2021 IEEE 3rd International Conference on Communications, Information System and Computer Engineering (CISCE 2021)

Multi-target Detection based on Multi-sensor Redundancy and Dynamic Weight Distribution for Driverless Cars

Qinghe Liu
School of Automotive Engineering
Harbin Institute of Technology at Weihai
Weihai, China
qingheliu@hitwh.edu.cn

Wenqi Zhou
School of Automotive Engineering
Harbin Institute of Technology at Weihai
Weihai, China
zhou_wenqi@126.com

Yankun Zhang
School of Automotive Engineering
Harbin Institute of Technology at Weihai
Weihai, China
20S030124@stu.hit.edu.cn

Xun Fei
School of Automotive Engineering
Harbin Institute of Technology at Weihai
Weihai, China
feixunbobo@163.com

2021 International Conference on Communications, Information System and Computer Engineering (CISCE) | 978-1-6654-0352-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/CISCE52179.2021.9446002

***Abstract*—The perception system of unmanned vehicles has developed rapidly in recent years, but in the process of multi-sensor fusion, a certain sensor's perception error often occurs which causes great errors in the perception results of the entire multi-sensor system and is harmful to human life and property. Therefore, it requires redundant fusion of the objects perceived by the sensors, and sets weights during the fusion process to eliminate sensors with poor detection results. This paper establishes a multi-sensor perception and a dynamic weight distribution and eliminate system. Multi-sensor weight distribution is performed by Kalman tracking variance. Such a system will greatly reduce the errors of the perception system.**

Keywords-multi-sensor redundancy; multi-target detection; dynamic weight distribution

I. INTRODUCTION

As the current mass-produced Advanced Driving Assistance System (ADAS) car accidents frequently occur, most of the problem lies in the sensor fails. When the sensor fails, it can perform fault self-diagnosis and system integrating. It is necessary to remove the sensor's perception. When the tracking deviation of a certain sensor is too much, the data of the sensor must be removed in the process of environmental perception. This requires the designed sensor system to be redundant in data. Multiple types of sensors perform multi-target sensing at the same time. A sensor failure is irrelevant to the entire system, and the system should be able to operate normally.

Through multiple source sensors redundancy and multiple targets tracking and fusion, smart cars were provided with stable and reliable fusion information. In this paper, a forward redundant multi-sensor system of Lidar-Camera-Radar is established.

The research content mainly includes the research of sensor fusion, and the estimation of the multi-target tracking state of each sensor. This is the basis for the subsequent redundant part. In addition, before the multi-sensor fusion, the perception of each sensor must be tracked. The sensor Multi-target tracking data is collected, and then the target information perceived by each sensor are fused into detected objects. While multi-sensor objects are associated, dynamic weight distribution is performed through the statistical data of sensor self-diagnosis, and this is done at the right time when the sensor data changes suddenly. The eliminate of the sensor ensures the reliable of the data source. The target tracking variance of each sensor is used to give different weights to the detection targets of each sensor, and the detection targets are fused according to the weights.

The contributions of this paper include:

- Establish a Lidar-Camera-Radar detection system. We use Cnn_seg algorithm on Lidar to detect targets. YOLOV4 algorithm is used on Camera detection.
- Multi-target tracking and state estimation based on Kalman filter.
- Multi-sensor target association and dynamic weight distribution.

The rest of this paper is organized as follows: Sec. IIintroduces Lidar detection algorithm and Camera detection algorithm used by the system and analyzes the causes of traffic accidents. Then this paper proposes a layout and effect of the system in Sec. III. Afterwards, experiments are performed to evaluate the proposed methods in Sec. IV. Finally, we conclude our work in Sec. V.

II. RELATED WORK

With the improvement of the level of autonomous driving, the requirements of automobiles on target detection are increasing gradually. Therefore, it is very important to study on target detection. Autonomous driving target detection relies on a variety of sensors. Camera, Radar and Lidar are the most important and widely used sensors. There is Apollo's open source Cnn_seg[1] algorithm for Lidar. The Camera uses the

978-0-7381-1215-2/21/$31.00 ©2021 IEEE

YOLO-V4[2] algorithm. As for Radar, the sensor itself will output the detection objects. The datasets used in this paper are nuScenes[3] and Kitti[4,5].

Through the analysis of the existing intelligent driving traffic accidents, the main reasons for the collision accidents come from two aspects. One is that traffic accidents occurrence due to the extremely complicated driving environment. The other is the perception system in the environment. There was a problem in the information extraction process: misrecognition of obstacles in the environment or insufficient contour recognition accuracy resulted in errors in distance calculations and ultimately resulted in collisions. Therefore, optimizing the defects of a single sensor and efficiently fusing multiple sensors at the same time is a subject that must be completed in the development of autonomous driving.

Before sensor fusion, the objects of each sensor need to be tracked using Kalman filter[6,7] and Hungarian matching algorithm[8]. Furthermore, SORT[9] algorithm is used for tracking.

When performing multi-sensor fusion, there are still a series of problems, such as sensor failure, recognition error, insufficient data accuracy, which make it impossible to use sensor data. So, it requires the on-board multi-sensor redundant design. Multi-sensor redundant fusion system refers to the processing and integration of data from multiple sensors to obtain more accurate and reliable useful information. Compared with a system that only using single sensor, the information from multiple sensors has redundancy, complementarity, and correlation. The multi-sensor redundant fusion system has strong fault-tolerant performance, it is necessary to dynamically allocate the weight of each sensor.

III. System Introduction

A. Framework of System

We establish a forward redundant multi-sensor system which contains Lidar, Radar and Camera. This research aims to achieve intelligence through multi-source heterogeneous sensors, multi-target tracking, fusion and sensor redundancy. The system provides stable and reliable fusion information. The research content mainly includes the research of multi-sensor data fusion, as well as the problems of multi-target tracking state estimation of each sensor.

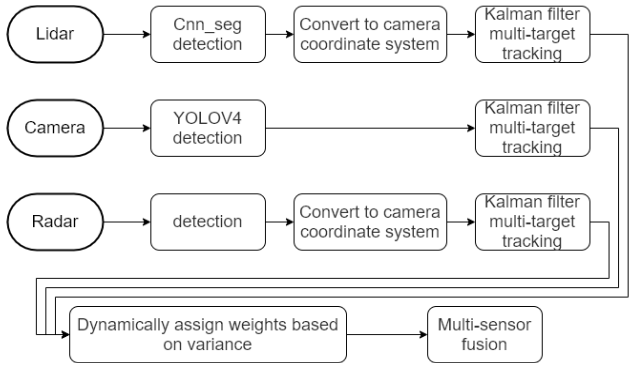

Figure 1. Framework of the system

The detection system is the basis of the redundant part. In addition, before multi-sensor fusion, the perceptual tracking of each sensor must be performed first. Sensor problems of multi-target tracking data association and the wrong target information perceived by each sensor are the problems of multi-sensor target association and fusion. While the sensors are associated, follow-up redundancy is carried out through the statistical data of sensor self-diagnosis. In the remaining research, different weights are given to the detection targets of each sensor with statistical data, and the forward redundancy of the sensors is carried out on the problem of sensor failure simulation and multi-sensor forward redundancy.

B. Obstacle detection and tracking with CNN_SEG algorithm based on Lidar

The point cloud data used in this test is nuScenes open source data. The sensor used in the training of cnn_seg algorithm is 64-line Lidar. However, 32-line Lidar is used for the nuScenes data, but it is still reliable for the detection effect.

Cnn_seg does not use traditional point cloud clustering methods for target detection, but uses semantic segmentation methods. By using semantic segmentation for target detection can avoid the problems on clustering methods for it will perform pixel-level classification. After semantic segmentation, how to cluster the detected pixels into a single target is the point. The most important thing is that cnn_seg returns a layer called center offset during semantic segmentation. Then clustering is performed according to the results of center offset and semantic segmentation to achieve the detection of a single target.

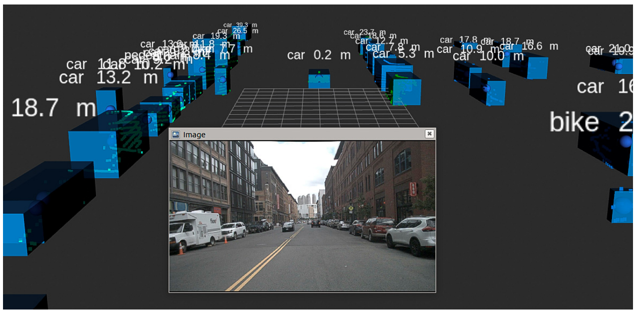

Figure 2. Lidar detection objects

Cnn_seg is actually divided into two steps: the first step is using a deep learning network to predict five layers of information. The second step is using these five layers of information to cluster. Specific steps are as follows:

- According to the objection layer information, the obstacle grid point object is detected;
- According to the center offset layer information, cluster the detected obstacle grid points to obtain clusters;
- Filter the background and the high points in each cluster according to the positiveness and object height layer information;
- According to class probability, classify each cluster to get the final target;

C. *Obstacle detection and tracking based on yolov4 machine vision algorithm*

The basic goal of Yolo-V4 is to improve the operating speed of neural networks in the production system, while optimizing for parallel computing, rather than optimizing for Billion Float Operations Per Second (BFLOPS).

Specifically, YOLO-V4 uses:

- Bag of Freebies (BoF) for backbone network: CutMix and Mosaic data enhancement, DropBlock regularization and class label smoothing;
- Bag of Specials (BoS) for backbone network: Mish activation, CSP and multi-input weighted residual connection (MiWRC);
- Bag of Freebies (BoF) for the detector: CIoU-loss, CmBN, DropBlock regularization, Mosaic data enhancement, self-antagonism training, Eliminate grid sensitivity, use multiple anchors for a true value, Cosine annealing scheduler, optimized hyper parameters and random training shapes;
- Bag of Specials (BoS) for the detector: Mish activation, SPP block, SAM block, PAN path aggregation block and DIoU-NMS.

This training uses Xeon W-2150B/3.00GHz CPU and two 2080Ti GPUs for training, 50,000 training times takes about 68 hours. There are 5 types of training (pedestrian, car, bicycle, truck, bus). Here is the loss function recorded during training:

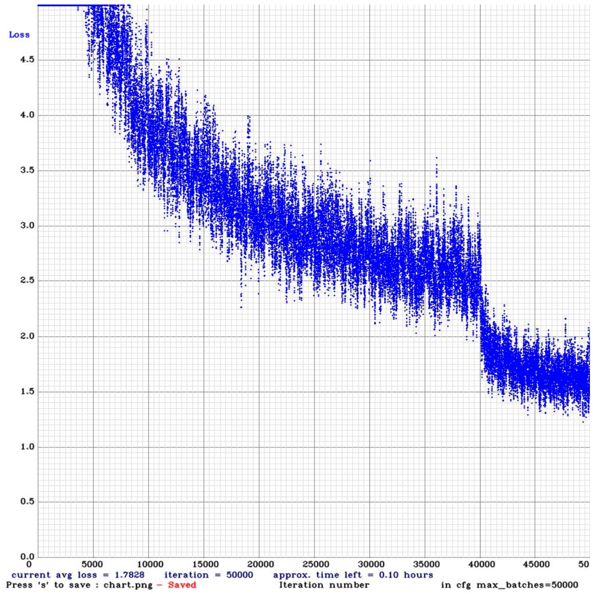

Figure 3. Loss value during YOLO-V4 training

The following figure shows the detection effect of YOLO-V4 algorithm. The input image used is the /cam_front/raw topic of the nuScenes dataset, and the detection result is obtained.

Figure 4. Camera detection objects

D. *Convert Lidar and Radar to Camera coordinate system*

In order to facilitate the fusion of various sensors, all sensors need to be transformed into the same coordinate system. In this paper, the Camera coordinate system is selected as the basic coordinate system, and the Lidar objects and Radar objects are uniformly transformed into the Camera coordinate system. The red part of the figure below shows the conversion of the point cloud of the Lidar detection objects to the Camera coordinate system.

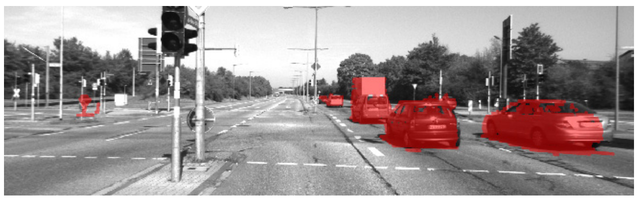

Figure 5. The point cloud of the Lidar object is converted to the Camera coordinate system.

E. Multi-target tracking and state estimation

Through the analysis of the basic principles of multi-target tracking, data association is the core link. Data association is the process of establishing a matching relationship between the current measurement of single sensor and the previous measurement to determine whether they belong to the same target source. The joint probability data association algorithm is applied to each sensor to associate each target, and then the target state is filtered and predicted. This is the basic element of multi-target tracking, and whether the target state space is linear is also an important issue.

The nature world is non-linear. For smart vehicles, the external world it perceives is even more so. The traditional multi-target tracking algorithms for intelligent vehicles are mainly aimed at linear systems, which are only a special existence in the real world. Therefore, it is very necessary to import the nonlinear state estimation method into the intelligent vehicle multi-target tracking, derive the representative method, and analyze the filtering accuracy to obtain the target tracking state of the sensor.

We use Kalman filter and Hungarian matching algorithm to track. Furthermore, SORT algorithm is used for tracking. Results are shown in the figure below:

- The blue box indicates the tracking prediction of the Lidar object;
- The red box represents the tracking prediction of the Camera object;
- The green box indicates the tracking prediction of the Radar object;

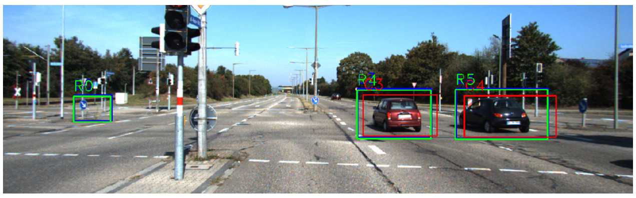

Figure 6. Lidar, Radar and Camera tracking prediction box

F. Multi-sensor target association and dynamic weight distribution

Among the fusion structures, distributed structures are commonly used in smart vehicles. It is the key issue in the application which is how to integrate different sensors of the same target and track the target for fusion. Here is the dynamic weight distribution and eliminate system:

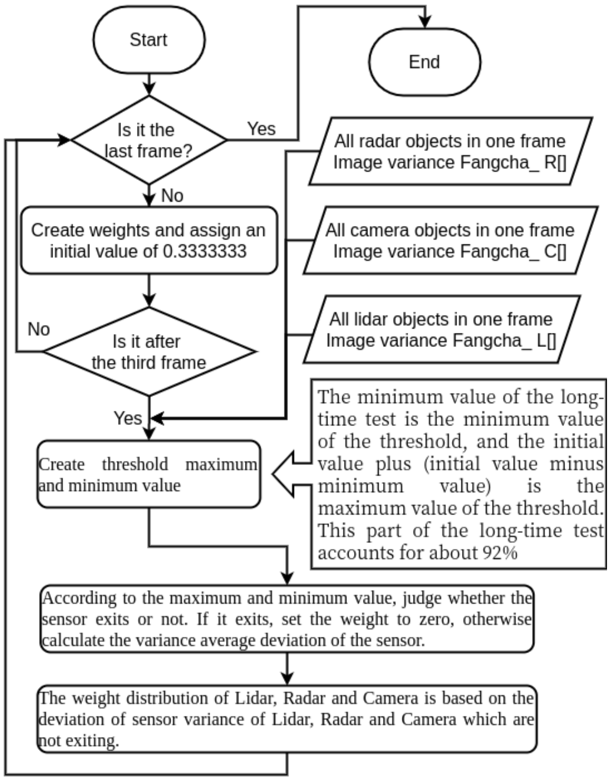

Figure 7. Flow chart of dynamic weight distribution and sensor eliminate system

The system of figure. 7 distributes weights based on the variance of each sensor and the maximum and minimum thresholds. Through this dynamic weight distribution and fusion system, we can obtain real-time weight distribution and real-time sensor fusion objects.

The multi-sensor target fusion issue is divided into two aspects. One is the multi-sensor target association issue, and the other is the multi-sensor target fusion issue. The multi-sensor target association algorithm and the multi-sensor target fusion algorithm respectively solve the above two problems.

On the basis of multi-sensor target association and information fusion, the statistical evaluation index of each sensor is obtained, and each sensor is given a corresponding weight. The failure is artificially simulated. In this case, the entire system can still perform the target with accurate perception.

IV. Experimental Results

This chapter mainly covers the dynamic weight distribution. Whether this part can run successfully ensures the stability of the system. The dynamic weight distribution of the system is based on a single frame. The weights of the sensors are allocated in each frame to ensure that the system can update the weights in time. Meanwhile, we should ensure that the system can process redundant data in time and make correct perception results.

A. Dynamic weight distribution

In order to avoid failure when single sensor detects objects, we must set an eliminate and a dynamic weight mechanism for the sensor. For sensors with poor tracking performance, reduce the weight in time or eliminate the fusion of the sensor, do not use the sensor to detect the object.

The figure below is the dynamic weight distribution curve. The black line is the Radar weight, the red line is the Lidar weight, and the green line is the Camera weight.

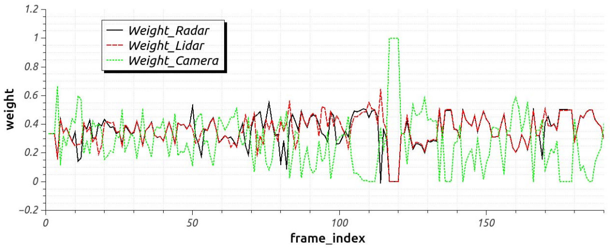

Figure 8. Weight curve for dynamic weight distribution

B. Test the weight reduction and eliminate of the sensors

This part is mainly to test whether the eliminate mechanism of the system and the weight reduction mechanism work. We take the Lidar as an example to explain the work of this part of the test:

By using a large Gaussian Noise processing each frame to check whether the weight of the Lidar will drop or eliminate, the actual test found that if the Lidar tracking do not work well, the weight of Lidar can be lowered.

All curves in this section are represented as follows: the black line is the Radar weight, the red line is the Lidar weight, and the green line is the Camera weight.

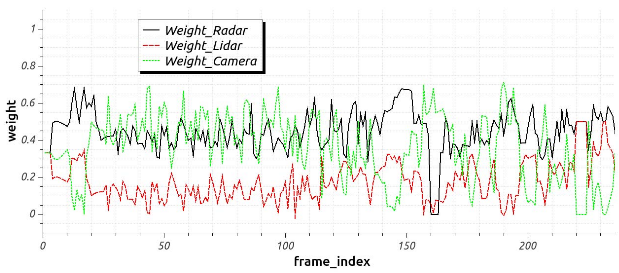

Figure 9. Weight curve of simulated Lidar failure

The Camera and Radar use the same method to test, and both can successfully reduce the weight of the sensor.

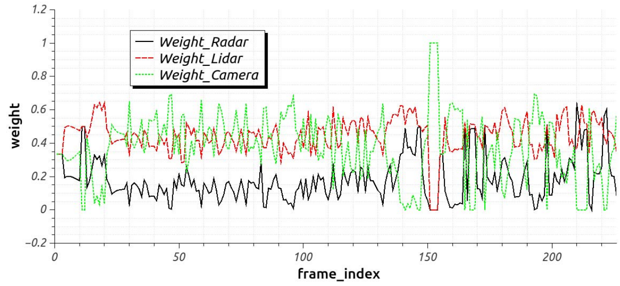

Figure 10. Weight curve of simulated Radar failure

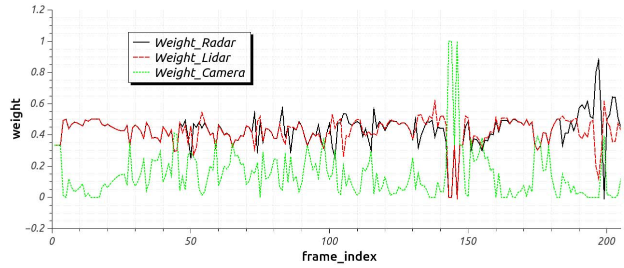

Figure 11. Weight curve of simulated Camera failure

It can be seen that the system can reduce the weight of sensors with poor tracking performance during operation, and there will be some frames with sensor eliminate. This is exactly what we need, which shows dynamic weight distribution and sensor eliminate mechanism.

C. Final sensor fusion results

Use the normal dynamic weights for fusion, and the results are shown in the following figure:

- The thin blue line represents the Lidar objects;
- The thin red line represents the Camera objects;
- The thin green line represents the Radar objects;
- The thick red line represents the fused objects.

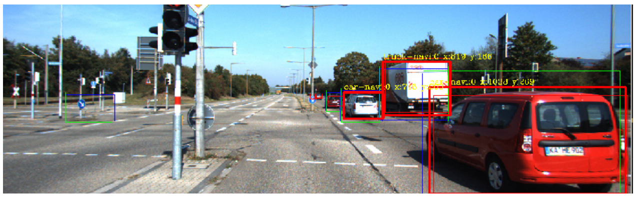

Figure 12. Sensor fusion result

V. CONCLUSION

Using dynamic weight distribution mechanism and sensor eliminate mechanism can greatly reduce the impact of sensor tracking failure on the multi-sensor sensing system, and thus minimize the traffic accidents caused by the failure of the sensor perception system. Through dynamic weight distribution, the weight of sensors with poor tracking performance is actively reduced during the operation of the multi-sensor fusion system. In addition, by the threshold setting, the objects of the sensor that have a particularly poor tracking effect are eliminated. The settings of the above two mechanisms ensure the stable of the fusion data and ensure the smooth operation of the fusion system.

ACKNOWLEDGMENT

The work described in this paper was fully supported by a grant from the Key R & D project of Shandong Province (No. 2019GGX104107).

REFERENCES

[1] B. Li, T. L. Zhang and T. Xia, “Vehicle Detection from 3D Lidar Using Fully Convolutional Network,” Workshop on Are the Skeptics Right Limits and Potentials of Deep Learning in Robotics at the 12th Conference on Robotics - Science and Systems (RSS), Univ Michigan, Ann Arbor, MI, Jun. 2016.

[2] A. Bochkovskiy, Ch. Y. Wang and H. Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv e-prints, Apr. 2020;arXiv:2004.10934.

[3] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan and O. Beijbom, “nuScenes: A multimodal dataset for autonomous driving,” arXiv preprint, 2019;arXiv:1903.11027.

[4] G. Andreas, L. Philip and U. Raquel, “Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, Jun. 2012, pp.3354-3361.

[5] G. Andreas, L. Philip, S. Christoph and U. Raquel, “Vision meets robotics: The KITTI dataset,” INTERNATIONAL JOURNAL OF ROBOTICS RESEARCH, 2013;32(11):1231-1237.

[6] R. E. Kalman, “A new approach to linear filtering and prediction problems,” Univ of North Carolina System, 1960.

[7] R. E. Kalman, “New Results in Linear Filtering and Prediction Theory,” Univ of North Carolina System, 1961.

[8] K. HW, “The Hungarian Method for the assignment problem,” NAVAL RESEARCH LOGISTICS, Feb. 2005;52(1):7-21.

[9] A. Bewley, Z. Ge, L. Ott, F. Ramos and B. Upcroft, "Simple online and realtime tracking," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 3464-346.