Research on Information Fusion Method of Millimetre Wave Radar and Monocular Camera for Intelligent Vehicle

Mingchi Feng\(^1\), Yunkun Chen\(^1\), Taixiong Zheng\(^1\), Ming Cen\(^2\) and Hong Xiao\(^1\)

\(^1\) College of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

\(^2\) College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

*Corresponding author’s e-mail: fengmc@cqupt.edu.cn

Abstract. The forward traffic environment cannot be fully perceived by a single sensor or multiple homogeneous sensors in the driving condition. Therefore, it is necessary to fuse the heterogeneous sensors to realize the mutual cooperation and compensation. Aiming at the advantages and disadvantages of different vehicle sensors, millimeter wave radar and monocular camera are selected as vehicle sensors to perceive the forward information. Firstly, the coordinate relationship between the sensors is calculated and sensor coordinates are mapped into the same vehicle coordinate system. Secondly, target information from the sensors are processed in parallel. Thirdly, in the case of time and space synchronization of two kinds of signals, the signals are processed by the matching algorithm of Global Nearest Neighbor (GNN) and the two matched targets are combined into one by the weighted average method. Finally, the unmatched targets and matched targets are tracked to determine the final state by Extended Kalman Filter (EKF) algorithm. Results in real car environment reveal that the fusion method can make up the deficiency of single sensor and improve the recognition rate of target.

1. Introduction

In the field of intelligent vehicle, a lot of researches on the environmental perception technology based on multi-sensor fusion has been studied in recent years. By weighting the advantages and disadvantages of these sensors, multi-sensor fusion is the best choice to gain the balance of each sensor’s features and achieve improved results. In the state-of-the-art, literatures\(^1,2\) have made an in-depth study on the development, existing problems and fusion algorithm of multi-sensor fusion. Some researchers\(^3,4\) project the point target detected by radar into the image plane, and generate a rectangular region of interest around the point. Then, search and verify the region to determine whether the radar target exists. This kind of method adopts region of interest can obviously shorten the detection area and that largely save the vision performing time. But it is difficult to verify the real obstacle that failed to detected by radar. Another method for multi-sensor fusion is based on uncertainty reasoning. Bayesian estimation method\(^5\) is presented and achieved relatively high recognition accuracy in vehicle identification. But the redundant information lack of consideration. Evidence theory\(^6,7\) can be regarded as the extension of subjective Bayesian estimation method. However, in \(^6\), the definition of data assignment is abstract and complex. In reference \(^8\), the target detected by LIDAR is used to generate region of interest. Then, radar objects and camera objects are recognized respectively in the region of interest. Finally, the D-S evidence theory algorithm is used to fuse radar targets and camera targets. Although the reliability of the algorithm is increased to a certain
extent with the increase of the number of sensors. It will also lead to the increase of redundant information and the complexity of the fusion system.

This paper proposed an advanced method to fuse the monocular camera and radar. In the detection stage, deep neural network based multi-object recognition and classification are applied to the vision detection. The MMW radar returns the position of objects and relative speed to the computer through a CAN bus. In the fusion stage, the vision outputs and radar outputs are associated, then the associated and not associated objects are delivered to track.

2. Sensor calibration

2.1. Camera calibration

The 3d world is projected into the 2d plane through the pin-hole model, but there will be geometric error (geometric distortion) in the actual image. In order to reduce the errors, the transformation can be expressed as Eq. (1)

\[
\begin{bmatrix}
    u_c \\
    v_c \\
    1
\end{bmatrix}
= 
\begin{bmatrix}
    \alpha & \gamma & u_0 & 0 \\
    0 & \beta & v_0 & 0 \\
    0 & 0 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
    R \\
    T \\
    0 \\
    1
\end{bmatrix}
\begin{bmatrix}
    x_w \\
    y_w \\
    z_w \\
    1
\end{bmatrix}
\tag{1}
\]

Where \([u_c, v_c, 1]^T\) and \([x_w, y_w, z_w, 1]^T\) stand for object’s image coordinates and world coordinates, \(z_c\) is the height of mounted camera, \((\alpha, \beta, \gamma, u_0, v_0)\) and \((R,T)\) are camera’s intrinsic parameters and extrinsic parameters.

2.2. Calibration between camera and radar

Knowing the transformation between radar and image coordinates is critical for accurate object fusion.

In order to unify the object information, we keep all the target information into the same vehicle coordinate system and convert the image plane into the bird’s eye view, so that the camera plane will coincide with the radar plane. The experimental environment in this paper is supposed to be flat road, so the height of the camera and radar is fixed in the vehicle coordinate system, which is 1.59 meters and 0.47 meters respectively. The horizontal distance from the radar to the camera is 2.382 meters. To determine the homographic matrix of the image captured by the camera into the bird’s eye view. A transition matrix \(P\) can be calculated through least square method to build the gap between radar plane and image frame. Figure 1 shows the experimental diagram of coordinate system transformation.

![Figure 1. Experimental diagram of coordinate system transformation](image)

In figure 1, \(O_c - X_cY_cZ_c\) is camera coordinate system, \(O_r - X_rY_rZ_r\) is radar coordinate system, \(XOY\) is vehicle coordinate system. A, B, C and D are the reference objects. Each reference object’s
homogeneous coordinates in radar plane and image plane are \([x_r \ y_r \ 1]^T\) and \([u \ v \ 1]^T\), the transformation is shown in Eq. 2.

\[
\begin{bmatrix}
  x_r \\
  y_r \\
  1
\end{bmatrix} = P
\begin{bmatrix}
  u \\
  v \\
  1
\end{bmatrix}
\]

(2)

3. Object detection

3.1. Monocular camera

Inspired by Single Shot Multi-Box Detector (SSD)[9], on the basis of YOLO[10] and faster-RCNN[11], multiple convolutional layers in different scales feed the judgment network are added, which makes the accuracy almost unchanged with the faster-RCNN, and the speed is faster than that of YOLO, we adopt it to train our own model.

A total of 8287 images are labelled, where the ratio of training set to test set is 4:1, and the detection accuracy of test set is 74.8%. Figure 2 is the detection result of a certain image.

![Camera detecting results](image)

Figure 2. Camera detecting results

The results are a list of box information \(C = [c_1 \ c_2 \ \cdots \ c_m]\), \(m\) denotes the number of objects, \(c_i = [x_c \ y_c \ w \ h \ class \ p], i \in [0, m]\), where \([x_c \ y_c]\) means the box’s top-left coordinates, \(w\) and \(h\) means widths and heights respectively, \(p\) presents the possibility of a certain class.

3.2. Millimeter wave radar

The list of unclassified observed radar objects is reported 20 times per second. In the actual measurement, part of the target signal acquired by radar is empty, invalid and static. The interference of these target signals should be removed first. These targets are excluded when they do not meet the threshold of distance or point of view or relative velocity. The remaining targets may also be an invalid signal, which has a short time and insufficient continuity. Therefore, it can be judged as an invalid signal if the change value of the same target parameter is too large for multiple consecutive times or in the adjacent sampling period.

Radar objects are a list of point information \(R = [r_1 \ r_2 \ \cdots \ r_n]\), \(n\) denotes the number of objects, \(r_j = [\rho \ \theta \ v], j \in [0, n]\), where \(\rho, \theta\) and \(v\) are distance, angle and relative velocity.

4. Fusion algorithm

When working with different sources of sensor data, it is critical to consider the similarity of each list of detections. First of all, camera objects and radar objects are projected to the same vehicle coordinate system. Then, the global nearest neighbour (GNN) approach is utilized to perform association of vision objects radar objects, when meeting with not associated objects, it is sometimes because of false alarms or misdetections. So, the next step, it’s necessary to keep track of these data in several frames.
We use Extended Kalman Filter (EKF) to confirm the detections and update the state of each object. Figure 5 shows the basic workflow of multi-sensor fusion.

5. Experiments
In order to verify the feasibility of the fusion algorithm, the experimental condition in this paper is on a flat road, and the vehicle speed is maintained at a speed of about 20km/h. During the running of vehicle, the radar and camera receive data and execute the fusion algorithm in parallel. The visualization results of radar target information, camera target information and fusion target information are saved to facilitate the analysis of experimental results. Then, several results are randomly selected and the ground truth of corresponding original images are labelled. Judging whether the fusion target is accurately positioned as shown in Eq. 3.

\[(x_i - x_j)^2 + (y_i - y_j)^2)^{\frac{1}{2}} < \Delta d\]  

Where \((x_i, y_i)\) and \((x_j, y_j)\) define the coordinates of the \(i^{th}\) fusion target and \(j^{th}\) labelled target. \(\Delta d\) is a fixed threshold.

As is shown in figure 4, which contains three parts of information, figure 4(a), 4(b) and 4(c). The left and right figure of each part represent the target information in the image and vehicle coordinate system respectively. In figure 4, two pedestrians and a car are detected by radar in figure 4(a) but failed to identify one of the pedestrians by camera in figure 4(b). However, the two pedestrians and one car in figure 4(c) are all correctly identified.
Figure 4. experiment results: (a) radar detection result; (b) camera detection result; (c) fusion result

The final saved results include total of 2500 images, we sample images at intervals of 5 frames and the final number of objects is 2428. The monocular camera correctly identified 1851 objects and the detection rate is 76.2%; The detection rate of radar is only 69.1%, which is poor relative to camera; However, 2122 targets are successfully identified based on our algorithm with lower false alarm and loss detection, the detection rate can up to 87.4%. Table 1 shows the detailed statistical results of the experiment.

| sensor | Total objects /2428 | Detection rate (%) |
|--------|---------------------|--------------------|
|        | correct detection   | false alarm        | loss detection    |
| camera | 1851                | 232                | 345               | 76.2 |
| radar  | 1678                | 339                | 411               | 69.1 |
| fusion | 2122                | 142                | 164               | 87.4 |

6. Conclusion

This article presented a fusion method to detect forward objects in the driving process of the intelligent vehicle. Sensors coordinate system and unified vehicle coordinate system are established. Effective object from the sensors are obtained in parallel. The whole fusion process is realized through object association and tracking. Results in real car environment achieve a better fusion function which increases the detection rate of target and makes up for the deficiency of single sensor. However, the future work will consider the complex road information, such as rough road.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (Grant No. 51505054), the Chongqing Science and Technology Commission (Grant No. cstc2015zdcy-ztzx60002,
cstc2015jcyjA70004, cstc2017rgzn-zdyfx0039), and the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. KJ1600430).

References
[1] Rashinkar P, Krushnasamy V S. (2017) An overview of data fusion techniques. In: 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). Bengaluru. pp. 694-697.
[2] Fung M L, Chen M Z Q, Chen Y H. (2017) Sensor fusion: A review of methods and applications. In: Control & Decision Conference. New York. pp. 165-178.
[3] Wang X, Xu L, Sun H, et al. (2016) On-Road Vehicle Detection and Tracking Using MMW Radar and Monovision Fusion. IEEE Transactions on Intelligent Transportation Systems, 17(7): 2075-2084.
[4] Zhang Y X, Feng K P. (2014) Research and Performance Analysis on Pedestrian Detection Based on Optimized Hog and SVM. Applied Mechanics & Materials, 511: 247-250.
[5] Chair Z, Varshney P K. (2007) Optimal Data Fusion in Multiple Sensor Detection Systems. IEEE Transactions on Aerospace & Electronic Systems, AES-22(1): 98-101.
[6] Morbee M, Tessens L, Aghajan H, et al. (2010) Dempster-Shafer based multi-view occupancy maps[J]. Electronics Letters, 46(5): 341-343.
[7] Oh S I, Kang H B. (2017) Object Detection and Classification by Decision-Level Fusion for Intelligent Vehicle Systems[J]. Sensors, 17(1): 207.
[8] Chavez-Garcia R O, Aycard O. (2016) Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 17(2): 525-534.
[9] Liu W, Anguelov D, Erhan D, et al. (2016) SSD: Single Shot MultiBox Detector. In: European Conference on Computer Vision. Amsterdam. pp. 21-37.
[10] Redmon J, Divvala S, Girshick R, et al. (2016) You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas. 779-788.
[11] Ren S, He K, Girshick R, et al. (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: International Conference on Neural Information Processing Systems. Istanbul. 91-99.