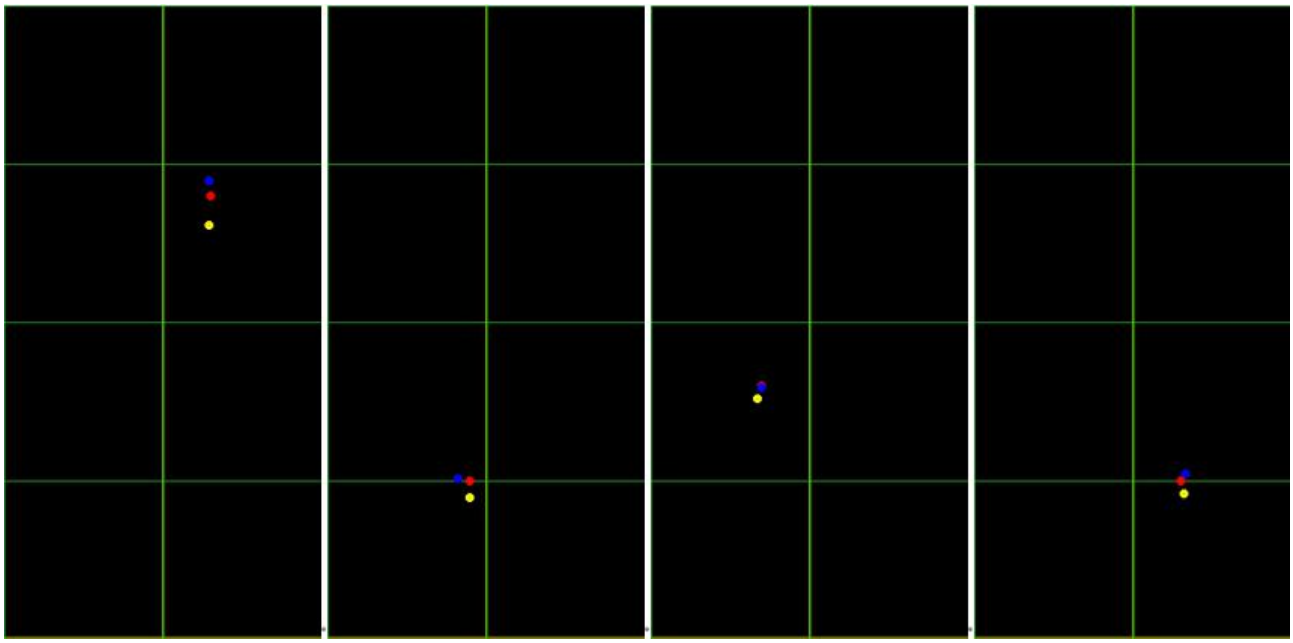

Figure 6: A bird's-eye view of a part of experimental results.

From Table 1, the detection rate of Radar is significantly lower than Camera. What's more, the miss rate of Radar is significantly higher than Camera. Since the columns and walls affects radar signals in transmission, radar cannot achieve equivalent results than camera. Comparing with these methods based on single sensor, sensor fusion method perform best in detection. What's more, miss rate is decreased via sensor fusion method. Although mistake rate of sensor fusion method if slightly higher than camera-based method, it is the most accurate method comprehensively.

Position measurement error in different distance level is analyzed in Figure 7. As shown in Figure 7, the position measurement error of radar-based method is higher than that of camera-based method, especially at long distance. It is worth noting that, sensor fusion method is able to further improve the position measurement accuracy at long distance. Thereafter, position measurement accuracy of sensor fusion method is more accurate and stable than single sensor method.

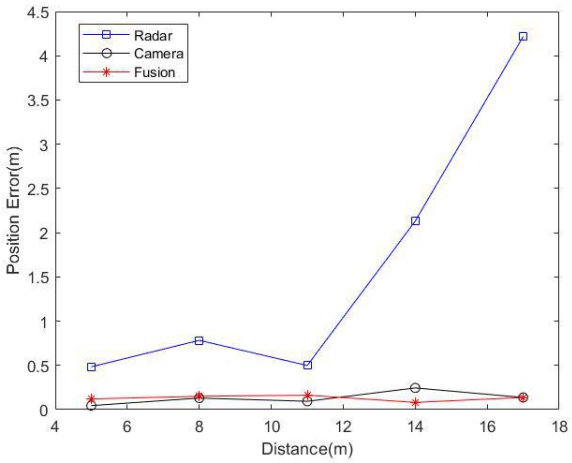

Figure 7: Position measurement error in different distance level.

Velocity measurement error in different distance level is analyzed in Figure 8. As shown in Figure 8, velocity measurement error of sensor fusion method is less than camera-based method and radar-based method in most cases. In general, sensor fusion method is able to improve velocity measurement accuracy.

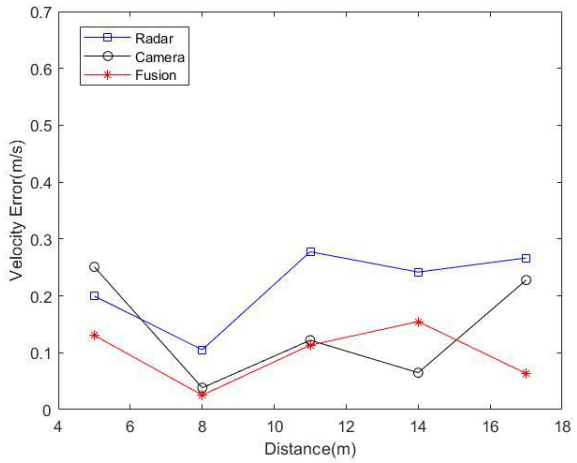

Figure 8: Velocity measurement error in different distance level.

4. Conclusions

This paper proposes a detection algorithm based on the IoM. The IoM is used to fuse radar and camera data to achieve the matching and fusion of homologous data. For cameras, this paper proposes a method for constructing the Total Object List for inter-frame object matching and implements position and velocity measurement based on a monocular camera. For radar, a combination filtering method is proposed in this paper, which improves the detection rate and reduces the false detection rate.

The experimental results show that the sensor fusion method used in the intersection object detection system can make up for the shortcomings of a single sensor. Compared with a single sensor, sensor fusion complements the measurement results of each sensor, and improves performance indicators such as detection rate, position measurement accuracy, and speed measurement accuracy. Under the complicated conditions of intersection scenes, this method can reduce the influence of interference, and the detection results are more reliable.

References

[1] R. G. Song, "Modeling and Simulation Research of Train-to-Wayside Communication System Based on Cross Induction Cable Loop," Beijing Jiaotong University, 2016.

[2] F. S. Dou, Z. Q. Long, C. H. Dai, D. P. Zhang, L. Luo, X. X. Zeng, "High Precision Vehicle Positioning Device Based on Induction Loop," ZL201420066341.4.

[3] X. D. Xiao, J. F. Li, "Vehicle Flux Detector Based on Pyroelectric Infrared Sensor," Laser & Infrared, vol. 35, pp. 93–95, February 2005.

[4] B. Hu, C. X. Zhao, "Vehicle Detection Method Based on MHT Model Using Millimeter-wave Radar," Journal of Nanjing University of Science and Technology, vol. 36, pp. 5–8, August 2012.

[5] B. Du, "Study of Vehicle Detection Method Base on Roadside MMW Radar," Shenyang Institute of Automation Chinese Academy of Sciences, 2010.

[6] I. Urazghildiiev, R. Ragnarsson, P. Ridderstrom, A. Rydberg, E. Ojefors, et al, "Vehicle Classification Based on the Radar Measurement of Height Profiles," IEEE Transactions on Intelligent Transportation Systems, vol. 8, pp. 245–253, July 2007.

[7] J. M. Munoz-Ferreras, F. Perez-Martinez, J. Calvo-Gallego, A. Asensio, B. P. Dorta-Naranjo, et al, "Traffic Surveillance System Based on a High-Resolution Radar." IEEE Transactions on Geoscience and Remote Sensing, vol. 46, pp. 1624–1633, July 2008.

[8] Y. Kuwata, J. Teo, G. Fiore, S. Karaman, E. Frazzoli, et al, "Real-Time Motion Planning With Applications to Autonomous Urban Driving," IEEE transactions on control systems technology: A publication of the IEEE Control Systems Society, vol. 17, pp. 1105–1118, September 2009.

[9] S. F. Tahir, A. Cavallaro, "Cost-Effective Features for Reidentification in Camera Networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, pp. 1362–1374, August 2014.

[10] Z. Y. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 1330–1334, November 2000.

[11] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," https://arxiv.org/abs/1804.02767, 2018.