Studies on Next-Gen Vehicular Detection: A Fusion of RF Signal and Fisheye Camera Technologies

A Thesis Submitted to the Department of Computer Science and Communications Engineering, the Graduate School of Fundamental Science and Engineering of Waseda University in Partial Fulfilment of the Requirements for the Degree of Master of Engineering

Submission Date: Jan 22nd, 2024

Dhindsa, Parneet Kaur

(5122FG10-3)

Advisor: Shigeru Shimamoto

Research guidance: Research on Wireless Access

Summary

This thesis presents the study on vehicle localization on an urban highway scenario, which employs a novel technique that investigates the fusion of RF signal and Fish-eye camera to improve object detection to further the future realization of Autonomous vehicles and their safety on roads within a Line-of-Sight(LOS) scenario. Firstly, various sensor technologies have been studied thoroughly along with their role in localization and assessed on different parameters such as low cost and computational resources and robustness to different weather conditions and are assessed to gain the insights of current dynamics and standards that primarily revolve around rules and regulations which are vital to achieve Vehicle-to-Vehicle communication. Additionally, a realistic 3D simulation environment is constructed, considering different parameters, designing a transmitter car(TC) and receiver car(RC) to study signal properties, receiver properties and the propagation model. And then, employing RF signals, the Direction of Arrival (DOA) of the signal is computed through the TOA-Trilateration method. The evaluation is done for the same-side DOA and opposite-side DOA on road. Then, this DOA value is utilised to select the Region of Interest (ROI) in the 360 degrees panoramic view made by the images captured by the Fish-eye camera from all the directions. This is the first work that talks about the fusion of RF signals with vision sensors. The selected images are fed to YOLO for object detection. THe results have confirmed that

this fusion will not only improve the overall classification but also reduce the computational time for the detection. The results have shown that the computation time has reduced by 10 percent , along with 17 percent improvement in confidence score for ROI-selected images resulting in accurate detection of targeted vehicles on the road.

Acknowledgement

I am deeply grateful to my academic advisor Professor Shigeru Shimamoto, for his valuable guidance, support and assistance throughout my research journey. His prompt responses to emails and feedback especially in the group meeting has been a great source of confidence and reassurance especially when I encountered challenges in the research. I would also like to thank Professor Zhenni pan for her continuous guidance throughout this journey. She provided me with a positive environment, where ideas can be openly discussed. She is always encouraging and assisting me with this research. Lastly, I would like to appreciate my family and friends for their endless support and encouragement throughout the Master's studies at Waseda University. Their love and continuous support acted as a pillar of strength for me.

CONTENT

FIGURES

Chapter-1.

Introduction

In this section, the introduction of the research background and the motivation behind the research is presented. After that, the outline and the structure of the thesis are outlined.

1.1.Research Background

With the advancements in the Intelligent transportation system (ITS), the Autonomous industry is inclining towards Autonomous vehicles(AVs) and their road safety. Perception and localization play a crucial role in determining the success of AVs by avoiding collisions. Because of the driver's negligence, around 94% of road accidents have been estimated for different reasons, such as poor visibility and over-speeding, which puts both their and others' lives on the roads in danger[1].

AVs have a great potential to reduce these accidents caused by the careless behaviour of human drivers. Vehicle-to-vehicle (V2V) communication plays an essential role in advancing these AVs. To do so, researchers have investigated vehicle positions and information about the surroundings to mitigate casualties. The deployment of these AVs publicly on roads, safety must be ensured. Hence, a plethora of onboard sensors, such as cameras, Lidar, and many more, have been equipped on the vehicle to understand better the surroundings and position of the car and road dynamics and generate data from these modalities, which assist the AVs in decision making, lane changing, or collision avoidance. Hence, it paved the way for research on emerging object identification and positioning techniques[2].

The basic idea is to utilise the sensory information from these sensors to localise and identify the vehicles autonomously for which these sensor-based systems find their application in various capacities [3] adequately. For example, radar signals are robust to lousy weather conditions. However, the performance degrades to the multipath effect[4]. Also, cameras give inaccurate results in adverse weather conditions or bumpy roads and demand high computational power. To tackle such issues, the data from different sensor modalities have been fused to achieve robustness in various conditions. In [5], Lidar and thermal infrared cameras detect and identify objects with poor visibility, such as fog, smoke, or severe glare. The authors discuss the extrinsic calibration algorithm between two sensors to get the extrinsic parameters using 3D calibration targets. In[6], the author uses radar and vision sensors for object detection. The author focuses on data coordinate calibration algorithms between radar and vision with actual sensors. These works focus on fusing sensors to achieve better performance by improving the calibration algorithms in real-world scenarios. In [7][8], the authors discuss the benefits of sensor fusion techniques in localization and identification over single modalities. In [9] and [10], authors demonstrated improvements in object detection when fusing radar and image sensors. In [11], the author combines birds-eye camera images and LiDAR views for 3D object detection. In [12], camera, LiDAR, and radar data fusion are presented for Moving object classification and tracking.

However, LIDAR is not adaptable to low visibility weather conditions and higher cost and computation requirements, which limits its useability in commonly occurring scenarios [13]. Radars also suffer from limitations such as noisy interference and limited distance range, which restricts the advancement of autonomous vehicles toward meeting safety requirements. Integrating these modalities with vision also leads to complex systems, more power requirements, and high costs.

Due to the above limitations, advances in RF signals with algorithmic-based accident mitigation must be encouraged to overcome the issues mentioned earlier while reducing financial and spatial limitations. Neighbouring vehicles can transmit information signals using V2V communication and optimise their routes to minimise collisions. These transmitting information signals can be utilised to accurately locate the position of the surrounding vehicles, which is crucial for real-time road environments. Moreover, when combined with other sensors, such as vision, the capabilities of RF signals can be improved.

Cameras offer significant visual data that assists in navigation and information about surroundings. Cameras can also be used for depth perception to recognize objects accurately and for distance measurements. In [14], cameras have been used as depth estimators and distance predictors using a transformer-based object detector framework. Despite being a cost-effective and versatile sensor, it is sensitive to light conditions and demands significant computational resources.

Furthermore, the existing works need a structure for combining the other sensor types with vision. Additionally, there remains uncertainty about how these different sensor modalities must be fused to improve situational perception.

1.2 Contribution of Thesis

The critical contribution of this paper is to provide a methodology to assist vision sensors by exploiting the advantages of RF signals. RF signals are robust to all weather conditions and can penetrate obstacles. These advantages, along with vision capabilities, are fused to not only reduce the computational complexity but also to improve the overall detection accuracy and robustness in all weather conditions.

1.3 Structure of Paper

The structure of the paper is as follows:

Chapter 2 introduces the Vehicle to everything communication as well as V2V. Firstly, a brief introduction to the wireless communication systems which revolves around vehicles will be introduced.

Chapter 3 talks about different sensor modalities. Firstly various sensors are introduced briefly and the merits and demerits were discussed. After that a brief introduction to the Sensor fusion and its significance is discussed.

Chapter 4 proposed Localization methods that have been commonly employed and which are utilised within this research. .

Chapter 5 extends its discussion on Object Detectors. The architecture and workflow for widely used detectors have been mentioned. A comparison is made between different detectors on the basis of different parameters.

Chapter 6 focuses on the system model and methodology. A detailed discussion about the way this research is conducted and the proposed scenarios along with details of various sub-part which are implemented in this research .

Chapter 7 discusses the Simulation study. The parameters used and the considerations made to conduct this research.

Chapter 8 puts the light on the results in which firstly different scenarios have been discussed for the DOA estimation and then the object detection results are shown and discussed thoroughly.

Chapter 9 Concludes the research along with the future works.

Chapter-2

Overview of V2X technology

2.1 Introduction to V2X Communication

The field of wireless communication in which the transmission of data is in between different vehicles or any entity that may or may be affected by vehicles on road is known as Vehicle-to-everything. This goal is to improve overall road safety, Fuel usage reduction, improve the overall traffic system and mass surveillance. V2X incorporate different types of communication:

- Vehicle-to-Vehicle(V2V),
- Vehicle-to-Infrastructure(V2I),
- Vehicle-to-Pedestrian(V2P),
- Vehicle-to-Devices(V2D),
- Vehicle-to-Network(V2N)

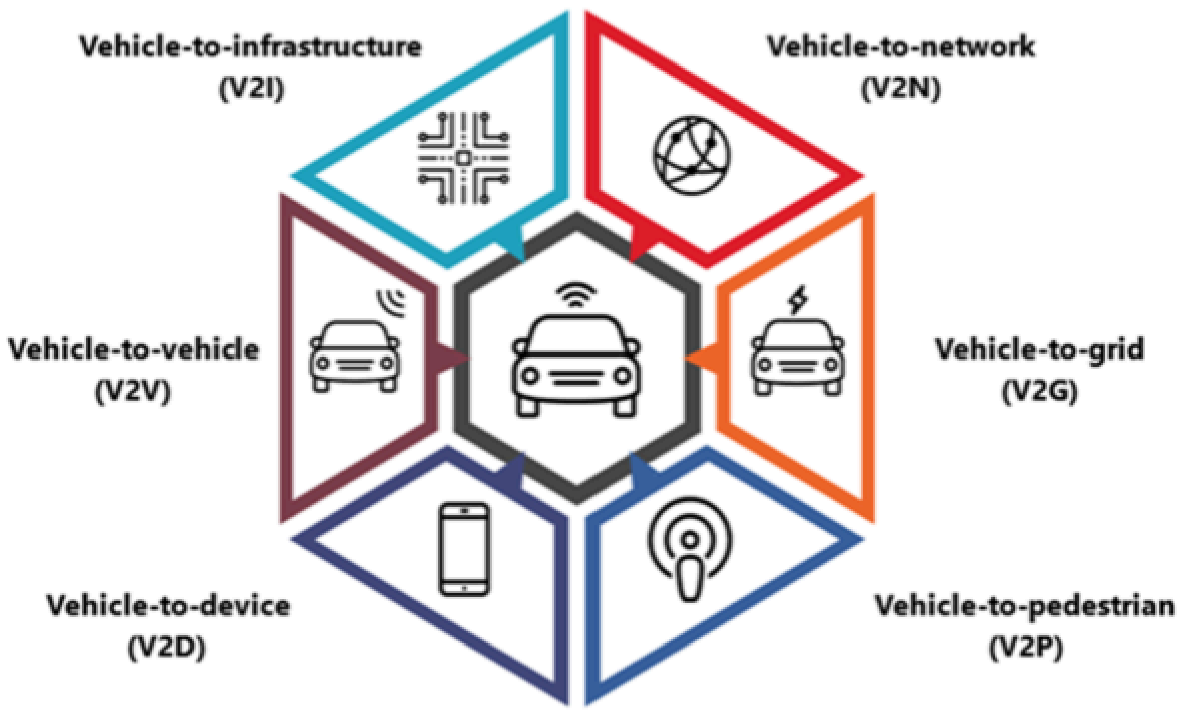

Figure 1. V2X Technologies[40]

2.2 V2V

Recently, V2V and V2I received a lot of attention from research and many use case scenarios can be seen on the road. Basically, the communication is in between the vehicles and vehicles and the infrastructure (traffic lights, base stations(BS)) to facilitate more safety on roads. As the number of vehicles are drastically increasing on roads, a global standard is necessary to keep in check the basic requirements for V2X applications and also to make a particular standard for the manufactures all across the world to avoid conflict and inconsistencies, hence mitigating fatal accidents.

In 2010, the Institute of Electrical and Electronics Engineers(IEEE) modified the standard 802.11 to 802.11p and introduced Wireless Access in Vehicle Environment (WAVE)[41]. With this standard, creation of a new management system which ensures rapid and reliable communication channels for seamless vehicle communication with minimal chances of failure.

Even before the amendments of the standards, the automotive companies were focused on integrating the different modalities to achieve a connected environment on the road. With V2V communication, the vehicle can make direct connections with other vehicles which further facilitates the increased data transmission, hence information sharing which ensures the future with increased efficiency and better driving assistance[42]. It also allows the vehicle to share information such as speed, position and other relevant data with each other which opens a new dimension of Vehicle localization hence assisting the vehicle in the close proximity with lane change, emergency braking and mitigating the collision that might be difficult for the human-drivers to predict[43]. This helped vehicles to be more intelligent and enhanced the overall awareness.

Despite being an advantageous technology, V2V has many challenges. Security and accurate detection are few of them. Since the technique requires absolute trust with V2V, any security breach can lead to hazardous results as the information can be manipulated for personal gain or terrorism which further results in fatal accidents rather than mitigating them[24].

Chapter-3

Sensor Technologies

Autonomous vehicles rely on a plethora of sensors which includes camera, Lidar, Radar and Artificial Intelligence(AI) to assist it for driving without any human operator. Experts have defined five levels in the evolution of autonomous driving. Each level describes the extent to which a car takes over tasks and responsibilities from its driver, and how the car and driver interact: 1)

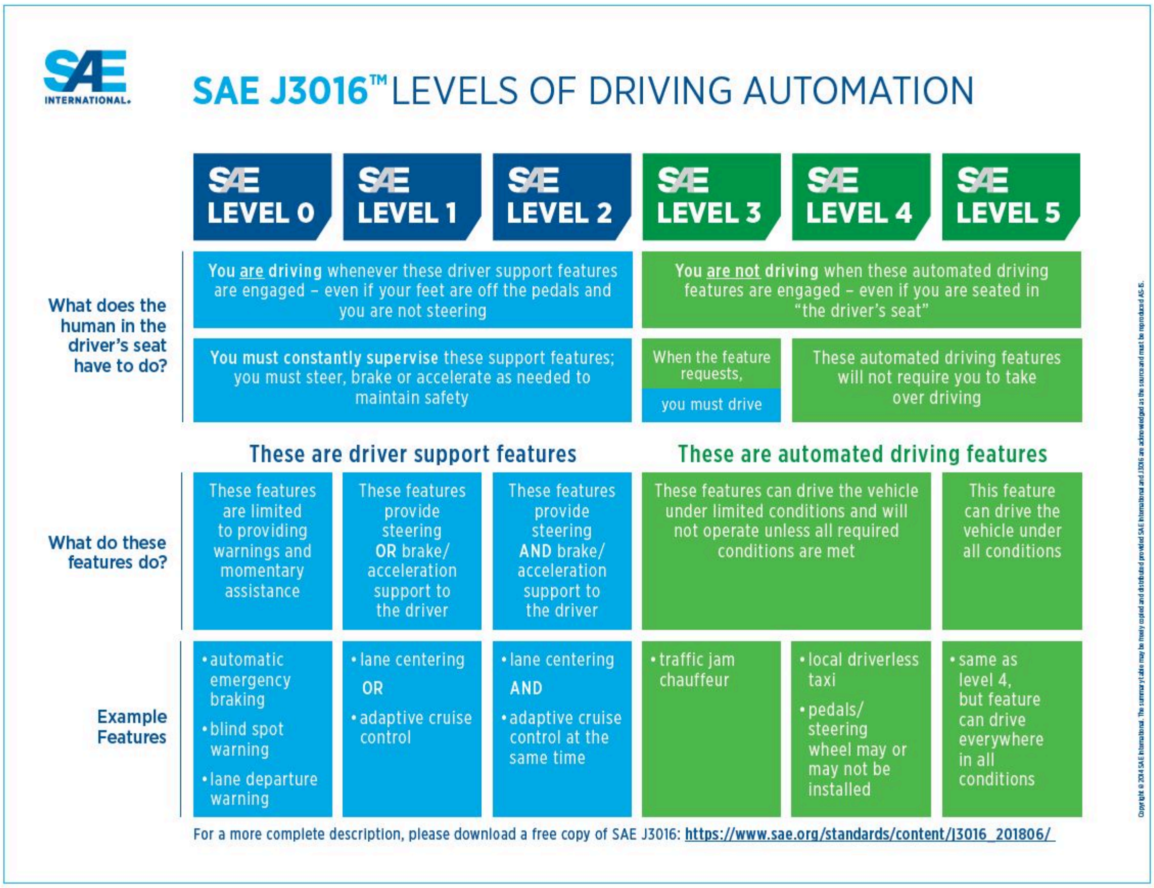

Figure 2: SAE Levels of Driving Automation[47]

Driver assistance, 2) Partly automated driving, 3) Highly automated driving, 4) Fully automated driving 5) Full automation.

These sensors collect the data about the surroundings which is then processed by a machine learning algorithm to detect or localise objects such as vehicles, pedestrians and do decision-making for the vehicle's action. Thus, Environment perception plays a crucial role in extracting the information to assist such decisions. Different Sensors have been discussed below:

3.1) LIDAR

LIDAR, or Light Detection and Ranging, was specifically originated for aeronautical and aerospace terrains in the 1960s. In further research, it was found the capability of Lidar can emit 2000 to 2500 pulses per second for topographic mapping[20] which paved its way towards autonomous driving and Advanced Driver Assistance Systems (ADAS). It is basically a laser scanning which uses a laser to create a 3-D representation of the environment which it has scanned, in the form of point cloud[21].

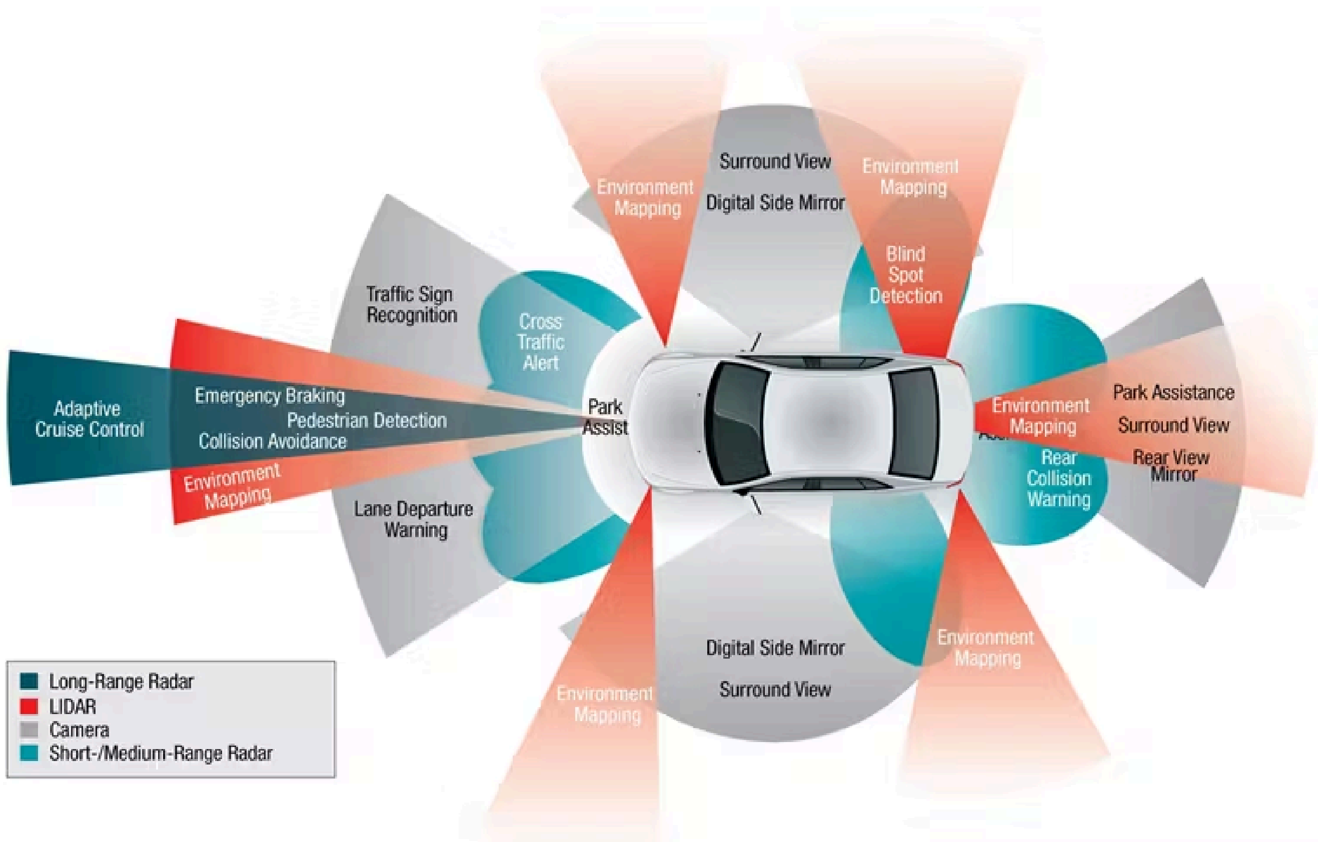

Figure 3. Type and position of sensors in Automotive Vehicles[50]

3.2) Radar

Radar stands for Radio Detection and Ranging and the operational principle is to analyse the wave that is scattered or reflected by objects in an environment where electromagnetic waves are emitted. It was first developed before World War II. Through these reflections Radar can determine the position and distance of the obstacle near to it, leveraging Doppler property [22]

Radar is a widely employed and well-known sensor for Autonomous vehicles both for night and day time. They are capable of mapping environments at different ranges and can identify the speed of the moving vehicle. However, Radar suffers from low resolution as compared to vision sensors. Hence, it can not be employed for Object detection. In[23], the road signs or guardrails were difficult to differentiate for radar sensors. In vehicles, the positioning of radar is crucial to avoid angular misalignment that might result in late or false detections [24,25].

3.3) Cameras

Cameras are one of the widely used sensors in precipitation nowadays especially in vehicular communication. The principle on which cameras work is that different photosensitive surfaces(image planes) emit light which is captured by the camera lens placed in front of the sensor, creating a clear image of the surroundings[22,26]. Both moving and static within the Field of view the surrounding can be detected. It is one of the inexpensive sensors. Cameras can assist an autonomous vehicle's perception system to identify the road signs, lane markings. This perception is also helpful in detecting the position of a vehicle in the surrounding, hence providing assistance with collision avoidance.

There are various types of camera systems, namely monocular cameras, binocular cameras and Fish-eye cameras. Monocular camera systems use a single lens which creates multiple images in series. Despite advancements, the monocular cameras have limitations. Depth calculations is the huge limitation and to overcome this, generally binocular cameras systems are preferred. Even though using dual-pixel autofocus hardware, the estimation of depth can be done; however, it will get computationally heavy [27,28]. The stereo or binocular cameras, two cameras, are used side-by-side and are separated by some distance(baseline). This system captures two images and covers a wider view as opposed to the monocular cameras and this assists with depth estimations.

Fish-eye cameras [29-31] are other cameras generally used for vehicle perceptions. It captures the 360-degree view of the surrounding which finds its application in parking, traffic jam assistance. Four cameras are required to cover the whole 360-degree view. There will be some overlapping area within the images which can be further stitched or combined together to get a bird’s eye view. Hence, in this research fish-eye cameras are considered. Compared to Standard cameras, the field of view is 180-degrees wider. These cameras usually suffer from radial distortion which is not considered within the research and is kept for future works.

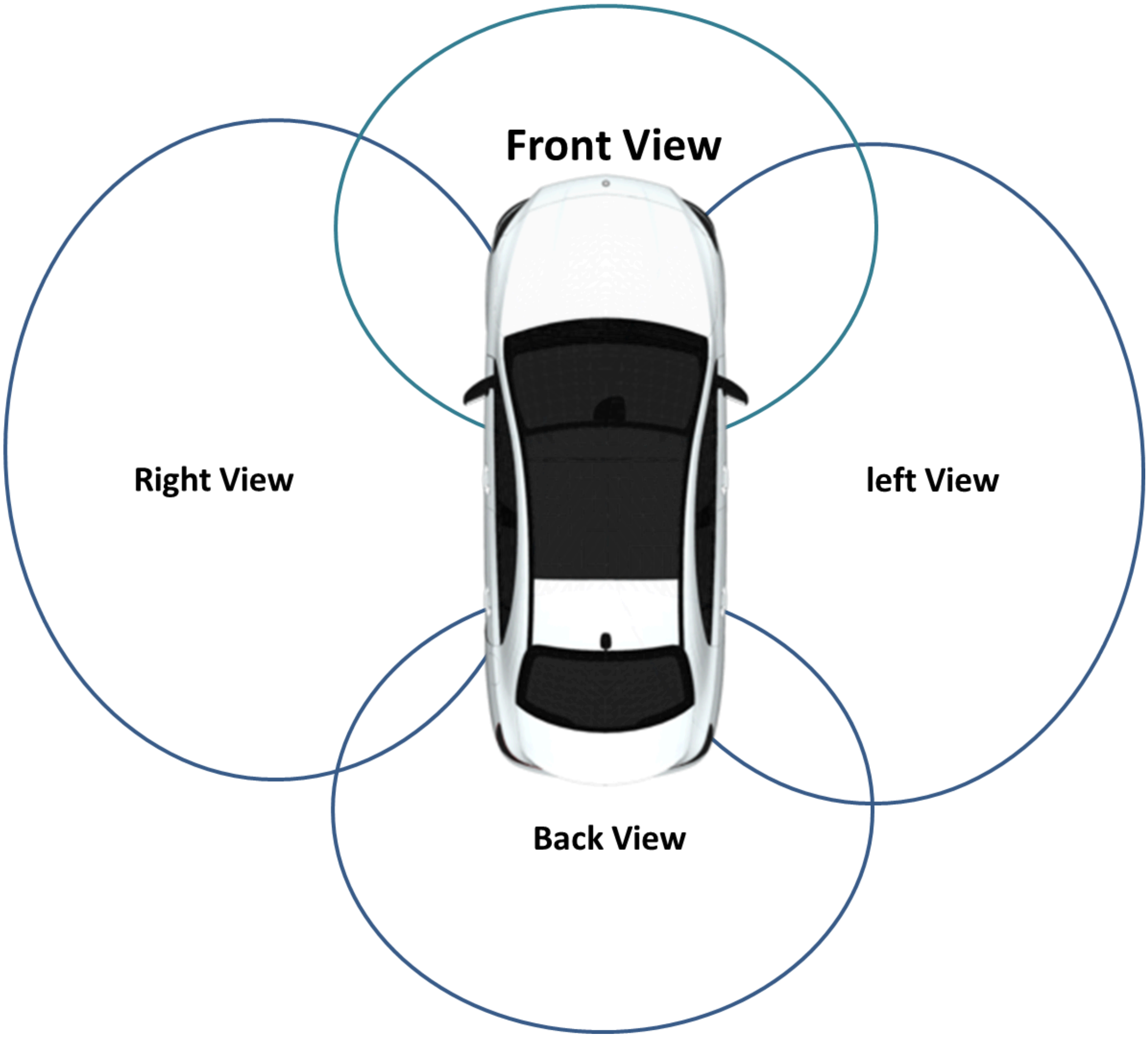

Figure 4: Fisheye camera view with overlapping area

To mitigate accidents, rear-view fisheye cameras were made compulsory in the United States in 2018[32]. These cameras were also employed in BMW to capture the panoramic view of surroundings for parking application [33].

Camera systems are computationally heavy and if the road is bumpy or any rain drop is on camera lens, it will provide inaccurate results.

Chapter-4

Localization Techniques

This chapter provides an overview of the techniques or methods that have been widely used in solving localization challenges. This chapter also explores the advantages and highlights the limitations associated with these methods.

4.1 Trilateration

The fundamental feature of this method is to compute the area under many circles. Trilateration is a very common method employed in localization and positioning systems especially in 2D plane.The calculations are done mathematically and are both simple and effective. To find the unknown point, this intersected area within the circle is pinpointed.

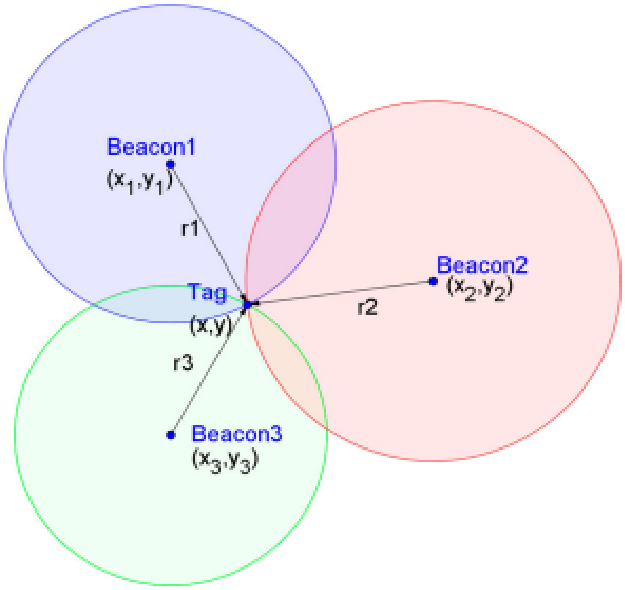

Figure 5: Example of Trilateration algorithm[34]

4.2 Time of Arrival(TOA):

Time of Arrival (TOA) is a localization method which computes the position of an object on the basis of time the signal takes to travel from the transmitter to multiple receivers. It is commonly utilised in indoor positioning systems and is a practical approach to trilateration[32]. Fundamentally, the distance between the receivers is calculated by computing the time at a known speed at which the signal propagates. In the trilateration method, the circle's centre points are replaced with the receivers of the vehicle which act as the nodes and detect the signal once arrived. After that, calculation of the transmission source is done[33]. Using this distance and keeping it as radius circles are constructed.

The signal propagates at the speed of light. The measurement of TOA [48] can be applied up to three reference points to facilitate the 2D positioning. Therefore, to achieve high accuracy in a communication system at least three nodes are required. This technique finds its practical application in Global Positioning systems[49].

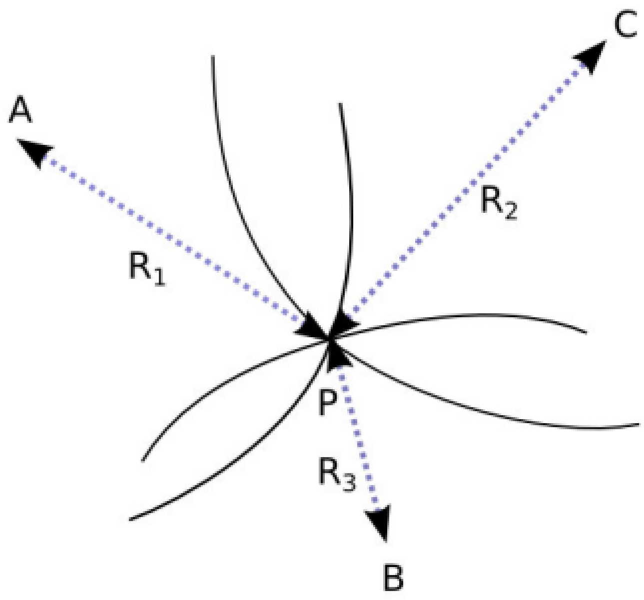

Figure 6: Localization through TOA measurements[35]

Since the synchronisation between the clocks is required, there is no need to synchronise different systems. However, the precision between the clocks in the TC and RC are considered to be synchronised to achieve high results, hence location estimation.

Chapter-5

Object Detection.

Every day our brain is engaging in the task of visual perception through different experiences such as textures, locations, size and distance. Our cognitive abilities empower us to comprehend and analyse the surroundings, examining it and evaluating and categorising them on the basis of position and characteristics. Computer vision, being a parallel concept deals with objects or multiple objects contained within a series of images. The basic idea is to classify and locate the object in which we are interested in the specific image set. These tasks of classification and detection can be performed by various object detectors. The common and widely employed object detectors are discussed below:

5.1 Faster R-CNN

Faster R-CNN is a Two-stage detector because it will read the image twice. The Regional Proposal Network(RPN) determines the existence of the object at different positions by utilising the sliding window technique to extract features from Convolutional Neural Network(CNN). Faster R-CNN employs a set of nine predefined anchors at every location, hence determining various sizes and ratios. These anchor boxes are characterised by six different parameters, out of which four are the bounding box anchors and the rest two are the label probabilities related to those anchor boxes which are related to ground truths. The structure of Faster R-CNN can be seen in Fig.7 [36]. The accuracy may decrease if some unimportant bounding boxes are excluded, which reduces the training time.

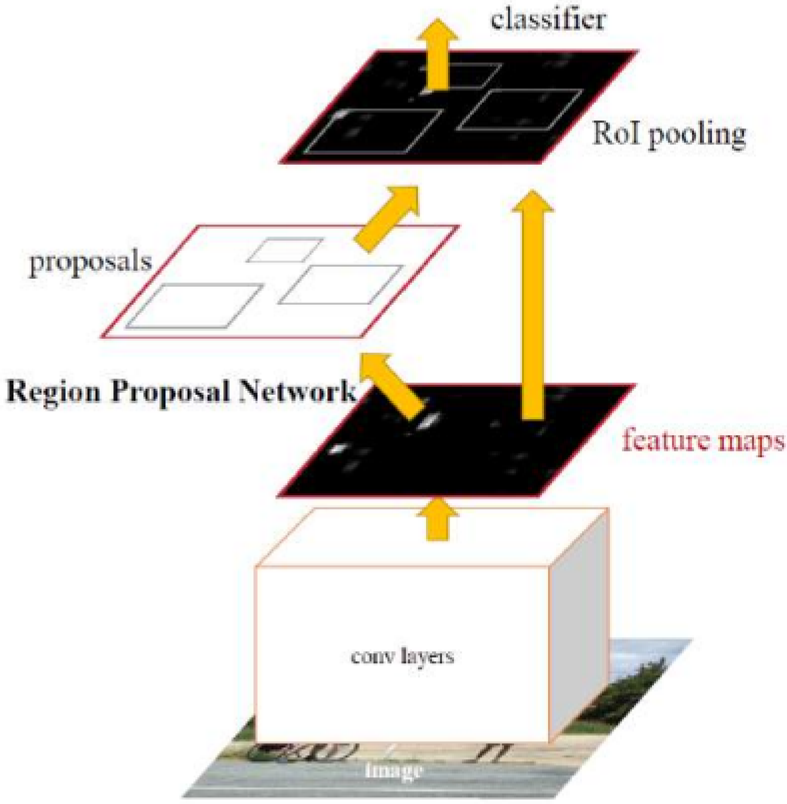

Figure 7: The structure of Faster R-CNN[36]

5.2 YOLO

YOLO stands for You Only Look Once exploits the advantages as a non-linear classifier which can handle much more complex problems as compared to Support Vector Machines(SVM). It basically transforms a standard classifier into an object detector by employing the strength of a CNN to accurately predict the locations along with the classification. It is a single forward pass through the network as the name itself suggests. YOLO v3 uses a convolutional layer with stride 2 to down sample the feature maps. It has 75 convolutional layers with up-sampling layers.

In term of operational mechanism, YOLO v3 divides the image which is input to it into S*S grid where each grid predicts the N bounding boxes which would have a confidence score that is directly linked to the presence of an object within that bounding box. Through this YOLO is able to extract not only the probability but also the position or location of the object within the image. The depiction of YOLO workflow can be seen in Fig.8[37]

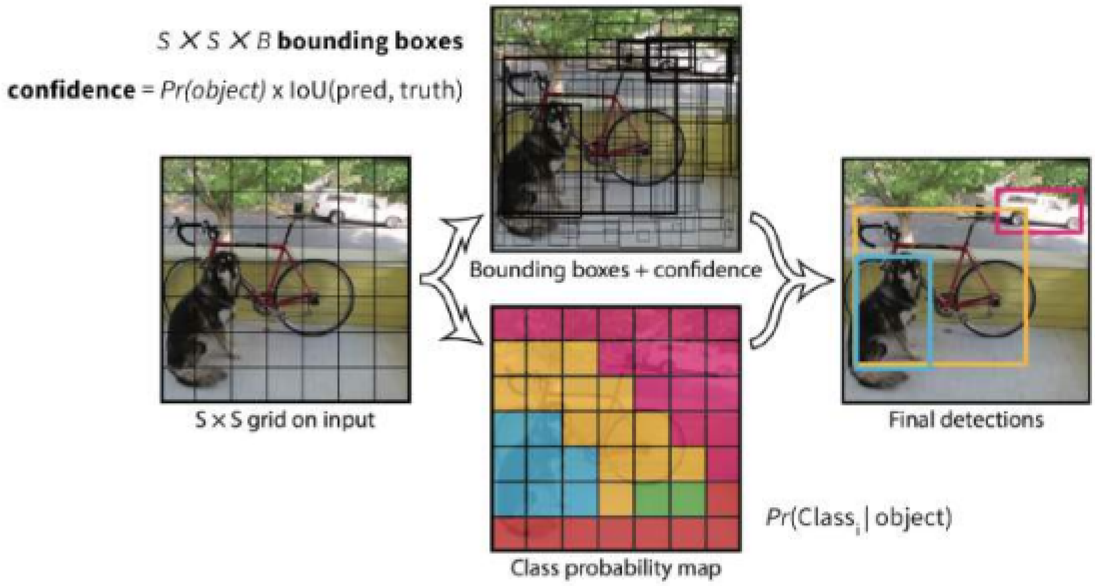

Figure 8: Depiction of YOLO workflow[37]

5.3 SSD

SSD stands for Single Shot Multibox Detector, uses VGG-16 Simonyan and Zisserman[38] and is a competitor of YOLO. In SSD, the values of the output for the bounding box offsets such as centre, width and height are computed relative to the default boxes and then imposing these default boxes to various feature maps varying with different resolutions. This assists the model to detect objects with different sizes and different scales.

By allowing various kinds of shapes and sizes of bounding boxes across the feature maps allows the SSD model to train on a variety of object shapes and sizes. Employing the ground truth with the default box by employing threshold, during training.

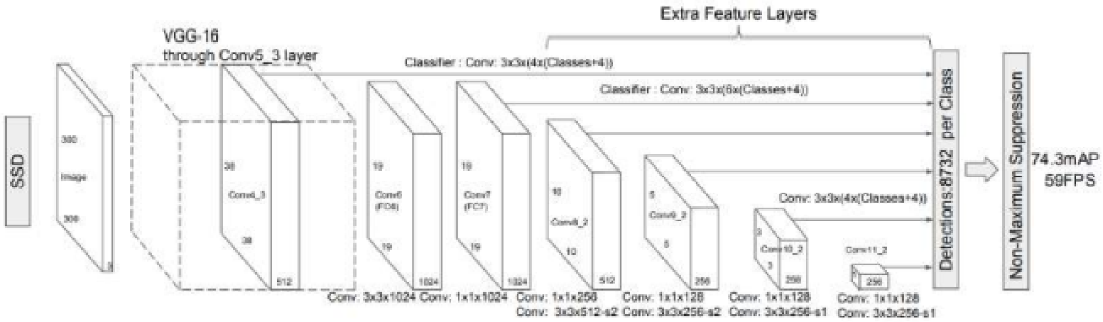

Figure 9: Example of SSD model with an input size of 300*300[39]

The default boxes with highest overlap are selected. This process ensures that the model predicts the bounding boxes accurately which aligns with the real objects within images. The Fig 9[39] shows the SSD model with 300*300 input size and different feature layers until the end which predicts the offsets to the default boxes of different sizes and ratios and related confidence score.

Chapter-6

System Model and Methodology

6.1 System Model

The proposed scenario is conducted on a 3D plane to investigate the model's effectiveness. The transmitter car(TC) and receiver car (RC) are both travelling along the straight road in a line-of-sight(LOS) scenario with other vehicles on the road travelling on the same and opposite sides. The RC car can be on the same side as the TC car or it can be on the opposite side as well. The conditions such as reflection have yet to be considered for ease of implementation as these conditions severely distort the DOA estimation. The RCcar has four receiving antennas placed on the four corners of the vehicle along with four Fisheye Cameras placed on the centre of the car's roof to capture the 360-degree surroundings. The receivers on the RC car receive the signal when it arrives, and the cameras will capture the images simultaneously. The positions of the Tand RC car are randomised using Uniform distribution. The speed of the TC and RC is kept at 60 and 80 km/hr. The sizes of the vehicles are kept similar for ease of implementation.

Ray tracing propagation has been used for the propagation model. There were two propagation methods within the model: the Shooting and bouncing rays(SBR) method and the image method. Being faster in computing the propagation, the SBR method has been opted for this research. To make the system realistic, a signal is transmitted from the transmitter antennas from the centre of the TC, which is isotropic and scatters the signal uniformly in all directions at a transmitting frequency of 5.9 GHz, which lies in the bandwidth for Intelligent Transport System[15].

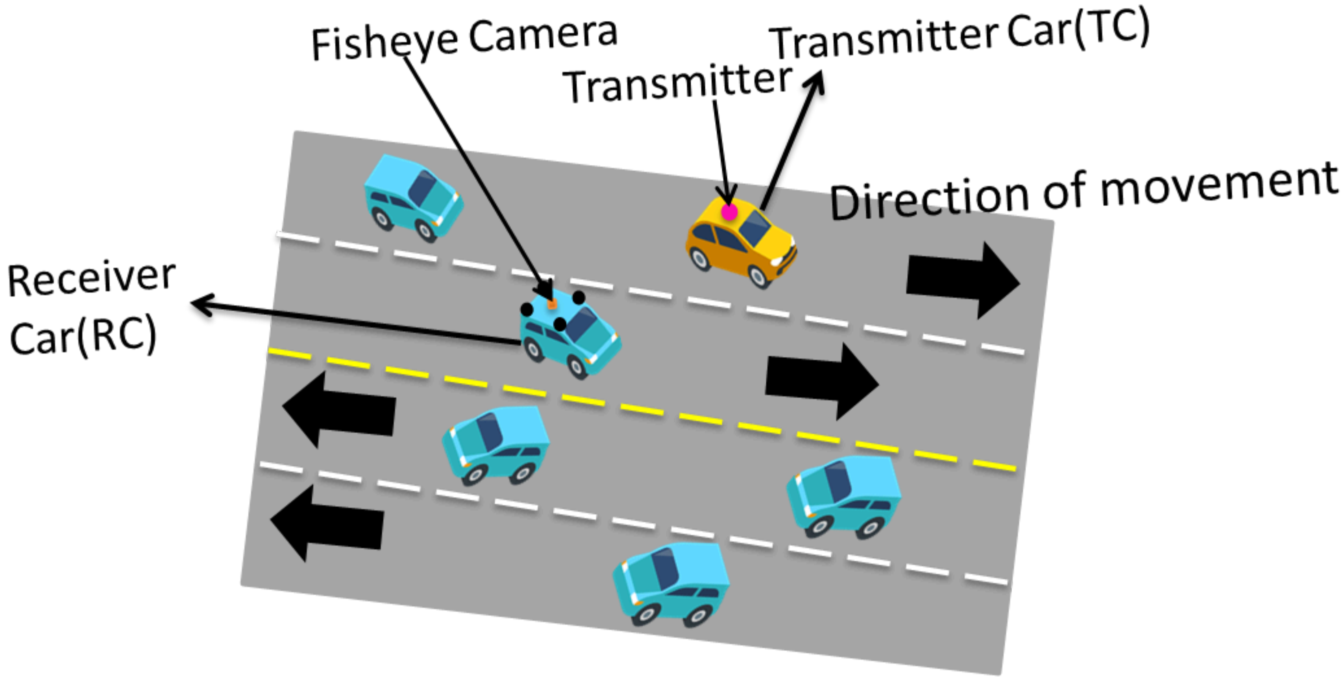

Figure 10: Proposed Scenario

6.2) Proposed Method:

The proposed methodology in this paper considers the advantages of both RF signals and cameras to detect the transmitting vehicle. The proposed methods are further divided into three parts, which are explained as follows:

i) Direction of Arrival(DOA) estimations

The transmitter on the TC starts propagating the RF signal as shown in Fig. 11 in all directions from the car's centre. The RC car will receive the signal, and the receivers will note the earliest time stamp at which the signal arrives. By leveraging the signal's Time of Arrival(TOA), the distance of the transmitting source is calculated using the Trilateration algorithm[16]. According to the TOA-Trilateration algorithm[17], once the distance of the transmitting source is known, it is considered a radii value, circles are constructed, and the trilateration process continues.

The precise location is calculated using the reference points(vehicle's receivers) and distance measurement given by the following equation[16]:

$$\begin{bmatrix} 1 & -2x_1 & -2y_1 & -2z_1 \\ 1 & -2x_2 & -2y_2 & -2z_2 \\ 1 & -2x_3 & -2y_3 & -2z_3 \end{bmatrix} \begin{bmatrix} x^2+y^2+z^2 \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} {s_1}^2 - {x_1}^2 - {y_1}^2 - {z_1}^2 \\ {s_2}^2 - {x_2}^2 - {y_2}^2 - {z_2}^2 \\ {s_3}^2 - {x_3}^2 - {y_3}^2 - {z_3}^2 \end{bmatrix}$$

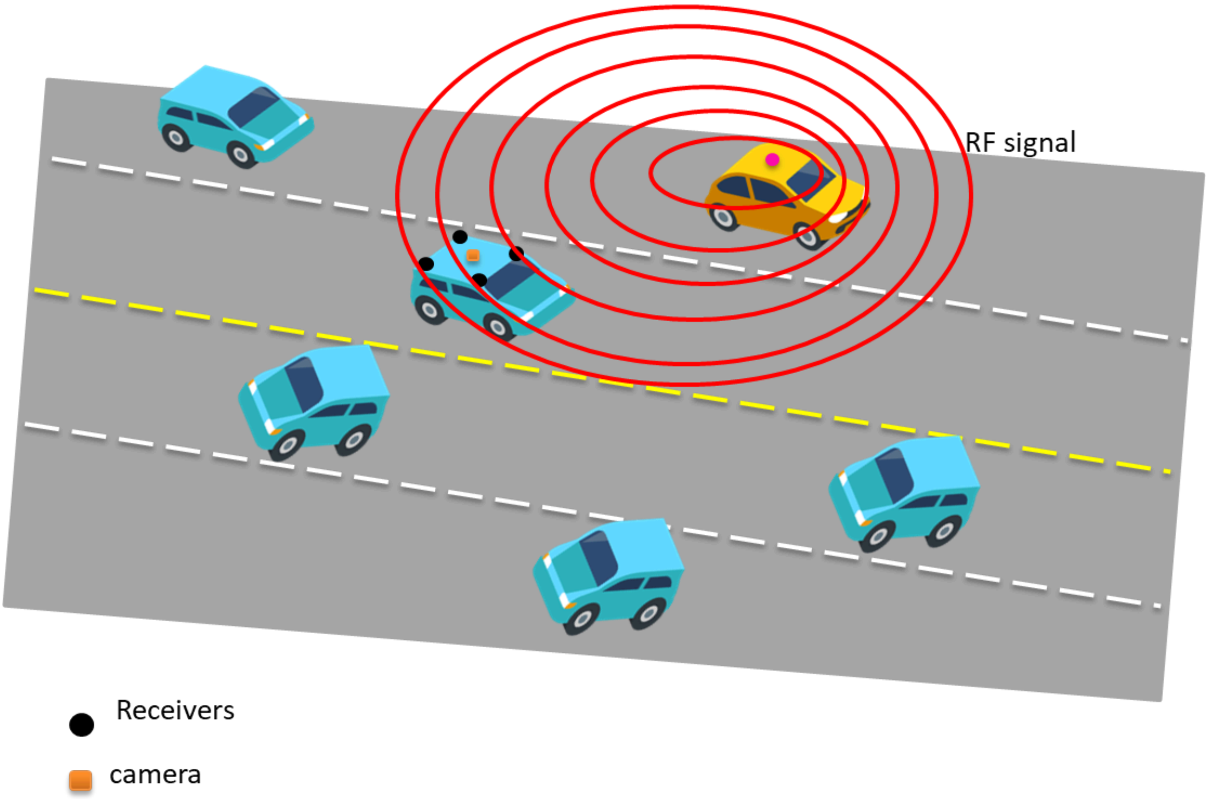

Figure 11: TC Propagating signal

Now, the direction of arrival (DOA) of the signal from which it is arriving can be calculated as shown in Fig. 12, as the angle between the triangle formed by the receivers and the transmitting source. The value for DOA is calculated for every 0.1ns.

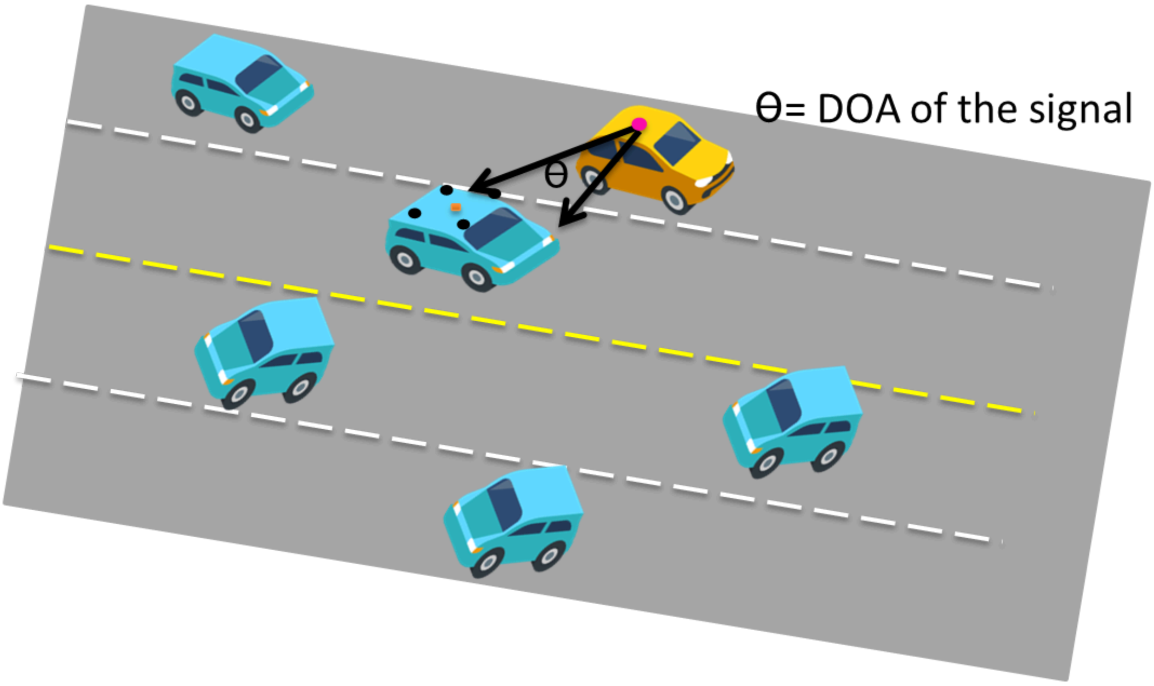

Figure 12: DOA Calculation

ii) Employing DOA for Region of Interest (ROI) selection

As the receivers are calculating the DOA of the signal, the Fisheye cameras on the roof, upon receiving the signal, will also start capturing the images from all four cameras. These four images from the four cameras will capture the surroundings from four different directions: front view, back view, left view, and right view as shown in Fig. 13. Then, the captured images were put as one frame in a signal image using the Montage function in Matlab can be seen in Fig. 14. This montage image is divided into four quadrants. The first quadrant ranges from -45(315) to 45 degrees, the second quadrant ranges from 45 to 135 degrees, the third quadrant from 135 to 225 degrees, and the fourth quadrant from 225 to 315 degrees, covering 360-degree surroundings.

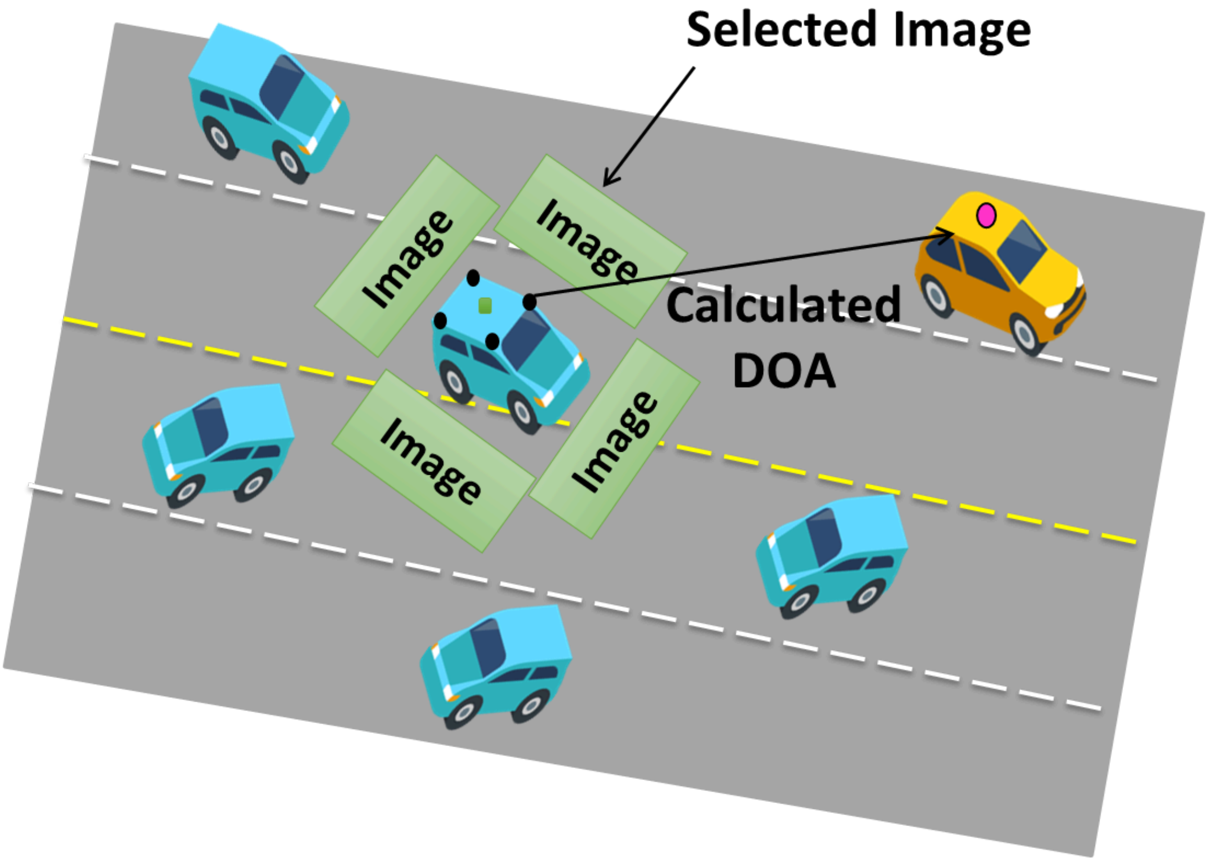

Figure 13: Image and Camera Selection

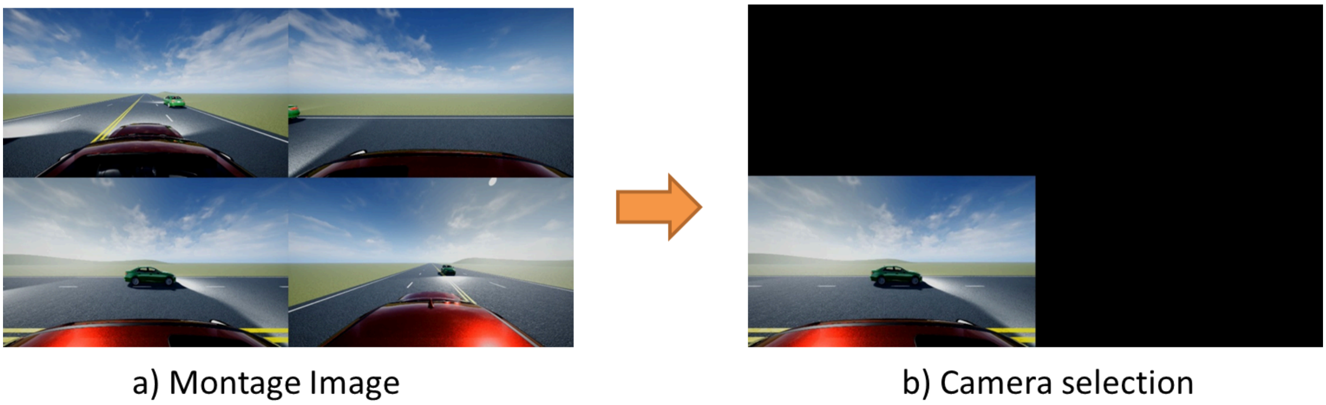

Figure 14: Montage Image and Camera Selection

Now, employing the calculated DOA of the signal, the quadrant and the corresponding image will be selected from the montage image, and all the other unselected images will be masked, as shown in Figure [6]. This selected image is then divided into an S*S grid as shown in Fig. 15(b)for The Region of Interest (ROI), which is the area within the image in which the transmitting vehicle may be present. The size of the selected image is 1280*720 pixels. The calculation of ROI is done theoretically. After selecting the quadrant, the DOA value is always re-adjusted between -45 to 45 degrees, and using this DOA value, the equation 1 is applied to calculate ROI in vertical direction(VROI) and equation 2 is used to calculate ROI in horizontal direction(HROI) within the image:

VROI= 200: 200+height

HROI= 540+6(x):540+6(x)+ width

where, the width and height of the ROI within this selected image are 200*200(height*width) pixels, respectively. As per the given formula, a 1-degree change in x will shift the ROI value by 6 pixels. For example, if the TC is at DOA of 110 degrees from the RC, then second quadrant will be chosen and x will be readjusted to 20 degrees, which will calculate HROI in the image which is going to be from 660 to 860 pixels and VROI is also calculated within the image from be 200 to 400 pixels. Hence, employing the calculated DOA value, this ROI is selected, and the remaining pixels of the image will be masked to 0 as shown in Fig. 15(c).

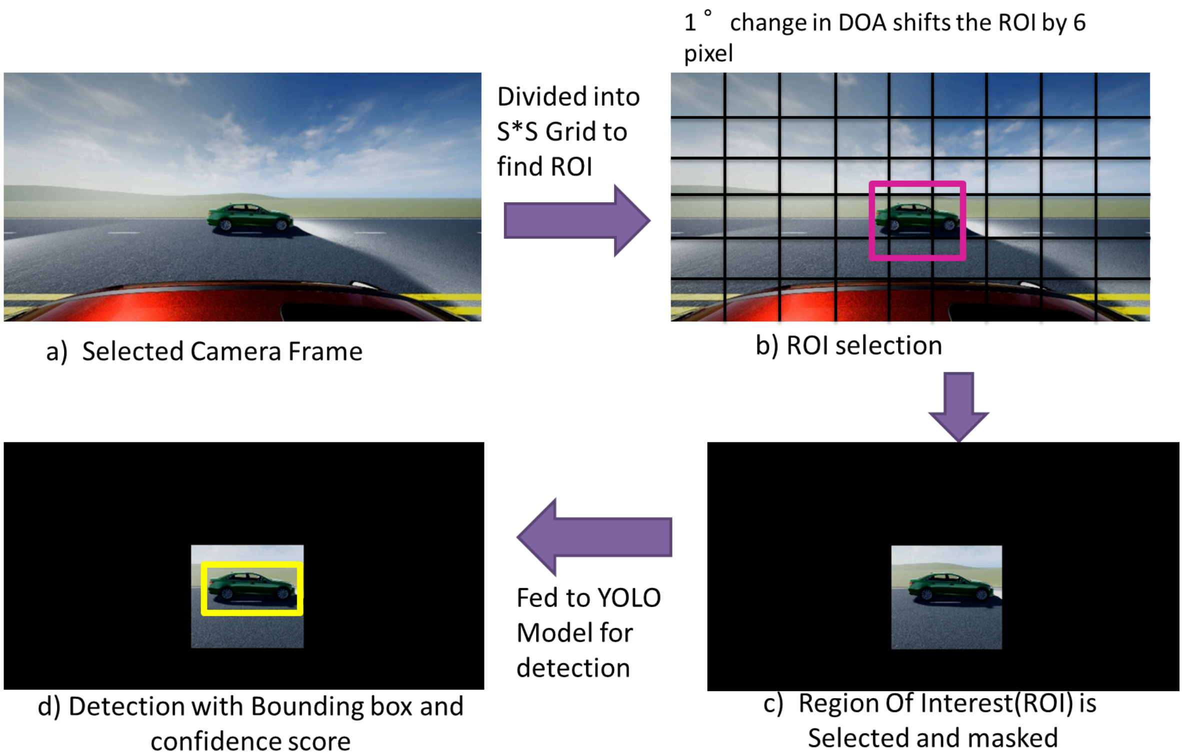

Figure 15: ROI selection and Object Detection

iii) Object Detections

This ROI-selected image is now fed to the YOLO model for Object detection. The YOLO [18] is a single-shot detector with excellent speed and accuracy. The model takes the ROI selected image as an input and outputs an image with a bounding box and confidence score, as shown in Figure 15(d). We have trained the detector from scratch by feeding it with standard and ROI-selected images of different scenarios from our dataset.

Chapter-7

Simulation Study

7.1 Simulation Parameters

Matlab and Simulink were used to facilitate the simulation process, and the following parameters written in Table 1 were used.

Description	Values
Time interval[ns]	1
Vehicle speed range[km/hr]	60-80
Vehicle Size[m]	**4.848** *2.107***1.517**
Transmission Frequency[GHz]	5.9
Road Dimensions[m]	913.85*20
Transmission Power[dbm]	30
Frames per second [fpm]	60
Fisheye camera (FoV)	171°

Table 1: Simulation Parameters

7.2 Dataset Generation:

The dataset used in this paper was generated by using a Fisheye camera in Simulink. The Fisheye camera's Field of View(Fov) is 171 degrees. Four Fisheye cameras were used to have a more comprehensive view and to capture the 360-degree panoramic view[19]. These cameras were placed in the center of the RC car, covering all four directions and output videos, which were then converted into images. In total, 2400 images were extracted from 10-second videos for every scenario.

Chapter-8

Results

8.1 DOA Estimation Results:

In Fig. 16, the DOA values demonstrate distinct behaviours in different scenarios. In Scenario 1, the DOA remains constant, as both the vehicles travel at constant speeds. A steady shift in values of DOA is noticeable as RC is moving faster than the TC. Scenario 3 and 4 depict a significant increase in DOA values as TC and RC, moving in opposite directions. The DOA values are stable towards the end, indicating that vehicles are far from each other, resulting in minimal changes in DOA values.

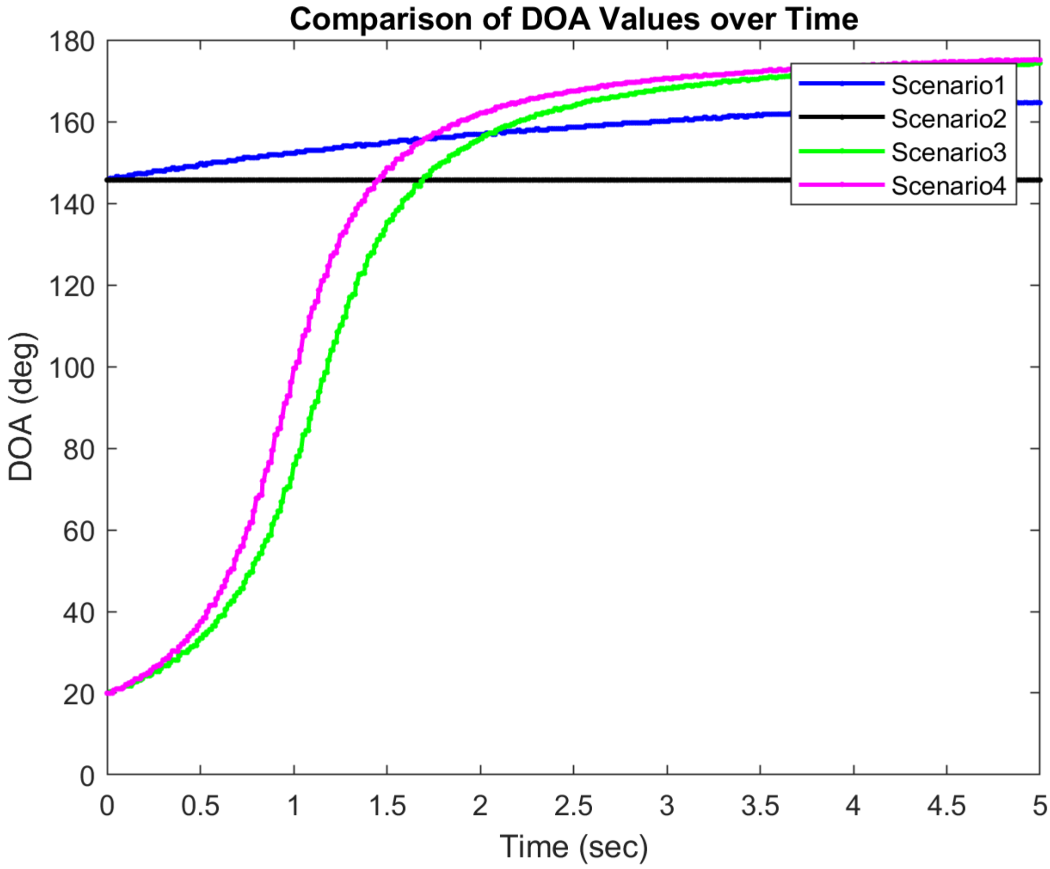

Figure 16: Comparison of DOA values over time

Table. II shows the average DOA error for the TOA-Trilateration method in both when the TC and RC car are on same-side and opposite-side scenarios at different time stamps. The DOA errors depict that the method of TOA-Trilateration was enough to provide accurate signal direction results. Integrating the other factors such as reflections, noise, and many more, will assist in providing more refined DOA values which are kept for future works.

Time stamp	Same-side DOA	Opposite-side DOA
1ns	3.279	5.356
0.1ns	2.325	4.345

Table 2 Average DOA error

8.2 Object Detection Results:

We trained the YOLOv4 model on our dataset in Matlab. The below graph shows the Precision-recall curve for the trained model. This trained model is saved as a .mat file which is further used to compare the computational time of normal image and ROI image.

Fig.17 provides the Precision-recall (PR) curve. The precision on the Y-axis ranges from 0.94 to 1 and the recall on the x-axis ranges from 0 to 1. Average precision is basically the overall performance of the model across various recall levels.

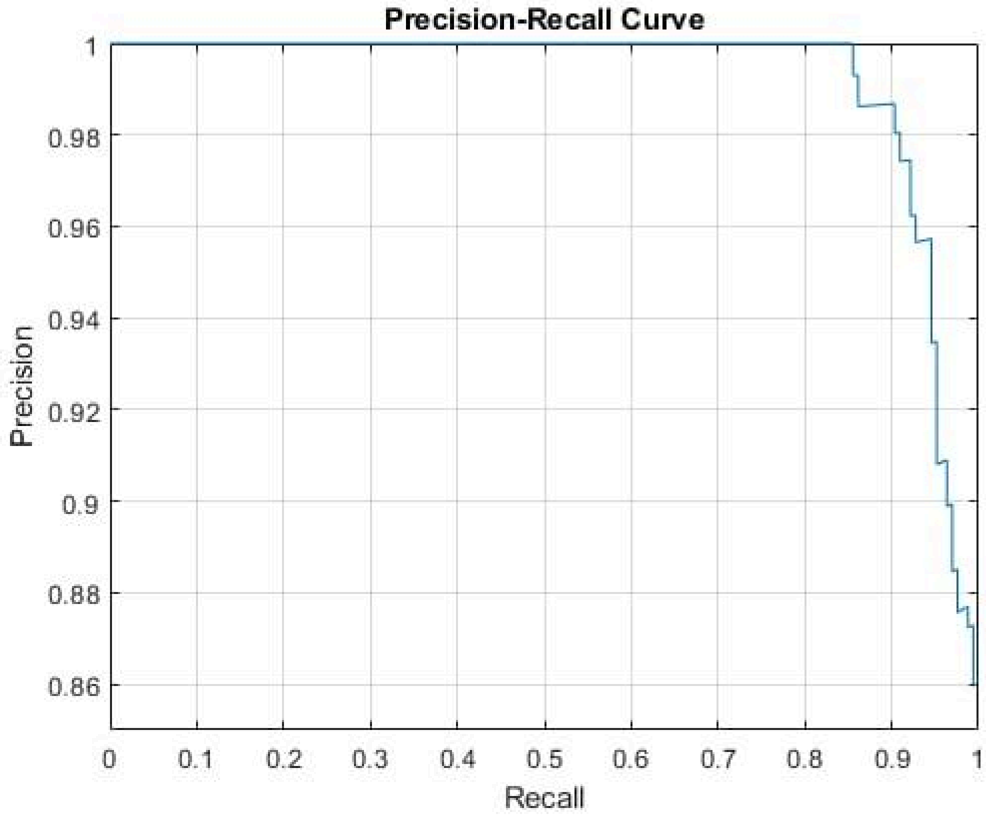

Figure 17: Precision-Recall Curve

It is concluded from Fig.17 that the model's performance achieves both high precision and recall. It means the model can detect the vehicle accurately while making a few errors.

The area under the curve(AUC) represents the average precision of the detector, which comes out to be 89 percent. The PR curve in Fig. 17, As the recall rate increases, false positive detections usually increase, and precision decreases. Hence, the closer the curve is to the upper right corner, the better the detection performance is.

The object detection result comparison between the ROI image as shown in Fig. 19 and the standard image as in Fig. 18 shows the following benefits of using DOA for choosing ROI in the images.

1. It helps in selecting the object(transmitter vehicle in our case) that interests us as shown in the image().
2. It decreases the computation time of the detector as it decreases the overall region of interest for the detector which means the detector has fewer features, objects, or pixels to process. Hence, decreasing the overall complexity and computational time in comparison to the standard image.
3. It helps in improving the confidence score of the detector as it removes overall complexity or objects with similar features that might confuse the detector that is present in the standard image.

Figure 18: Object Detection in standard image

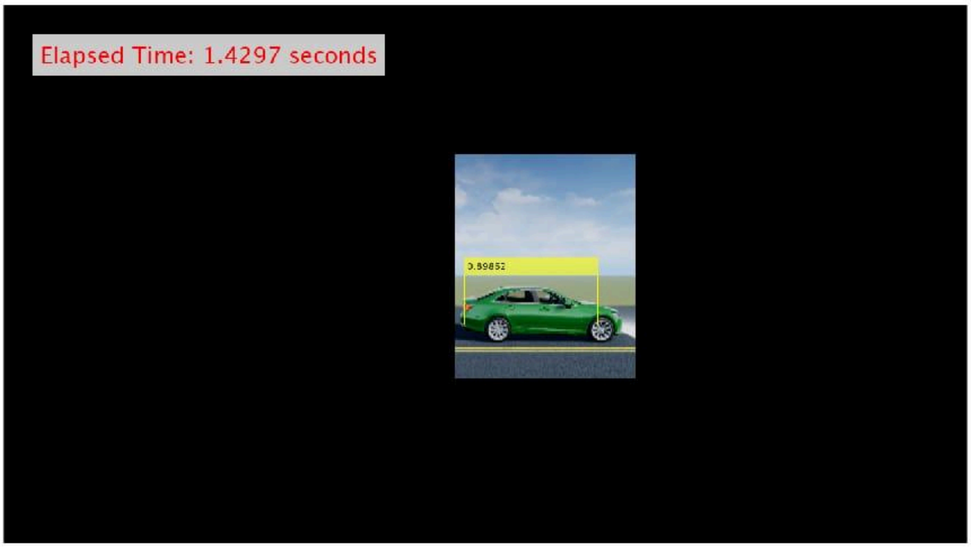

Figure 19: Object Detection in ROI image

We also iterated the ROI-selected images and the Standard images to compare the average computational time. The comparison can be seen in Table 3.

Number of Images	Standard Image	ROI-selected Image
1	1.4207	0.8858
10	0.90223	0.854480
100	0.83198	0.76747

Table 3: Average Elapsed Time

Number of Images	Standard Image	ROI-selected Image
1	0.79939	0.97676
10	0.810488	0.98967
100	0.8210	0.9853

Table 4: Average Confidence Score

From Table3, it is deduced that the average computational time for both the ROI-selected image and Standard images has a significant difference which is approximately 10 percent.

Table 4, summarises the average confidence score of vehicle detection for both Standard and ROI-selected images. It is concluded from the Table IV, that the confidence score has improved by 17 percent for ROI-selected images as compared to Standard images.

Chapter-9

Conclusion and FutureWorks

9.1 Conclusion

Sensor fusion is a great way to achieve high reliability for AVs. Extensive research in this field is vital to improve the accuracy and reliability of various sensors fused, especially in systems critical to user safety. This paper explores the merits of the fusion of the RF signal and vision sensor to detect the transmitting vehicle. The proposed method is successful in reducing the computational time by 10 percent with an improved confidence score for detection by 17 percent.

9.2 Future work

To further enhance the system, Non-Line of Sight(NLos) scenarios will be incorporated to test the effectiveness of the whole system which further can find its potential application in collision avoidance by improving the reliability and accuracy through the fusion of on-board sensors. Additionally, we are also considering introducing interference, noise and the complexity of the signal environment in the future.

References

1. Singh, S. Critical reasons for crashes investigated in the national motor vehicle crash causation survey Tech. Rep. 2015
2. R. Ali, R. Liu, Y. He, A. Nayyar and B. Qureshi, "Systematic Review of Dynamic Multi-Object Identification and Localization: Techniques and Technologies," in *IEEE Access*, vol. 9, pp. 122924-122950, 2021, doi: 10.1109/ACCESS.2021.3108775
3. S. H. Siadat and A. Selamat, ‘‘Location-based system for mobile devices using RFID,’’ in Proc. 2nd Asia Int. Conf. Model. Simul. (AMS), May 2008, pp. 291–296, doi: 10.1109/AMS.2008.44.
4. S. Chadwick, W. Maddern, and P. Newman, “Distant vehicle detection using radar and vision,” in Proc. Int. Conf. Robot. Automat., 2019, pp. 8311–8317
5. J. D. Choi and M. Y. Kim, "A Sensor Fusion System with Thermal Infrared Camera and LiDAR for Autonomous Vehicles: Its Calibration and Application," 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), Jeju Island, Korea, Republic of, 2021, pp. 361-365, doi: 10.1109/ICUFN49451.2021.9528609.
6. J. Kim, D. S. Han and B. Senouci, "Radar and Vision Sensor Fusion for Object Detection in Autonomous Vehicle Surroundings," 2018 Tenth

International Conference on Ubiquitous and Future Networks (ICUFN), Prague, Czech Republic, 2018, pp. 76-78, doi: 10.1109/ICUFN.2018.8436959.

7. D. Feng et al., “Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges,” IEEE Trans. Intell. Transp. Syst., vol. 22, no. 3, pp. 1341–1360, Mar. 2021.
8. R. Ali, R. Liu, Y. He, A. Nayyar and B. Qureshi, "Systematic Review of Dynamic Multi-Object Identification and Localization: Techniques and Technologies," in IEEE Access, vol. 9, pp. 122924-122950, 2021, doi: 10.1109/ACCESS.2021.3108775.
9. K. Kim, C. Lee, D. Pae, and M. Lim, “Sensor fusion for vehicle tracking with camera and radar sensor,” in Proc. 17th Int. Conf. Control, Autom. Syst., 2017, pp. 1075–1077.
10. H. Jha, V. Lodhi, and D. Chakravarty, “Object detection and identification using vision and radar data fusion system for ground-based navigation,” in Proc. 6th Int. Conf. Signal Process. Integr. Netw., 2019, pp. 590–593
11. X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, “Multi-view 3D object detection network for autonomous driving,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 6526–6534.

12. R. O. Chavez-Garcia and O. Aycard, “Multiple sensor fusion and classification for moving object detection and tracking,” IEEE Trans. Intell. Transp. Syst., vol. 17, no. 2, pp. 525–534, Feb. 2016.

13. S. Saleh, A. S. El-Wakeel, S. Sorour, and A. Noureldin, "Evaluation of 5G Cell Densification for Autonomous Vehicles Positioning in Urban Settings", ICCSPA, pp. 1-6, 2021.

14. Lee, S.; Han, K.; Park, S.; Yang, X. Vehicle Distance Estimation from a Monocular Camera for Advanced Driver Assistance Systems. Symmetry 2022, 14, 2657. https://doi.org/10.3390/sym14122657

15. International Telecommunications Union ITU, "Comparison of time-difference-of-arrival and angle-of-arrival methods of signal geolocation," International Telecommunications Union ITU, Geneva, 2018.

16. 1.A. Norrdine, "An Algebraic Solution to the Multilateration Problem," *International Conference on Indoor Positioning and Indoor Navigation,* pp. 1-4, 2012.

17. Pradhan, S.; Bae, Y.; Pyun, J.-Y.; Ko, N.Y.; Hwang, S.-s. Hybrid TOA Trilateration Algorithm Based on Line Intersection and Comparison Approach of Intersection Distances. *Energies* **2019**, *12*, 1668. https://doi.org/10.3390/en12091668.

18. Bochkovskiy, Alexey & Wang, Chien-Yao & Liao, Hong-yuan. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection.

19. E. Plaut, E. Ben Yaacov and B. El Shlomo, "3D Object Detection from a Single Fisheye Image Without a Single Fisheye Training Image," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, TN, USA, 2021, pp. 3654-3662, doi: 10.1109/CVPRW53098.2021.00405.
20. Campbell, S.; O'Mahony, N.; Krpalcova, L.; Riordan, D.; Walsh, J.; Murphy, A.; Conor, R. Sensor Technology in Autonomous Vehicles: A review. In Proceedings of the 2018 29th Irish Signals and Systems Conference (ISSC), Belfast, UK, 21–22 June 2018.
21. Petit, F. The Beginnings of LiDAR—A Time Travel Back in History. Available online: https://www.blickfeld.com/blog/thebeginnings-of-lidar/#:~{}:text=Lidar%20technology%20emerged%20already%20in,such%20as%20autonomous%20driving% 20today (accessed on 20 December 2020)
22. Shahian Jahromi, B.; Tulabandhula, T.; Cetin, S. Real-Time Hybrid Multi-Sensor Fusion Framework for Perception in Autonomous Vehicles. Sensors 2019, 19, 4357.
23. Wang, Z.; Wu, Y.; Niu, Q. Multi-Sensor Fusion in Automated Driving: A Survey. IEEE Access 2019, 8, 2847–2868.
24. Determining the Mounting Position of Automotive Radar Sensors | Rohde & Schwarz. Available online: https://www.rohdeschwarz.com/applications/determining-the-mounting-p

osition-of-automotive-radarsensors-application-card_56279-661795 .html (accessed on 28 December 2020)

25. Walling, D.H. The Design of an Autonomous Vehicle Research Platform. Master's Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, 3 August 2017.
26. Campbell, S.; O'Mahony, N.; Krpalcova, L.; Riordan, D.; Walsh, J.; Murphy, A.; Conor, R. Sensor Technology in Autonomous Vehicles: A review. In Proceedings of the 2018 29th Irish Signals and Systems Conference (ISSC), Belfast, UK, 21–22 June 2018
27. Joglekar, A.; Joshi, D.; Khemani, R.; Nair, S.; Sahare, S. Depth Estimation Using Monocular Camera. IJCSIT 2011, 2, 1758–1763
28. Garg, R.; Wadhwa, N.; Ansari, S.; Barron, J.T. Lerxarning Single Camera Depth Estimation using Dual-Pixels. arXiv 2019, arXiv:1904.05822v3
29. Yahiaoui, M.; Rashed, H.; Mariotti, L.; Sistu, G.; Clancy, I.; Yahiaoui, L.; Yogamani, S. FisheyeMODNet: Moving Object Detection on Surround-view Cameras for Autonomous Driving. In Proceedings of the IMVIP 2019: Irish Machine Vision & Image Processing, Technological University Dublin, Dublin, Ireland, 28–30 August 2019.
30. Yogamani, S.; Hughes, C.; Horgan, J.; Sistu, G.; Varley, P.; O'Dea, D.; Uricar, M.; Milz, S.; Simon, M.; Amende, K.; et al. WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving. arXiv 2019, arXiv:1905.01489v2.

31. Heng, L.; Choi, B.; Cui, Z.; Geppert, M.; Hu, S.; Kuan, B.; Liu, P.; Nguyen, R.; Yeo, Y.C.; Geiger, A.; et al. Project AutoVision: Localization and 3D Scene Perception for an Autonomous Vehicle with a Multi-Camera System. arXiv 2019, arXiv:1809.05477v2.
32. C. R. Sunstein, "Rear visibility and some unresolved problems for economic analysis (with notes on experience goods)," Journal of BenefitCost Analysis, vol. 10, no. 3, p. 317–350, 2019.
33. C. Hughes, M. Glavin, E. Jones et al., "Wide-angle camera technology for automotive applications: a review," IET Intelligent Transport Systems, vol. 3, no. 1, pp. 19–31, 2009
34. E. Teoman and T. Ovatman, ''Trilateration in indoor positioning with an uncertain reference point,'' in Proc. IEEE 16th Int. Conf. Netw., Sens. Control (ICNSC), May 2019, pp. 397–402, doi: 10.1109/ ICNSC.2019.8743240.
35. J. Shen and A. F. Molisch, ''Passive location estimation using TOA measurements,'' in Proc. IEEE Int. Conf. Ultra-Wideband (ICUWB), Sep. 2011, pp. 253–257, doi: 10.1109/ICUWB.2011.6058839.
36. Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99.

37. Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788.
38. Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for largescale image recognition. arXiv preprint arXiv:1409.1556.
39. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In European conference on computer vision, pages 21–37. Springer
40. C. Queiroz, "Access Partnership," Access Partnership, 9 March 2021. [Online]. Available: https://www.accesspartnership.com/how-vehicular-communications-will-revolutionise-mobility/. [Accessed 10 April 2022]
41. e. E. Team, "everythingRF," everythingRF, 22 September 2020. [Online]. Available:https://www.everythingrf.com/community/what-is-ieee-802-11 p. [Accessed 28 May 2022].
42. "Avnet Abacus," Avnet Abacus, [Online]. Available: https://www.avnet.com/wps/portal/abacus/solutions/markets/automotive-and-transportation/automotive/communications-and-connectivity/v2v-communication/. [Accessed 17 May 2022]
43. S. K. a. P. R. P. S. Darbha, "Benefits of V2V Communication for Autonomous and Connected Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 5, pp. 1954-1963, 2019.

44. R. P. A. V. J. P. M. J. G. v. d. M. a. M. S. G. J. L. Naus, "String-Stable CACC Design and Experimental Validation: A Frequency-Domain Approach," IEEE Transactions on Vehicular Technology, vol. 59, no. 9, pp. 4268-4279, 2010.

45.R. P. F. Y. T. P. V. Jennie Lioris, "Platoons of connected vehicles can double throughput in urban roads," Transportation Research Part C: Emerging Technologies, vol. 77, pp. 292-305, 2017.

46.K. S. D. F. S. S. J. P. P. R. Zeinab El-Rewini, "Cybersecurity challenges in vehicular communications," Vehicular Communications, vol. 23, 2020. [26] E. &. A. I. Mitsakis, "Recent Developments on Security and Privacy of V2V & V2I Communications: A Literature Review," Periodica Polytechnica Transportation Engineering, vol. 48, no. 4, 2020.

47.https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic

48.J. Shen and A. F. Molisch, “Passive location estimation using TOA measurements,” in Proc. IEEE Int. Conf. Ultra-Wideband (ICUWB), Sep. 2011, pp. 253–257, doi: 10.1109/ICUWB.2011.6058839.

49. J. Shen, A. F. Molisch, and J. Salmi, “Accurate passive location estimation using TOA measurements,” IEEE Trans. Wireless Commun., vol. 11, no. 6, pp. 2182–2192, Jun. 2012, doi: 10.1109/TWC. 2012.040412.11069

50. Manuel Silverio, 2019, How to make a vehicle autonomous <https://towardsdatascience.com/how-to-make-a-vehicle-autonomous-16edf164c30f>

Achievement

International Conference

Parneet Kaur Dhindsa, Zhenni Pan, Shigeru Shimamoto; **Studies on Next-Gen Vehicular Detection: A Fusion of RF Signal and Fisheye Camera Technologies,** 2024 IEEE Vehicular Technology Conference 2024, Singapore. — Under Review

International Workshop

Parneet Kaur Dhindsa, Zhenni Pan, Shigeru Shimamoto; **RF Signal and Fisheye Camera Fusion for Robust Vehicular Detection,** Global Information and Telecommunication Workshop, December 16th 2023, Tokyo, Japan