International Journal of Advanced Smart Convergence Vol.10 No.4 263-272 (2021)
http://dx.doi.org/10.7236/IJASC.2021.10.4.263

IJASC 21-4-31

A Vehicle Recognition Method based on Radar and Camera Fusion in an Autonomous Driving Environment

Mun-Yong Park*, Suk-Ki Lee**, Dong-Jin Shin***

**General Manager, Novacos Co., Ltd., Anyang, South Korea*
***Research Fellow, Korea Institute of Civil Engineering and Building Technology, South Korea*
****Assistant Manager, Novacos Co., Ltd., Anyang, South Korea*
pmy@novacos.co.kr, oksk@kict.re.kr**, djshin@novacos.co.kr****

Abstract

At a time when securing driving safety is the most important in the development and commercialization of autonomous vehicles, AI and big data-based algorithms are being studied to enhance and optimize the recognition and detection performance of various static and dynamic vehicles. However, there are many research cases to recognize it as the same vehicle by utilizing the unique advantages of radar and cameras, but they do not use deep learning image processing technology or detect only short distances as the same target due to radar performance problems. Radars can recognize vehicles without errors in situations such as night and fog, but it is not accurate even if the type of object is determined through RCS values, so accurate classification of the object through images such as cameras is required. Therefore, we propose a fusion-based vehicle recognition method that configures data sets that can be collected by radar device and camera device, calculates errors in the data sets, and recognizes them as the same target.

Keywords: *Camera, Radar, Radar and Camera Fusion, Vehicle Recognition, Autonomous Driving*

1. Introduction

With the development of artificial intelligence, IoT, and big data, vehicle recognition technology using radar and camera equipment used to monitor autonomous driving of vehicles and traffic on roads is gradually increasing [1, 2]. Radar and camera device are mainly used to recognize vehicles in the field, cameras derive two-dimensional images, and radars derive three-dimensional positions and speeds [3].

Since the device used for vehicle recognition derives different types of data, a technology of converging and recognizing different types of data is required for more sophisticated vehicle recognition [4]. Radar is continuously tracked by receiving a vehicle ID at the time of initial recognition, and even if multiple vehicles are recognized at the same time, location information is stored based on the ID, allowing the vehicle to be identified and the size of the vehicle can be classified through the Radar Cross Section (RCS) value [5]. In addition, cameras show high performance in the vehicle recognition field by precisely classifying six types

Manuscript Received: December. 5, 2021 / Revised: December. 8, 2021 / Accepted: December. 9, 2021
***Corresponding Author: djshin@novacos.co.kr
Tel: +82-31-423-2300, Fax: +82-31-423-2308
Assistant Manager, Novacos Co., Ltd., Anyang, South Korea

(large/medium/small/large'/medium'/small') based on deep learning-based image processing, but there is a limit to vehicle recognition due to factors occurring in night/all-weather environments [6].

Therefore, more accurate vehicle recognition is possible by combining deep learning-based image processing technology and radar vehicle recognition technology, which are advantages of cameras. In the previous studies, research on the development of a target vehicle selection technique using the fusion of radar and camera information has fused radar and camera using distance information, but it is difficult to use on general roads [7]. Research on the front vehicle collision warning system using radar/camera sensor fusion is a method of matching image information of cameras while mainly using radar, and has high accuracy using relative speed and relative distance, but it is difficult to achieve high accuracy due to the recent active deep learning-based image processing technology [8]. In addition, research on camera-radar coordinate system matching-based data fusion technology improved the accuracy of recognizing objects using external parameters of camera calibration, an image processing technology, but deep learning image processing technology was not applied as in the previous study [9]. Finally, although the target detection study using stereo vision sensor and radar sensor fusion reduced errors by correcting the projected coordinates of the image, there is a data distance problem of radar that can be detected, and deep learning image processing technology is not applied, so it is insufficient to recognize and classify vehicles [10]. Therefore, in this paper, an open source-based neural network model is used by learning data and image data extracted from radar and camera device. It is expected that high accuracy will be achieved by matching coordinate values indicating the location collected from radar, area size RCS with corrected coordinates and object size of images collected from camera, and reducing errors through correction.

Starting with the current introduction, this paper investigates and introduces cases of other studies that combine radar and camera, and section 3 proposes radar and camera fusion methods for accurate vehicle recognition. Finally, in section 4, the thesis is concluded with a conclusion.

2. Related Work

2.1 Development of a Target Vehicle Selection Technique using the Fusion of Radar and Camera Information

When using radar, there is a problem in that the vehicle recognition rate is lowered when driving on a steep curved road, complex traffic conditions, or frequent lane movement of the vehicle occurs. To improve this, the lane detected using the camera contains road environment information irrelevant to the driver's will, so if the driver maintains the lane, it is easy to predict the trajectory of the long-distance vehicle regardless of the dynamic movement of the vehicle. In this study, in order to solve the inaccurate trajectory prediction problem, a fusion technology was proposed using the long-range prediction performance of the camera and the near-field prediction performance of the radar. In this technology, short-distance radar path information, long-distance camera lane information, and path-lane data fusion method with an interconnected line structure of two trajectories using a tertiary curve in the middle region are composed.

Figure 1 shows the structure of the designed path-lane data fusion and the vehicle trajectories consist of path-only range, path-lane range, and lane-only range. And the overall fusion distance has a structure that is determined according to the road curvature [7].

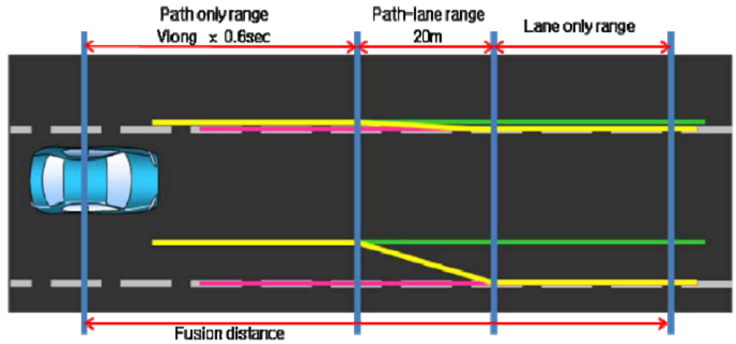


Figure 1. Path-Lane date fusion structure

As a result of the experiment, the fusion-based in-path target selection performance was significantly improved compared to the general radar system. However, since the detection distance of the lane detection system is basically 55m, it is insufficient to be used in all ranges.

2.2 Front Vehicle Collision Warning System using Radar/Camera Sensor Fusion

In this study, sensor fusion techniques were applied in a way that focuses on radar and auxiliary cameras. The radar can measure the exact relative distance and relative speed for the detected object regardless of the relatively external environment. On the other hand, the camera can measure lane information and characteristics of objects that cannot be measured by radar, but is relatively more vulnerable to changes in the external environment than radar. Therefore, both radar and camera sensor advantages were used, and sensor fusion techniques were used to compensate for each limitation.

Figure 2 shows a conceptual diagram of radar and camera target identification proposed in this study, and compares attributes such as the position and speed of objects detected by each sensor to determine the presence or absence of identities between individual objects detected by individual sensors [8].

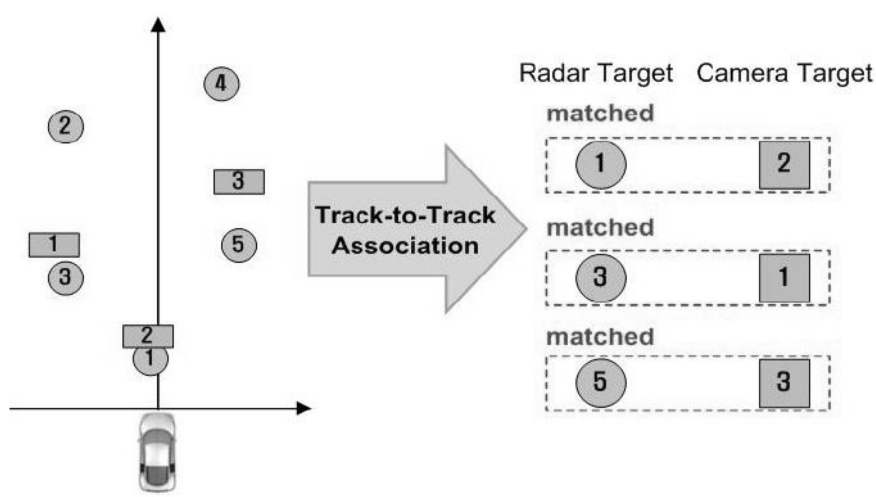


Figure 2. Conceptual diagram of the identification method

Although mis alarm was minimized by setting parameters using radar and camera sensor fusion technology and driver driving data analysis results, it is difficult to ensure the accuracy of vehicle recognition because deep learning-based image processing, which has been actively handled recently, has not been applied.

2.3 Data Fusion Technique based on Coordinate System Matching Between Camera and Radar

In this study, it is mentioned that in order to express different data of radar and camera in a specific coordinate system, the transformation matrix of camera-radar and the distance between sensors should be measured to calculate the transformation relationship between coordinate systems. However, since the distance between sensor coordinate systems is physically difficult to measure accurately, and there is a high possibility of an error due to measurement, a method of calculating an external parameter which is a conversion relationship between a camera and a radar coordinate system is proposed to solve the above problem.

In the case of the camera is shown in Figure. 3, a conversion relationship from a three-dimensional coordinate system to a two-dimensional image coordinate system appears through an internal parameter and an external parameter according to the pinhole camera model. In addition, the existing camera calibration method was used to calculate the internal parameters of the camera, and the external parameters were calculated by applying the proposed point-based matching method [9].

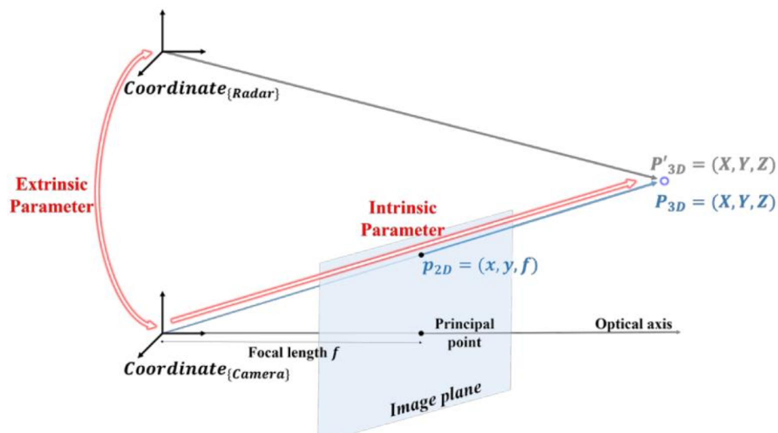


Figure 3. Camera-Radar coordinate system conversion relationship

Through the camera-radar fusion method proposed in this study, the matching error was confirmed from a minimum of 0.24m to 0.41m, confirming the result of an average of 0.3m. However, if the range of radar that can be detected is wide because only the coordinate system is used, large matching error results can occur, and it is difficult to show high accuracy because deep learning-based image processing, which has not been used much recently, has not been utilized.

2.4 Target Detection using Stereo Vision Sensor and Radar Sensor Fusion

In this study, a method of increasing the accuracy of distance information of targets detected by the camera using radar with relatively high accuracy of distance information was proposed. Multi-channel radar with high distance accuracy has a complex and expensive disadvantage in the system, so distance accuracy complements the camera with a low-cost, single-channel radar without angle information. The information on the target obtained from the camera is the coordinates, size, and approximate distance and angle of the target in the image plane, and the target information obtained from the radar is the distance and speed. Each of these information is represented by a matrix, of which a column containing distance information is extracted, numbered, and listed in order of distance to obtain respective column vectors v and r. Calculating the inter-distance error between all the targets (m) obtained from the camera and all the targets (n) obtained by radar gives a matrix D representing the difference between each term of the v and r vector as shown in the following equation.

$$D_{ij} = \left(v_i - r_j\right)^2 \tag{1}$$

$$D = \begin{bmatrix} D_{11} & D_{12} & \cdots \\ D_{21} & \cdots & \cdots \\ \cdots & \cdots & D_{11} \end{bmatrix} \tag{2}$$

In Equation (1), *(i, j)* means target *i* obtained from vision and target *j* obtained from radar. As mentioned above, in Equation (2), *m* is the number of targets detected by vision, and *n* is the number of targets detected by radar [10]. If the matrix selects elements with the minimum value of each row, i.e., the minimum error in the distance information of the two sensors and arranges them in *m:n*, the matching of the two sensor information is performed in the order of the fewest errors. Since this study aims to supplement the distance information of the vision, the distance accuracy is increased by selecting maximum m matching and replacing the distance information obtained from the vision with radar distance information with less error.

Figure 4 shows the target information obtained using vision and radar through the polar coordinate system. The unit of distance was expressed up to 5m in meters, and it unfolded up to 55° left and right according to the azimuth of the radar used. The actual location of the target is indicated by red dots, the location information of the target obtained only by vision is indicated by blue dots, and the location information obtained by the fusion system of vision and radar is indicated by green dots. In the case of Object #1 close to Figure 4, both the result obtained only by vision and the result obtained by the fusion system were almost the same as the actual value. On the other hand, in the case of long-distance Object #2, there is a large difference in distance obtained by the vision and fusion system, and it can be seen that the results supplemented through the target matching algorithm proposed in this study are consistent with the actual value.

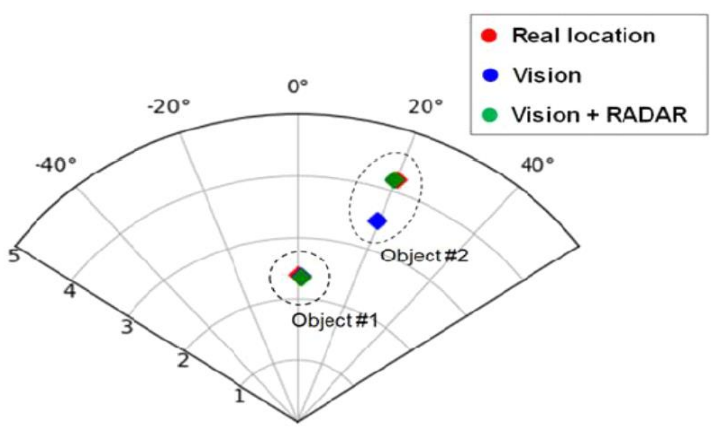


Figure 4. Target locations detected with vision only and vision data assisted by radar

Since the radar and the camera are fused to correct the distance information error, more accurate recognition is possible, and it is efficient because angle information of the target is obtained from camera image data. However, the distance at which objects can be detected on the road should be measured farther when using industrial radars, so performance may be measured differently depending on the radar device model.

3. Proposed Method

Section 3 proposes a method that enables more accurate vehicle recognition by fusing data into one data by using a radar and a camera. The proposed method first introduces data from radar and cameras used for fusion-based vehicle recognition, and introduces matching and fusion methods.

3.1 Radar Data for Fusion-based Vehicle Recognition

When the raw data processing and purification process collected from the radar device is completed, data that may be used for radar vehicle recognition may be extracted.

Table 1. Example of extracted radar data

ID	Time	Lane	Point x	Point y	Velocity X	Velocity Y	RCS
1	2021-09-24 13:35:53.646	1	-12.288	207.488	0.36	71.64	49
2	2021-09-24 13:35:53.646	1	-18.176	114.944	12.48	-54.72	117
3	2021-09-24 13:35:53.646	2	18.944	132.096	10.32	-40.43	85
...	...	...	...	...	...	...	...

Table 1 shows consist of ID, Time, Lane, Point x, Pointy, Velocity X, Velocity Y, and RCS as attributes of data extracted from radar device. ID means the vehicle's unique order, and Time means the date and time measured when the vehicle is detected. Lane is the lane number of the road where the vehicle is detected, Point x is the X coordinate position value where the vehicle is detected at the coordinates of the two-dimensional form based on the radar device, and Point y is the Y coordinate position value where the vehicle is detected at the coordinates of the two-dimensional form based on the radar device. Velocity X refers to the X coordinate speed value of the vehicle in a two-dimensional form based on the radar device, and Velocity Y refers to the Y coordinate speed value of the vehicle in a two-dimensional form based on the radar device. RCS refers to a measure of the size measured by reflecting a vehicle detected as a radar reflective area on the radar.

Based on the detected time of the vehicle, six types that can be recognized according to the size of the vehicle through ID, Time, Lane, Point x, Point y, Velocity X, and Velocity Y, which are attributes of radar data extracted from the lane of the road, are expressed using the radar reflection area, which is an RCS.

Figure 5 shows the picture expressed using the collected radar data. Depending on the RCS value, the size of the circle black colored circle varies, which is a criterion for dividing large/medium/small by vehicle size.

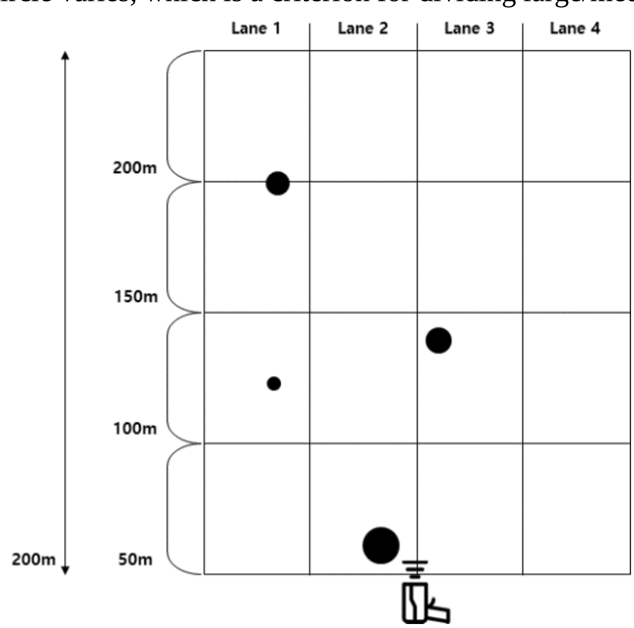


Figure 5. Vehicle detection using radar data

3.2 Camera Data for Fusion-based Vehicle Recognition

Image data captured from camera device should be converted into images, the meaning of images should be expanded through image labeling and annotation processes, and high-quality learning data should be secured by configuring metadata. There are six types of vehicles drawn as bounding boxes in labeling and annotation, and six types of vehicles in the image are labeled and annotated using an open source-based program to display the vehicle's bounding boxes in the image to build a training dataset.

Figure 6 shows an open source-based program for displaying a boundary box of an object in an image for Yolo neural network model learning. After converting the extracted image into a JPG form, it shows that data can be labeled by designating the area of the object by hand, which designates the ObJectID for each object to be detected.

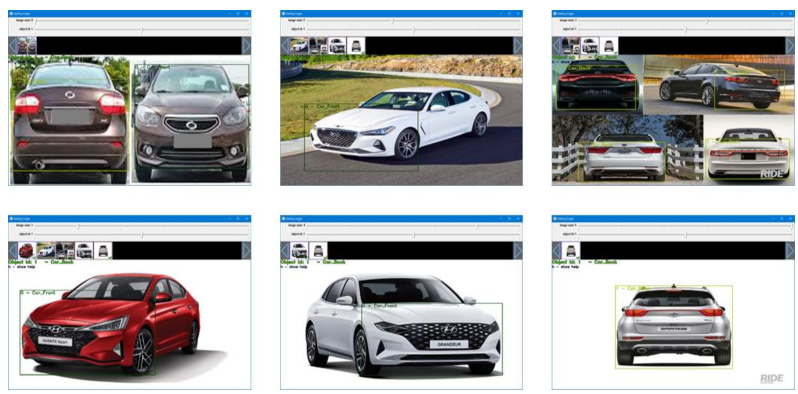


Figure 6. Labeling and annotation process

The vehicle recognition model using deep learning-based image processing utilizes an open source-based Yolo neural network to learn learning datasets that have completed the image labeling and annotation process. By inputting image data into the neural network model that has been trained, six types of vehicles are recognized, and accuracy (IOU) of which of the six types of vehicles are included is expressed.

Figure 7 shows predict the probability of correct answer by calculating the difference between the prediction box (combination) consisting of blue and red at the bottom of the top picture and the correct answer box (intersection) consisting of white and red, and the calculation method is as shown in Equation (3).

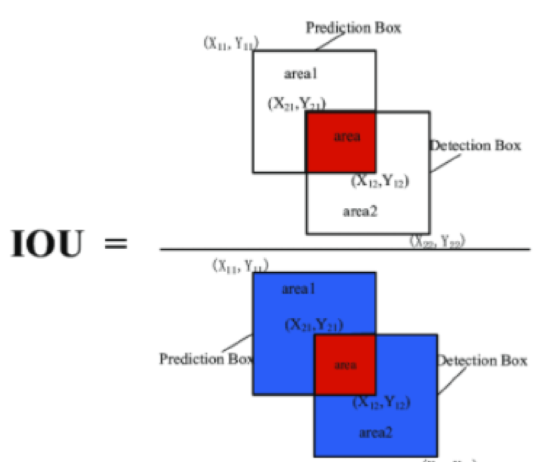


Figure 7. A deep learning-based vehicle detection and accuracy expression method.

$$IOU = \frac{PredictionBox \cap DetectionBox}{PredictionBox \cup DetectionBox} \tag{3}$$

3.3 Vehicle Recognition Method based on Radar and Camera Fusion

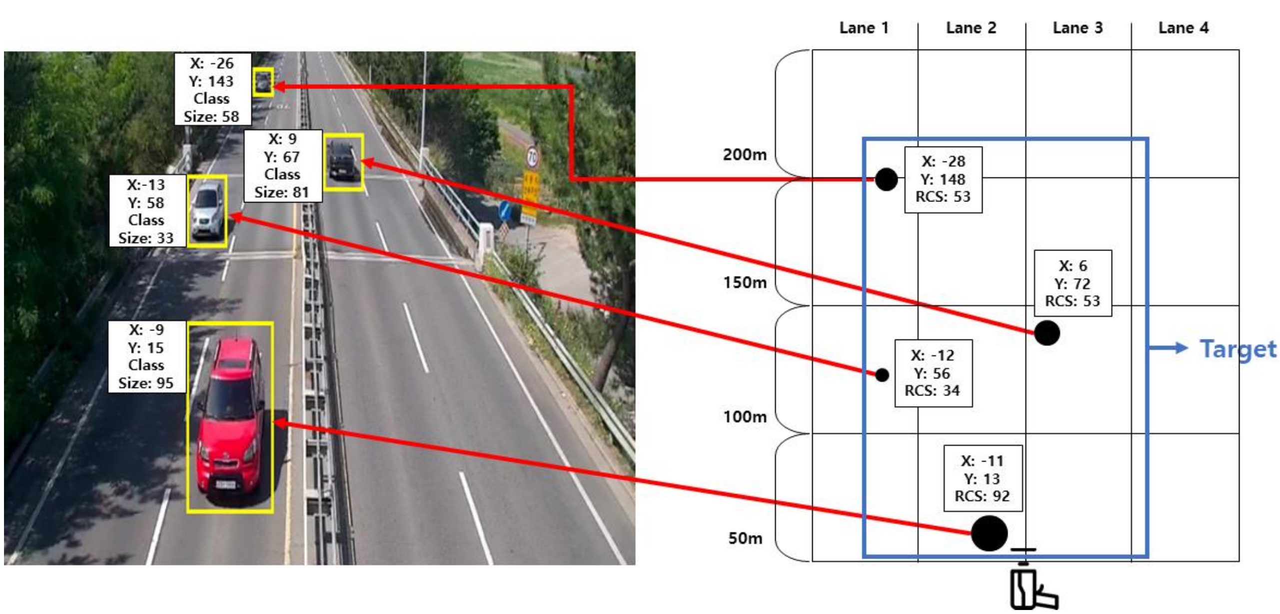
The radar-camera fusion-based vehicle recognition method consists of three stages, and the first stage is camera image position correction, and the image extracted from the camera is improved by applying maximum brightness processing to secure vehicle visibility. And as shown in Figure 8(a) in Step 1, the image is projected into two-dimensional *(X, Y)* coordinates, and the vehicle size of the metadata generated through the labeling and annotation process of image processing, X_{width} and Y_{height}, is calculated to construct the dataset as shown in Equation (4).

$$\begin{gathered} Camera\ Calibration = [X, Y, ClassSize] \\ where\ is\ ClassSize = X_{width} \times Y_{height} \end{gathered} \tag{4}$$

As shown Figure 8(b) In Step 2, data measured by radar vehicle detection is divided into six types of vehicles and noise data, and the target is set, and the coordinates *(X, Y)* and *RCS* values are composed through the extracted radar data Point X and Point Y.

$$Radar\ Detection = [X, Y, RCS] \tag{5}$$

In the last Steps, the range of values of the Camera Calibration dataset and the Radar Detection dataset is normalized between 0 and 1 by matching and converging radar and camera vehicles. In addition, it is the process of specifying and calculating the threshold error value of the *ClassSize* calculated through the coordinates of the radar/camera and the camera image position correction and the *RCS* value generated by radar vehicle detection. Figure. 8(a) and Figure 8(b) shows the information about the radar data of each area and the same camera target performed in Steps 1 and 2 with arrows.



(a) **(b)**

Figure 8. Vehicle recognition plan based on radar and camera fusion

A calculation method for recognizing a vehicle of the same target is shown in Equation (6). First, through the normalization process of camera datasets and radar datasets configured through Steps 1 and 2, the elements of the dataset are reduced one by one. And the error between the coordinate values (*CC.X'*, *RD.X'*, *CC.Y'*, *RD.Y'*) of the normalized radar and camera datasets and the size of the object area (*CC.ClassSize'*, *RD.RCS'*) to calculate If it is less than 0.9 after calculating the error, it returns True and recognizes it as the same target.

$$\begin{gathered} Perform\ Normalization\ of\ CM\ Set\ and\ RDSet\ \left(e.g.\ CC.X' = \frac{CC.X - CC.X_{min}}{CC.X_{max} - CC.X_{min}}\right) \\ (CC.X' \varDelta\ RD.X' \times CC.Y' \varDelta\ RD.Y' \times CC.ClassSize' \varDelta\ RD.RCS') < 0.9 \equiv True \\ where\ is\ CM = Camera\ Calibration, RD = Radar\ Detection \end{gathered} \tag{6}$$

4. Conclusion

In camera device, high-accuracy object detection and classification is possible through deep learning image processing technology, which is being actively used recently, but it has the disadvantage that detection is impossible at night and in bad weather environments. Radar device has higher accuracy to detect an object than a camera, but it is difficult to recognize which object it is. Therefore, in this paper, we proposed a method for recognizing the same target through the fusion of data sets of radar and camera device, which are widely used in autonomous driving and traffic. We proposed method reduces the range of values through normalization of the data set generated by each device, calculates the error of the values collected from each device, and determines whether the target is the same according to the threshold value. In particular, it is expected to realize safer autonomous driving AI technology by providing high-accuracy vehicle recognition rate and detection rate even in night/all-weather environments through radar and camera fusion technology.

Acknowledgement

This work was supported by Korea Institute of Police Technology (KIPoT) grant funded by the Korea government (KNPA) (No.202100200201, Development of Advanced Technology for Perception Improvement on Traffic Entities and Risk Mitigation on Adverse Driving Conditions for Lv.4 Connected Autonomous Driving)

References

[1] D. J. Shin, J. H. Park, J. H. Kim, K. J. Kwan, J. M. Park and J. J Kim, "Big Data-based Sensor Data Processing and Analysis for IoT Environment,", *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, Vol. 19, No. 1, pp. 117-126, Feb 2019.
DOI: https://doi.org/10.7236/JIIBC.2019.19.1.117

[2] S. Y. Kim, Y. S. Yun, S. B. Eun, S. C and J. M. Jung, "Implementation of Autonomous IoT Integrated Development Environment based on AI Component Abstract Model," *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, Vol. 21, No. 5, pp. 71-77, Oct 2021.
DOI: https://doi.org/10.7236/JIIBC.2021.21.5.71

[3] J. S. Lee, M. G. Kim and H. K Kim, "Camera and LiDAR Sensor Fusion for Improving Object Detection," *Journal of Broadcast Engineering*, Vol. 24, No. 3, pp. 580-591, July 2019.
DOI: https://doi.org/10.5909/JBE.2019.24.4.580

[4] M. C. Lee, J. H. Han, C. H. Jang and M. H. Sunwoo, "Information Fusion of Cameras and Laser Radars for Perception Systems of Autonomous Vehicles," *Journal of Korean Institute of Intelligent Systems*, Vol. 23, No. 1, pp.

35-45, Feb 2013.
DOI: https://doi.org/10.5391/JKIIS.2013.23.1.35

[5] C. Marco, A. Giorgetti and E. Paolini, "Sensor radar for object tracking," *Proceedings of the IEEE*, Vol. 106, No. 6, pp. 1022-1041, May 2018.
DOI: https://doi.org/10.1109/JPROC.2018.2819697

[6] H. Sun, C. Wang and B. Wang, "Night vision pedestrian detection using a forward-looking infrared camera," *In 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, pp. 1-4, Jan 2011.
DOI: https://doi.org/10.1109/M2RSM.2011.5697384

[7] J. Y. Hwang and S. B. Park, "Path-Lane Data Fusion Method for In-Path Target Selection," *KSAE Annual Conference*, pp. 1807-1809, Nov 2010.

[8] S. W. Moon, I. K. Moon and K. K. Shin, "Forward Collision Warning System based on Radar driven Fusion with Camera," *Journal of Auto-Vehicle Safety Association*, Vol. 5, No. 1, pp. 5-10, June 2013.
DOI: https://doi.org/10.22680/kasa.2013.5.1.005

[9] C. H. Song, H. G. Kim, G. J Son and Y. D. Kim, "Data fusion technique based on coordinate system matching between camera and radar," *The 30th workshop on video processing and understanding*, Vol. 30, No. 01, pp. 2-6, Feb 2018.

[10] J. S. Kim, W. H. Park, P. H. Yoon, B. S. Kim and R. Song, "Object Detection Using Stereo Vision and Radar Sensor Fusion," *The Journal of Korean Institute of Electromagnetic Engineering and Science*, Vol. 31, No. 11, pp. 986-989, Nov 2020.
DOI: https://doi.org/10.5515/KJKIEES.2020.31.11.986