Vehicle Detection and Width Estimation in Rain by Fusing Radar and Vision*

Jian-Gang Wang, *Senior Member, IEEE*, Simon Jian Chen, Lu-Bing Zhou, Kong-Wah Wan, *Senior Member, IEEE*, and Wei-Yun Yau, *Senior Member, IEEE*

Abstract—**While much effort has been devoted to deep learning object detection, relatively limited attention has been paid to object detection in bad weather, e.g. rain, snow or haze. In heavy rain, the raindrop on the front windshield can make it difficult to detect object from an in-car camera. The conventional way to cope with this has been to use radar as the main detection sensor. However, radar is highly susceptible to false positives. Furthermore, many entry level radar sensors only return the centroid of each detected object, rather than its size and extent. In addition, due to lack of texture input, radar cannot discriminate a vehicle from a non-vehicle object, e.g. roadside pole. This motivates us to detect vehicle by fusing radar and vision. In this paper, we first calibrate the radar and camera with respect to the ground plane. The radar detections are then projected to the camera image for target width estimation. Empirical evaluation on a large database shows that there is a natural synergy in both sensors, as the image based estimation is found to be greatly facilitated by the accuracy of the radar detection.**

I. INTRODUCTION

Autonomous vehicles have attracted more attention in recent years [10, 23-25] because of the requirement for traffic safety, as well as significant progress in autonomous vehicle (AV) hardware and software. In the context of AV, the basic requirement is to follow the front vehicles safely and to prevent rear-end collision. Vehicle detection is important for a driver to understand the behavior and intention of the vehicles ahead. The detection can be achieved by using different sensors, including camera [10], IR camera [27], lidar [15], radar [3], or their fusion [11, 26]. A good survey on on-road vehicle detection can be found in [21].

While much effort has been devoted to deep learning object detection and classification, relatively limited attention has been paid to practical object detection in bad weather, e.g. rain, fog or snow. As an AV typically performs detection via an in-car camera, the quality of the images is affected seriously when it is raining, particularly by the raindrops that remain on the windscreen. De-raining methods [4, 6, 22] may help to improve the image quality. However, their applicability is severely limited because their underlying image models may be very different from the rain images captured by our in-car camera. Furthermore, de-rain methods are very compute intensive. Although the raindrop detection and removing from in-vehicle camera have been discussed in [28, 13, 9], the proposed approaches are not applicable to our application which requires to be executed in real time under heavy rain conditions.

The conventional way to address objection detection in rain has been to use radar as the main detection sensor. However, it has several significant drawbacks that motivate us to fuse it with computer vision. Firstly, although radar can discriminate target from raindrop, false positive could happen due to the rain reflection on the road. With vision, we expect that the false positive rate can be reduced. Secondly, radar cannot discriminate a still vehicle on the road from a non-vehicle object, e.g. roadside pole, a task that is likely to be easier for vision. Thirdly, most entry level radar can only return the distance, velocity and angle, but without any size and extent of the target. The latter is especially important for AV to avoid on-road obstacles. When fused with vision, these vehicle attributes can now be more easily and accurately estimated.

In this paper, we describe a practical AV system to detect vehicles in rain. To effectively avoid rear-end collision, we accurately localize all front vehicles by fusing detections from radar and vision. We first calibrate the two sensors such that the data provided by the radar can be projected to image and vice versa. The centroid of target provided by the radar includes a range and a bearing angle but not the vertical height. Hence, it is difficult to calibrate the two sensors directly because the image coordinates of the centroid provided by the radar are unknown. In this paper, a new calibration method is proposed to solve this problem based on the fact that the radar plane is parallel to the ground plane. You Look Only Once (YOLO) version 2 [18], one of state-of-the-art object detectors, is adopted to detect vehicles from an image. By comparing the detection results on obtained by vision and radar on the image, some false positives can be removed. The width of the detected vehicle is estimated from the bounding boxes. By using the accurate longitudinal distance and bearing angle, provided by the radar, the vehicle width estimation error caused by the image blur can be reduced. Finally, a tracking is developed to remove false positive.

* This research is supported in part by the A*STAR Grant for Autonomous Systems project, Singapore.

Jian-Gang Wang, Kong-Wah Wan and Wei-Yun Yau are with the Robotics Department, Institute for Infocomm Research, 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632 (e-mail: {jgwang, kongwah, wyyau}@i2r.a-star.edu.sg).

Simon Jian Chen is with Desay SV Automotive, 3A International Business Park #09-13 Tower B, ICON@IBP Singapore 609935. This work was done when he was a research scientist with the Autonomous Vehicle Department, Institute for Infocomm Research (email: simon.chen@desay-svautomotive.com).

Lu-Bing Zhou is with nuTonomy, 77 Ayer Rajah Crescent, Singapore 139954. This work was done when he was a research scientist with the Autonomous Vehicle Department, Institute for Infocomm Research (e-mail: zlubing@gmail.com).

In the rest of the paper, we first list some related works in Section II. We then describe the fusion of vision and radar in Section III, and the real time object detection in Section IV. For robust detection, we describe a tracking approach in Section V. The evaluation method is given in Section VI. The experimental results are discussed in Section VII. In the final, the conclusion and future work are given in Section VIII.

II. RELATED WORKS

In the literature, detecting objects in rainy condition has attracted much attention [3, 5]. The authors in [3] disclose a method for better discriminating targets in rain conditions. A range-frequency map is calculated from the radar data, and cells associated with signals having a higher spectral power than noise, corresponding to reflections from raindrops, are identified on the map. The signals of rain-cells are replaced by thermal noise signals, thus reducing the false alarms associated with the rain signal. In [5], the authors disclosed a method for obtaining data associated with an object of interest under rainy conditions using a millimeter-wave radar detection system, while mitigating the effect of rain clutter in the reflected radar signal. The disclosed technique is applied to a W-band, repetitive linearly frequency modulated continuous-wave (FMCW) type radar. FMCW radar facilitates processing techniques which derive simultaneously the range and velocity of the detected objects, displayed as cell patterns in two-dimensional cell arrays known as range-velocity maps. A cell pattern associated with rain is used to apply an object detection scheme that is minimally affected by the radar clutter resulting from the rain.

There have been a few papers [26, 14, 12, 2, 1, 11] on fusion of radar and vision for vehicle detection. Wang et al [26] applied active contour detector within the region-of-interest provided by a radar which has been calibrated with the camera. Song et al [19] presented an obstacle detection approach in which radar data is combined with stereo vision. The obstacle detection range is limited by the working range of the stereo camera. Langer and Jockem [14] describe an integrated radar and vision sensor system for on-road navigation. Range and angular information of targets from the radar are obtained by Fast Fourier transform. Detected targets are kept in an object list, which is updated by successive data frames from the radar sensor. Target information is fused with road geometry. To the best of our knowledge, we are the first to investigate vehicle detection in the rain condition by fusing radar and vision. The vehicle detection in the rain condition is much more challenging than the case of the normal weather conditions.

III. SENSOR FUSION

Radar have an important role to play in field of robotics for applications that require reliable perception in challenging environmental conditions, e.g. in the presence of fog, airborne dust. Radar and camera compensate for each other in terms of feature. As we mentioned above, radar provides distance, angle and velocity information without texture. On the contrast, camera is rich in texture. In order to improve the accuracy and robustness of vehicle detection, radar and camera are combined in this paper.

The Delphi's multimode Electronically Scanning RADAR (ESR) [20], used in this paper, provides information on the estimated centroids of the targets including the range to the centroid, its bearing angle, its longitudinal and lateral speeds, its acceleration and power of the returned signal. It does not provide any direct measurement of the size of the target which is necessary for autonomous navigation. The sensor data are transmitted from the radar to the computer using a CAN communication protocol at a normal bus speed, and the measurement rate is 50Hz. ESR combines a wide field of view at midrange with long-range coverage to provide two measurement modes simultaneously. In this paper, we use long-range which is able to detect objects as far as 180m ahead. In particular, it provides accurate range and speed data with powerful object discrimination that can identify up to 64 targets in the vehicle's path, with longitudinal Root-Mean-Square (RMS) error of 1.8m and lateral RMS error of 0.4m [20].

Three ESRs are installed in the left, frontal, and right of the AV, respectively, to cover the full field-of-view. In this paper, we discuss the frontal radar only.

Meanwhile, a high resolution (1600 × 1200) camera is installed inside the autonomous vehicle, below the rearview mirror. The fusion of the radar and ESR will be discussed in the following section.

A. Calibration

As the radar and camera are rigidly fixed to the vehicle, the rigid relationship between them does not change when the vehicle moves. Hence, the fusion of radar and camera can be done by pre-calibrating them. Although complex 3D calibration, e.g. [16], can be applied, we develop a simple approach to calibrate the relationship between radar and camera.

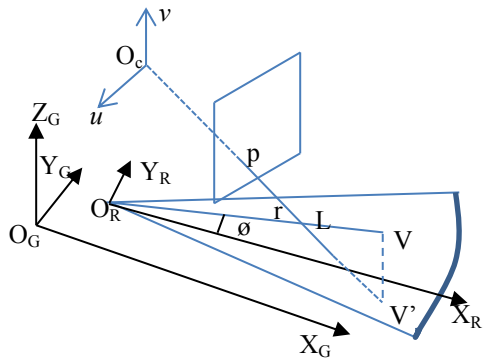

Fig. 1 The calibration of the radar and camera. O_c and O_r are the origins of the camera and radar system, respectively. The radar is installed inside the autonomous vehicle's frontal bumper such that the radar plane (X_RY_R plane) and ground plane (X_GY_G plane) parallel to each other. Polar coordinates, V(r, ø), of the target are provided by radar can be converted to Euclidean data, (x_r, v_r), using equations (1) and (2) but not the elevation angle. V' is the vertical projection of V on the ground plane. The Light ray L and the projected point p in the image O_C are shown together with the horizontal radar plane.

The radar is installed inside the frontal bumper such that the radar plane and the ground parallel to each other. The relationship of radar, camera and ground coordinate systems

is shown in Fig. 1 where O_C , O_R and O_G are the origins of the camera, radar and ground system, respectively. Polar coordinates V(r, ø) of the target are provided by radar can be converted to Euclidean data but not the elevation angle with the following formulas:

$$x_r = r \cos ø \quad (1)$$

$$y_r = r \sin ø \quad (2)$$

V' is the vertical projection of V on the ground plane. The Light ray L and the projected point p in the image are shown together with the horizontal ground plane.

The transform matrix, **A**, between ground plane and image plane is calibrated. The polar coordinates provided by the radar are converted to Euclidean (x_r, y_r). The coordinates (x_r, y_r) can be converted to the ones on the ground by a translation (d_x, d_y) which is pre-calibrated because the radar plane is parallel to the ground. By reading the coordinates of corners (on the ground, at least four corners are needed) on the calibration objects in both camera and radar, a transform matrix **A** (3 × 3) is computed which can transform radar coordinates to the image and vice versa.

$$\begin{bmatrix} ut \\ vt \\ t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix} \quad (3)$$

where (u, v) are the coordinates of a pixel on the image, (x_r, y_r) are the coordinates of a point in the radar coordinate system.

In order to find the corresponding corners between the radar and vision accurately, a calibrating object, shown in Fig. 2, is designed. It is made of metal material which is easily measured by radar. The positions of the centroid in the radar coordinate system can be computed from the distance to the centroid and the angle. The positions of the centroid (ground) can be read from the image. By putting the calibration object to some premeasured positions on the ground, a set of pair of radar and image coordinates can be collected and a transform matrix can then be computed.

Fig. 2 The calibration object used in our experiment. As the radar plane is parallel to the ground and the calibration object is perpendicular to the ground, the ground coordinates of the point marked in red, (Xg, Yg) are the same with the radar coordinates of the centroid provided by the radar, marked in yellow, (Xr, Yr). Hence, the relationship between the radar and camera can be calibrated by using the centroid, (Xr, Yr), provided by the radar and image coordinate, (u, v), of the point (Xg, Yg).

In order to solve the equation (3) which has 8 unknown parameters, we have to provide at least four pairs of image and radar coordinates. In the experiments, we use more than four pairs to cover more ground region. Hence, the parameters are solved by a least-square error algorithm.

A. Fusion of radar and vision for vehicle detection

There are three kinds of sensor fusion schemes: (1) sensor-level which integrate two or more sensors together by special hardware; (2) feature-level which extract feature from each sensor, respectively and then fuse the features; (3) decision-level which make decision separately and the final decision is made by fusing the decisions.

In this paper, we adopt decision-level fusion because radar we used provides detection results directly rather than raw data. By projecting the radar detection results to image using above-mentioned calibration, we can improve the width estimation accuracy by fusing the results obtained by using radar and vision.

B. Measure the width of the vehicle

The width of the detected vehicle is likely to have significant impact on the vehicle detection for two reasons. (1) A particular radar we used in this paper provides the positions of the centroids with no size of the target. The accuracy of the positions of the centroids provided by the radar is expected to vary with the width of the target; (2) The width information is important for autonomous vehicle's navigation, e.g. obstacle avoidance.

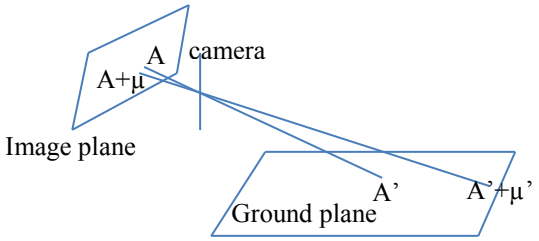

Fig. 3 A small disturbance on the image could cause a large error for its projection on the ground. A' is the projection of A. The projection is changed to A'+μ' as the measurement of A on the image is A+μ. The projection error μ' is large even μ is small.

In this paper, the width of the target is computed using the 3D distance between the bottom left and bottom right corners of the target. The corners on the ground can be located by projecting the corresponding corners of the bounding box to the ground. However, the corners of the bounding box on the image could be higher or lower than the true positions because the image is blurred by rain. Unfortunately, the width estimation accuracy is very sensitive to the corners' image coordinate especially when the distance to the target is large. The reason for this is explained as follows. The projection of a pixel from image to ground is the intersection of a line, passing through the pixel and camera centre, and the ground. A small disturbance could cause a big error when the target is far from the host. We explain this in Fig. 3. When a pixel change from position A to A+μ on the image, the projection on the ground will be changed from A' to A'+μ'. We can see that the difference μ' cannot be ignored even the difference μ is minor.

In order to estimate the width accurately, radar data is used to correct the corners on the image before they are projected

to the ground. The accuracy of the distance and orientation provided by radar is high enough [20] to be used for this correction. An example of the width measurement is shown in Fig. 4. The radar detection is projected to the image (represented as a circle with the centre corresponding to the centroid of the object, marked in yellow). The green rectangles are the image detection results. We can see that the bottom lines of the rectangle could not lie on the ground because the detection could be affected by rain. On the contrary, the centroid provided by the radar is more accurate than the one obtained from image vehicle detection. Hence, the bottom lines of the bounding box are corrected to the position provided by radar, red lines in Fig 5, which passing through the centroid of the object and parallel to the original bottom lines of the bounding box. Finally, the new bottom left and right corners will be used to compute the width of the target.

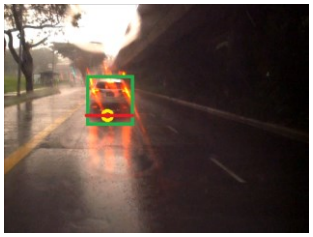

Fig. 4 Correcting the bottom left and right corners of the bounding box with the centroid provided by the radar. The red line which passes through the centroid is used to correct the bottom corners of the bounding box obtained by vision-based vehicle detection. The intersections of the red line and the left and right sides of the bounding box will be projected to the radar plane with equation (3). The vehicle's width is then computed as the distance between the two projected corners.

By doing so, the 3D information of the detected vehicle can be reconstructed because the distance and orientation information of the vehicle can be obtained from radar directly.

C. Synchronization of the radar and vision

In order to fuse radar and vision, a synchronization method is required because the processing time is not simultaneous. One of ways to do so is to synchronize them by using timestamp. In our approach, the radar is much faster than vision. Hence, the synchronization method is that radar is kept when an image is captured.

IV. Deep learning vehicle detection from Images

Object detection is one of the benchmarks for deep learning algorithms. A significant progress has been achieved to detect object from images [7, 10]. However, only a few deep learning algorithms can detect object in real-time which is required for many real applications.

In this paper, we aim at a practical vehicle detection approach. We train real-time deep learning vehicle detection with our own data, especially rainy data.

Open source, e.g. You Look Only Once (YOLO) [18], makes it possible to apply publicly available deep convolutional neural networks, to our application. In order to achieve real-time system, YOLO version 2 is adopted in our system.

The input to YOLO is resized to a fixed-size 224 × 224 RGB image during training process. The only pre-processing we do is subtracting the mean RGB value, computed on the training set, from each pixel.

YOLO is a single unified network for performing object detection. It does this by predicting multiple bounding boxes and the class probabilities for those boxes. In its original implementation, it divides an image into a grid of 13 × 13 cells, and each cell predicts 5 different bounding boxes. It also outputs a confidence score on how certain it is that there is an object in that bounding box. This is defined as: $P(Object) \times \text{IOU}$, where IOU is the intersection over union. YOLO also predicts the classification score for each box for every class in training, as a conditional class probability: $P(Class|Object)$. At test time, the results are combined, by multiplying the classification score and the box confidence, in the form:

$$P(Class|Object) \times P(Object) \times IOU = P(Class) \times IOU \quad (4)$$

V. Tracking

In order to improve the accuracy and robustness of radar, tracking technology is needed. By projecting the radar to image, we can do vehicle tracking on the image instead of radar data. This makes the tracking more intuitive. In this paper, we develop a traffic light tracking approach based on temporal trajectory analysis.

A high speed in-vehicle camera ensures that the relative position and size of a target in the captured image constantly change. Temporal spatial analysis is a process to examine previous frames and determine whether a candidate detected in the current frame has been found in the same area earlier. Temporally, locations of the vehicle in the image plane in consequence frame are spatially continuous. Proper temporal spatial tracking can greatly benefit the task in two aspects: (1) improve the result's smoothness by filling detection of the middle frames with missing or low confident detections; (2) improve the detection confidence and reduce isolated false alarms. Overall, this module is trying to track two entities: spatial location of detected instances and history of vehicles.

Specifically, we term the whole tracking history of a vehicle instance as a trajectory. A trajectory has a few components: type, a vector of points storing history locations, age, and discontinuity. The trajectory is categorized in terms of bounding box and one Boolean age indicating the trajectory stability: resulting in two type of trajectory: stable vehicle and temporary vehicle. For instance, a stable vehicle trajectory means the trajectory has been confirmed as a tracking of a vehicle. The age depicts the existing period of the trajectory since the first detection of a vehicle instance at the beginning of the trajectory, and the discontinuity of trajectory records the number of passed frames since last detection of the instance.

The trajectory pool updates after every frame. At the very beginning, every trajectory is initialized as a temporary trajectory. It requires a minimal age of 1 second and minimal 5 detections of the instance for changing a temporary trajectory to stable trajectory. A trajectory is removed from the trajectory pool when its age is above a threshold (e.g. 70 second in our experiments). Given an input bright image, vehicles are first detected using the aforementioned dual

channel fusion method. Then these newly detected points are added into the trajectory pool. As an example, suppose a vehicle is detected in current frame, these trajectories with vehicle in the pool are traversed and verified by calculating the distance between the point and the latest point stored in a trajectory. If the minimal distance among all vehicle trajectories is below a certain value (60-pixel distance is used), the new point will be added into that trajectory. If no valid trajectory is found, then a new temporary red trajectory will be created with the new point. When a stable trajectory is found, it would be very confident to say the new point is a stable vehicle. If only a temporary trajectory is found, then the new point is considered as a temporary vehicle, which occasionally is a false alarm.

VI. EVALUATION METHOD

In an object detection task, the deep learning model can either make a correct prediction (true-positive), a wrong prediction (false-positive), or a missed prediction (false-negative). To categorise the predictions into these categories, an overlap criterion is first defined. The overlap criterion is usually defined as 0.5, i.e. 50% or more of a predicted bounding box must overlap with the ground truth box for the detection to be considered a true positive or correct detection. Any predicted box that does not meet this criterion is a false positive. If 50% or more of a ground truth box does not overlap with any predicted bounding box, it is considered a false negative. In this paper, precision and recall are used to evaluate the accuracy of the proposed approach.

Precision refers to the proportion of all examples above a certain rank. Its mathematical definition is shown below:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

where TP is the number of True-Positives, while FP is the number of False-Positives of a given prediction. Recall refers to the proportion of all positive examples ranked above a certain rank:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

where TP is the number of True-Positives, and FN is the number of false-negatives of a given prediction.

Fig. 5 Vehicle detection and width estimation results obtained by fusing radar and camera. The yellow circles are the radar detection results and the green rectangles are the camera detection results. In our experiments, only the objects, which can be detected both radar and camera, are kept. In order to estimate the width of the vehicles accurately, the bottom lines of the rectangles are corrected by the red lines corresponding to the yellow circles provided by the radar.

VII. EXPERIMENTAL RESULTS

We have implemented the algorithms presented in this paper in our autonomous vehicle. A HDR camera, pointgrey Zebra2 [17], is installed in the AV under the interior rearview mirror. Three electronically scanning radars (ESR) are installed in the frontal of the AV to cover left, right and frontal, respectively. In this paper, only frontal ESR is used.

We divided vehicle into five categories: car, truck, bus, lorry and van. We have collected and manually labelled a large database, 59,019 images, using our autonomous vehicle for vehicle database. For training YOLO vehicle detector, 40,188 images are used for training while 18,231 images for testing. The remaining 600 images are used for evaluate the performance of the detector in terms of the precision and recall. There are 2,187 labeled bounding boxes and nearly same number bounding boxes for each of five categories mentioned-above.

Table I Detection rate for each category

	car	truck	bus	lorry	Van
Ground truth	441	432	421	432	452
Detected from images	433	401	399	401	423
Detection rate (%)	98.2	92.3	94.8	92.8	93.6

Table II Confusion matrix for vehicle detection

	car	truck	bus	Lorry	van
Car	433	2	0	0	0
truck	0	401	0	6	0
Bus	0	11	399	0	0
lorry	0	23	0	401	9
Van	1	0	0	7	423

Table III Precision and recall for each categories

	car	truck	bus	lorry	van
Precision (%)	99.5	98.5	97.3	94.6	98.1
Recall (%)	99.8	91.8	100	96.9	97.9

As discussed in introduction section, long-range coverage ESR provides accurate range and speed data with powerful object discrimination that can identify up to 64 targets in the vehicle's path. The Delphi ESR [20] is able to detect 99.995% of the cars as far as 180m ahead, with a longitudinal Root-Mean-Square (RMS) error of 1.8m and a lateral RMS error of 0.4m. As we combine ESR with camera at decision level, we can achieve nearly perfect accuracy even in rain. However, we have found that the radar is highly susceptible to false positives in rain conditions. In this paper, we compute the accuracy by considering the cases such that vehicles can be found from images and the width can then be estimated.

Table IV Accuracy of the width measurement without/with sensor fusion (m)

Distance	Ground truth	Absolute error of the width using image only	Absolute error of the width using both image and radar
10	1.4	0.2	0.05
30	1.4	0.2	0.03
40	1.4	0.3	0.07
75	1.4	0.5	0.08
Average error	NA	0.3	0.06

Some of the vehicle detection results are shown in Fig. 5. The bounding boxes in green are the detection results with vision and the circles in yellow are the results with radar.

The detection rates for each category are given in Table I. The confusion matrix for the vehicle detection (only the vehicle can be detected in the image are counted) is given in Table II. The precision and recall for each category, computed based on Table I and equations (5) and (6), are listed in Table III.

In our experiments, we measure the width errors by keeping another vehicle in frontal of our autonomous vehicle at different distances. The comparison of the image estimation and the corrected one by using centroid provided by radar is given in Table IV. We can see that the width measurement error can be reduced by five times when the fusion of the radar and vision is applied.

The algorithm (implemented in C++) presented in this paper has been integrated into our autonomous vehicle via Data Distribution Service (DDS). The vehicle detection and width estimation on a GIGABYTE Mini-PC (2.5Ghz CPU + GTX 760) can run in 30-40 fps depending on the number of the targets in an image.

VIII. Conclusion and Future Work

In heavy rain, the raindrop on the front windshield can make it difficult to detect object from an in-car camera. In this paper, a practical AV system to detect vehicles in rain has been described. To effectively avoid rear-end collision, we accurately localize all front vehicles by fusing detections from radar and vision. Although radar can cope with this problem and has been used as the main detection sensor, radar is highly susceptible to false positives. Furthermore, many entry level radar sensors only return the centroid of each detected object, rather than its size and extent. By fusing vision and radar, we have reduced the false positives and localized target in 3D. With the accurate distance provided by the radar, the vehicle width accuracy has been improved significantly. The experimental results on a large database collected by our autonomous vehicle have shown that the proposed approach can achieve high vehicle detection/localization accuracy in rainy conditions. The algorithm has been integrated into our autonomous vehicle via Data Distribution Service (DDS) and can run in real-time in a GIGABYTE Mini-PC.

Although the region of interest (ROI) of vehicles on the images can be located with the help of radar detection results, it is still a challenging problem to obtain the accurate bounding box of the vehicles under heavy rain conditions. The use of the Generative and Adversarial Network (GAN) [8] could be studied to solve this problem in the near future. As the drivers will turn on the head lightings when it is raining, we could investigate vehicle detection based on vehicles' lightings by using high dynamic range camera, similar to the approach proposed in our traffic light and vehicle signal recognition research [23-25].

Reference

[1] M. Bertozzi, L. Bombini, P. Cerri, P. Medici, P. C. Antonello and M. Miglietta, Obstacle detection and classification fusing radar and vision, IEEE Intelligent Vehicle Symposium, pp. 608-613, 2008.

[2] J. Burlet and M. Dalla Fontana, Robust and efficient multi-object detection and tracking for vehicle perception systems using radar and camera sensor fusion, IET and ITS Conference on Road Transport Information and Control , pp.1–6, 2012

[3] I. Collazo and D. D. Dean, Rain versus target discrimination for Doppler radars, U.S. Patent No. US6937185B1, 2005.

[4] H. Dong and X Zhao, Detection and removal of rain and snow from videos based on frame difference, The 27th Chinese Control and Decision Conference, 2015.

[5] F. Ehud, A method for mitigating rain clutter interference in millimeter-wave radar detection, WO/2014/108889, July 2014.

[6] K. Garg and S. K. Naya, Detection and removal of rain from videos, CVPR 2004.

[7] R. Girshick, Fast R-CNN, IEEE International Conference on Computer Vision, 2015.

[8] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks, NIPS, 2014.

[9] J. C. Halimeh and M. Roser, Raindrop detection on car windshields using geometric-photometric environment construction and intensity-based correlation, IEEE Intelligent Vehicles Symposium, 2009.

[10] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. C. Yue, F. Mujica, A. Coates, A. Ng, An empirical evaluation of deep learning on highway driving, https://arxiv.org/abs/1504.01716, April 2015.

[11] Z. Ji, D. Prokhorov, Radar-vision fusion for object classification, 11th International Conference on Information Fusion, June 2008.

[12] T. Kato, Y. Ninomiya, I. Masaki, An obstacle detection method by fusion of radar and motion stereo, IEEE Transactions on Intelligent Transportation Systems, Volume: 3, Issue: 3, Sep 2002, pp. 182-188.

[13] H. Kurihata, T. T. Hashi, I. Ide, Y. Mekada, H. Murase, Y. Tamatsu and T. Miyahara, Detection of raindrops on a windshield from an in-vehicle video camera, International Journal of Innovative Computing, Information and Control, Volume 3, Number 6(B), December, 2007.

[14]D. Langer and T. Jochem, Fusing radar and vision for detecting, classifying and avoiding roadway obstacles, Fusing radar and vision for detecting, classifying and avoiding roadway obstacles, IEEE Intelligent Vehicles Symposium, pp. 333-338, 1996.

[15]A. Mendes, L. C. Bento and U. Nunes, Multi-target detection and tracking with laser range finder, IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004.

[16] G. E. Natour, O. A. Aider, R. Rouveure, F. Berry, P. Faure, Radar and vision sensors calibration for outdoor 3D reconstruction, IEEE International Conference on Robotics and Automation, pp. 2084-2089, 2015.

[17]Pointgrey zebra2 camera (ZBR2-PGEHD-28S4C), https://www.ptgrey.com/zebra2-28-mp-color-gige-hd-sdi-sony-icx687-camera

[18]J. Redmon, A. Farhadi, YOLO9000: Better, Faster, Stronger, https://arxiv.org/pdf/1612.08242.pdf, Dec 2016.

[19] W. Song, Y. Yang, M. Fu, F. Qiu and M. Wang, Real-time obstacle detection and status classification for collision warning in a vehicle active safety system, IEEE Transactions on Intelligent Transportation Systems, Volume: 19 , Issue: 3, pp. 758-773, March 2018.

[20]L. Stanislas and T. Peynot, Characterisation of the Delphi Electronically Scanning radar for robotics applications, Australasian Conference on Robotics and Automation, 2-4 December 2015, Canberra.

[21] Z. Sun, G. Bebis, R. Miller, On-road vehicle detection: a review, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 5, May 2006.

[22] M. Wahab, C.-H. Su and N. Zakaria, Review on raindrop detection and removal in weather degraded images, 5th International Conference on Computer Science and Information Technology, 27-28 March 2013.

[23] J.-G. Wang and L.-B. Zhou, Traffic light recognition with high dynamic range imaging and deep learning, IEEE Transactions on Intelligent Transportation Systems, to appear.

[24] J.-G. Wang, L.-B. Zhou, Z.-W. Song and M.-L. Yuan, Real-time vehicle signal lights recognition with HDR camera, IEEE International Conference on Internet of Things (iThings), 15-18 Dec. 2016, Chengdu, China.

[25] J.-G. Wang, L.-B. Zhou, Y. Pan, S. Lee, Z.-W. Song, B. S. Han and B. V. Saputra, Appearance based brake lights recognition using deep learning and vehicle detection, IEEE Intelligent Vehicles Symposium, June 19-22 2016, Gothenburg, Sweden.

[26] X. Wang, L. Xu, H. Sun, J. Xin and N. Zheng, On-road vehicle detection and tracking using MMW radar and monovision fusion, IEEE Transactions on Intelligent Transportation Systems, Vol. 17, Mo. 7, July 2016, pp.2075-2084.

[27]Y. Wasaki, M. Misumi and T. Nakamiya, Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring, Sensors, 13(6), Jun 2013, pp.7756–7773.

[28]S. You, R.T. Tan, P. Kawakami and K. Ikecuchi, Adherent raindrop detection and removal in video, CVPR 2013.