POLITECNICO DI TORINO

Master of Science in Automotive Engineering
Department of Mechanical and Aerospace Engineering
Academic Year 2023-2024

MASTER OF SCIENCE THESIS

Research on Scenario Comparison for Driving Simulations in Automotive Applications.

Supervisors:
Prof Guido Albertengo [a]
Prof. Dr.-Ing. H.-C. Reuss [b]
Lukas Lang [b]

Candidate:
Vishnu Vijayan (s289871)

[a] Politecnico di Torino
[b] University of Stuttgart

Acknowledgement

This endeavor would not have been possible without the supervision and support of Prof. Guido Albertengo from Politecnico di Torino along with Prof. Dr.-Ing. H.-C. Reuss and Lukas Lang from the University of Stuttgart. I am extremely grateful for the guidance and motivation from my supervisors in achieving my dream of doing a Master's Thesis program, in autonomous driving, in Germany. I would like to extend my sincere thanks to both the universities, Politecnico di Torino and the University of Stuttgart, for providing me with this opportunity to do the Master's Thesis program abroad and helping me to learn about the professional working cultures in both Italy and Germany.

Additionally, this journey would not have been possible without the generous support from the FKFS Research Institute Stuttgart, which provided me with a great working space and all the necessary resources to complete my research activity. I am extremely thankful to Dr.-Ing. Bernhard Bäuerle, Academic Advisor for Automotive Engineering at the Institute of Automotive Engineering Stuttgart, who proposed the thesis opportunity and helped me through all the official procedures even after reaching Germany.

I would like to mention my special thanks to Daniel Heidle, my friend and colleague at FKFS, for his constant support and the great friendship that helped to learn a lot professionally and as a person. I had the pleasure of working with Valeo, which helped me to grow professionally and I am thankful to my manager and my colleagues for allowing me to achieve practical knowledge in the domain of ADAS.

Words cannot express my gratitude to my parents and family for their moral support and financial assistance without which I would not have achieved this dream. Thanks for all the prayers and blessings that kept me going during the difficult times. Finally, I would like to thank all my friends and the people who supported me during this wonderful journey.

Contents

Acronyms

6LM 6 Layer Model

A2D2 Audi Autonomous Driving Dataset

AIO All in One

DOF Degree of Freedom

FCM Fuzzy Cognitive Map

FMCW Frequency Modulated Continuous Wave

FOV Field of View

GPS Global Positioning System

h Height of the Bounding Box

IMU Inertial Measurement Unit

l Length of the Bounding Box

POI Point of Interest

SAE Society of Automotive Engineers

UAV Unmanned Aerial Vehicle

VRU Vulnerable Road User

w Width of the Bounding Box

Symbols

Symbol	Unit	Description
v	m/s	Velocity of the ego car
x	m	Translation on X-Axis
y	m	Translation on Y-Axis
z	m	Translation on Z-Axis
t	m	Time in seconds
d	m	Distance in metres
d_{max}	m	Maximum distance in metres
θ	$degree$	Angle in degree
$Cosine_{distance}$	m	Distance metric in metres
$Euclidean_{distance}$	m	Distance metric in metres
$Normalized_{distance}$	m	Distance metric in metres
$Manhattan_{distance}$	m	Distance metric in metres

List of Figures

List of Tables

Kurzfassung

Die Masterarbeit konzentriert sich auf Methoden des Szenenvergleichs für Fahrsimulationen und auf das Finden einer ähnlichen Szene aus zwei verschiedenen Fahrdatensätzen. Die Forschung wird für die Anwendungen in autonom fahrenden Autos durchgeführt, bei denen das Verständnis der aktuellen Fahrszene sehr wichtig ist. Die Fähigkeit dieser autonomen Systeme, ihre Umgebung genau zu erfassen und zu verstehen, so dass sie in Echtzeit präzise Urteile fällen können, ist ausschlaggebend für ihren Erfolg. Das Ziel dieser Arbeit ist es, eine Fahrszenenbeschreibung zu definieren und quantifizieren wie auch die Ähnlichkeit zwischen zwei Szenen aus Datensatz A und Datensatz B zu finden.

Die Forschungstätigkeit begann mit der Untersuchung bestehender Methoden zur Beschreibung von Fahrszenen, die in anderen Anwendungen wie unbemannten Luftfahrzeugen eingesetzt werden, da es nicht viele Forschungsergebnisse im Bereich der Automobilindustrie gab. Die Literaturrecherche wurde genutzt, um verschiedene Parameter der Fahrszene zu verstehen und die Beziehungen zwischen ihnen herzustellen. Die von den Sensoren vorgenommenen Messungen zur Bewertung der Parameter waren als verschiedene Fahrdatensätze verfügbar, so dass auch eine Recherche zu den vorhandenen Fahrdatensätzen durchgeführt werden konnte. Für die Implementierung der Vergleichsmethoden wurde ein Satz von fünf verschiedenen Fahrparametern und zwei Fahrdatensätzen ausgewählt. Die ausgewählten Fahrdatensätze waren nuScenes und View of Delft, und die Parameter wurden nur für drei verschiedene Objektklassen extrahiert: Autos, Fußgänger und Fahrräder.

Das Tool *Python* wurde verwendet, um die erforderlichen Fahrparameter aus den ausgewählten Fahrdatensätzen zu ermitteln. Die Korrelationsanalyse der Parameter wurde durchgeführt und eine detaillierte Studie der Datenverteilung der Parameter in jedem Fahrdatensatz wurde in dieser Masterarbeit durchgeführt. In dieser Arbeit wurde mit einer sogenannten Ähnlichkeitsanalyse eine neue Idee vorgeschlagen, um die beiden Fahrszenen zu vergleichen. Diese Ähnlichkeitsanalyse wurde mit dem Tool *MATLAB* durchgeführt.

Abstract

The research activity focuses on the scene comparison methods for driving simulations and on finding a similar scene from two different driving datasets. The research is done for the applications in autonomous driving cars, where understanding the current driving scene is very important. The ability of these autonomous systems to precisely sense and comprehend their surroundings, enabling them to make precise judgments in real-time, is what determines their success. The goal of the thesis is to answer two important questions, the first one being how to define and quantify a driving scene description and the second one about how to find the similarity between two scenes from dataset A and dataset B.

The research activity commenced with the study of existing driving scene description methods used in other applications like unmanned aerial vehicles since there were not many results of the research in the automotive domain. The literature review has been used to understand different driving scene parameters and to establish the relations between them. The measurements made by the sensors, for evaluating the parameters, were available as different driving datasets, so research on the existing driving datasets was also conducted. A set of five different driving parameters and two driving datasets were selected to implement the comparison methods. The driving datasets selected were the nuScenes and the View of Delft and parameters were extracted only for three different classes of objects; cars, pedestrians, and bikes.

The *Python* was the tool that was used to find the required driving parameters from the selected driving datasets. The correlation analysis of the parameters was conducted and a detailed study of the data distribution of the parameters in each driving dataset was done in this research activity. A new idea has been proposed in the thesis to compare the two driving scenes and this research activity on similarity analysis was performed using the tool *MATLAB*.

1 Introduction

This section introduces the thesis report subject. Initially, there is a brief discussion about the background and scope and then the chapter proceeds to have a look at the research gap and the research questions answered by the thesis report.

1.1 Motivation

In recent years, the advancement of autonomous driving technology has captured the attention of both researchers and industries alike. The promise of safer roads, reduced traffic congestion, and increased mobility has propelled the development of self-driving vehicles. Central to the success of these autonomous systems is their ability to accurately perceive and interpret the surrounding environment, allowing them to make informed decisions in real-time. Achieving this level of sophistication requires extensive training and testing using diverse datasets that encompass a wide array of driving scenarios.

In this thesis, the research focuses on the crucial task of comparing scenes from two distinct open-access driving datasets. By performing a comprehensive comparative analysis, this research aims to uncover patterns, disparities, and potential pitfalls that might arise during the deployment of autonomous vehicles in complex real-world environments.

As acknowledged by [1], knowing the current scene is an increasing part of automotive applications due to the growing emphasis on real-time situational awareness for advanced driver assistance systems and autonomous vehicles. The ability to accurately recognize and interpret the surrounding environment is pivotal for ensuring safe and efficient operation in various driving scenarios. However, the existing driving datasets predominantly remain sensor-driven, as pointed out by [2], which poses a significant challenge in effectively searching for relevant scenes within these datasets. The development of enriched datasets that not only include sensor data but also contextual information such as road characteristics, traffic situations, weather conditions, and spatial interactions between objects would be necessary to close this research gap. This work attempts to fill that gap by conducting a thorough and detailed comparison of two open-access driving datasets, giving useful insights for autonomous driving researchers and developers.

Research Questions

The complete research activity focuses on the two main questions as discussed below and the ways to find an effective solution;

1. How can scene description be effectively defined and quantified when searching for similar driving scenes in two different datasets, A and B?
2. How to evaluate the similarity between two scenes from dataset A and dataset B in the context of driving scenarios?

2 Literature review

This chapter will discuss the state of art methods available in the domain of autonomous driving and scene comparison methods. A detailed explanation of the concept of driving scenes and its comparison methods, the datasets available, and different parameters researched till now by different automotive enthusiasts will be discussed in this section. Let's start the discussion by understanding the concept of autonomous driving in the section 2.1.

2.1 The concept of autonomous driving

Autonomous driving, commonly referred to as self-driving, is a ground-breaking idea that seeks to completely change the automotive industry by allowing vehicles to navigate and run without the need for human input. According to a report by [3], autonomous vehicles have the potential to reduce traffic accidents by up to 90 percent, revolutionizing the automotive industry. This technology has the potential to transform transportation infrastructure, raise driving standards, and increase mobility for many societal groups.

Furthermore, the study conducted by [4] demonstrates that autonomous driving systems, through real time data processing and adaptive control strategies, can significantly enhance vehicle performance and response to dynamic road conditions, paving the way for a more efficient and reliable transportation ecosystem. The sophisticated sensors that provide vehicles with the ability to precisely see and understand their environment are essential to the success of autonomous driving.

This study of the literature explores the idea of autonomous driving, underlines its importance, and concentrates on significant technological developments. The main goals of autonomous driving are to increase traffic flow, improve road safety, and provide practical transportation options. The fusion of numerous technologies that enable vehicles to detect their environment, make instantaneous decisions, and carry out actions that resemble human driving behavior is necessary to accomplish these goals. Sensor technology is one of the essential elements enabling autonomous driving.

2.1.1 Levels of automation

Autonomous driving systems are categorized into levels based on their increasing automation and decreasing reliance on human intervention. The classification follows the Society of Automotive Engineers (SAE) J3016 standard [5].

- Level 0 - No Automation: At this level, the human driver is responsible for all aspects of driving, including control and monitoring. There might be warnings or momentary interventions, but the vehicle itself does not perform any driving tasks autonomously.
- Level 1 - Driver Assistance: Driver assistance systems provide limited automation, typically in a single aspect of driving, such as adaptive cruise control or lane-keeping assistance. The driver remains engaged and responsible for overall vehicle control. According to the SAE J3016 standard, at Level 1, the human driver performs the dynamic driving task.
- Level 2 - Partial Automation: Level 2 automation allows the vehicle to control both steering and acceleration/deceleration simultaneously under certain conditions. However, the driver must monitor the environment, remain engaged, and be ready to take control at any moment. SAE J3016 states that the automated system performs the entire dynamic driving task, but the human must monitor the system at all times and perform the rest of the driving tasks.
- Level 3 - Conditional Automation: At this level, the vehicle can manage most aspects of driving within specific conditions and environments. The driver can disengage from actively monitoring the vehicle, but must still be available to take control if requested by the system. The SAE J3016 standard states that the human is a fallback, and has the right to intervene; however, the automated system does not require the human to intervene.
- Level 4 - High Automation: Level 4 vehicles are capable of fully autonomous operation within predefined conditions and environments. These vehicles can operate without human intervention or oversight in specific scenarios, such as a designated geographic area or favorable weather conditions. The SAE J3016 standard defines Level 4 as a system that can perform the dynamic driving task and monitor the driving environment but only under certain conditions.
- Level 5 - Full Automation: Level 5 represents complete autonomy in all driving scenarios and conditions, with no human intervention required. These vehicles lack traditional controls like steering wheels and pedals, as there is no need for human driving. The SAE J3016 standard defines Level 5 as full self-driving automation, where the vehicle an perform all driving tasks and monitor the driving environment without any human intervention or oversight.

2.1.2 Sensors in autonomous driving

Sensors constitute the foundation of autonomous driving systems, facilitating the perception of the vehicle's environment. LiDAR (Light Detection and Ranging), Radar, Cameras, and Ultrasonic sensors work in tandem to collect and process data, enabling the vehicle to detect obstacles, pedestrians, road markings, and other relevant information. LiDAR, in particular, employs laser pulses to create detailed 3D maps of the surroundings, offering a high level of accuracy and reliability. According to [6], the ability of LiDAR to generate precise depth information significantly enhances object recognition and localization accuracy, thereby playing a pivotal role in autonomous navigation. Additionally, cameras, as highlighted by[7], contribute to visual scene analysis and interpretation, enabling the vehicle to understand traffic signs, signals, and human gestures. The integration of multiple sensor types ensures redundancy and robustness, crucial for addressing various environmental conditions and scenarios. Ultrasonic sensors aid in detecting nearby objects for precise maneuvering, especially during parking.

2.1.3 Technological advancements

The development of autonomous driving systems has been accelerated by advances in sensor technology. Originally big and expensive, LiDAR sensors have undergone significant cost and size reduction. For example, solid-state LiDAR decreases the form factor and gets rid of moving elements, improving affordability and dependability. Additionally, enhanced object detection and tracking skills have been made possible by advancements in sensor resolution and accuracy.

With the use of cutting-edge signal processing methods, radar sensors have also made significant advancements. These developments make it possible for radar systems to distinguish between various things, such as cars, bikes, and pedestrians, improving safety in intricate urban areas. Camera technology has made significant strides in object detection and scene understanding because of advancements in computer vision and deep learning. Modern cameras are essential for autonomous navigation because they can recognize and track objects, read traffic signals, and read road signs.

The use of sensor fusion techniques, which combine data from several sensors to produce a more complete and accurate sense of the environment, has also become a major trend. The car is better able to handle a variety of situations, such as bad weather, dim lighting, and complicated road geometry, thanks to sensor fusion. The fusion and integration of sensor data represent a significant leap in autonomous driving. LiDAR, radar, and camera data are combined to improve overall perception accuracy and reduce the limitations of individual sensors. As described by [8], Kalman filters and Bayesian frameworks are frequently used to merge various heterogeneous data sources. The vehicle can produce a thorough and accurate depiction of its surroundings thanks to this fusion method, which is essential for making wise driving judgments.

Deep learning algorithms have changed autonomous vehicle object recognition and scene comprehension. Convolutional Neural Networks (CNN) have proven to be incredibly effective at spotting people, cars, and other road elements in sensor data. AlexNet, a CNN design that set the path for later advancements in image analysis, was introduced by [9]. Vehicles can now recognize intricate details in real time thanks to the deep learning revolution, making navigation safer and more effective. [9]

High-definition (HD) maps offer accurate and current information regarding infrastructure, lane markings, and roads. These maps, along with real time sensor data, are used by autonomous cars to provide precise localization. According to [10], high-precision maps are essential for obtaining centimeter-level localization accuracy, which is essential for safe navigation, particularly in complicated urban contexts. [10]

2.1.4 Impact and challenges

Numerous advantages could result from the development of autonomous driving technology, including fewer traffic accidents, better traffic flow, more accessibility for those with disabilities, and lower energy use. However, there are still several issues to be resolved.

Firstly, regarding safety concerns, the task of ensuring the safety of driverless vehicles is still very important. While the perception and decision-making capabilities of vehicles have increased because of technology breakthroughs, the unpredictable nature of real-world surroundings necessitates the use of reliable fail-safe devices and redundant systems.

The next concern is regulations and standards, widespread deployment of autonomous vehicles has been delayed by the lack of uniform laws and norms. Liability, safety certification, and the integration of autonomous vehicles into current traffic networks all require clear regulations.

Finally, ethical considerations, when harm is inevitable, autonomous cars may face moral quandaries. For instance, deciding how to balance pedestrian safety with that of vehicle occupants presents difficult ethical issues that demand careful thought.

2.2 Terms scene, situation, and scenario

In the rapidly advancing field of automated driving, clear definitions of fundamental terms are crucial to foster effective communication and promote a shared understanding. This paper [11] aims to define and substantiate the terms scene, situation, and scenario within the context of automated driving systems. The authors begin by acknowledging the necessity of standardized definitions for these terms to facilitate meaningful discussions among researchers, engineers, and stakeholders in the field. They emphasize that a precise understanding of these terms is pivotal for the development and evaluation of autonomous driving technologies.

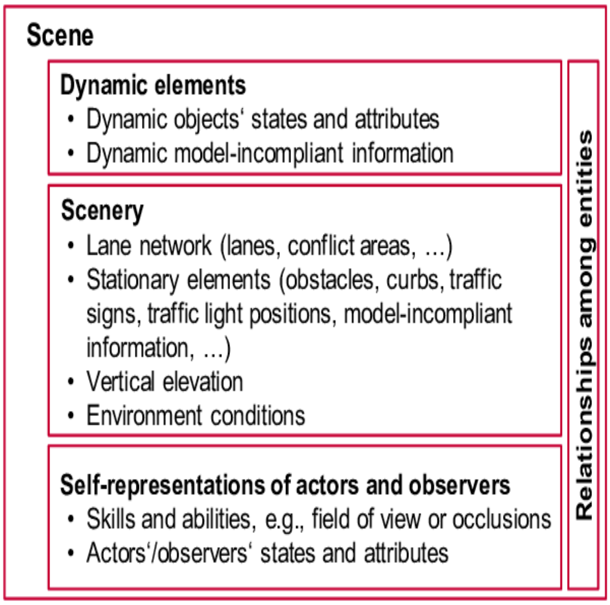

Scene
Dynamic elements
• Dynamic objects‘ states and attributes
• Dynamic model-incompliant information
Scenery
• Lane network (lanes, conflict areas, ...)
• Stationary elements (obstacles, curbs, traffic signs, traffic light positions, model-incompliant information, ...)
• Vertical elevation
• Environment conditions
Self-representations of actors and observers
• Skills and abilities, e.g., field of view or occlusions
• Actors‘/observers‘ states and attributes
Relationships among entities

Figure 2.1: Example of a scene representation [11]

2.2.1 Scene

The [11] defines the term scene in the following way:

"A scene describes a snapshot of the environment including the scenery and dynamic elements, as well as all actor's and observer's self-representations, and the relationships among those entities. Only a scene representation in a simulated world can be all-encompassing (objective scene, ground truth). In the real world, it is incomplete, incorrect, uncertain, and from one or several observer's points of view (subjective scene)."

The term scene refers to the immediate surroundings of an automated vehicle, encompassing all objects, entities, and environmental factors that can be perceived by the vehicle's sensors at a given moment. This definition will be followed in Chapter 3 to explain the methodology. The fundamental purpose of a scene is to act as an interface between modules that deal with the environment and one's perception on the one hand, and modules and tasks that are unique to an application or mission on the other. These elements include road infrastructure, traffic signs, other vehicles, pedestrians, and any other relevant features within the sensor's range. A scene is essentially a snapshot of the vehicle's sensory inputs at a particular point in time. The **Figure 2.1** depicts the elements of a scene. A scene is made up of dynamic elements, spatially static scenery, and a collective representation of all performers and viewers. [11]

2.2.2 Situation

The [11] defines the term situation in the following way:

"A situation is the entirety of circumstances, which are to be considered for the selection of an appropriate behavior pattern at a particular point of time. It entails all relevant conditions, options and determinants for behavior. A situation is derived from the scene by an information selection and augmentation process based on transient as well as permanent goals and values. Hence a situation is always subjective by representing an element's point of view."

Moving beyond the raw sensory data of a scene, a situation involves the interpretation and understanding of the relationships and dynamics between the various elements present in the scene. It encompasses the ability of the automated system to process the sensory inputs and recognize patterns and interactions. For instance, recognizing a red traffic light, or a pedestrian waiting to cross contributes to defining a specific driving situation. The situation is made up of several situational factors that situation assessment modules must interpret or comprehend. Such modules simultaneously input and output a scenario. There is a significant overlap between a scene and a situation to contain, all pertinent scenery, all pertinent dynamic components, and all pertinent self-representational features.

2.2.3 Scenario

The [11] defines the term scenario in the following way:

"A scenario describes the temporal development between several scenes in a sequence of scenes. Every scenario starts with an initial scene. Actions & events as well as goals & values may be specified to characterize this temporal development in a scenario. Other than a scene, a scenario spans a certain amount of time."

A scenario is a broader and more complex context that encompasses a sequence of scenes and situations. It represents a particular driving setting or challenges that an automated vehicle might encounter. The definitions of situation and scenario will be useful in understanding the section 2.3.4. Scenarios vary in complexity, ranging from routine situations like highway driving to intricate and demanding situations such as navigating through crowded intersections or adverse weather conditions. A scenario, unlike scenes, lasts for a specific period. To completely characterize a path, a scenario must contain at least one (starting) scene as well as actions and events. A scenario can alternatively be described by a full collection of scenes, with the actions and events just spanning the passage of a given amount of time. Test cases are given to simulate and test an automated vehicle or its modules. A scenario and pass/fail criteria are included in each of them for evaluation. In addition, the system's use case, or functional description, must be specified in the system's early design stages by the V-model. **Figure 2.2** depicts the elements of a scenario implementation. At least one scene, actions and events, and goals and values all make up a scenario. [11]

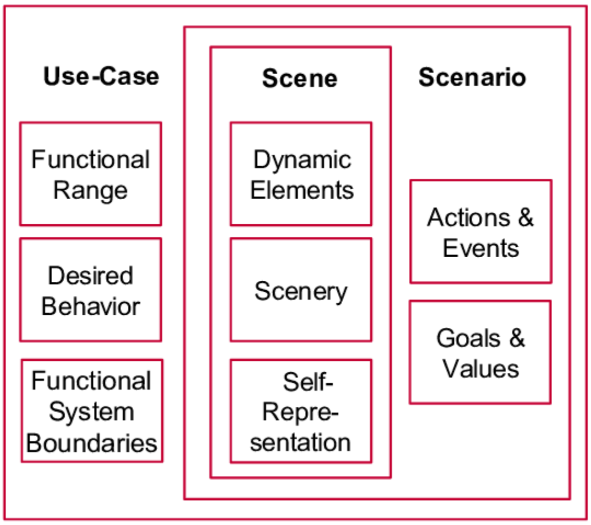


Figure 2.2: Example of a scenario representation [11]

2.3 Understanding the driving scene

Intelligent driving, advanced driver assistance systems (ADAS), and human driving all depend on the ability to comprehend driving scenes. Interpreting and making sense of the complex and continuously shifting environments in which vehicles operate is included in it. To enable cars to navigate safely, make wise decisions, and adapt to changing circumstances, a combination of sensors, data processing, and artificial intelligence is used to interpret the world. It is evident how difficult it is to fully comprehend a driving situation due to issues with urban environments, data fusion, real time processing, robustness, and ethical issues. More advanced and dependable solutions that improve road safety and convenience should be on the horizon as technology develops [12]. In this discussion, it is possible to understand the fundamentals of understanding driving scenes and how it serves as a basis for scene comparison.

2.3.1 Components of understanding driving scenes

The essential factors that are needed to understand a scene are discussed in detail in this section. The first one is the data collection, data gathered from several sensors, including cameras, LiDAR, radar, Global Positioning System (GPS), and ultrasonic sensors, is the first step towards understanding driving situations. An abundance of information about the environment around the car is provided by these sensors. The second factor is perception and object detection, sophisticated computer vision algorithms use sensor data to create an environment perception. Objects including cars, pedestrians, road signs, and obstructions can all be detected and categorized in this way. For object recognition, deep learning methods like convolutional neural networks (CNN) are frequently employed.

Semantic segmentation is a significant component in driving scene understanding. Understanding a scene through driving involves more than just identifying objects; it also entails examining how those objects behave. For example, safe and efficient driving requires knowing the trajectory of a person or the intents of another vehicle. Road infrastructure and sign recognition is another factor that focuses on recognizing and understanding road infrastructure, such as stop signs, traffic lights, and lane markers, which is necessary to comprehend road scenes. This is essential for following traffic laws and driving safely. Finally, environmental awareness is a very important component since assessing the surrounding environment, such as the lighting, weather, and state of the road surface, is part of scene understanding. These elements have an impact on the safety and behavior of the car.

2.3.2 Driving scene comparison

Driving scenes are dynamic and complex, comprising diverse environments, road conditions, and traffic situations. Comparing driving scenes is a useful technique that may be applied to many situations, including enhancing traffic control, assessing the effectiveness of autonomous vehicles, and enhancing road safety [12]. It makes it possible to evaluate various driving situations and create plans of action to improve traffic safety and the effectiveness of transportation. The primary factors that set driving scenarios apart are examined in this comparative analysis, emphasizing the significance of these factors in raising road safety and transportation effectiveness. Some of the commonly used ways of scene comparison are as follows:

Urban vs rural scenes: Both human drivers and autonomous vehicles operate in different and diverse conditions in urban and rural driving scenes. Urban environments necessitate more awareness and quick decision-making due to their high traffic density, intricate crossings, and frequent pedestrian encounters. On the other hand, rural areas tend to have faster speeds but also offer simpler road layouts, less traffic, and more predictable driving conditions. To construct adaptive vehicle systems and improve traffic management, it is essential to comprehend these variations since doing so helps to develop technologies that can solve the unique problems and safety concerns associated with each location. [13]

Daytime vs nighttime scenes: Driving sequences throughout the day and at night depict two different times of the day, each with its own set of challenges and difficulties. Daytime settings benefit from an abundance of natural light, which makes visibility better and makes it easier for drivers to see their surroundings. On the other hand, scenes shot at night provide more difficulties because of the decreased visibility, changing lighting, and more reliance on artificial lighting. To ensure safe and effective driving in both daytime and nighttime conditions, adaptive lighting systems, cutting-edge night vision technologies, and improved driver assistance systems are essential. These scenes differ, necessitating customized approaches to scene understanding and vehicle operation. [14]

Adverse weather vs clear weather scenes: According to [15], bad weather, such as rain, snow, and fog, has a big impact on how well people comprehend a scene. Scenes with adverse weather and scenes with clear weather are two different driving scenarios, each with unique consequences for road safety and scene comprehension. Clear skies provide drivers with the best vision and road conditions, making for reliable driving experiences. When there is bad weather, like rain, snow, or fog, visibility is greatly reduced and road surfaces become dangerous, necessitating increased vigilance and modified driving techniques. In order to help vehicles adapt to unfavorable weather conditions by modifying speed, following distance, and other parameters to ensure safe and dependable transportation, weather adaptive driving systems which ultimately improve road safety and reduce weather related accidents need to be developed. [15]

Intersections vs highways: According to [16], Intersections are complicated hubs that are frequently seen in cities, with plenty of moving cars and potential places for confrontation. When driving at an intersection, drivers have to make snap decisions and deal with a lot of uncertainty. Highways, on the other hand, offer a comparatively less complicated and predictable driving environment, with fewer collisions and a constant pace. Designing efficient traffic management systems and autonomous driving technologies requires an understanding of the distinctions between intersections and highways. This is because these distinctions influence safe and effective navigation strategies within each context, which enhances overall road safety and optimizes traffic flow.

2.3.3 Parameters for driving scene comparison

The particular standards or qualities that are employed to assess and compare various driving situations are known as parameters for driving scene comparison. For one to fully understand the special qualities and difficulties connected with any kind of driving environment, these parameters are necessary. These criteria offer a framework for assessing and contrasting various driving scenarios, enabling a more thorough comprehension of their distinctive qualities and difficulties. These factors are used by researchers, planners of transportation, and developers of autonomous vehicles to maximize overall transportation systems' safety, efficiency, and effectiveness.

To compare driving scenes, the following important parameters are frequently utilized:

1. Traffic density:

   The quantity of cars on the road at any given moment is referred to as traffic density, and it has a big influence on the traffic flow and safety. High traffic density, which is frequently seen in metropolitan environments, can cause congestion and reduced average speeds, making it difficult for cars to maneuver around congested streets. Because there is generally less traffic in rural areas, traffic can go more smoothly and at higher speeds. [17]

2. Speed:

   One important factor influencing the dynamics of driving scenarios is speed. It shows the maximum and average speeds of cars in a certain area. Highway scenes are typified by high-speed driving because vehicles can maintain steady speeds there, but urban scenes have more frequent speed fluctuations and slower average speeds because of traffic congestion. For the purpose of managing traffic and maintaining road safety, speed evaluation is essential. [13]

3. Road conditions:

   For both comfort and safety when driving, the condition of the road surface is essential. Urban roads can range in condition from well-kept to broken, which could have an impact on a car's suspension and driving experience. Driving on rural roads can be unpredictable due to their unpredictability and poor maintenance since some of them are unpaved. [18]

4. Complexity:

   A scene's complexity may be affected by several elements, including the number of lanes, crossroads, and possible dangers. Compared to rural settings, where road layouts are simpler and traffic interactions are less frequent, urban scenarios are usually more complicated due to their numerous junctions and diverse road users.

5. Safety:

   Accident rates, difficult driving conditions, and the availability of safety equipment are examples of safety criteria. Due to their larger traffic densities, urban locations frequently have higher accident rates, whereas rural areas may have fewer accidents overall but have particular difficulties because of their slower emergency response times. [19]

6. Road signs and markings:

   Navigation and communication depend heavily on the existence and caliber of road signs, lane markers, and other signage. Rural roads could have fewer and less thorough markings than urban ones, which are usually well-designated with signs. In addition to posted traffic signs and traffic lights, road markings sometimes include painted directions such as speed restrictions, stopping lines, or turn arrows. [20]

7. Driver behavior:

   Different driving situations might call for different driving behaviors from drivers, such as aggressive driving in city traffic to more careful and laid-back driving in rural locations. According to the AAA Foundation for Traffic Safety (2021), among three measures examined in this survey—perceived danger, risk of apprehension, and social disapproval—people's engagement in unsafe driving behaviors was associated with the perceived danger and social disapproval. Driving in urban areas can be stressful due to excessive traffic, which might result in hostile behavior. In contrast, driving in rural areas may be less stressful. [21]

8. Vehicle type:

   Different driving situations may call for different kinds of vehicles, such as trucks, motorbikes, passenger cars, and specialty vehicles. Smaller passenger cars are more common in urban areas than bigger vehicles like trucks and SUVs, which are more common in rural regions.

9. Dynamic objects:

   The things that may move are mentioned in this section. These things might be motionless, resting in one place, like parked cars, motionless pedestrians, trash cans positioned in the middle of the road, etc. Observe that the term contains not just entities that are intended to move, but also entities that move regularly or in response to an outside stimulus. Every traffic participant gives a good representation of a dynamic object. They include automobiles (including trailers), motorbikes, bicycle riders, pedestrians, and rail transportation like trams. Moreover, moving items and animals are included in the category of objects. [22]

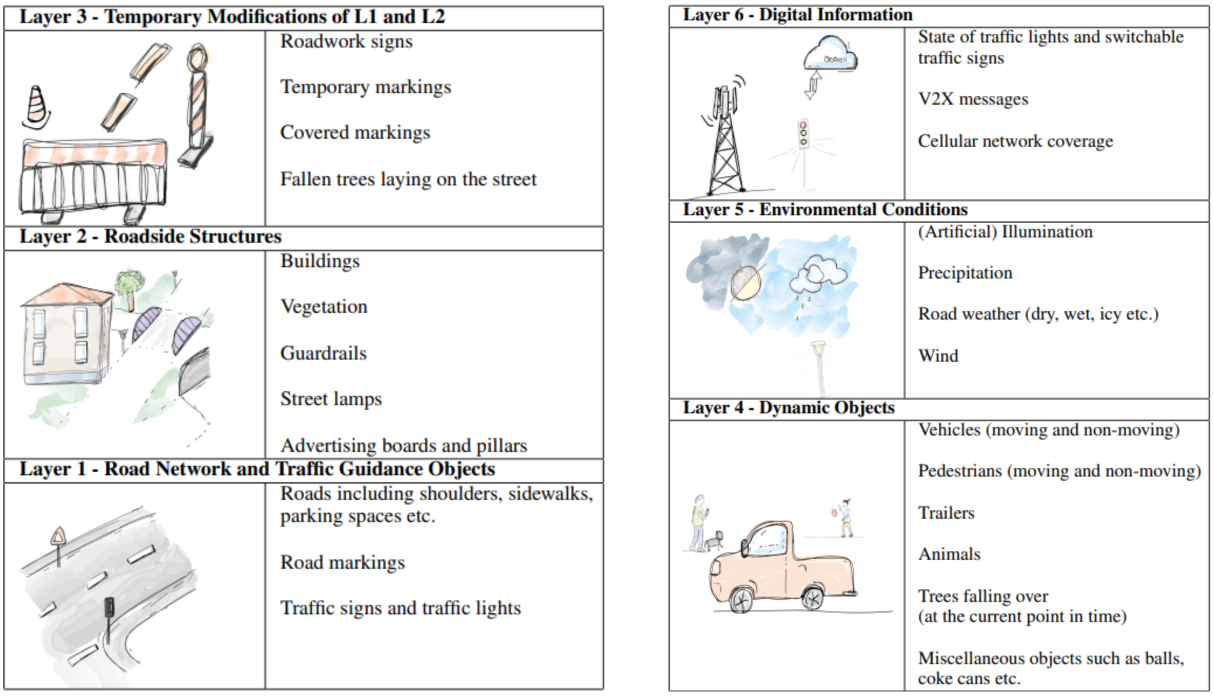
10. Roadside structures:

    According to [22] includes every stationary item that is typically positioned next to the road rather than directly on it. Buildings, shrubs and trees, walls, fences, street lights, above-ground hydrants, bollards, and other permanent poles are a few examples of such static items. This layer also includes adjacent structures like tunnels and bridges, as well as bus shelters with benches. The same is true with so-called vehicle restraint systems, which lessen the severity of crashes by keeping cars on the road. Impact barriers, concrete step barriers, and guardrails are a few examples of safety structures.

2.3.4 State of art in parameter relations and driving scene description

An organized description of the environment is required to establish the scene descriptions and function in a complicated real-world design domain. The 6 Layer Model (6LM), in **Figure 2.3**, was developed for the PEGASUS research project to describe roadway situations [22]. This includes the environment and transportation in cities. According to PEGASUS, the 6LM offers the ability to classify the surroundings and, as such, serves as an organized foundation for the description of a scenario that follows. The model makes it possible to describe and classify the overall environment in an organized manner without assuming any roles for actors or adding any information. The road network, traffic guiding objects, roadside buildings, temporary alterations to the former, dynamic objects, environmental circumstances, and digital information are all described in the 6LM.

A spatially-based description is carried out by layers 1, 2, and 3. There are no elements of them that vary with time. From Layer 4 onward, time-based descriptors are included. Temporary modifications to Layers 1 and 2 can be found in Layer 3. During the whole scenario, these modifications remain unchanged. In the meaning of Layers 1 and 2, they are not fixed. Changes in state are implemented starting at Layer 3. Moreover, state modifications might rely on time starting at Layer 4, Layer 4 is where an entity belongs if its attributes are time-dependent (which could be flexible during a scenario) which implies that not all of its features must depend on time. An entity's attributes aren't always found in the same layer. On the other hand, an entity's properties shouldn't be found on distinct tiers. When in dispute about where to put a property, it goes in the layer where it most closely fits the layer's description and has the most effect. All layers' characteristics have the ability to affect those of other layers. There isn't just one influence direction.



Layer 3 - Temporary Modifications of L1 and L2
Roadwork signs
Temporary markings
Covered markings
Fallen trees laying on the street

Layer 2 - Roadside Structures
Buildings
Vegetation
Guardrails
Street lamps
Advertising boards and pillars

Layer 1 - Road Network and Traffic Guidance Objects
Roads including shoulders, sidewalks, parking spaces etc.
Road markings
Traffic signs and traffic lights

Layer 6 - Digital Information
State of traffic lights and switchable traffic signs
V2X messages
Cellular network coverage

Layer 5 - Environmental Conditions
(Artificial) Illumination
Precipitation
Road weather (dry, wet, icy etc.)
Wind

Layer 4 - Dynamic Objects
Vehicles (moving and non-moving)
Pedestrians (moving and non-moving)
Trailers
Animals
Trees falling over
(at the current point in time)
Miscellaneous objects such as balls, coke cans etc.

(a) Layer 1 to 3

(b) Layer 4 to 6

Figure 2.3: Layers of 6LM for Structured Description [22]

In order to create a multi-perspective mental landscape of the situation, the authors of the study [23] combine cognitive science methodologies with semantic technologies to create a unique model for detecting genuine dynamic scenarios happening in Unmanned Aerial Vehicle (UAV) recordings. The spatio-temporal context of the scene is provided by the semantic, ontology-based description of the scenario, which also facilitates the identification of the objects and their primary interactions. A multi-perspective Fuzzy Cognitive Map (FCM), which is constructed by combining many FCM on individual scenario objects, their interactions, and general scenario features, may be automatically generated thanks to the spatio-temporal context. Evaluations of potential scenario evolution are provided by the FCM. Semantic technologies play a critical role in coding raw data from video analysis into a high-level description of the objects in the scene. The constructed spatio-temporal context initiates the dynamic construction of a FCM, which combines object and basic scene information to generate an overall scenario description as well as an evaluation of the scenario's potential risks.

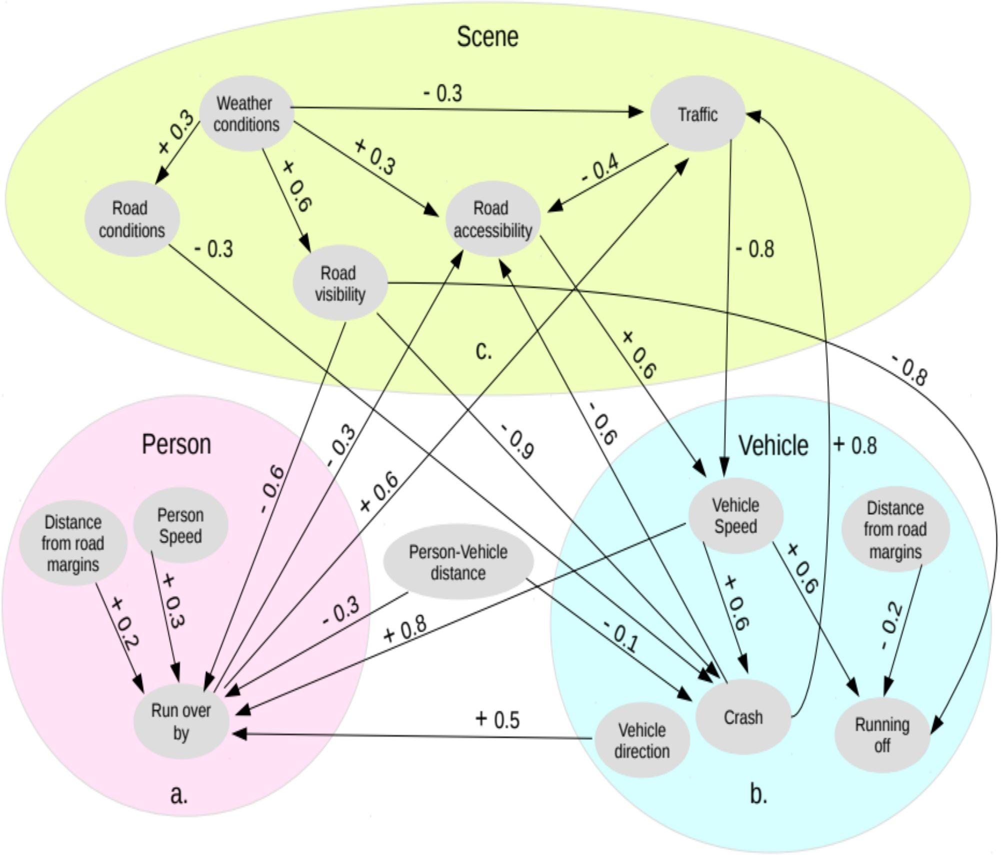


Figure 2.4: FCM about a Road Scene [23]

To provide light on the model, let's examine the FCM on the individuals depicted in **Figure 2.4a**. The Person map explains the key elements (concepts) of a single person's movements in the context of a route. notions include the tracked person's speed, distance from the edge of the road, and other related notions. The causal relationships between these ideas and additional concepts that reflect the potential alerting events for an individual are also depicted in the map. A comparable map that incorporates ideas for the primary alerting events that might affect vehicles is created for cars (see Vehicle map in Figure 2.4b). Based on the spatio-temporal relationships identified by the ontological reasoning, these two maps are combined. It is now necessary to consider the FCM that was created by combining the two separate FCM from a broader angle, taking into account additional scene and environmental details. A generic FCM on the road scenarios is therefore added to the FCM that was previously acquired. In Figure 2.4c broad map depicts notions that stand in for outside variables that may result in different track reactions. These ideas mostly relate to the weather, traffic, and how such factors affect the state of the roads, visibility, and accessibility. These causal relationships are quantified by the weight edges. By adding edges that indicate the causal relationship between the general ideas and track characteristics, or between the general concepts and alerting events directly, the latter FCM is eventually fused with the two FCM on people and vehicles, and vice versa.

An input video frame from the example situation is displayed in **Figure 2.5**. Every recovered item in a frame is identified by a unique ID that identifies the same object across frames as well as a bounding box, which appears as a red rectangle in the output image. The spatiotemporal relations between tracks and Point of Interest (POI), as well as between tracks and track data, are produced by the Ontology-based spatiotemporal context module, which combines track data with contextual information based on POI. The constructed FCM is subjected to reasoning. For instance, there is a greater chance of a vehicle running you over in this situation than in an automobile collision.

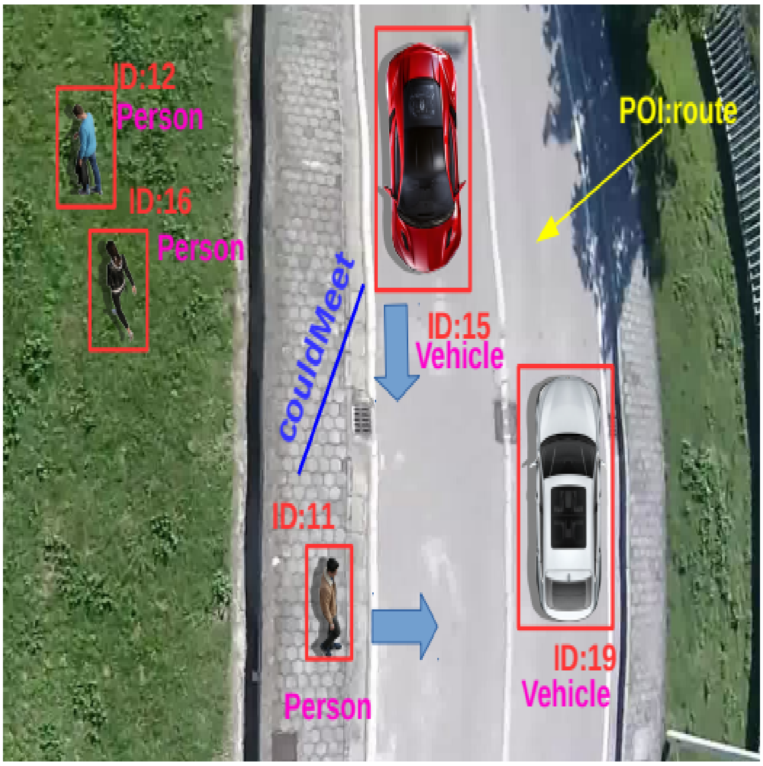


Figure 2.5: The Scene Study [23]

The paper [24] eliminates the need for extra, occasionally costly sensors by using the footage from a monocular video camera as its only input. The main elements of the road scene, such as cars, people, environmental items, and so on, are arranged in an ontology that describes their relationships, interactions, and hierarchy. The ontology tool uses information from video-based attributes associated with the key entities to estimate the level of risk in a particular scene. The ontology assesses the level of danger in the road scene by utilizing features that are taken from the key entities. When determining the level of danger in a particular road scenario, it is essential to identify significant items in the surrounding area. However, because the behavior of these things is equally significant, object recognition alone does not give enough information to assess the scenario in terms of safety. The process of extracting these attributes from frames taken by a monocular camera is illustrated in **Figure 2.6**. An estimate of the separation between a pedestrian bounding box's centers in frames t and t-1 is made. This distance, which is measured in terms of pixels per frame, indicates the pedestrian's movement between two successive frames.

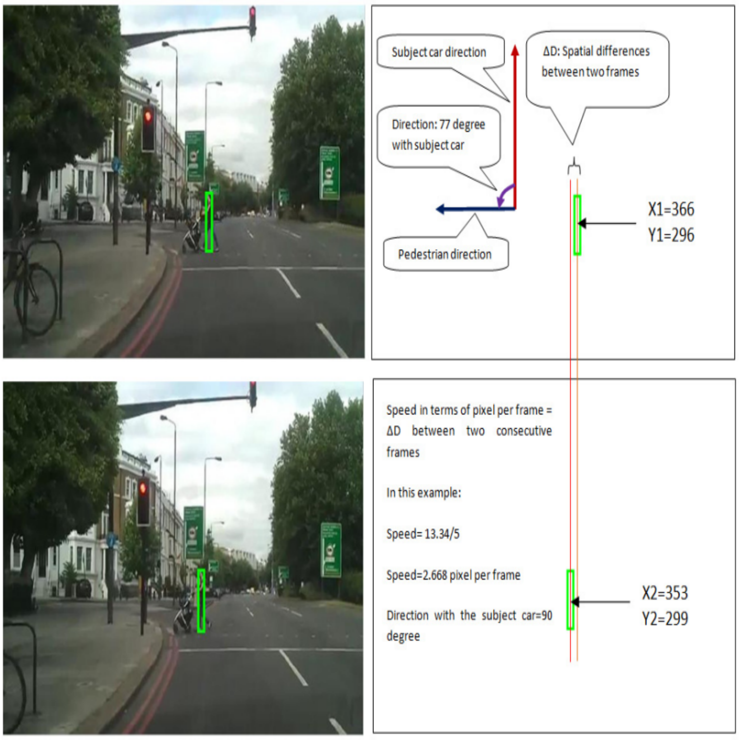

Subject car direction
ΔD: Spatial differences between two frames
Direction: 77 degree with subject car
Pedestrian direction
X1=366
Y1=296
Speed in terms of pixel per frame = ΔD between two consecutive frames
In this example:
Speed= 13.34/5
Speed=2.668 pixel per frame
Direction with the subject car=90 degree
X2=353
Y2=299

Figure 2.6: The Speed and Direction Estimation [24]

The [25] creates the semantic descriptor from three levels: road types, scene types, and challenging conditions, taking into account the intrinsic semantic features of the Autonomous cars test site and the elements impacting the safety of autonomous cars and algorithm performance. Different types of roads include high-speed, rural, and metropolitan locations. varied road types represent varied scene contents, and the main pattern of the scene is reflected in the road kinds. There are differences in the types and amounts of difficult characteristics in different environments. The unique road type is identified using an n-dimensional vector with a value of 0 or 1. This vector describes the many types of roads. Typical driving, intersections, up/down viaducts, through charges, tunnels, turntables, steep slopes, bridges, railroads, and so forth are examples of scene types. The various scene categories are independent of one another and represent the semantic content of the scene. A m-dimensional vector with a value of 0 or 1 describes the scene type and is used to identify the distinct scene type. Bend, overtaking, pedestrian avoidance, construction, huge automobile flow, haze, darkness, road surface traces, lane line blurring, light effect, and so on are all challenging road conditions that influence the environment cognition algorithm in the scene data frame. The quantity and scope of test data within the scene data frame determine the scene's level of complexity. The degree of the distinct demanding situations is represented by an o-dimensional vector with the values 0, 0.2, 0.4, 0.6, 0.8, and 1, which describes the type of problematic conditions.

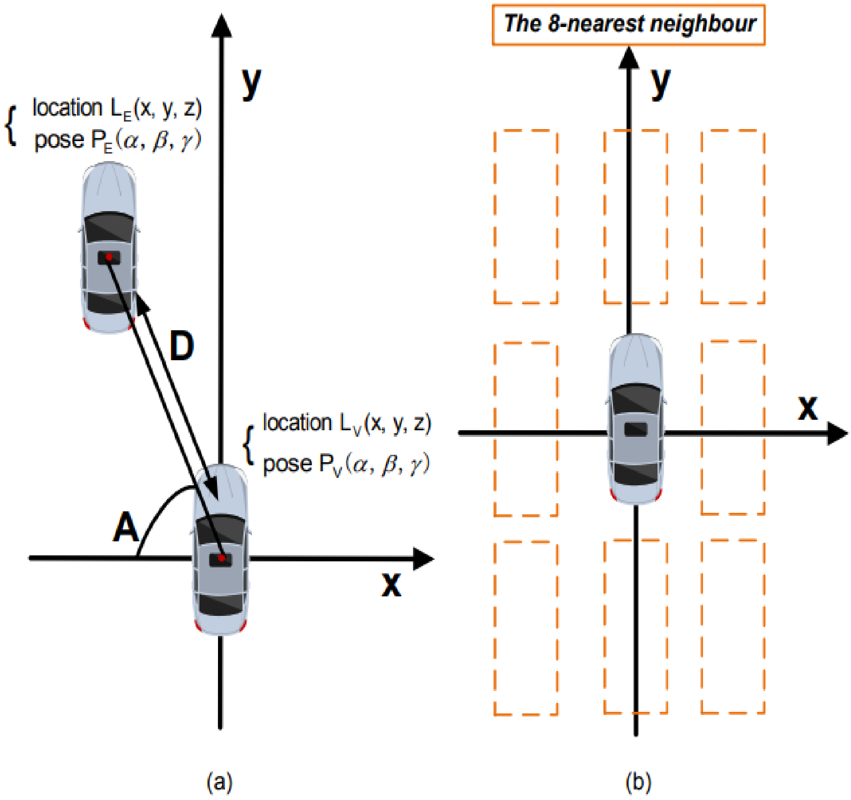


Figure 2.7: A schematic diagram for estimating the complexity of traffic components. [25]

When considering unmanned vehicles from the driving perspective, the distance and angle of other vehicles from unmanned vehicles affect how challenging it is to assess their performance [25]. As a result, this paper creates a traffic element description matrix that provides two dimensions—distance and angle—that characterize the traffic components represented in semantic data. **Figure 2.7(a)** shows computing the length and angle based on the position, orientation, and view of the traffic components. In Figure 2.7(b) it is possible to observe the eight-nearest neighbor of the viewpoint. The information about each automobile is represented by a row in the traffic element's description matrix, assuming that there are N vehicles total on the road except the unmanned vehicle. The first column indicates the distance between the cars, while the second column indicates the cars' point of view. The point cloud that LiDAR gathers determines the length and angle of each traffic element from the viewpoint.

2.4 Datasets for Autonomous Driving

Comprehensive collections of generated or real-world data that capture a range of driving situations and environmental circumstances are known as driving datasets for autonomous driving. These datasets are essential sources for developing, evaluating, and verifying algorithms used in self-driving cars. They usually contain data gathered from a variety of sensors, including cameras, radar, and LiDAR, to give a comprehensive picture of the road ahead. These datasets simulate various circumstances like as urban traffic, highway navigation, pedestrian interactions, bad weather, and more, with the goal of simulating the complexities and difficulties faced when driving. These datasets may be used by academics and developers to improve autonomous vehicles' perception, interpretation, and response to a variety of dynamic scenarios, therefore advancing and improving the safety of autonomous driving technology.

Even though these datasets have proven crucial in the development of autonomous driving technology, there are still issues with making sure the datasets accurately reflect a wide range of situations and complexity. Research is being conducted to create more varied datasets that include important and uncommon occurrences to alleviate the biases and limitations present in the datasets. Furthermore, the creation of simulators such as CARLA and Apollo Scape enhances real-world datasets by offering virtual settings for scaled and controlled algorithm testing. The ongoing development of driving datasets is essential for improving the stability and security of autonomous vehicles while also stimulating new ideas and advancements in the industry.

Regarding the dataset, research participants and dataset producers have distinct concerns [26]. Data providers strive to make their datasets more noteworthy than previous ones in certain ways, even though they frequently gather data for one or more specific tasks when supplying datasets. They focus more on the equipment and data itself, including equipment selection and assembly, data types and quantities, label types and quality, scenario coverage and complexity, etc. Less of them, though, concentrate on whether the data is better suited for these kinds of activities. Users of the datasets focus more on autonomous driving tasks in the comparison, as well as which datasets are best suited for study, which datasets are usable, and what benefits they offer. Data types, label types, and quality, code availability, whether to give a training/testing set, benchmark, and widely used techniques for this task are some of the factors to be taken into account. The primary goal of our work is to use investigation and analysis to close the gap between the disparate concerns of these two groups.

2.4.1 Driving Datasets used in the Research

This research investigates 9 prominent driving datasets that have made substantial contributions to the advancement of autonomous car technology. Each dataset has a distinct set of problems, including a wide range of climatic conditions, traffic situations, and sensor modalities. This dataset analysis gives vital insights into the complexity of autonomous driving and serves as a basis for constructing strong and safe autonomous systems. Also, they provide a link between theory and reality, allowing computer systems to learn and adapt to a wide range of driving conditions. As the need for safer and more efficient self-driving vehicles rises, researchers and developers are turning to datasets that not only represent regular driving circumstances but also push the boundaries of problems, such as dense urban traffic and harsh weather conditions. As autonomous cars improve, these datasets will be key resources for improving perception, decision-making, and navigation systems, hence contributing to the development of safer and more reliable autonomous driving technologies. The datasets used for the research, in this thesis, and some of their key characteristics are as follows:

1. The nuTonomy scenes/nuScenes Dataset:

   The nuScenes dataset is an extensive collection of urban driving situations. This dataset, created by nuTonomy, which is now part of Aptiv, stands out for its comprehensive annotations, which include precise information on item categories, trajectories, and sensor calibration. It includes a wide range of situations, including complicated junctions, pedestrian interactions, and adverse weather conditions. The nuScenes dataset is widely used for training perception and prediction models, establishing it as a gold standard in the field of autonomous driving. This is the first dataset to include the whole autonomous vehicle sensor suite: six cameras, five radars, and one LiDAR, all with a full 360-degree field of view. nuScenes is made up of 1000 scenes, each of which is 20 seconds long and completely annotated with 3D bounding boxes for 23 classes and 8 characteristics. It has 7 times as many annotations and 100 times as many photos as the original KITTI dataset. This dataset proposes new measures for 3D detection and tracking and offers rigorous dataset analysis and baselines for LiDAR and image-based detection and tracking. The data was gathered in Boston (Seaport and South Boston) and Singapore (One North, Holland Village, and Queenstown), two cities noted for their congested traffic and difficult driving conditions. [27]

2. The View of Delft Dataset:

   In the field of autonomous driving research, the Driving Dataset from TU Delft, the Delft University of Technology, provides a unique viewpoint. The comprehensive sensor data, which includes LiDAR and camera pictures, depicts the complexities of navigating through a European urban setting, which includes tiny streets, various road users, and distinct traffic patterns. The TU Delft Driving Dataset offers a great resource for academics looking to improve urban driving algorithms, giving real-world insights into the issues faced by autonomous systems in European city scenes. It includes 8693 frames of 64-layer LiDAR, 3+1D radar, and (stereo) camera data that were synced and calibrated during complicated urban traffic. It includes 123106 3D bounding box annotations of static and moving items, such as 26949 automobile labels, 10800 cyclists, and 26587 pedestrians. The dataset was captured throughout the demonstrator vehicle's journey around the city of Delft's old town, suburbs, and campus in the city of Delft, Netherlands. Preference was given to recordings of situations involving Vulnerable Road User (VRU), such as bicycles and pedestrians. [28]

3. The WAYMO Open Motion Dataset:

   The Waymo Open Dataset is a large set of information that Waymo's driverless cars have collected. This collection contains high-resolution sensor data from cameras and LiDAR that covers a variety of urban and suburban locations. The Waymo dataset is renowned for its extensive and varied scenarios, including details on detailed driving scenarios, complicated road geometry, and complex traffic circumstances. Researchers and developers may use the dataset as a useful tool to train and verify algorithms for perception, object identification, and decision-making in practical settings. The dataset has over 100,000 scenes, each lasting 20 seconds at a frequency of 10 Hz. This results in approximately 570 hours of unique data covering 1750 km of roads. It was gathered by searching six American cities for noteworthy encounters involving cars, pedestrians, and bikes. It generates high-quality 3D bounding boxes for each road agent and provides matching high-definition 3D maps for each scene using a high-accuracy 3D auto-labeling method. [29]

4. The Cityscapes 3D Dataset:

   A dataset called Cityscapes was created especially for the purpose of recognizing urban settings semantically. It consists of high-resolution pictures annotated at the pixel level, grouping individual pixels into groups like sidewalks, roads, and cars. Cityscapes, which was created for the purpose of segmenting urban scenes, is useful for teaching algorithms how to comprehend the composition of urban landscapes. Research on semantic segmentation, a crucial component of autonomous cars' comprehension and navigation of intricate urban environments, makes extensive use of it. A total of 5000 images—2975 for training, 500 for validation, and 1525 for testing—make up the Cityscapes dataset. All eight semantic classes, car, truck, bus, on rails, motorcycle, bicycle, caravan, and trailer, in the vehicle category of the Cityscapes dataset are covered by the 3D bounding box annotations. [30]

5. The KITTI Dataset:

   A benchmark in the domain of autonomous driving, the KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) dataset is well-known for its contribution to the advancement of computer vision and robotics research. The KITTI dataset was gathered in Karlsruhe, Germany, and consists of GPS data obtained from a moving platform, high-resolution pictures, and LiDAR point clouds. It provides a wide range of driving situations with obstacles including shifting weather, pedestrians, and dynamic traffic, including urban, interstate, and country settings. The KITTI dataset is a key resource for the assessment and benchmarking of algorithms in the creation of autonomous driving systems as it is frequently used for tasks like object detection, tracking, and 3D scene interpretation. Using a range of sensor modalities, including high-resolution color and grayscale stereo cameras, a Velodyne 3D laser scanner, and a high-precision GPS/Inertial Measurement Unit (IMU), it includes six hours of traffic situations at 10–100 Hz. [31]

6. The Audi Autonomous Driving Dataset:

   Audi produced a dataset called the Audi Autonomous Driving Dataset (A2D2), which focuses on urban driving situations. It contains sensor data from cameras, radar, and LiDAR, which offers rich information for tasks involving perception and scene interpretation. Because of its precise annotations and superior sensor data, A2D2 is an excellent choice for developing sophisticated algorithms that will help autonomous cars navigate intricate urban settings. The dataset tackles several issues related to driving in metropolitan areas, such as fluctuating traffic patterns, varied road configurations, and complex situations. The sensor suite comprises of six cameras and five LiDAR units and provides a complete 360° coverage. The captured data is synced in time and mutually registered. Non-sequential frames have annotations, 41,277 frames with semantic segmentation picture and point cloud labels, including 12,497 frames with 3D bounding box annotations for objects in the front camera's field of vision. Furthermore, the dataset gives 392,556 sequential frames of unannotated sensor data from three cities in southern Germany. [32]

7. The Argoverse 2 Dataset:

   Argoverse 2 is a collection of three datasets for self-driving perception and forecasting research. The Sensor Dataset is annotated with 1,000 sequences of multimodal data, including high-resolution images from seven ring cameras and two stereo cameras, as well as lidar point clouds and 6-Degree of Freedom (DOF) map-aligned posture. Sequences provide 3D cuboid annotations for 26 item categories, all of which are sufficiently sampled to facilitate 3D perception model training and assessment. There are 20,000 sequences of unlabeled lidar point clouds and map-aligned posture in the Lidar Dataset. This dataset contains the most lidar sensor data ever collected and allows self-supervised learning as well as the developing challenge of point cloud forecasting. Finally, the Motion Forecasting Dataset comprises 250,000 scenarios that have been mined for intriguing and difficult interactions between the autonomous vehicle and other characters in each local setting. Each scenario in all three datasets has its own HD Map with 3D lane and crosswalk geometry generated from data taken in six different cities — Austin, Detroit, Miami, Palo Alto, Pittsburgh, and Washington D.C, as well as varied seasons, ranging from snowy to sunny. [33]

8. The RadarScenes Dataset:

   Particularly, real-world radar point cloud data for automotive applications is the emphasis of the RadarScenes dataset. Radar systems perform better in bad weather and in low light than optical sensors, which is why they are essential for improving the reliability of autonomous cars. RadarScenes provides an extensive set of radar point cloud data that was taken in various driving situations, such as city, highway, and rural areas. This dataset is a useful tool for developing and testing algorithms that use radar data to recognize objects, perceive scenes, and comprehend them. This data collection aims to facilitate the creation of innovative radar perception algorithms based on machine learning, with a particular emphasis on road users who are in motion. A documentary camera was used to take pictures of the scenes that were recorded. [34]

9. The All In One Drive Dataset:

   To innovate strong multi-sensor multi-task perception systems in autonomous driving, it is crucial to provide datasets covering complete sensors, annotations, and out-of-distribution data. Even though a lot of datasets have been made available, they are intended for various use cases, including large-scale training and assessment (Waymo), radar data (nuScenes), and 3D segmentation (SemanticKITTI). The need for a dataset that unites the different qualities of already-existing datasets is very important according to the authors of this dataset. They have the All in One (AIO) Drive dataset, a synthetic large-scale dataset with extensive sensors, annotations, and environmental variations, to help with this difficulty. In particular, this dataset offers eight different sensor modalities

(RGB, Stereo, Depth, LiDAR, SPAD-LiDAR, Radar, IMU, GPS), annotations for every common perception job, and driving scenarios that fall outside of the distribution, such bad weather, illumination, congested areas, fast driving, and collisions with vehicles. [35]

The state of the art methods and all the latest advancements in the domain of scene comparison and driving datasets were explained in this section. The knowledge gained here would be a great foundation for the work done in this thesis program. This chapter also helped the thesis activity in coming up with the possible methods to further continue this research work using new methodologies which will be discussed in the next chapter.

3 Research and analysis on the parameters in driving datasets

The research methodology and the main research contributions are explained in this section. A lot of research was done as part of the thesis including the selection of the driving scene parameters, comparison of datasets, collecting information, and the final similarity analysis to find similar scenes. The thesis work starts with an overview of the activity that was conducted as part of the thesis study and then goes into more detail about each of these steps. Two programming tools were mostly used in this thesis, Python and Matlab, and the whole application part was implemented using these tools. The thesis work will also discuss in detail the tools and the methods used to get the required results in each section.

3.1 Overview

Most of the initial part of the thesis research was done by collecting data from the websites of datasets and also from other research papers in order to do a descriptive analysis on a wide range of related topics. The next stage was the selection of required parameters based on the understanding of the state of the art and logic to develop an algorithm to find the similarity between different driving scenes. The research activity comprises five steps/sections that help to understand the driving scenes and compare the similarity between them using different parameters in a driving scene. The sections 3.2 describe the research on possible driving parameters that can be used for the thesis and 3.3 explains the research on the nine different driving sets based on the above parameters. The selection of the five parameters and two possible datasets for the final analysis is discussed in the next section. The section 3.5 shows the methodology used to find the parameter values in each dataset and section 3.6.1 illustrates the correlation analysis of the parameters and data distribution in two different datasets. The final section of this chapter explains the algorithm and the similarity measures used to find the similarity between two scenes.

Now that the thesis work has discussed the contents of each section, the research will look into the details and concepts behind each section.

3.2 Research on the possible driving parameters

Keeping in mind that the thesis work discussed about the possible parameters for driving scenes in the section 2.3.3 on page 12, a detailed study was conducted on each of the following parameters. This helped to find the suitable parameters for the thesis taking into account the possible limitations of time and resources for the thesis work. The twelve different parameters that were used in this research to get to the final five parameters are discussed below:

1. Weather conditions:

   The weather has a big influence on driving performance and safety. Unfavorable weather can cause problems including decreased visibility, slick roads, and changed driving dynamics. Rain, snow, fog, and storms are examples of bad weather that can impair both human and autonomous driving. Rain, for example, can reduce traction, while snow and fog can reduce vision. This highlights the necessity for strong algorithms that can adjust to a variety of weather conditions [36]. Ensuring the dependability and security of autonomous driving systems requires the capacity to manage these fluctuations in meteorological circumstances. This thesis work could categorize the weather conditions as sunny, rainy, foggy, day and night for the evaluation in this thesis.

2. Speed of the vehicle:

   Driving speed affects both the overall flow of traffic and the probability of accidents, making it an important element in traffic dynamics and safety. Legal speed restrictions, traffic density, and road conditions all affect speed. To maintain efficiency and safety, autonomous vehicles must dynamically adjust their speed to the surrounding environment. According to research, controlling speed appropriately is essential for lowering collision risks and improving general traffic safety [37]. In order to function, autonomous systems need to be able to evaluate and modify their speed in real time based on several criteria such as the kind of road, traffic volume, and pedestrian presence. In this research, driving speed can be studied to get an idea about an urban, highway, or rural scene and how this speed varies across different sets of scenes in various datasets.

3. Lane of the ego vehicle:

   For autonomous driving systems, one of the most important parameters is the lane that the ego vehicle drives. This parameter controls things like tracking, lane detecting, and maintaining the car on the intended course. Lane-keeping is especially important while driving on highways and in metropolitan areas where the vehicle's path is guided by distinct lane markers. In the paper [38], lane annotations in 3D space are provided by OpenLane-V2 to represent their characteristics in the actual environment. In order to facilitate downstream activities, the directed lane center lines and their relationship to one another act as map-like perception outcomes. They also build linkages between center lines and traffic elements in addition to the annotations of the traffic elements. In other words, a lane and a traffic element are considered to be corresponding if and only if the traffic element has control over the lane. Self-driving cars are able to comprehend the current driving conditions and make decisions about where to go and whether to accelerate thanks to these representations. The thesis work could use the concept of single lane, double lane, or multiple lanes with or without markings to classify scenes as rural, urban, or highway in this research.

4. Bounding box dimension:

   In an image or point cloud, bounding box dimensions refer to the measurement of the smallest enclosing rectangle around an object of interest, generally a car or pedestrian. Accurately predicting bounding box dimensions is critical in autonomous driving for comprehending the size and geographical extent of nearby objects. This parameter is required for activities such as object detection and localization, which enable the exact identification and tracking of things in the vehicle's environment. Bounding boxes that are properly specified add to the dependability of perception algorithms, allowing autonomous systems to make educated judgments based on the size and position of nearby objects [39]. In this research, the length, width, and height of the bounding box of different objects are used as a parameter to compare each scene.

5. Yaw Orientation of objects relative to the ego vehicle:

   The yaw orientation of objects, defined as an angular divergence from the ego vehicle's heading direction, is a significant metric in autonomous driving awareness. This parameter indicates the orientation of nearby objects, such as automobiles or people, with the forward direction of the ego vehicle. Accurate yaw orientation assessment is critical for forecasting object behavior and making educated judgments, especially in scenarios requiring turns, lane changes, or complex traffic maneuvers. This information is used by autonomous cars to interpret the relative motion of objects and enable safe navigation, highlighting the need for exact yaw orientation estimates in perception algorithms [40]. This parameter plays an important role in this thesis in finding similar scenes which will be explained in detail in the last section of this chapter.

6. Steering wheel angle and rate of change:

   The steering wheel angle and its rate of change are critical characteristics in understanding the vehicle's direction and the mechanics of its motion. The steering wheel angle shows the angle at which the steering wheel is rotated from its center point, indicating the desired direction of the vehicle. The rate of change of the steering wheel angle, also known as steering wheel angular velocity, indicates how rapidly the driver turns the wheel. These factors are critical for autonomous cars to effectively interpret and forecast the vehicle's course. Monitoring the steering wheel angle and its rate of change allows autonomous systems to model and adapt to the driver's purpose, assuring safe and precise navigation [41]. The steering wheel angle at each instant would be a parameter that can be used in this research as it focuses on the scene rather than the scenario.

7. Speed of the object:

   The speed of nearby objects, such as cars or humans, is an important metric for self-driving systems. Estimating the speed of objects in the vehicle's vicinity accurately is critical for making educated judgments about vehicle dynamics and potential accident hazards. This metric is often obtained from an object's motion over time as measured by sensors such as Radar or LiDAR. Understanding the speed of adjacent objects enables autonomous cars to maintain safe following distances, identify possible dangers, and adjust their speed in response to traffic flow. Precise speed estimation improves the overall efficacy and safety of self-driving algorithms [42]. The relative speed and the direction of the objects with respect to the ego vehicle could be a potential variable to find the similarity across different scenes.

8. Road conditions and visibility:

   Road conditions and visibility are critical environmental elements that affect the safety and performance of self-driving systems. Road conditions include wet or dry roads, ice spots, and uneven terrain, all of which impact the vehicle's traction and control. The clarity of the road environment, as impacted by circumstances such as fog, rain, or low-light conditions, is referred to as visibility. Accurate assessment of road conditions and visibility is required for self-driving cars to alter their driving behavior. Sensor fusion, which combines data from cameras, LiDAR, and Radar, improves vision by detecting and adjusting to environmental difficulties. Addressing these characteristics ensures that autonomous cars may operate in a variety of environments safely and reliably [43]. A possible method in this work is the assignment of numerical values (for example, 0 to 4 with 4 being the highest visible scene) to different scenes based on the visibility and their evaluation to find the relations.

9. Distance of objects from the road margins:

   Object distance from road margins is an important statistic for autonomous cars, helping to safe and precise navigation. The lateral location of surrounding objects in relation to the road borders is determined by this property. An accurate estimate of object distances from road margins is critical for good lane-keeping and ensuring that items are at a safe distance from the road to avoid emergency situations. Perception algorithms examine the lateral location of objects and determine their distance from road edges using sensor data from LiDAR and cameras. This data helps drivers make more educated judgments about lane changes, overtaking tactics, and general trajectory planning, thereby improving the vehicle's capacity to negotiate complicated road scenarios [44]. The distance of each object, pedestrians, cars, and so on, from the nearest road margins can be estimated and would be a great way to study the type of scene for this thesis. This would also help to identify a potential danger for the VRU's and critical scenes.

10. Distance of objects from the ego vehicle:

    One crucial factor in autonomous driving is the object distance from the ego vehicle, which gives vital information for preventing collisions and ensuring general safety. This parameter entails precisely measuring the distance in space between the ego vehicle and other objects, including other cars, people, and obstructions. These distances are crucially measured by sensors like Radar and LiDAR, which enable perception algorithms to build a comprehensive picture of the surroundings. Accurate distance assessment from objects enables the autonomous car to make decisions in real time, allowing it to safely drive through traffic, change its speed, and plan moves depending on the closeness of nearby objects. According to [45], this metric is essential for guaranteeing the overall efficacy and dependability of autonomous driving systems. In this work, the thesis work will use the distance from the ego vehicle reference frame to the center of the bounding box as a parameter. The distance parameter has a significant role in contributing to some interesting results, which will be discussed in the coming sections.

11. Angle of objects with respect to the ego vehicle:

    One of the most important parameters for autonomous driving is the angle formed by objects with the ego vehicle; this angle indicates the surrounding objects' angular location with respect to the orientation of the ego vehicle. When making decisions about changing lanes, passing, or dodging obstacles, this characteristic is essential for comprehending the relative spatial arrangement of objects. The information required to precisely estimate these angles is provided by sensors such as LiDAR and Radar. Using this data, perception algorithms build a complete picture of the surroundings, enabling the autonomous vehicle to judge an object's relative location and make judgments based on how far away it is from the ego vehicle. Accurate understanding of these angles improves autonomous driving systems' overall safety and effectiveness [46]. This parameter is equally important as the distance parameter discussed above since the thesis work will be using both of these parameters to exactly locate the position of the objects in the scene, thus forming the backbone of the algorithm.

12. Headway - distance from the leading car:

    Headway is a term used to describe the separation in time and space between the leading vehicle and the ego vehicle in a traffic flow. Accurately calculating and maintaining a suitable headway is essential for autonomous driving to guarantee efficient and safe traffic flow. For adaptive cruise control and other autonomous driving functions that require following other cars, this value is very crucial. Perception algorithms use sensors like Radar to measure progress continually and calculate the relative speed and distance from the leading vehicle. The autonomous car can react to changes in traffic circumstances, apply the proper amount of braking or acceleration, and maintain a safe following distance when the headway is precisely controlled [47]. The main idea behind using this parameter would be to get an idea about the urban and highway traffic scenes compared to the rural driving scenes where the number of cars in the street will be comparatively less as compared to the the former two cases.

In the above section 3.2, the parameters that were used to do the research for studying the scene description were discussed in detail. Since the usage of all these variables won't be possible in this thesis the work will see how these parameters are available or estimated in each of the datasets and finalize these parameters for the evaluation. So, the next section 3.3 will discuss the datasets and their parameters available in each dataset.

3.3 Research on the driving datasets based on the parameters

A detailed discussion about the different datasets and their characteristics has been already done in section 2.4 on page 20. In this section, the thesis will focus on the estimation or availability of the twelve different parameters in each different dataset. A detailed summary of the twelve parameters on each dataset is shown in Table 3.1 and A.9. The section 3.3 along with the section 3.2 were studied to select the final two datasets and five parameters for the detailed work in this thesis. Now, the thesis will focus on a detailed explanation of how these parameters are available in each dataset.

The nuScenes dataset consists of 1000 scenes with a duration of 20 s each. It has day and night scenes in different weather conditions but there are no clear weather attributes for each scene, so this parameter is not so reliable. The average driving speed is 16 km/h and there is a well-defined CAN bus data from which velocity data of the ego vehicle can be used. The dataset also has detailed map expansion data, which helps to estimate the lane of the ego vehicle and also the distance of the objects from the nearest road margins. The steering wheel

Datasets/Parameters	nuScenes	Delft	Waymo	Cityscapes	KITTI
1. Weather	Yes	No	Yes	No	No
2. Speed of vehicle	Yes	Yes	No	Yes	Yes
3. Lane of ego vehicle	Yes	No	Yes	No	No
4. Bounding Box dimension	Yes	Yes	Yes	Yes	Yes
5. Yaw orientation	Yes	Yes	Yes	Yes	Yes
6. Steering wheel angle	Yes	No	No	No	No
7. Speed of object	Yes	Yes	Yes	No	Yes
8. Road conditions & visibility	Yes	Yes	No	Yes	Yes
9. Distance from road margins	Yes	No	Yes	No	No
10. Distance from ego car	Yes	Yes	Yes	Yes	Yes
11. Angle with respect to ego car	Yes	Yes	Yes	No	Yes
12. Headway	Yes	No	Yes	No	No

Table 3.1: Comparison of parameters in each dataset

angle can also be found in the CAN bus data. Every keyframe has a cuboid represented as x, y, z, width, length, height, and yaw angle, along with characteristics (visibility, activity, and posture) that have annotations for each of the 23 object classes. The heading direction and object's velocity vector are available in this dataset and visibility has 4 labels between 0 and 100 percent. Thus the nuScenes dataset is one of the best datasets with almost all the possible parameter evaluations possible for this research.

View-of-Delft automotive dataset contains 8693 frames acquired in urban traffic and it includes 123106 3D bounding box annotations of both stationary and moving objects, such as labels for 26949 cars, 10800 cyclists, and 26587 pedestrians. This dataset has a lot more annotation for VRU's including bicycles as compared to other datasets. There are no attributes for the weather data, lane of the ego car, steering angle, headway, and road margins. The data has the radial velocity of the ego car but absolute velocity can be estimated from frames. The velocity data is recorded as radial velocity from the 3+1 D Radar data. The yaw orientation of the objects is available in the dataset and the visibility data is available in three occlusion levels. This dataset has a considerably good number of parameters available for comparison especially if the work wants to analyze urban scenes with more pedestrians and bicycles.

The Waymo dataset has 100,000 scenes with a duration of 20s each also providing a high definition 3D maps for each scene. This makes it possible to analyze the lane data, road margins, and headway parameters effectively. Even though there are weather data it does not provide a clear attribute for each scene similar to the nuScenes. The bounding box (3D center point, direction, length, width, and height) and velocity vector of the item are included in each scene for interesting interactions between vehicles, pedestrians, and cyclists. The heading and object's velocity vector make it easy to analyze the speed and the orientation of the objects. The dataset does not provide the CAN bus data, so the velocity of the ego car has to be estimated from frames. The steering angle and visibility data are not available in this dataset. This dataset has a lot of similarities to the nuScenes dataset since it's the data collected in the United States but a few parameters as compared to the nuScenes.

A total of 5000 images make up the Cityscapes dataset, which is divided into 2975 pictures for training, 500 images for validation, and 1525 images for testing [30]. All eight semantic classes in the vehicle category of the Cityscapes dataset—car, truck, bus, on rails, motorcycle, bicycle, caravan, and trailer—are covered by the 3D bounding box annotations. Additionally, the dataset has complete 3D orientation annotations for yaw, pitch, and roll angles, covering all nine (position, extent, and orientation) degrees of freedom of a rigid object, which means data is available as rotation matrices. The dataset has no data about the weather and CAN bus data, hence no steering angle data. There are no HD maps to analyze the lanes, headway, and road margins making it useful mainly for semantic segmentation purposes. The speed of the ego vehicle can be found but the speed of the annotated objects is not available in this dataset. The dataset mentions the availability of the parameter visibility as a percentage of occlusion and truncation. The parameters mentioned are far less compared to other datasets and might not give a good possibility to analyze different scenes. There are no pedestrian annotations in this dataset which makes it difficult to study the VRU's and the critical scenes for this research.

There are five categories in the raw KITTI dataset: Road, City, Residential, Campus, and Person. It comprises 22 scenes with each having a duration of 100 seconds. Thirty distinct GPS/IMU values are saved for each frame of data in a text file: the geographic coordinates, which include satellite data, global orientation, altitude, velocities, accelerations, and angle rates. Two coordinate systems are used to specify accelerations and rotation rates: one is affixed to the vehicle body (x, y, z), while the other is mapped to the tangent plane of the earth's surface at that point (f, l, u) [31]. The dataset contains annotations in the form of 3D bounding box cuboids, encoded in Velodyne sensor coordinates, for each dynamic item inside the field of view of the reference camera. In this dataset, they establish the classifications Car, Van, Truck, Pedestrian, Person (sitting), Cyclist, Tram, and Misc (such as trailers and segways). The height, width, and length of each object along with the translation and rotation is specified in 3D which helps in the analysis of yaw orientation and distance from the ego car. The GPS/IMU unit helps in finding the velocity as well as the exact position of the ego car. The range of driving speed in the KITTI dataset is 0 to 90 km/h with most of the images in the 0 to 30 km/h range. The visibility is specified as four levels of occlusion and truncation but there is no data about the steering angle, headway, and distance from road margins. The dataset can be used for a potential result in this research because of the availability of some good parameters.

The A2D2 dataset contains annotations for non-sequential frames: 12,497 frames include 3D bounding box annotations for items within the front camera's field of view, out of 41,277 frames with semantic segmentation picture and point cloud labels. Additionally, it contains vehicle bus data, which offers more details on the condition of the automobile (such as steering wheel angle, throttle, brake, and transnational/rotational speed and acceleration). The dataset contains various weather data (sunny, rainy, and cloudy) and a combination of urban, highway, and rural riding conditions. Even though it helps to find the distance and angular orientation of the objects, the velocity of the annotations is not available. The headway, road margins, and lane data are also not available. The visibility parameter is also not mentioned properly in the dataset to be considered for evaluation in this research. The A2D2 is a potential dataset with almost all the important parameters that can be utilized for the thesis research. [32]

1,000 multimodal data sequences of 15 seconds duration with 3D cuboid annotations for 26 different item categories are available in Argoverse 2. Every scenario has a unique HD map with a three-dimensional crosswalk and lane geometry. Argoverse and the nuScenes sensor dataset are most comparable in this case since they both contain 1,000 scenarios and HD maps, however, Argoverse is distinct as it also includes ground height maps. A local vector map and 11s (10 Hz) of trajectory data (two-dimensional location, velocity, and orientation) for every track the ego-vehicle detected in the immediate surroundings are included in each scenario. The weather data is available with no clear attributes for an efficient comparison. The range of driving speed is 0 to 54 km/h and quaternions are used to give the rotation coordinates of the objects. The HD maps give a detailed representation richer than nuScenes, providing the 3d lane geometry and the possibility to calculate the road margins with 1 cm resolution. The dataset does not mention the steering angle and the visibility levels but has detailed and clear attributes for all the other parameters making this a very good dataset for the scene comparison analysis. It is interesting to see that the Argoverse 2 has the maximum number of moving vehicle annotations.

A total of 118.9 million Radar points collected over 100.1 km of both urban and rural roads make up the data collection in RadarScenes. The data were gathered throughout a range of circumstances and times, from 13 seconds to 4 minutes, totaling 4.3 hours. There are no 3D bounding box annotations since the Radar point clouds are given and the majority class of data is the cars and pedestrians group. The ego vehicle's motion state including position, orientation, velocity, etc are recorded but it is difficult to analyze the velocity and yaw orientation of objects due to point-wise labeling having less accuracy. There are no clear ideas or labels about the weather conditions, visibility, and steering angle data. Also, there are no HD maps so the lane and road margins cannot be evaluated. The dataset has a rich Radar point cloud with the focus being mostly on the Radar data as compared to other datasets and can be used for research in the domain of Radar-based scene classifications. Hence RadarScenes is not an effective dataset for this thesis topic where the work needs other parameters with equal importance.

AIODrive dataset is a synthetic large-scale dataset with 100 different sequences with each having a duration of 100 seconds. The major highlight of this dataset is that it is not a real world dataset, so this might not be very effective in comparing scenes of real world data. The dataset has a wide variety of weather conditions and a wide range of vehicle speeds ranging from 0 to 120 km/h. Measurements for truncation and occlusion are given, and the dataset uses (x, y, z, Length of the Bounding Box (l), Width of the Bounding Box (w), Height of the Bounding Box (h), θ) to represent 3D boxes, where (x, y, z) is the object center, (l, w, h) is the box size, and θ is the heading direction. This dataset is comparable to the KITTI dataset. Together with the vehicle control signals, such as the brake, steer, and throttle, all agents' motion data, including linear velocity, acceleration, and rotational velocity, are also delivered. The lane, headway, and road margin data are not discussed but still, this dataset has a good number of useful parameters. The reliability of this dataset is not good enough for the thesis since it is using synthetic data which may or may not promise a better result for the comparison analysis with a real world dataset.

After the distribution of different parameters in each dataset is discussed in detail in this chapter, the research discussion will move on to the section 3.4 where the focus will be on the final parameters and datasets for the further analysis of this thesis.

3.4 Selection of final parameters and datasets

In this section, the final five parameters and two datasets which will be used to study the data distribution and to find similar scenes will be discussed. A short explanation of why each of these parameters is chosen and why these particular datasets were selected will also be discussed. Furthermore, the data collection methods and sensor setup of the final datasets will be explained, so that it is possible to understand how the data will be distributed in both the datasets.

3.4.1 Why these datasets and parameters?

Table 3.1 shows the final five parameters, highlighted in green color, and the two datasets that will be used to study the different scenes and find similar scenes between these datasets. In fact, there was an additional parameter, making it a total of six final parameters for the evaluation in this thesis program, speed of the objects which was later not evaluated due to some missing data in the View of Delft dataset, but the thesis will discuss the methodology later in this chapter. Let's begin the discussion with why these datasets were chosen out of the nine total datasets discussed in the section 3.2 on page 28.

The first focus of this thesis is to study the ways to analyze scene descriptions and found that almost 90 percent of the annotations were made up of three main classes - cars, pedestrians, and bikes. So taking these three classes into evaluation will help in getting a good picture of the scene and also help to study the Vulnerable Road Users as well. The first dataset chosen was nuScenes because it was the perfect dataset with all the parameters and also it was practically possible to evaluate almost all the parameters within the desired time frame of this research activity. The second dataset was chosen in a way to evaluate the European urban roads and study the effect of pedestrians and bikes on the European roads. Since the Delft dataset has a very good number of annotation numbers for the bikes, which is considerably less in nuScenes and all the other datasets, it provides us an opportunity to study the role of bikes in each scene. Also, the Delft dataset has most of its data captured in crowded urban scenarios with a lot of pedestrians and bikes, and narrow roads which typically represent the European driving conditions as compared to other datasets like Argoverse, Waymo, etc which even though has more advanced data, represents the scenes in the USA. The Delft dataset proposes one of the toughest scenarios for the implementation of autonomous driving because of the factors discussed above and this work really wanted to study and research more on how could come up with possible solutions in the future to tackle these problems and evaluate similar scenes from this.

Now that this work has finalized two datasets, the focus should be on the possible five parameters to be evaluated in this thesis. Since it is not feasible to evaluate all these parameters in the limited time frame of the research activity, it was decided to shortlist a few important ones. The most important thing to analyze two similar scenes is to get the position of the objects of interest in each scene and then compare them. So in this thesis, the research needs the exact position of cars, pedestrians, and bicycles from the ego vehicle to start with the analysis. So, this analysis fixed the two parameters - Distance of Objects from the Ego Vehicle and Angle of Objects with Respect to the Ego Vehicle for the analysis, this was essential to pinpoint where exactly the objects are located from the ego vehicle. The second thing this work was looking for was to look more into each scene to understand how these annotations are oriented. For example, in a highway scene, the chances are most of the cars will be aligned either in the direction of the ego car or facing the opposite way, but in an intersection, the objects will be at an angle perpendicular to the ego car. Thus this work finalized the third parameter - The Angle of Objects with Respect to the Ego Vehicle, and this along with the distance and the angle forms the backbone of the algorithm which the work will explain in detail later.

Since the driving speed plays an important role in understanding the traffic and the road scenes the analysis chooses the fourth parameter - Speed of the Vehicle, which helps in understanding the distribution of data in both datasets and hence helps us to study the type of scenes in each dataset. The final parameter chosen for the research activity is the Bounding Box Dimension, which was easy to calculate compared to other parameters like distance from the road margins and also feasible to extract from both datasets. It also gives us an idea about the dimensions of different classes in both the datasets and studies if the parameter has some effect on the driving scene comparison since this research work finds the position, angle, and many other parameters using the bounding box. The additional parameter which was included in the analysis initially and was later removed was the Speed of the Objects since it was possible to estimate them in nuScenes but the Delft dataset did not have enough data to correctly provide this parameter value for each annotation in every frame. A small estimation analysis was done using Delft frames and since the data was not accurate for research, it was decided to remove this parameter later. The speed of the object coupled with the yaw orientation or the heading would have been a great combination to analyze the different scenes since it provides a detailed view of the participants in the scene.

So the explanation about the reason for selecting the datasets and respective parameters for the analysis is discussed in this section and now the discussion will move to the data collection and sensor setup part in both the datasets which will be useful in the next section of this chapter to understand how the work is evaluating each parameters and the methodology behind the evaluation.

3.4.2 Sensor setup and data

In this section, the thesis work will discuss the sensor setup on the vehicle for data collection in both nuScenes and Delft datasets. The proper understanding of this section is very essential for the rest of the research work as these sensor coordinates serve as the origin or the benchmark for all the measurements made.

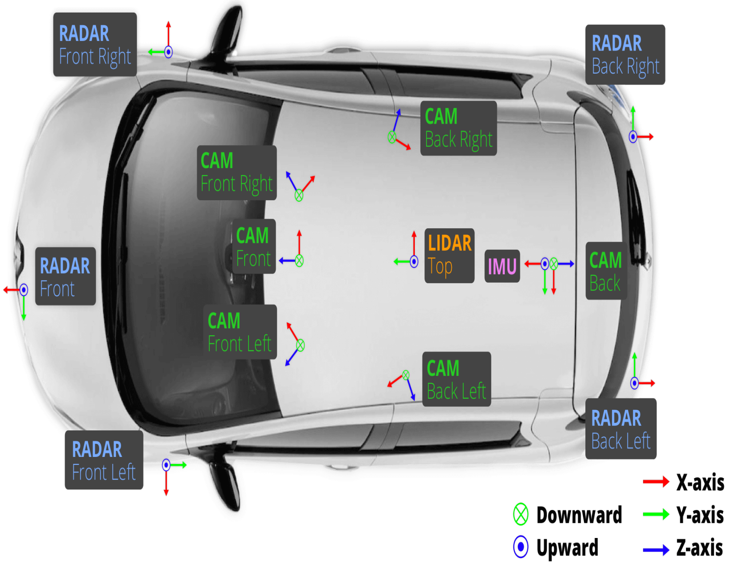


Figure 3.1: Car setup in nuScenes [48]

Two Renault Zoe vehicles with the same sensor configuration are used by the nuScenes for driving in Singapore and Boston. The information was obtained from a study platform and does not correspond to the configuration seen in Motional products. The **Figure 3.1** gives an overview of the sensor coordinate system and position of different sensors mounted in the car. Data from the following sensors are released by them; 1x spinning LiDAR (Velodyne HDL32E) with 20 Hz capture frequency, 360° Horizontal Field of View (FOV) and 80 m-100 m range with usable returns up to 70 metres; 5x long range Radar sensor (Continental ARS 408-21) with 13 Hz capture frequency, 77 GHz frequency signals and independently measures distance and velocity in one cycle using Frequency Modulated Continuous Wave up to 250 m distance; 6x camera (Basler acA1600-60gc) with 12Hz capture frequency and 1x IMU and GPS (Advanced Navigation Spatial) with a position accuracy of 20 mm, heading accuracy of 0.2 degrees with global navigation satellite system and roll and pitch accuracy of 0.1 degrees. [48]

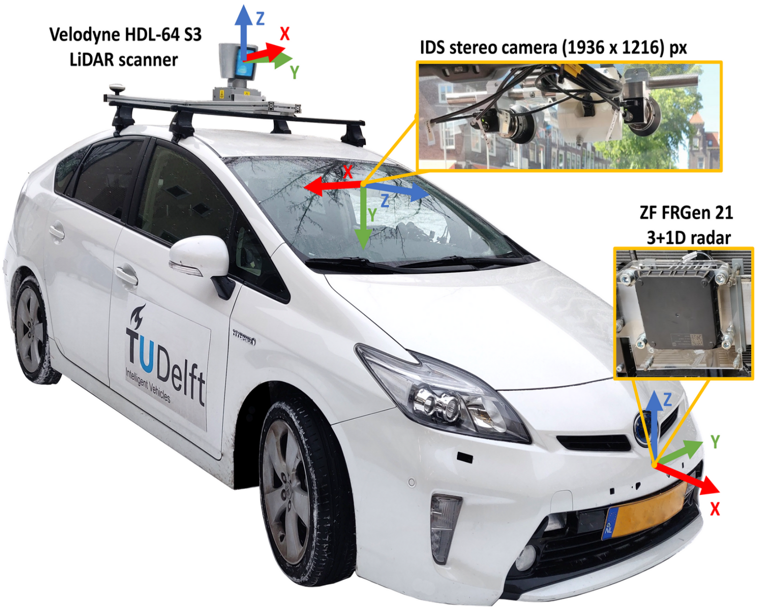


Figure 3.2: Car setup in View of Delft [49]

A Toyota Prius 2013 platform included in the Delft dataset has a stereo camera arrangement, a ZF FRGen 21 3+1D Radar, a spinning 3D LiDAR sensor, and an integrated GPS/IMU inertial navigation system. The **Figure 3.2** demonstrates the recording platform and sensor configuration used for the data collection with this car. The following sensor outputs were recorded; a Velodyne HDL-64 S3 LIDAR scanner on the roof operating at 10 Hz, a stereo camera provides colored, rectified images of 1936 × 1216 pixels at around 30 Hz mounted on the windshield, a ZF FRGen21 3+1D Radar (13 Hz) mounted behind the front bumper, and the ego vehicle's odometry (filtered combination of Real-time kinematic positioning GPS, IMU, and wheel odometry with a frame rate around 30 Hz). Bin files are used to store LiDAR point clouds. An Nx4 array, with N being the number of points and 4 being the number of features (x, y, z, reflectance), is used to represent a 360° scan in each bin file. Bin files are used to store the Radar point clouds. A collection of points in the form of an Nx7 array, where N is the number of points and 7 is the number of features, is contained in each bin file:(vr, vrcompensated, time, x, y, z, RCS) where vrcompensated is the point's absolute (ego motion compensated) radial velocity and vr is its relative radial velocity, time is the point's time id, showing the scan from which it came and x,y,z are the coordinates of the points which are the objects in the field of view of the sensor. [49]

The work has a detailed idea about the sensor setup and data collection platform in each of the datasets. Now the discussion can move to the next section 3.5 where the work will explain the methodology used to find each parameter from both the datasets.

3.5 Methodology to evaluate the parameters

In this section, the thesis will discuss the methods adopted to find the final five parameters from both the nuScenes and the Delft dataset. Each section will also point out the difficulty and the techniques that this work adopted to overcome this difficulty at each point of the research activity. The research will also discuss the field of view, that was chosen for this particular research taking into account the limitations of the sensor data in both datasets.

3.5.1 The reference point and the field of view

The work will be measuring all the parameters with respect to the ego vehicle, so first of all this work needs to define a reference point within the ego vehicle which will be the first point of interest. After that, the discussion is needed to define a reference point on the bounding box because all the measurements from the ego vehicle to each object will be made with respect to the reference in the bounding box. According to the sources [48] and [50] it is found that in nuScenes the ego vehicle body reference frame is in the rear axle, at the centre of the rear axle, and all the measurements to the bounding box are made from the centre of the rear axle of the car to the centre of the bounding box. So the bounding box reference frame is exactly the centre of the cuboid and at the ego vehicles, the reference frame is at the centre of the rear axle. But according to the source, it can be found that in the Delft dataset, all the measurements are made from the camera coordinates which act as the reference frame so there is a need to shift this reference frame to the middle of the rear axle to make it equivalent to the nuScenes reference frame. It is also found in this source [51] that the Delft dataset follows the KITTI coordinate transformation so this work could have a look at this particular reference [52] which explains the KITTI coordinate systems (also see Table 3.2) and how the measurements are made in the KITTI coordinates. Also, it can be found that in the Delft all the measurements of the bounding box are made to the bottom centre of the cuboid of the bounding box but there is a need to transfer these coordinates to the centre of the bounding box to make it equivalent as nuScenes dataset measurement which makes all the measurements from the centre of the cuboid in the bounding box.

Sensor	X-axis direction	Y-axis direction	Z-axis direction
Camera	Right	Down	Forward
LiDAR	Forward	Left	Up
GPS/IMU	Forward	Left	Up

Table 3.2: Sensor coordinate frames in KITTI

Having a look at Figure 3.2, it is possible to understand that the thesis work needs to shift the reference frame of the bounding box in the camera coordinate system by a distance of half the height of the bounding box in the direction of the y-axis to make it similar to the nuScenes, which is in the middle of the bounding box. Now the research will have to transform the coordinates of that Delft dataset from the camera coordinates to the centre of the rear axle, so the thesis can use the source [53] and take the dimensions of the car to get an approximate idea about the distances that need to be shifted to get the origin at the center of the rear axle. There is a need to shift the coordinates in the direction of the Z-Axis and also in the direction of the Y-axis by units of two metres in the positive Z direction and one metre in the negative Y direction to make it compatible with the nuScenes reference frame that is at the centre of the rear axle of the car.

The nuScenes dataset has a full 360 degree field of view annotations which means that all the objects that are within the 360 degrees angle of the car have a bounding box. The Front and side cameras have a 70 degree FOV and are offset by 55 degree. The rear camera has a FOV of 110 degree which can be seen in the **Figure 3.3**. According to [51] any object of interest within 50 metres of the LiDAR sensor and partially or fully within the camera's field of view (horizontal FOV: +32° to -32°, vertical FOV: +22° to –22°) was annotated. So the Delft dataset only has annotations or bounding boxes for the objects which are exactly in front of the car and also at a distance within 50 metres from the car. This means that the thesis work can only use this field of view which is compatible in both the data sets for further analysis in this research work so in the later sections the experiment will be discussing for the comparison only this field of view. In the coming sections the thesis will also study how the data is distributed in the nuScenes for the full field of view which is 360 degrees around the car but all the comparisons will be made only within the field of view which is compatible with the Delft FOV.

In the next section, the discussion will be about how each parameter is calculated from both datasets in detail.

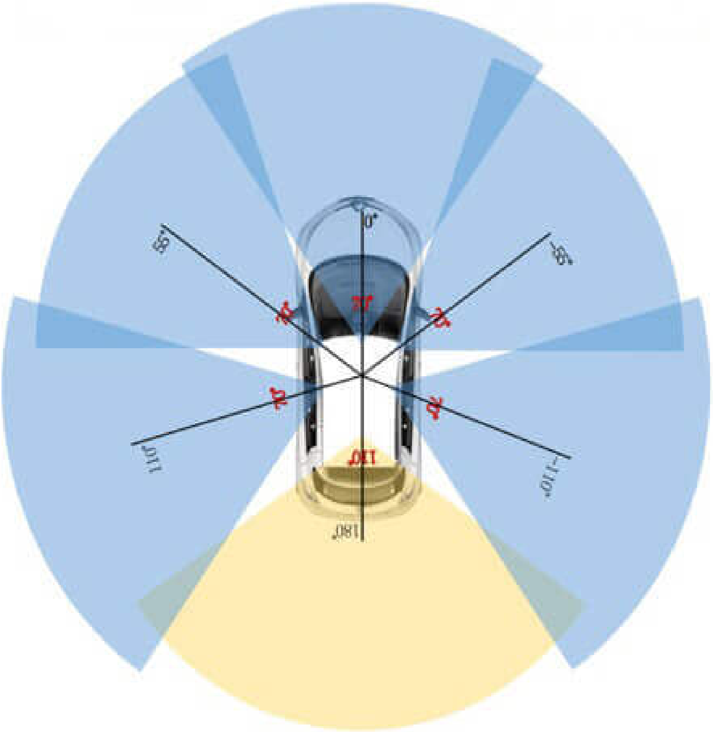


Figure 3.3: Camera sensor FOV in nuScenes [48]

3.5.2 Evaluation of the parameters

The thesis will start the discussion by finding the velocity of the ego vehicle from the nuScenes dataset. With the help of the nuScenes CAN bus tutorial [54] the thesis work found the velocity of each scene using the sample number of the corresponding scene. In the CAN bus data, the experiment extracted the parameter velocity from the vehicle monitor data for each corresponding scene in metres per second. Since there are some scenes in the nuScenes dataset having no CAN bus information, the thesis work did not consider those scenes in this evaluation and used all other scenes which are not blacklisted by the nuScenes for further analysis in this thesis. In the case of the Delft dataset, since they do not publish the ego vehicle velocity or do not have CAN bus data, this research used the smoothed velocity estimation method as suggested by the source [55]. So from the dataset, the research work took the x,y,z coordinates at each frame and estimated the corresponding coordinates of the objects in the preceding frame and calculated the distance of the coordinates from the ego vehicle, and then divided with the time between frames which is 0.1 seconds in this case to get an estimated velocity of the vehicle. The equation used is as follows [56]:

$$v = \sqrt{(x^2 + y^2 + z^2)}/t \tag{3.1}$$

In the above equation 3.1, v is the velocity of the car in m/s, x, y, and z are the position coordinates of the object and t is the time between two frames which is 0.1 seconds in this case.

For finding the size of the bounding box from the nuScenes dataset the thesis work used the annotation data and from the sample annotation, this work used the parameter size which specified the length, width, and height of the bounding box. The same methodology was used to get the bounding box dimensions in the Delft dataset where they specify the length, width, and height of the bounding box of the objects in the parameter raw labels for each frame. The measurements are recorded in metres for both datasets.

The third parameter, the yaw orientation of the objects with respect to the ego vehicle, was evaluated in this work using the rotation coordinates of the sample annotations from the nuScenes dataset. The rotation coordinates give the quaternion and this quaternion was used with the function, box.orientation, to estimate the yaw parameter in nuScenes [57]. For the Delft dataset, the yaw orientation of the annotations is specified in the labels of the frames so the thesis work used that data to estimate the orientation of each object with respect to the ego vehicle. The angular measurements were taken in degrees to make it suitable for the analysis in this research.

The angle made by the objects with respect to the ego car was found in the nuScenes by estimating the yaw angle of the objects with respect to the ego vehicle reference frame first, then since the ego vehicles already have a yaw angle with respect to the global reference frame in the nuScenes this work had to compensate for this. So the research analysis subtracted the yaw of the ego vehicle with respect to the global reference frame from the yaw angle made by the objects with respect to the ego vehicle coordinates to get the exact angular position of the objects. In the Delft, the thesis work uses the same principle but the ego vehicle yaw angle and object yaw angle are available for each frame. This work has to compensate for the ego vehicle rotation from the observation angle of the objects in Velodyne coordinates to get the exact angular position of the objects in each frame. **Figure 3.4** gives the representation of the object coordinates to better understand this concept.

The distance of the objects from the ego vehicle in the nuScenes is found by using the translation data from the sample annotation. The ego vehicle translation gives the x, y, and z coordinates of the ego vehicle from the global reference frame in nuScenes which is the origin of every map. The object translation is also specified in the x, y, and z coordinates of the object with respect to the global reference frame so the research work gets the coordinates of both the ego vehicle and the object and then uses the following formula to estimate the distance between the two points [56]:

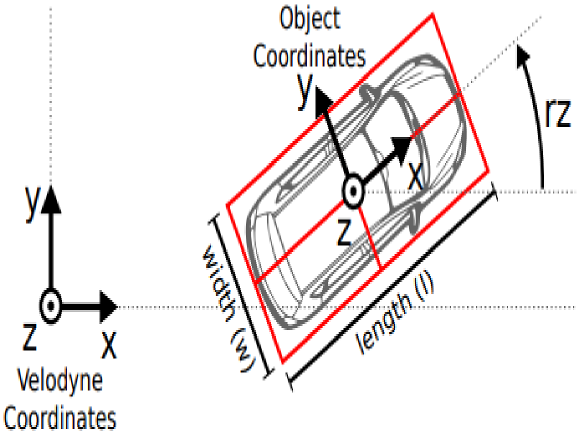


Figure 3.4: Object coordinates in Delft [31]

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{3.2}$$

In the above equation 3.2, d d is the distance of the object from the ego vehicle, x_2, y_2, and z_2 are the ego vehicle coordinates and x_1, y_1 and z_1 are the object coordinates with respect to the global reference frame.

In the case of the Delft dataset since the thesis work already has the x, y, and z coordinates of the objects from the vehicle camera coordinates the thesis work can directly estimate the distance using the above formula but the only difference is that here there are no x_2, y_2 and z_2 coordinates.

So the thesis work has explained in detail about the parameter estimation in this section and could use these concepts to proceed with the further analysis of this research. The next section 3.6.1 will focus on the correlation analysis of these parameters and the data distribution of these parameters, this work will use some plots to understand data distribution in both the datasets in detail.

3.6 Correlation analysis and data distribution

In this section, the research activity will be checking the dependency of one parameter with respect to the other so that the work could analyze the possibility of eliminating any parameter in future analysis. After finishing the correlation analysis the thesis work will see in detail how the parameter data is distributed in both the datasets using some plots. In the following section 3.6.1 the discussion is started by looking at the correlation part.

3.6.1 Correlation study on the parameters

A statistical method for determining the direction and degree of a linear relationship between two quantitative variables is correlation analysis. In statistics and research, it is an essential technique for figuring out how changes in one variable could affect changes in another. Measuring how much two variables move with respect to one another is the primary goal of correlation analysis. According to [58], correlation is a statistical measure that quantifies the extent to which two variables are related. It does not imply causation; correlation only indicates a relationship, not that one variable causes the other to change. There are three main types of correlation; positive correlation when two variables rise in tandem with an increase in one; negative correlation occurs when one variable rises while the other falls and zero correlation when variables don't seem to have a linear relationship with one another.

The strength and direction of the link are expressed numerically by the correlation coefficient which falls between -1 and 1. There is no connection when the correlation coefficient is 0, perfect negative correlation is represented by a value of -1, and perfect positive correlation is denoted by a value of 1. The greater the connection, the closer the correlation coefficient is to 1 or -1. There may be little to no linear association indicated by a coefficient close to 0. In this study the thesis work uses three measurement parameters to evaluate the correlation coefficient which are Covariance, Pearson's Correlation Coefficient, and Spearman's Correlation Coefficient. Covariance is a measure of how much two variables change together. It indicates whether an increase in one variable corresponds to an increase or decrease in another [59]. Pearson's correlation coefficient, often denoted as r, measures the strength and direction of a linear relationship between two continuous variables. It ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no linear correlation [58]. Spearman's correlation coefficient is a non-parametric measure of the strength and direction of the monotonic relationship between two variables. It assesses whether there is a consistent increase or decrease in one variable corresponding to the increase or decrease in the other [60]. **Figure 3.5** shows the linear and monotonic relations of general data.

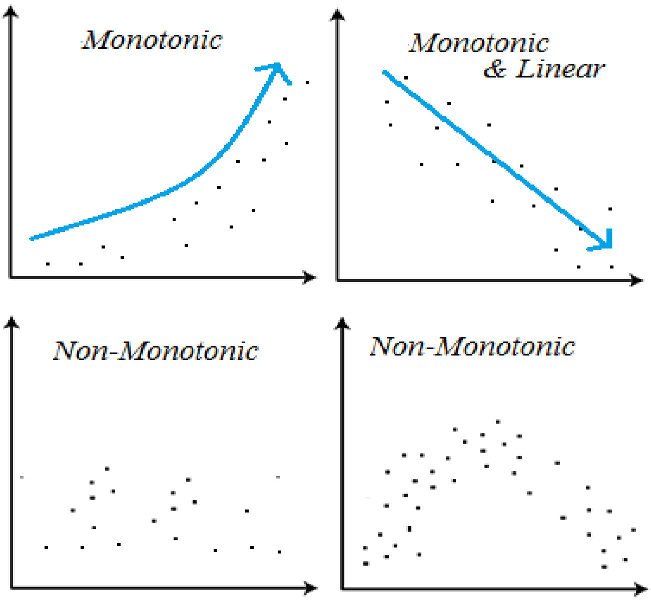


Figure 3.5: Monotonic and Linear Correlation [61]

A common tool for visualizing the connection between two variables is the scatter plot. Individual data points are represented as points on the scatter plot, and the arrangement of these points might provide information about the correlation. In this research, the thesis work uses the scatter plot to plot the relation of the variables between datasets of Delft and nuScenes but this work split the nuScenes dataset into two parts, the first one which is aligned with the field of view of the Delft dataset and the second one with the complete full dataset of the nuScenes having the 360 degree field of view.

Figure A.1 illustrates the scatter plot and the relationship between the parameters, distance from the ego vehicle, and the height of the bounding box, for the class car for 3 categories of datasets. A summary of the statistical values is given in Table A.1. These values show that there is neither a linear relationship nor a monotonic relationship between the two parameters in all the datasets because the correlation coefficient and covariance values are almost near zero. The further analysis results and figures are shown in Chapter A.

Figure 3.7 shows the scatter plot and the relationship between the parameters, distance from the ego vehicle, and the angular position, for the class car for three categories of datasets. A summary of the statistical analysis is done in this figure and the values are given in Table 3.3. There is no evidence of any kind of linear and monotonic correlation between the parameters in all three cases of the datasets. However, it is interesting to observe that the Delft dataset has some kind of relationship between the parameters which needs to be studied further in detail. There is also a similar pattern in the nuScenes dataset, but it is neither a linear relationship nor a monotonic one. **Figure 3.6** represents this sign convention to easily understand the concept.

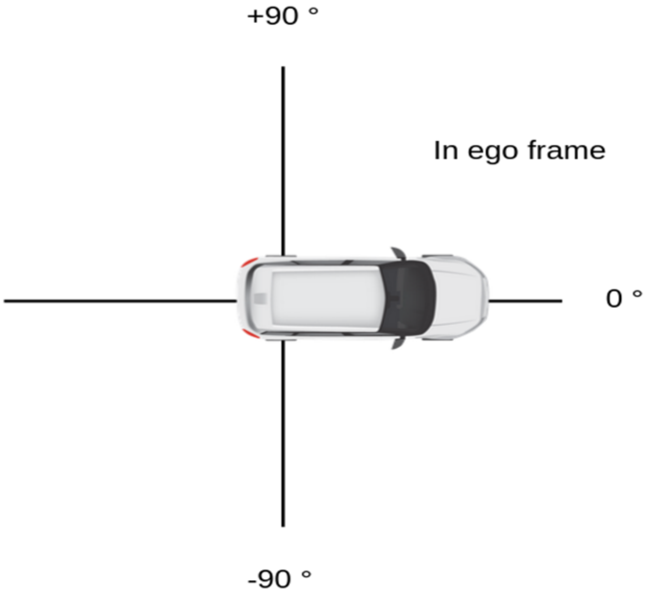
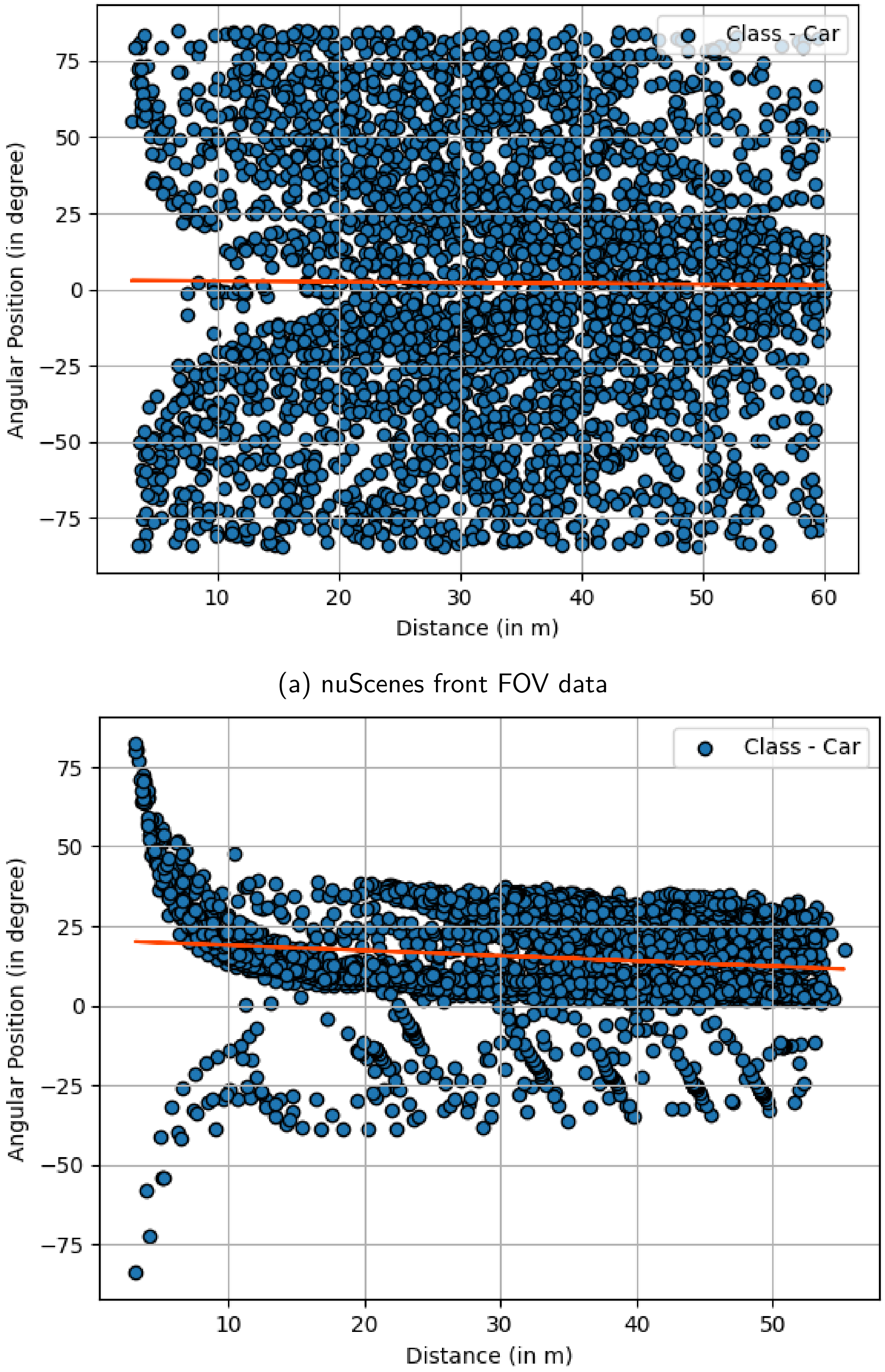


Figure 3.6: Sign convention for angular measurements [48]

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	-28.8	150.4	-28.8
Pearson's Correlation	0.06	-0.009	-0.138
Spearman's Correlation	0.06	-0.01	-0.14

Table 3.3: Statistical values from Figure 3.8

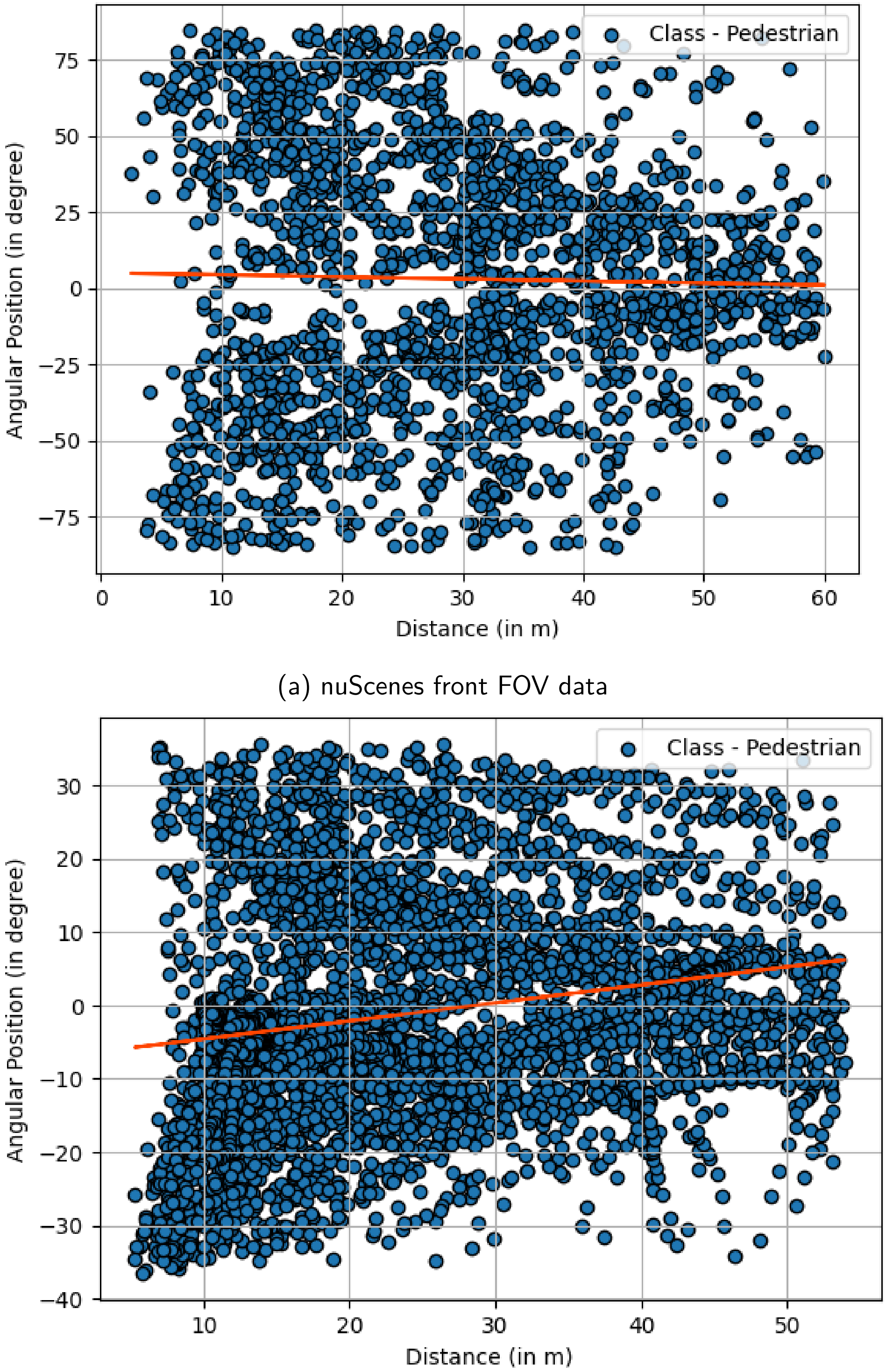
This section evaluated the correlation between the parameters for different classes of objects and the thesis work found no evidence of any linear or monotonic relationships between the datasets of Delft and nuScenes to eliminate any parameters. It was quite interesting to see that even though this research used distance to estimate the angular position in the nuScenes dataset, there are no direct relations in the correlation analysis. This means that the thesis work has a total random distribution of data in the datasets and this research could use all the parameters as needed for further evaluation. It is also noticed that the bounding box parameters do not have much effect on the analysis of this thesis since it always has a defined range of values for each class in both datasets. So further evaluations will not consider the bounding box dimensions. It is evident that all the other parameters will have a significant effect on finding similar scenes. So the research work will continue the discussion with all the other parameters. In the next section, the data distribution trends in both datasets will be discussed.



(a) nuScenes front FOV data

(b) Delft data

Figure 3.7: Scatter plot of Distance from the Ego Car and Angular Position for Car



(a) nuScenes front FOV data

(b) Delft data

Figure 3.8: Scatter plot of Distance from the Ego Car and Angular Position for Pedestrian

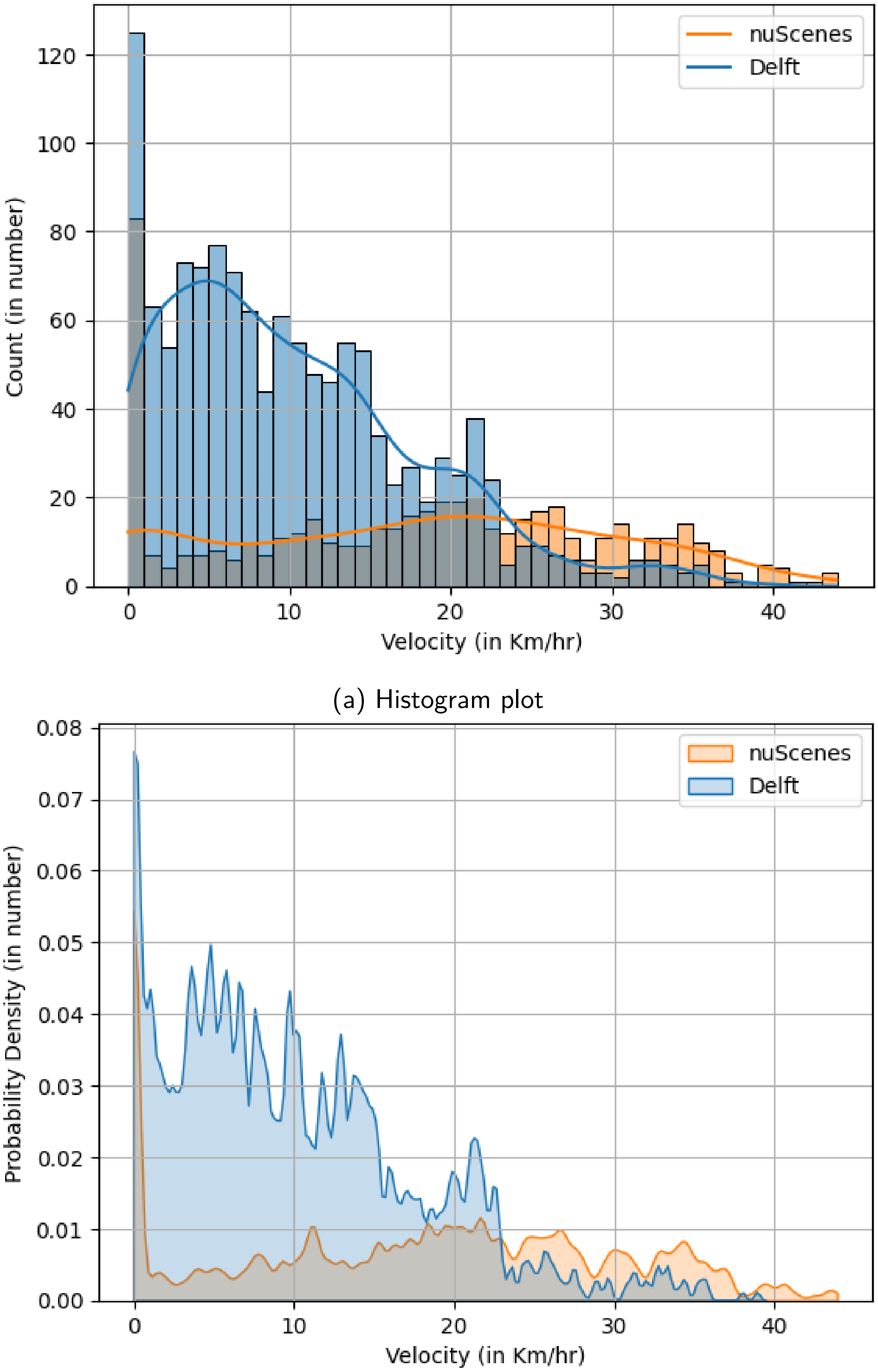
3.6.2 Data distribution in datasets

In this section, the thesis work will study the plot of each parameter versus its probability distribution data to get a detailed idea of how the data is distributed among all the parameters in both datasets. The research will also do the consideration for the nuScenes front and nuScenes full field of view datasets along with the Delft dataset. The area of attention will be on how the nuScenes front field of view dataset is going to be compared against the Delft dataset since this work will be using these two datasets for the algorithm to find similar scenes in the next section.

For the comparison, the nuScenes front FOV dataset will have the distance and the angular measurements in the same range which is the distance from the ego car less than 60 metres. However, the nuScenes front FOV and nuScenes full dataset will be compared with all the annotations within the distance and angle range of the nuScenes dataset. The angular measurements follow the same sign convention as discussed in the section 3.6.1. The probability density values are expressed in numbers and these numbers are the percentage value calculated with respect to the amount of particular data in the total dataset.

Figure 3.9(a) gives the histogram and Figure 3.9(b) gives the probability distribution of the velocity of the ego car for the nuScenes front FOV and Delft dataset. It is very interesting to see that the average driving velocity of the Delft dataset is less than 10 kilometres per hour which is the ideal speed for real urban driving conditions, so this supports the point that the Delft dataset has a lot of urban scenarios within the crowded city centres. Whereas the nuScenes dataset has a uniformly distributed velocity from zero to 40 kilometres per hour and the average velocity in this nuScenes dataset is somewhere around 15 to 20 km per hour which again proves that it has a balanced number of scenes from urban to rural and highway driving conditions. It is mentioned in [62] that the average driving speed in the full dataset is 16 m/hr making it almost the same as the nuScenes dataset with front FOV.

The probability distribution of the parameter distance for both datasets of the class car is shown in Figure A.5. When the nuScenes front field of view is compared with the Delft dataset it is observed that both of them follow almost a similar trend except for the fact that the data distribution curve in the nuScenes front dataset is slightly shifted towards the right which means it has an increased number of annotations for the class car at a distance of 30 to 55 metres. This could mean that the nuScenes dataset has more highway scenes or scenes having long stretches of roads as compared to the Delft dataset. Since the Delft dataset is a pure urban dataset it usually consists of traffic scenes and scenes in urban areas, which explains the balanced distribution of data. However, it is really interesting to see that the nuScenes front field of view and the nuScenes full field of view dataset follows the exactly same trend with an average distance of cars at around 25 to 30 metres.



(a) Histogram plot

(b) Probability density plot

Figure 3.9: Velocity of the Ego Car

In Figure A.5, the research observed that the Delft dataset has an increased distribution of pedestrians at a distance of 10 to 20 metres from the car. This supports the statement that the Delft dataset is a pure urban dataset and it means that it has a lot of scenes with the crowded pedestrians somewhere around the vehicle. While comparing the nuScenes distance curve, the result was a balanced distribution of the pedestrians at a distance from zero to 50 metres around the car which means that nuScenes has more or less a balanced distribution of all types of scenes such as urban to the highway. The Delft dataset is an excellent dataset for analyzing vulnerable road users because it has an increased number of annotations for the class pedestrians and bicycles. However, the Delft dataset follows a balanced distribution of the bike data with a slight shift towards the left, meaning there are more bicycles around five to twenty metres around the car. In the nuScenes dataset, no clear conclusion could be drawn from this data as it does not have a significant number of bicycle annotations.

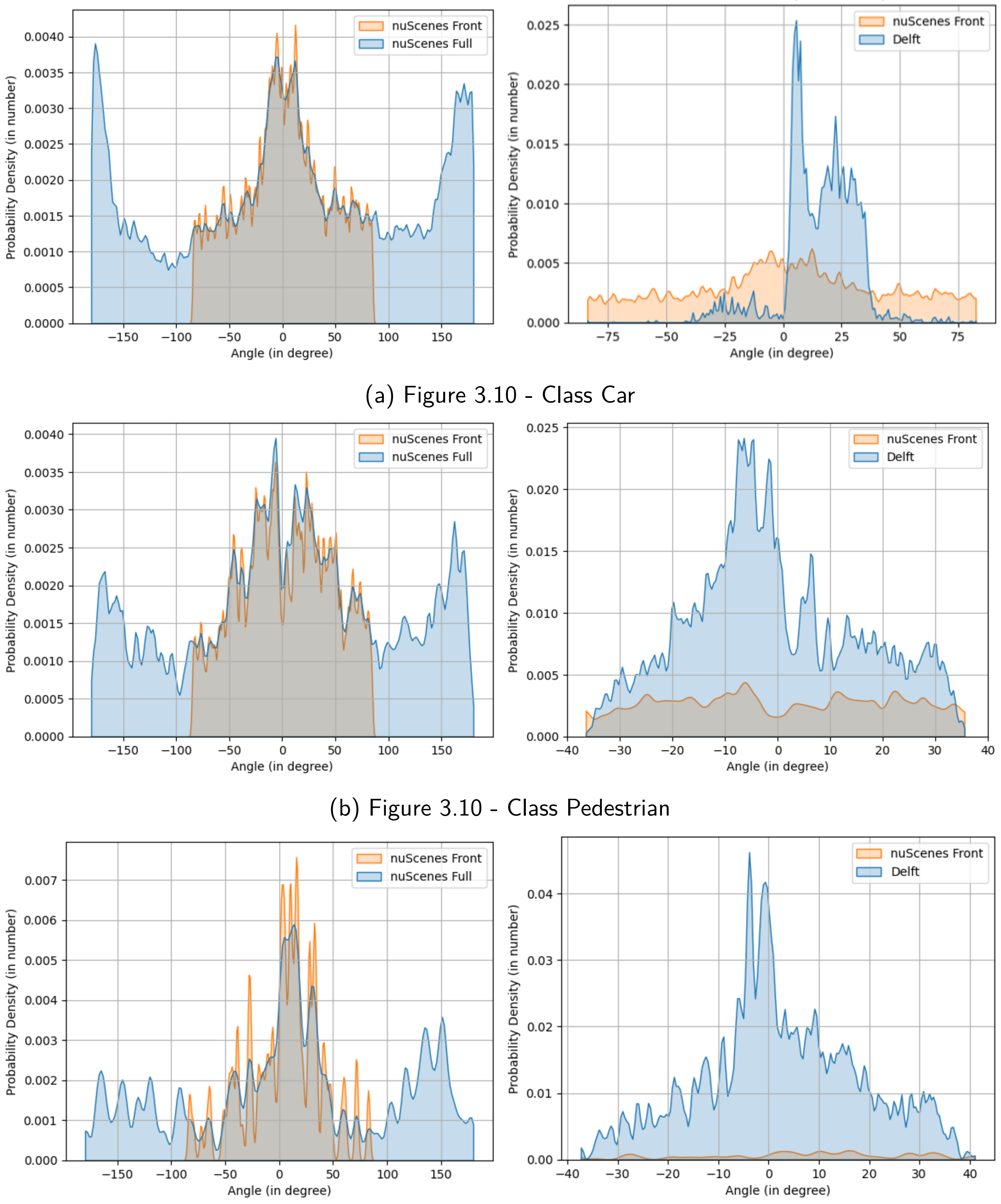
While comparing the nuScenes front field of view with the Delft dataset, in Figure A.6, it could be seen that the Delft dataset has comparatively more scenes with vehicles that are oriented either in the direction or in the direction opposite to the ego car, so more straight road scenes and very less intersections. However, the nuScenes front dataset has balanced data with also cars oriented at +90 and -90 degrees to ego car direction, which means more intersection scenes. These data also help to get to a conclusion that the nuScenes dataset has a significant number of intersection and road junction scenes where vehicles are oriented perpendicular to the ego car direction. It also has some considerable scenes with pedestrians crossing the roads which could explain the angles around +90 and -90. However, the nuScenes front dataset has balanced data with also pedestrians oriented in all different directions. The Delft dataset has more bikes oriented around the same direction as the ego car direction. The number of bikes oriented perpendicular to the ego car direction, which could mean crossing or intersections, is comparatively less in number. In the nuScenes front field of view and the nuScenes full field of view dataset, they follow the same trend with most of the scenes with bikes oriented either along the direction of the ego car or perpendicular to the ego car.

Figure 3.10 shows the probability distribution of the parameter angular position for both datasets for the category of class cars. The research was able to witness the fact that the Delft dataset has most of the cars located on the left side at an angle between zero and 35 degrees to the ego vehicle. This can be explained from the Delft dataset as it mostly contains single-lane roads in an urban area and most of the cars that will be traveling in opposite directions are on to the left of the ego vehicle. The only cars which will be on the right side will be the parked ones but still, it would be less as the roads in Delft are narrow and urban parking scenes are less. However, the nuScenes front follows a balanced distribution of data with an almost equal number of cars on both sides. Even though the nuScenes front field of view and the full field of view dataset follow the same distribution there is an increased number of cars in the front and back of the ego vehicle, in the full dataset which is quite normal in ordinary driving conditions. The nuScenes dataset also contains both the left and right-hand driving scenes, with multiple lanes which could be another reason for the balanced distribution. The Delft dataset is limited to just left-hand driving scenes and mostly single lanes which explains the shift of data towards a positive angle which is on the left side of the car from Figure 3.6.

While having a look at Figure 3.10, the curve for the angular position for the class pedestrians, the Delft dataset has a peak at around zero degrees and -10 degrees. This means that there is an increased number of pedestrians directly in front of the car and on the sidewalk which could be either in the direction in which the car is traveling or in the direction opposite to the car. This result explains the fact that the Delft dataset has a lot of crowded city scenes with pedestrians trying to cross the roads and also a lot of pedestrians in front of the car on the sidewalks. However, the nuScenes front dataset follows a balanced distribution of data without any peaks.

In Figure 3.10, the angular position curve for class bikes, the Delft dataset has a peak at around zero degrees and minus five degrees this means that there is an increased number of bike scenes directly in front of the car. These numbers explain the fact that the Delft dataset has a lot of urban scenes with bikes in it but the nuScenes front dataset does not have a significant number of bike scenes for a meaningful comparison with the Delft dataset. These data also help us to get to a conclusion that the nuScenes dataset is not good enough to study the VRU's taking into consideration the number of scenes involving the bikes.

In this section, the research activity focused on the data distribution of datasets for each class of cars, pedestrians, and bikes. This helped to understand what types of scenes can be expected more in each of the datasets and reach a conclusion on how to proceed with the algorithm part in Chapter 4. In the following chapter, the development of the algorithm to find similar scenes from the nuScenes and Delft datasets is explained. Chapter 4 will focus on the similarity analysis and the final results of the thesis work.



(a) Figure 3.10 - Class Car

(b) Figure 3.10 - Class Pedestrian

(c) Figure 3.10 - Class Bike

Figure 3.10: Probability Distribution of Angular Position

4 Similarity analysis and results

This Chapter will focus on the scene comparison methods and development of techniques to find the similarity between two scenes in the nuScenes and Delft datasets. A lot of discussions and studies are going on in the domain of machine learning to compare similar scenes for autonomous driving. This research will focus on much simpler and more logical ways to identify the effect of important parameters on scene comparison and methods to compare scenes more effectively. The discussion in this chapter will consist of mainly three sections, the first section will give an introduction to different similarity measures that will be used in this research, the next section will emphasize the methodology adopted in this research to make the dataset compatible with the similarity analysis and the final section will discuss the different algorithms or methods adopted in this thesis and its final results. The methodology and the algorithms used in this chapter are new ideas that are being experimented with in this domain, and also not borrowed from other resources, so the research will be providing the very first results of this idea that would need further improvements and detailed study in the future.

4.1 Similarity measures

Similarity measures are essential in many domains, such as information retrieval, machine learning, and data analysis, since they offer a numerical evaluation of the degree of similarity or difference between two sets of data. A similarity coefficient indicates the strength of the relationship between two data points. The more the two data points resemble one another, the larger the similarity coefficient [63]. When examining patterns, structures, or distributions within datasets, these metrics are crucial tools. The idea is to represent similarity or dissimilarity between data points mathematically so that it may be used for comparison in this research. This will allow meaningful comparisons between data points. Different similarity metrics may be more suited depending on the specific situation at hand and the nature of the data. The similarity metrics that were found to be best suited for this research on datasets are Cosine similarity, Manhattan distance, and Euclidean distance.

A popular metric for measuring the similarity of two vectors is Cosine similarity, which is especially useful when dealing with high-dimensional data. It evaluates the cosine of the angle that separates two vectors, highlighting the direction of the vectors instead of their magnitude. This makes it especially helpful in applications involving natural language processing where the orientation of word vectors in a high-dimensional space is the primary concern and when the size of the data is not critical. From -1 (completely unlike) to 1 (perfectly similar), the Cosine similarity scale indicates orthogonality [64]. The use of this concept for scene similarity analysis in this research will be discussed in detail in the next section.

A traditional way to estimate the straight-line distance in a multidimensional space between two places is to use the Euclidean distance. It emphasizes the size and direction of contrasts between data points, offering a geometric viewpoint on similarity. The square root of the sum of squared differences between related coordinates must be calculated in order to find the formula. Based on the spatial relationships between points, Euclidean distance is frequently employed in clustering and classification tasks to assist in finding groups or patterns in data. It is, nonetheless, dependent on the volume of data, and in some situations, normalization can be necessary. [56]

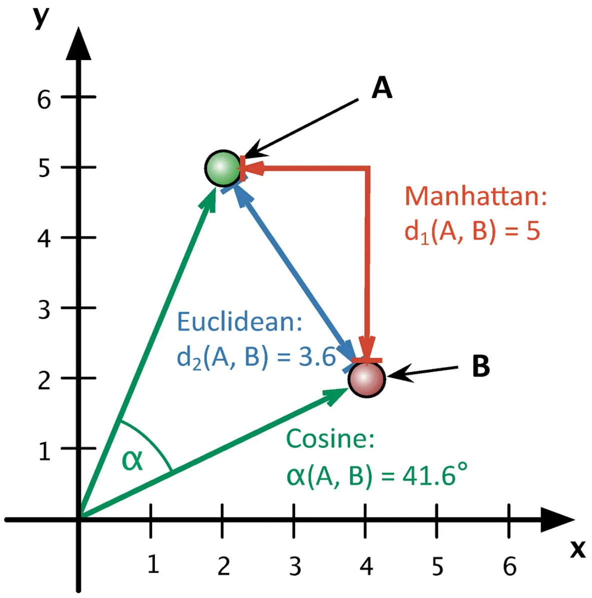


Figure 4.1: The Similarity Measures [65]

Another basic similarity metric that adds the absolute differences between the corresponding coordinates of two vectors is the Manhattan distance, sometimes referred to as the L1 distance. Manhattan distance is a measure of "distance" that is based on the pathways along the grid lines rather than the straight line between points, making it less susceptible to outliers than Euclidean distance. This makes it appropriate in situations when the route traveled matters more than the actual distance traveled. The Manhattan distance is frequently used to solve optimization issues, network analysis, and image processing issues [66]. **Figure 4.1** gives the visual representation of the above mentioned three similarity measures.

The similarity measures were discussed briefly in this section and the research activity will demonstrate what methods were implemented on the data to make this analysis possible in the section 4.2.

4.2 Methodology and implementation

The research has reached the stage where it is possible to describe if two scenes are similar or not. To do this, the experiment takes into account the most important parameters, what is already the existing similarity between two datasets, and what all can be compared more in this research. So starting with what is already known, it is evident that both the datasets have the same field of view of the senors and the same category of classes which are cars, pedestrians, and bikes. So the only thing that will be changing is the value of these parameters and how it is distributed amongst the three classes of objects. These thoughts resulted in coming up with the idea of generating grids or dividing the field of view of the sensors of both datasets into a certain number of small sections and checking for the classes of objects and their parameter values in each grid, hence using this for the similarity analysis later in this section.

The first step is to divide the field of view of both datasets into small sections also called grids. To do this first of all it is necessary to understand how the sensors of the vehicle see the objects in front of them. Since the Radar sensors would be most important in all weather conditions, this research uses the technical data sheet from the Radar sensor model used in nuScenes, which uses the Continental ARS 408-21 Long Range Radar Sensor 77 GHz. This rugged sensor from Continental measures the distance and velocity (Doppler's principle) of objects independently, without a reflector, in one measurement cycle. This is accomplished by using Frequency Modulated Continuous Wave (FMCW) with very fast ramps, with real-time scanning of 17 scans per second. A special feature of the device is the simultaneous measurement of great distances up to 250 m, relative velocity, and the angle relation between 2 objects [67]. **Figure 4.2** shows the sensor ranges of the Radar sensor used for collecting the data in nuScenes dataset.

Looking at the technical data sheet from [67] it is observed that the sensor has a near range of 70 to 100 m for an angular field of view of ±45 degree and a very near range of 20 m with an angular field of view ±60 degree. The resolution in a smooth angle for the sensor at the very near range is 12.3 degrees and the near range is 4.5 degrees this means that at the very near range, a 20-metre distance or less, the sensor won't be able to differentiate two targets which are within 15 degrees angle apart from each other. So considering the technical limitations of the sensors it was decided to split our field of view of two datasets from 50 degrees and at a distance of a maximum of 60 m which is compliant with the delft dataset. To understand this concept better consider this example, for the positive 50 degree FOV in both the datasets, this work divides the FOV angle as 15, 15, and 20 degrees to be compliant with the resolution angle of the sensor. The final 20 degrees was considered as sensors might deviate from ideal conditions at extreme angles. The distance of 60 m is divided into 12 parts, each part at 5 m so that each grid could have a size more than the length of an average car. So, in this research

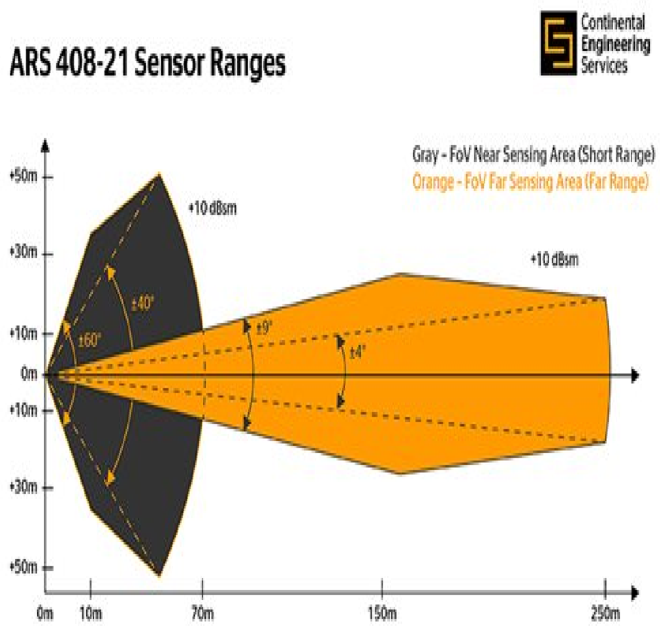

ARS 408-21 Sensor Ranges
Continental Engineering Services
Gray - FoV Near Sensing Area (Short Range)
Orange - FoV Far Sensing Area (Far Range)
+10 dBsm
+10 dBsm
±40°
±60°
±9°
±4°
+50m
+30m
+10m
0m
+10m
+30m
+50m
0m
10m
70m
150m
250m

Figure 4.2: ARS 408-21 Sensor Ranges [67]

the whole FOV of both datasets is divided into 12 segments in distance and 6 segments in angle, resulting in a total of 72 grids.

Figure 4.3 shows the final result of the total number of grids in the field of view of both the data sets. In the above Figure 4.3, (0,0) is going to be the origin point which is the ego vehicle and all the objects will be placed accordingly in a grid based on the two parameters which are the distance from the ego car and the angular position from the ego car. This concept discussed in the section 4.3 forms the backbone of the further similarity analysis and the results of this thesis. The research in the next section will focus mainly on the presence of each class of objects in these grids, analyzing their parameter values and comparing the similarity between two different scenes based on the combination of the above mentioned factors.

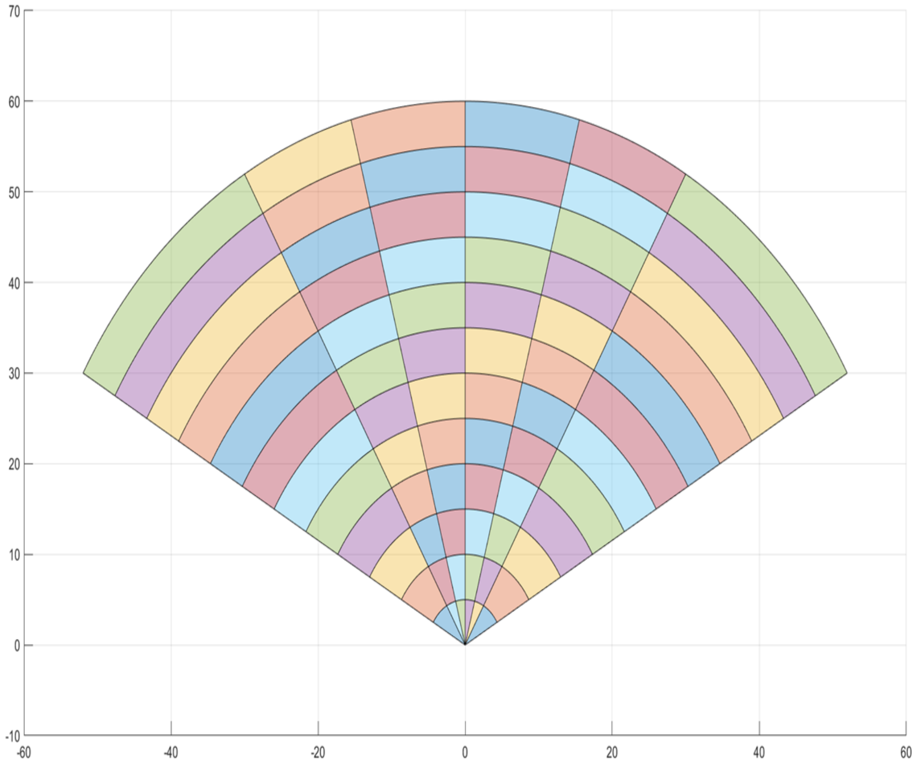


Figure 4.3: The Field of View in 72 Grids.

4.3 Methods for scene comparison and results

In this section, the experiments that were conducted using the concept discussed in the section 4.2 will be explained in detail and the results of each method will be presented. A detailed discussion of the advantages and limitations of each method will be also explained. The similarity analysis conducted using the three similarity measures explained in the section 4.1 on each of the methods can be also found in this section.

The algorithms that will be discussed in all the cases will have two concepts in common. The first one is that similarity will be analyzed for the three different classes of the scenes separately, which means the result will have three separate similarity scores which are car, bike, and pedestrian similarity scores, for each scene. The bike similarity score will not be used for scene comparison between the two datasets since the number of bike annotations in the nuScenes dataset is not significant enough to provide a good result. The second concept is that for every scene, this research takes the average value of the parameters in the grids occupied by the objects and uses this average value of those parameters in each grid for different methods of similarity analysis.

$$Parameter_{grid} = \frac{\sum_{i=1}^{n} P_i}{n} \tag{4.1}$$

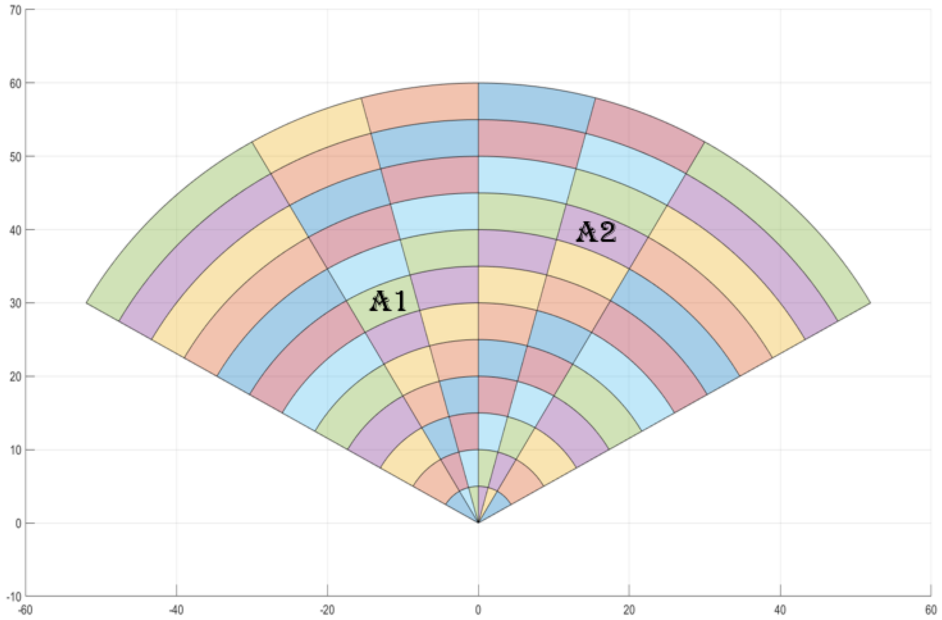
The equation 4.1, summarises the concept discussed before, where;

- $Parameter_{grid}$ is the mean value of the parameters in each grid.
- $grid$ is the number of the object occupied grid (0 to 72).
- n is the total number of objects in each grid.
- i is the number of the object in the grid.
- P_i is the parameter value of the object in the grid.

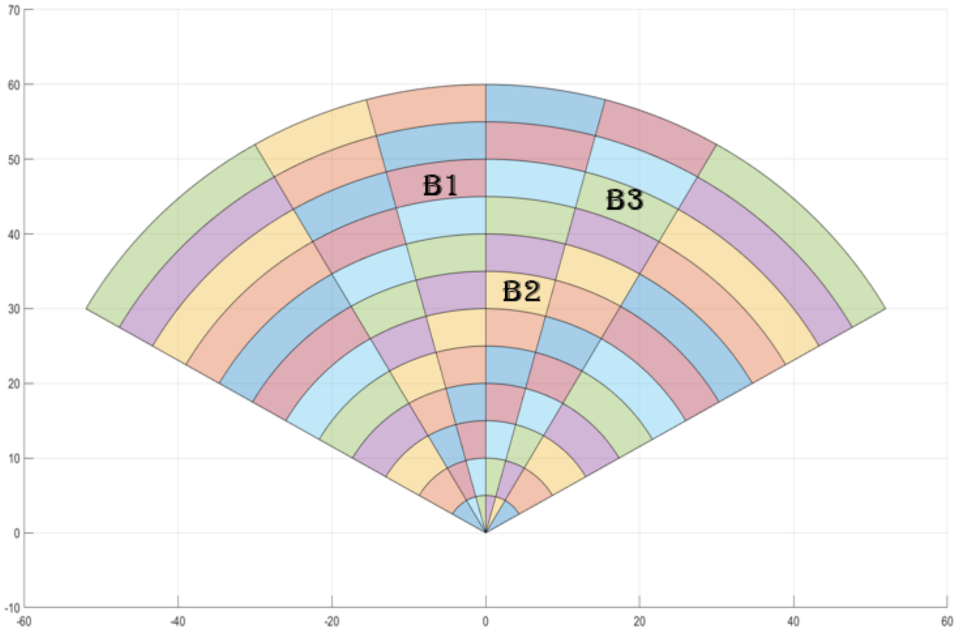
Consider the example of a scene with two pedestrians in grid two, three in grid five, and five in grid fifteen. In each method, the first step will be to identify the occupied grids by pedestrians which are two, five, and fifteen in this case. The second step will be evaluating the mean value of the parameters in each grid, let's say for grid five this research method calculates the mean of the three pedestrian parameter values since grid five has three pedestrians in it. This will be repeated for all the occupied grids for each different class of objects separately. Thus the research will be using only the mean of the parameter values in each grid, based on the number of occupants in that grid, which will give just one solid value for each parameter in every occupied grid. The same method is applied for the object class cars, resulting in two classes with mean parameter values for each scene.

4.3.1 Cosine similarity method

In the above section 4.3, the thesis already discussed the first step and the second step, now the focus will be on the final step in comparing two scenes. In this method, the parameters that are selected for the comparison are the distance of the object from the ego car and the angle made by the object with the ego car. In a nutshell, these two parameters form the coordinate within the FOV and can be tracked down to each grid. Then each occupied grid will have a mean value for distance and angle, which will be treated similarly to the x and y coordinates. This method will be used for every 492 scenes in nuScenes and 1248 scenes in Delft, so the nuScenes dataset will have a 492 x 1 matrix and the Delft dataset with 1248 x 1 matrix, each matrix having 2 vectors representing each class, which are the car and the pedestrian. So the research now has all the mean values of the two parameters, in occupied grids of each scene, for two different classes of objects.



(a) nuScenes Scene A



(b) Delft Scene B

Figure 4.4: Grid occupancy by pedestrians in Scene A and Scene B

The next part will be using the Cosine similarity measure between a scene in nuScenes and Delft. From the **Figure 4.4**, consider an example of class pedestrian which means grids represent pedestrian occupancy in each scene, it is possible to demonstrate how the method is working. Scene A from nuScenes has two grids A1 and A2 with coordinates as the mean value of distance and angle while Scene B has three grids with respective coordinates. The algorithm takes each coordinate (mean parameter values) from A1 and estimates Cosine similarity with each coordinate of the Delft dataset, which is B1, B2, and B3 in this case. Then the same method is followed for A2 with B1, B2, and B3 hence resulting in 6 similarity scores. The final similarity score will be the mean value of these 6 similarity scores. In the general case, if there are n grids in Scene A and m grids in Scene B, the final similarity score for the class pedestrians from these two scenes will be the average value of m x n similarity values. The process is repeated for the other class, car, and the final result will be having three similarity scores for two classes of objects.

Once the similarity score is estimated for each class, a 492 x 1248 matrix is formed having a similarity value between all scenes in the nuScenes and the Delft dataset. The tool used for the analysis is Matlab and it uses the following equation to find Cosine similarity between two coordinates;

$$Cos(\theta) = 1 - Cosine_{distance} = 1 - \frac{A_i \cdot B_j}{|A_i||B_j|} \tag{4.2}$$

In the above equation 4.2 [64];

- $Cos(\theta)$ is the similarity score for a pair of Scene A and Scene B.
- θ is the angle between the two vectors representing the coordinates in Scene A (nuScenes) and Scene B (Delft).
- A_i and B_j are vectors representing coordinates of the grid from the ego car.
- i and j are numbers representing grids in Scene A and Scene B.
- $Cosine_{distance}$ is the distance metric evaluated by Matlab.

The value of $Cos(\theta)$ or the similarity score is in the range [-0.2, 1] where -0.2 means the least similarity score (maximum value of θ is 100 degree in the selected FOV) and $+1$ means maximum similarity, for each class of objects, between each pair of a scene in nuScenes and Delft dataset. The results of this method were filtered by scenes which are having similarity scores greater than 0.7 for both the classes of cars and pedestrians. A few results of this method after the final filtration are shown in Figure A.7.

Figure A.7 gives promising results because this research used only cars and pedestrians for the comparison of scenes based on their distance and angle from the ego car. The similarity score of cars and pedestrians, for each pair of scenes from nuScenes and Delft, are also shown in the figure in the following format; Score = [Similarity score car, Similarity score pedestrian]. Even though this method gives some good results there are two things which need to be improved. The first one is the fact that $Cos(\theta)$ gives the same similarity score for positive and negative angles since cosine is an even function, which means it does not matter if the coordinates of the second grid are to the right or the left of the first grid. The second one is that it does not matter if two coordinates of grids in Scene A and Scene B are making the same angle and are located at different distances from the ego car the similarity score will be the same since the angle between the coordinates is the same. The next methods discussed in this research try to eliminate these problems of the analysis and obtain better results.

4.3.2 Euclidean similarity method

The Euclidean method is expected to solve the second problem discussed in the above section 4.3.1, here the concept of Euclidean distance is used to evaluate the distance metrics for the similarity analysis. The whole procedure in this method is the same as the one explained in section 4.3.1, except the final step where this method uses Euclidean distance to evaluate similarity instead of Cosine distance. So, considering Figure 4.4, this method evaluates the Euclidean distance between each pair of coordinates from Scene A and Scene B. The normalization of Euclidean distance is done to make the values in the range [0, 1], which will be the final similarity score. The method followed for the average value of similarity score is the same as the one discussed in section 4.3.1. Once the similarity score is estimated for each class, a 492 x 1248 matrix is formed having similarity between all scenes in nuScenes and Delft dataset. The tool used for the analysis is also Matlab and it uses the following equation to find Euclidean similarity between two coordinates;

$$Euclidean_{distance} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} \tag{4.3}$$

$$Normalized_{distance} = Euclidean_{distance}/d_{max} \tag{4.4}$$

$$Euclidean_{similarity} = 1 - Normalized_{distance} \tag{4.5}$$

In the above equation 4.3, 4.4 and 4.5 [56];

- $Euclidean_{distance}$ is the distance metric for a pair of coordinates in Scene A and B.
- x_B, y_B, x_A, y_A are the coordinates in Scene B (Delft) and Scene A (nuScenes).
- d_{max} is the maximum distance between two coordinate points which is 110 m.

- $Normalized_{distance}$ is the normalized value of $Euclidean_{distance}$.
- $Euclidean_{similarity}$ is the similarity score

The results shown in Figure A.8 are the final comparison results of this method and solve the second problem from the previous section 4.3.1. The similarity score of cars and pedestrians, for each pair of scenes from the nuScenes and Delft dataset, are also shown in the figure in the following format, Score = [Similarity score car, Similarity score pedestrian]. This method has the advantage of finding similar scenes with objects at an equidistant space between two pairs of scenes. However, there is a significant drawback, the two coordinates at the same distance will have the same similarity score irrespective of the angle between the coordinates. The drawbacks discussed in both these methods, Cosine similarity and Euclidean similarity method, were considered and a new method which solves these problems is discussed in the next section of the thesis.

4.3.3 Cosine and Euclidean Method

The Cosine and Euclidean method uses the combination of both Euclidean and Cosine similarity to eliminate the major drawbacks discussed in the previous sections. This method helps to tackle the two major drawbacks, the independence of the Cosine method on the distance between the coordinates and the independence of the Euclidean method on the angle between the coordinates. The initial steps in this method are the same as the one discussed in the section 4.3.1, only change will be at the final similarity analysis part. In this algorithm, both the Euclidean similarity score and Cosine similarity score are estimated in Matlab, using the equations 4.2 and 4.5, and the final similarity score will be the average value of these two similarity scores as shown in the 4.6.

$$Similarity = \frac{Euclidean_{similarity} + Cos(\theta)}{2} \tag{4.6}$$

In conclusion, the output will be the final similarity score for each class of objects, for each pair of scenes from nuScenes and Delft. Thus a 492 x 1248 matrix is formed having similarity between all scenes in nuScenes and Delft dataset. This method has so far the best result obtained from the thesis and it can find two scenes with cars and pedestrians aligned almost at the same location in the FOV of both the scenes, hence solving the drawbacks of previous methods. Some of the few results obtained from this method are shown in **Figure 4.5**, it is possible to observe that the cars and pedestrians are almost at the same distance and angle from the ego car in both scenes.

4.3.4 Cosine and Euclidean method on nuScenes

Since the Cosine and Euclidean methods gave the best results, it was decided to test this method on the nuScenes dataset to verify the findings. The main motive was to eliminate the errors that might happen because of the variability of both datasets, like more scenes with bikes in Delft and fewer annotations of bikes in nuScenes and many others as discussed in previous sections. Also, the Delft dataset has no visible bounding box annotations to properly visualize the objects in the scene.

For this analysis, the nuScenes dataset was split into two sets of Scenes based on the map or the location where the data was collected. The first dataset is the Singapore map with 181 scenes and the second dataset is the Boston map with 311 scenes. The only difference will be Boston map will be left-hand drive and Singapore will be right-hand drive conditions, but everything else in these datasets is equivalent including the way the data was collected, and annotated, and even the markings of the bounding box. The results from this method are shown in Figure A.10. Looking at the results it is possible to see that the algorithm can work effectively in the nuScenes dataset and the pair of scenes are pretty much similar considering the position of the cars and pedestrians. However the nuScenes dataset has a drawback in having very few pedestrian objects in the scene or comparatively less crowded urban scenes, so most of the comparisons look like highway scenes. The Delft dataset has the advantage of having a lot of pedestrians on the road, hence it is possible to find urban similar scenes as well.

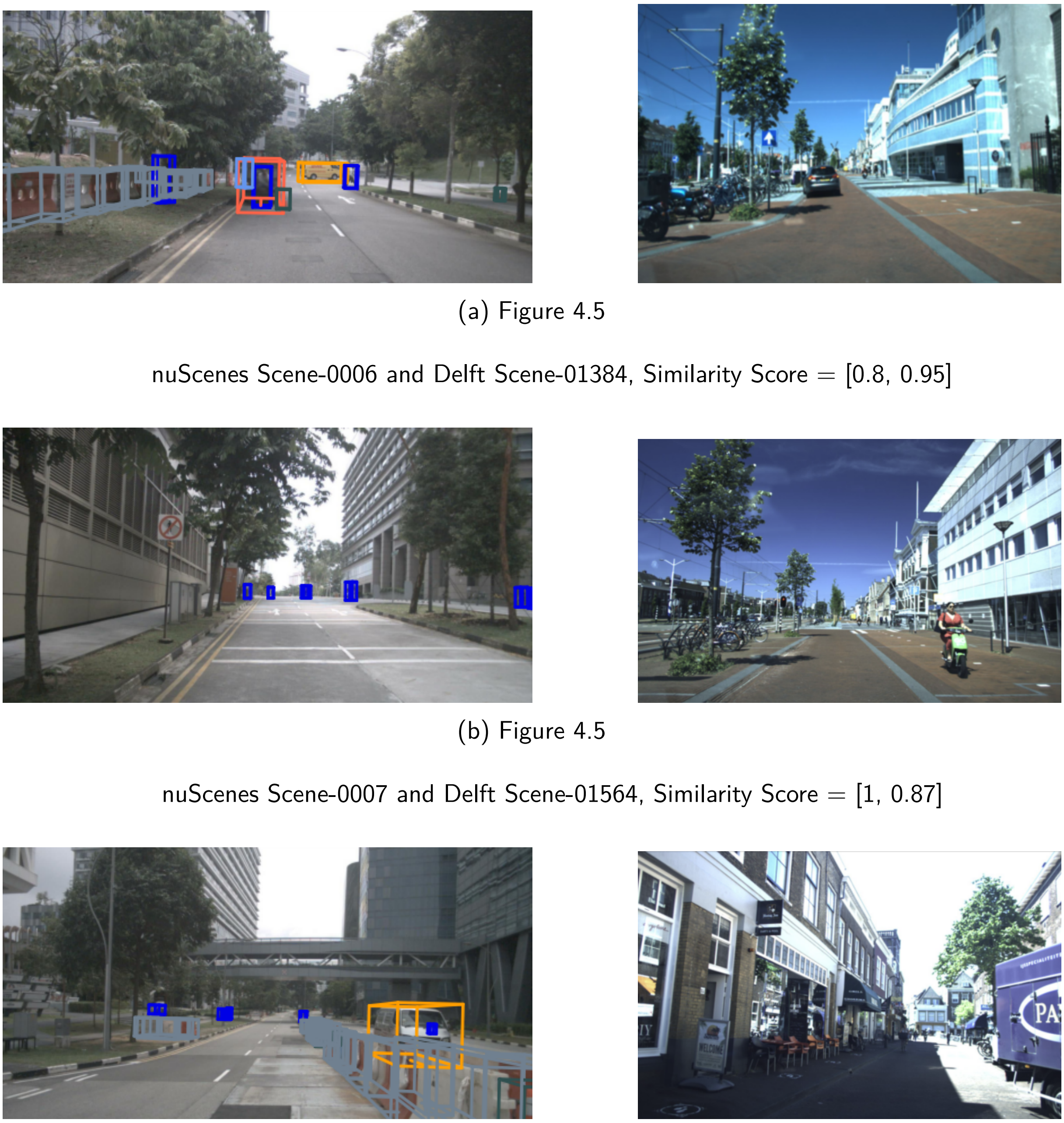
4.3.5 Manhattan similarity method

The Manhattan method was just used for another trial and to compare the results with the Euclidean method. The procedures in this method are exactly similar to the one explained in section 4.3.2 except that the Manhattan distance is used for comparison instead of the Euclidean distance. The Manhattan distance is defined in the equation 4.7 and the normalization part is also the same as equation 4.4 [66].

$$Manhattan_{distance} = |(x_B - x_A)| + |(y_B - y_A)| \tag{4.7}$$

The results shown in Figure A.9 are the final comparison results of this method and also have the same drawback as the Euclidean method, it does not consider the angle of the coordinates for the evaluation. It is quite interesting to see that it has almost the same results as that of the Euclidean distance but the similarity scores have different values for the same two images.

This Chapter explained all the similarity methods that were used to compare the scenes and also explained the methodology used to achieve the results. The method which gives the best results was found to be the Cosine and Euclidean method as compared to the other three methods used to evaluate the similarity between two scenes. The next Chapter will be the final chapter of this thesis discussing the summary and future scope of the research activity.



(a) Figure 4.5

nuScenes Scene-0006 and Delft Scene-01384, Similarity Score = [0.8, 0.95]

(b) Figure 4.5

nuScenes Scene-0007 and Delft Scene-01564, Similarity Score = [1, 0.87]

(c) Figure 4.5

nuScenes Scene-0004 and Delft Scene-01028, Similarity Score = [0.9, 0.79]

Figure 4.5: Similar scenes from Cosine and Euclidean Method

5 Conclusion

This Chapter summarizes the results of the research activity done in this thesis to find the similarity between two scenes from two different datasets - nuScenes and View of Delft. There are mainly two sections in this chapter, the first section will summarize all the research activities and the second section will focus on the future scope of the research activities and some suggestions for improving the results in the future.

5.1 Summary

The aim of this thesis was to find the similarity between two scenes from two different data sets and also to study the different scene description methods possible for driving simulations. To find the similarity between the two scenes however, it was important to understand why this research was being done and what were the prerequisites that are required to finish this thesis. So the first chapter was the introduction to the thesis giving a basic idea about the background and scope of this thesis, and then the main research questions were defined in this chapter.

In the second chapter, the concept of autonomous driving was explained in detail and also the main technological advancements that are happening in this domain were discussed. The five SAE levels of automation needed to understand the concept of autonomous driving were discussed and the main sensors that are in the vehicle that make this concept of autonomous car possible were also introduced. The next step in this thesis was to understand the difference between a scene and a scenario since the main aim of the thesis is to find the similarity between two scenes. Later in this chapter, the main components that are essential for understanding and comparing the driving scenes were briefly explained. The main research activities that are going on in this domain and also the parameters that are used for the driving scene comparison were discussed in detail along with establishing the relations between the parameters for describing a driving scene. A basic idea about different driving data sets that are used for research was included in the final section of the chapter since it helps to understand the different driving datasets and the sensors that are used in these datasets to obtain the driving data.

The third chapter focused mainly on the research activities that were done in this thesis, starting from the selection of the five driving scene parameters and two proper driving datasets to be analyzed in this thesis. A total of 12 different driving parameters and 9 different driving data sets were evaluated and from these 12 different parameters, five final parameters were chosen for the further research activity in this thesis. Also, two different datasets were finalized, the nuScenes and the View of Delft dataset for comparing the scenes between the two datasets. The reason for selecting each of these parameters as well as both of these datasets where also explained in this chapter. The next section of Chapter 3 focused on establishing a reference point and a proper field of view that was compatible with both datasets for further comparisons in the future. A correlation study was done on the five parameters to check if any two parameters have any direct or monotonic relationship but it was found that none of the parameters were directly correlated, so all the four parameters were selected except the bounding box dimension for further analysis in the thesis. The last part of this chapter studied the data distribution of different classes of objects namely the cars, the pedestrians, and the bikes in the two different datasets. All the major observations were recorded which helped to understand what kind of scenes were expected in each dataset and this also helped to come up with better ideas for the next section of the chapter, which was the similarity analysis of two scenes.

The fourth chapter was the similarity analysis and results, in which the four main similarity analysis methods that were used in this chapter were briefly discussed. The four main similarity analysis methods were the Euclidean distance, Manhattan distance, Cosine similarity, and the combination of Cosine and Euclidean methods. This chapter also introduces the new idea of dividing the whole field of view of the driving scene into 72 grids based on the angle and distance from the ego car. This idea of grids was later used in the research to find the mean value of each parameter, belonging to different classes of objects, and then mark a mean parameter value to each grid which were called coordinates in this research. The coordinates used in this research activity were the average value of distance from the ego car and the angle made with respect to the ego car, in each grid, for all the similarity measures. Then each similarity method that was used to compare two different scenes was explained in detail with the methodology and the results. Out of all the methods the Cosine and the Euclidean method gave the best results for this research activity, so this method was also validated on the nuScenes dataset by splitting the dataset into two different datasets, one from the Boston map and the other one from the Singapore map.

This research resulted in innovating a new method of comparing two scenes without using the machine learning algorithm or any complex data comparison methods by using simple logic. However, since this is a new approach it might need some improvements and a lot of further research work in the future to improve the results.

5.2 Future scope

The research activity conducted in this thesis gave some positive results and it can give even better results in the future. Since it was a new method, it took some time to first identify the required parameters and also select the proper datasets. Now that this research activity gives an idea about how each parameters affect the final results of the similarity measures, it is possible to try different combinations of these parameters in the future to improve the results. This research activity selected only five parameters that could be improved in the future by selecting or adding new parameters to the same algorithm or also finding new similarity measures to compare these parameters and hence coming up with new algorithms. A set of parameters that could be experimented with, in the future, is the combination of the yaw orientation of objects relative to the ego car and the velocity of the objects since it gives a clear indication of the direction in which the objects are headed and at what velocity, in each scene. In this research, the velocity of the objects was not possible to evaluate because of the limitation of the Delft dataset, where the velocity of the objects had to be evaluated from two consecutive frames but still, the results were not accurate for the research. In the future, this might be possible, when the Delft dataset publishes the updated dataset with more parameters. It would be a better option to start the evaluation on a similar but more mature European driving dataset since the European driving datasets have a lot of pedestrian and bike annotations to study the crowded urban scenes. Also, it would be possible to select a different set of parameters and try out the same algorithm on entirely different driving scene datasets to see how the parameters are affecting the similarity analysis in the scene. Another possible improvement is a better similarity measure than the combination of the Euclidean and Cosine methods, which gave a better result in this thesis, where multiple parameters could be used for the analysis at the same time.

A Appendix

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	0.65	0.058	0.146
Pearson's Correlation	0.12	0.016	0.067
Spearman's Correlation	0.1	0.008	0.072

Table A.1: Statistical values from Figure A.1

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	-28.5	13.9	-19
Pearson's Correlation	-0.01	0	-0.01
Spearman's Correlation	-0.02	0	-0.01

Table A.2: Statistical values from Figure A.2

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	0.67	0.239	0.048
Pearson's Correlation	0.06	0.016	0.008
Spearman's Correlation	0.018	0.008	-0.015

Table A.3: Statistical values from Figure A.1

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	0.44	0.03	0.66
Pearson's Correlation	0.009	0.002	0.09
Spearman's Correlation	0.008	0.009	0.13

Table A.4: Statistical values from Figure A.3

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	1.98	2.3	-1.8
Pearson's Correlation	0.04	0.05	-0.03
Spearman's Correlation	0.03	0.03	-0.02

Table A.5: Statistical values from Figure A.3

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	71.8	-38	-350
Pearson's Correlation	0.006	-0.008	-0.19
Spearman's Correlation	0	-0.007	-0.13

Table A.6: Statistical values from Figure A.4

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	3.3	-45.5	7.34
Pearson's Correlation	0.002	-0.09	0.03
Spearman's Correlation	-0.02	-0.14	0.04

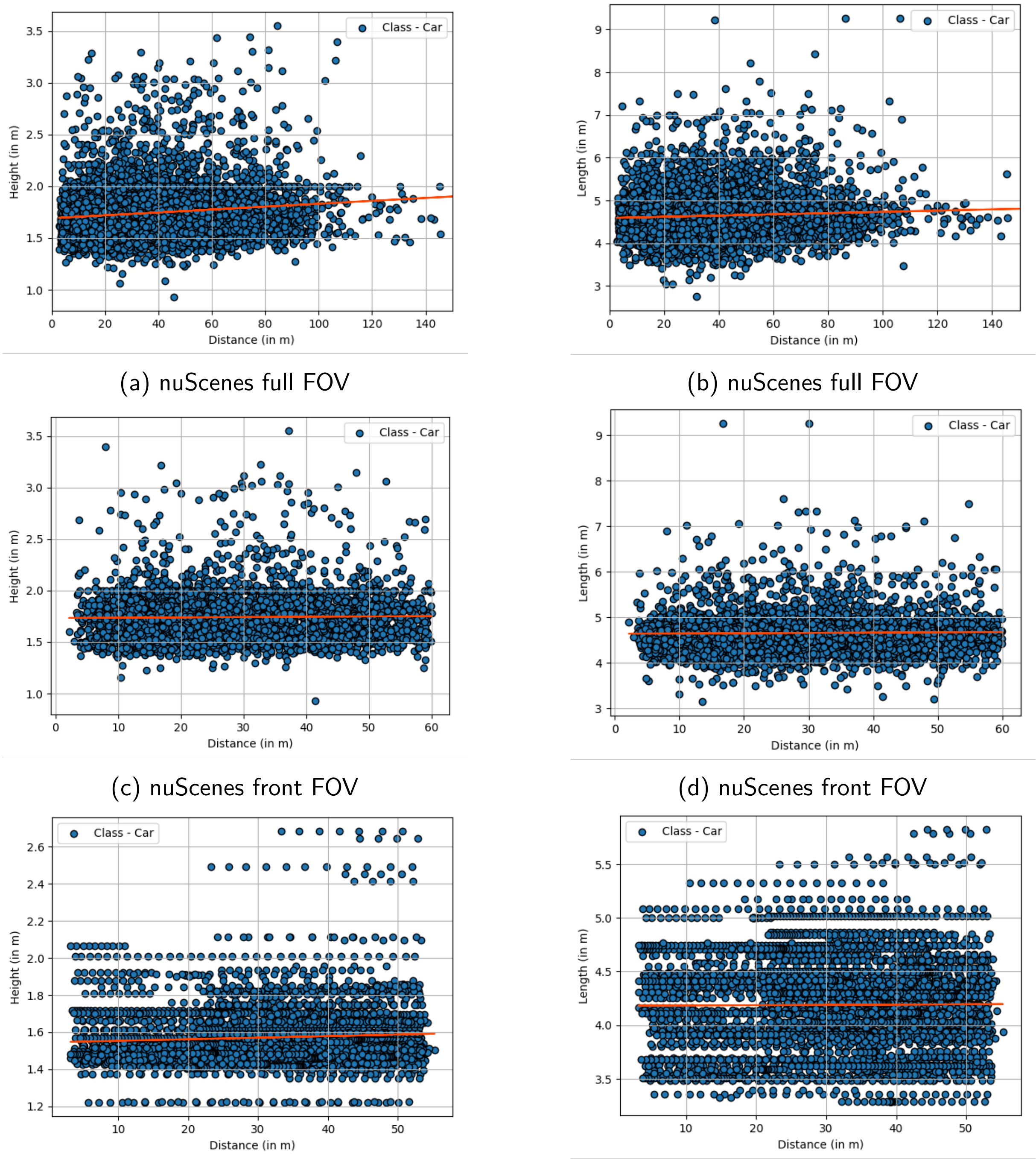
Table A.7: Statistical values from Figure A.2

Datasets/Parameters	nuScenes full FOV	nuScenes front FOV	View of Delft
Covariance	83.2	-13	37.6
Pearson's Correlation	0.05	-0.02	0.19
Spearman's Correlation	0.05	-0.02	0.24

Table A.8: Statistical values from Figure 3.8

Datasets/Parameters	A2D2	Argoverse 2	RadarScenes	AIO Drive
1. Weather	Yes	Yes	No	Yes
2. Speed of vehicle	Yes	Yes	Yes	Yes
3. Lane of ego vehicle	No	Yes	No	No
4. Bounding Box dimension	Yes	Yes	No	Yes
5. Yaw orientation	Yes	Yes	No	Yes
6. Steering wheel angle	Yes	No	No	Yes
7. Speed of object	Yes	Yes	Yes	Yes
8. Road conditions & visibility	Yes	No	No	Yes
9. Distance from road margins	No	Yes	No	No
10. Distance from ego car	Yes	Yes	Yes	Yes
11. Angle with respect to ego car	Yes	Yes	No	Yes
12. Headway	No	Yes	No	No

Table A.9: Comparison of parameters in each dataset



(a) nuScenes full FOV

(b) nuScenes full FOV

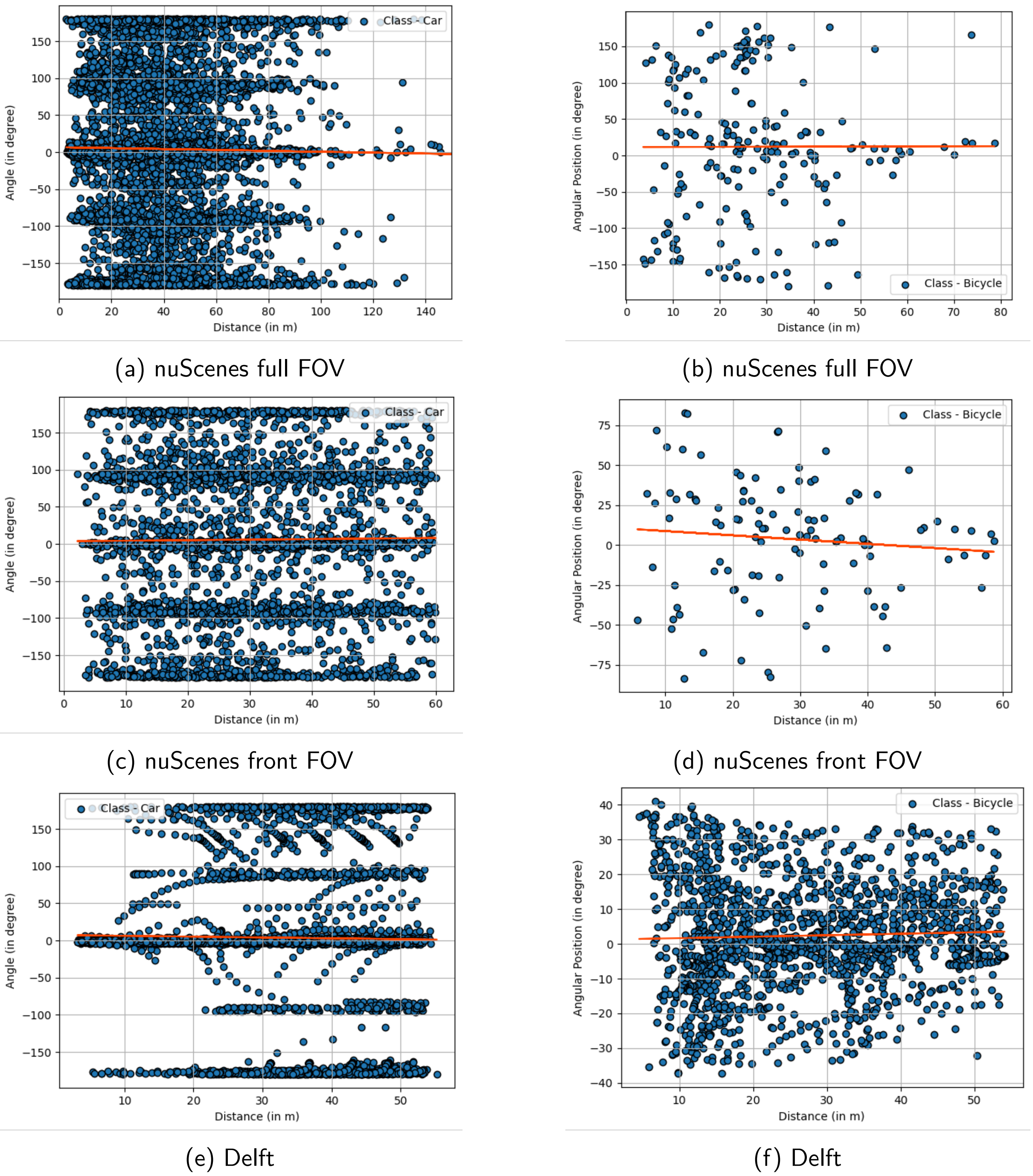
(c) nuScenes front FOV

(d) nuScenes front FOV

(e) Delft

(f) Delft

Figure A.1: Scatter plot of Distance from the Ego Car and Bounding Box dimension



(a) nuScenes full FOV

(b) nuScenes full FOV

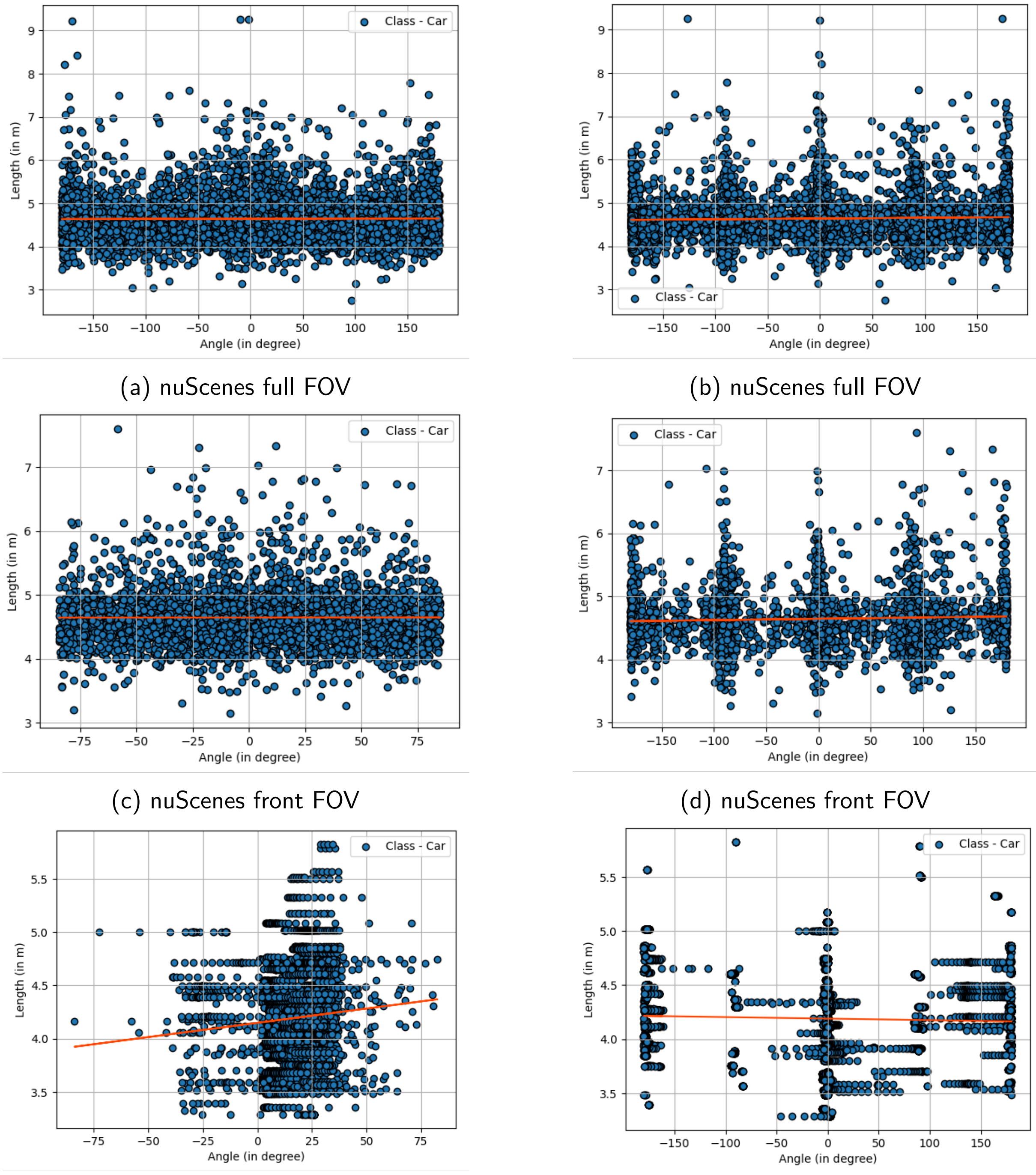
(c) nuScenes front FOV

(d) nuScenes front FOV

(e) Delft

(f) Delft

Figure A.2: Scatter plot of Distance from the Ego Car with Yaw Orientation [a,c,e] and Angular Position [b,d,f]



(a) nuScenes full FOV

(b) nuScenes full FOV

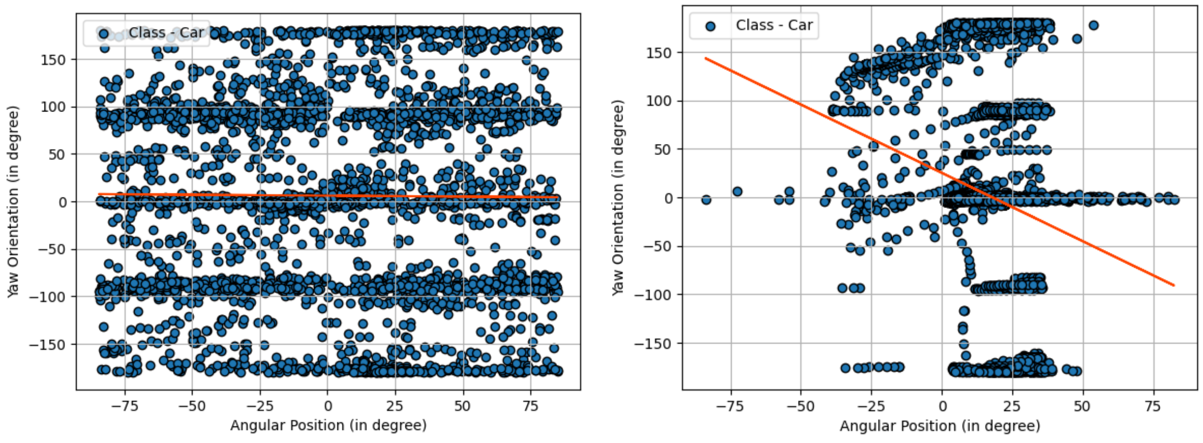
(c) nuScenes front FOV

(d) nuScenes front FOV

(e) Delft

(f) Delft

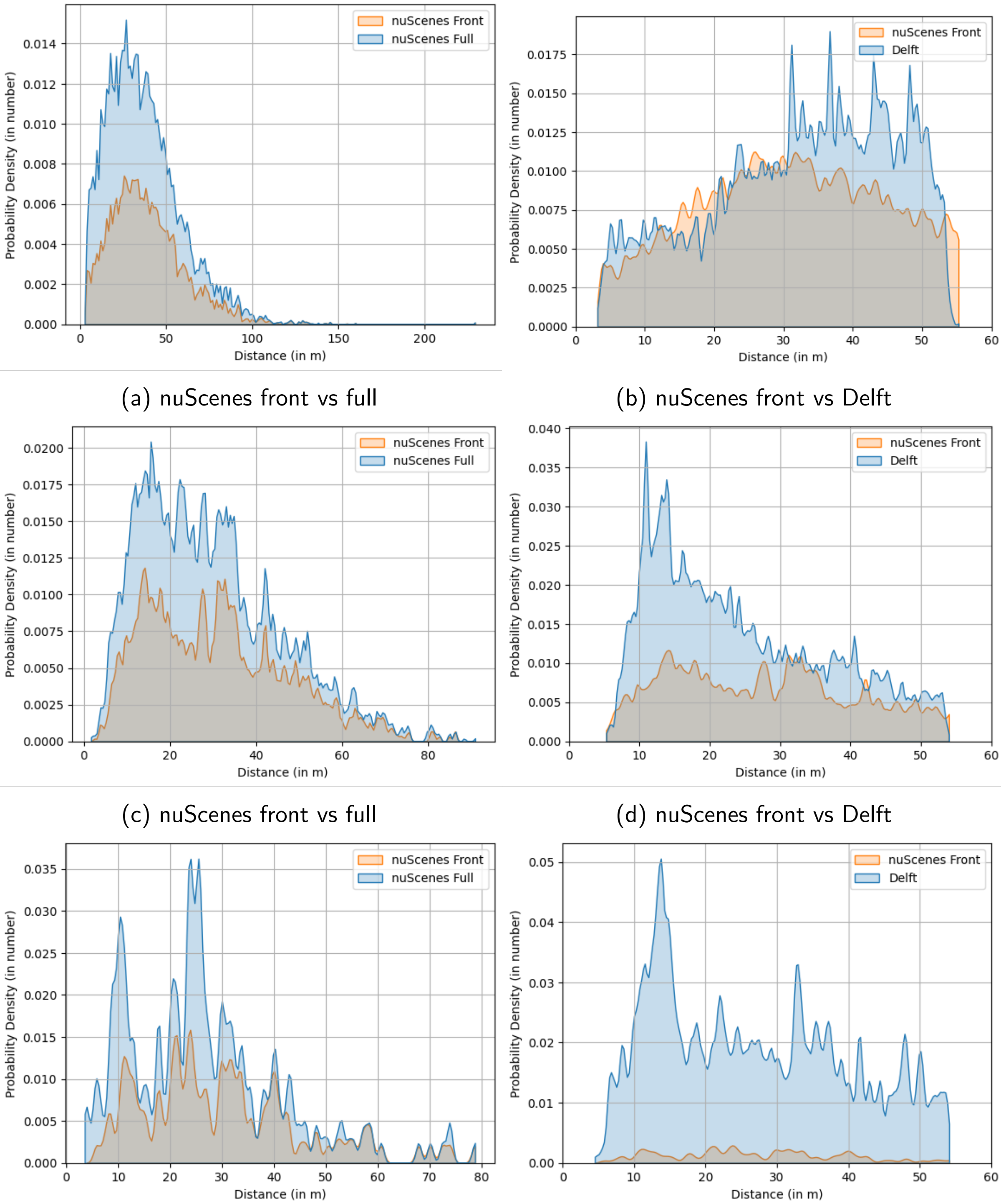
Figure A.3: Scatter plot of Angular position [a,c,e] and Yaw orientation [b,d,f] and Length of Bounding Box



(a) nuScenes front FOV

(b) Delft

Figure A.4: Scatter plot of Angular Position and Yaw Orientation



(a) nuScenes front vs full

(b) nuScenes front vs Delft

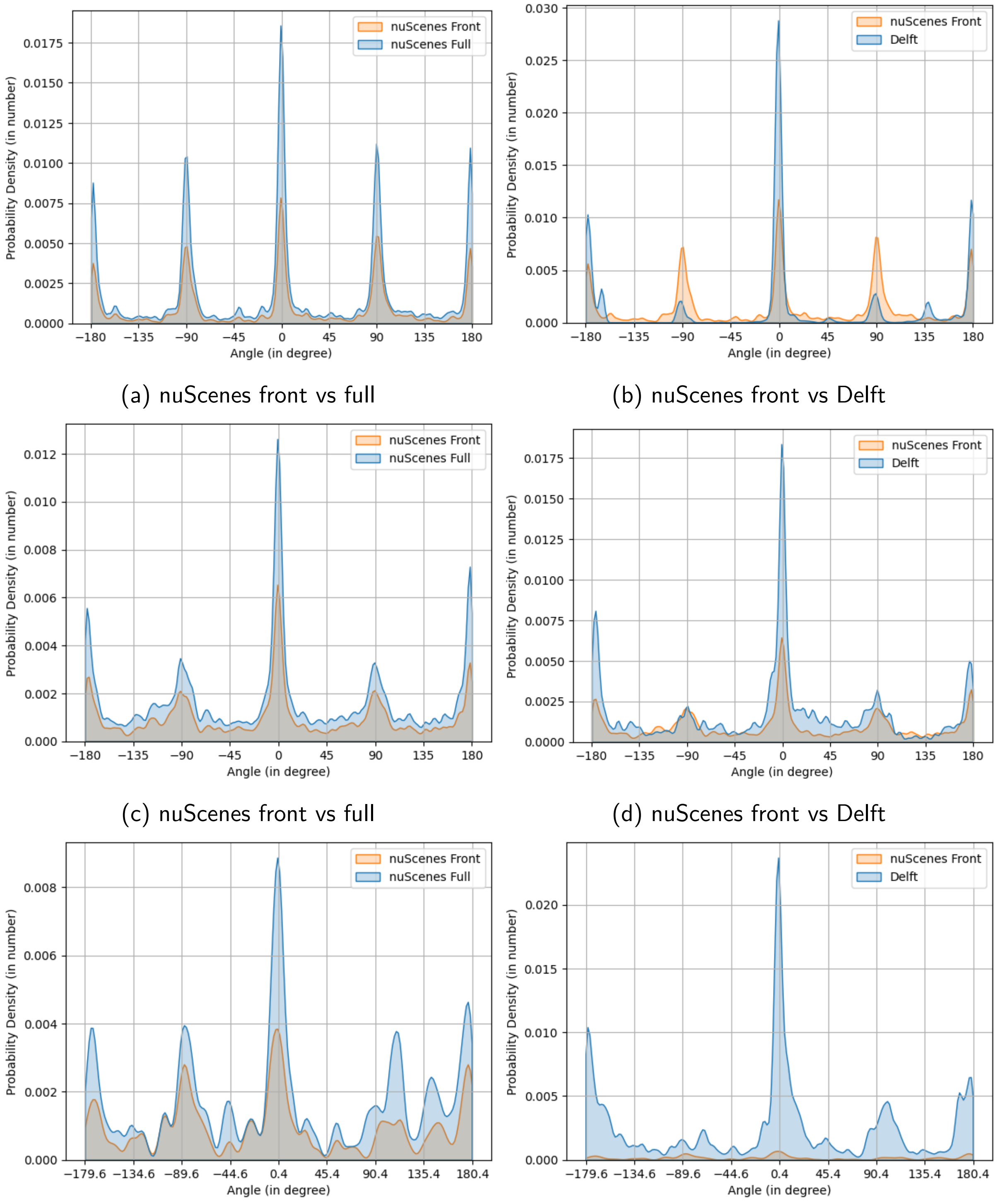
(c) nuScenes front vs full

(d) nuScenes front vs Delft

(e) nuScenes front vs full

(f) nuScenes front vs Delft

Figure A.5: Probability Distribution of Distance [Car(a,b), Pedestrian(c,d) and Bike(e,f)]



(a) nuScenes front vs full

(b) nuScenes front vs Delft

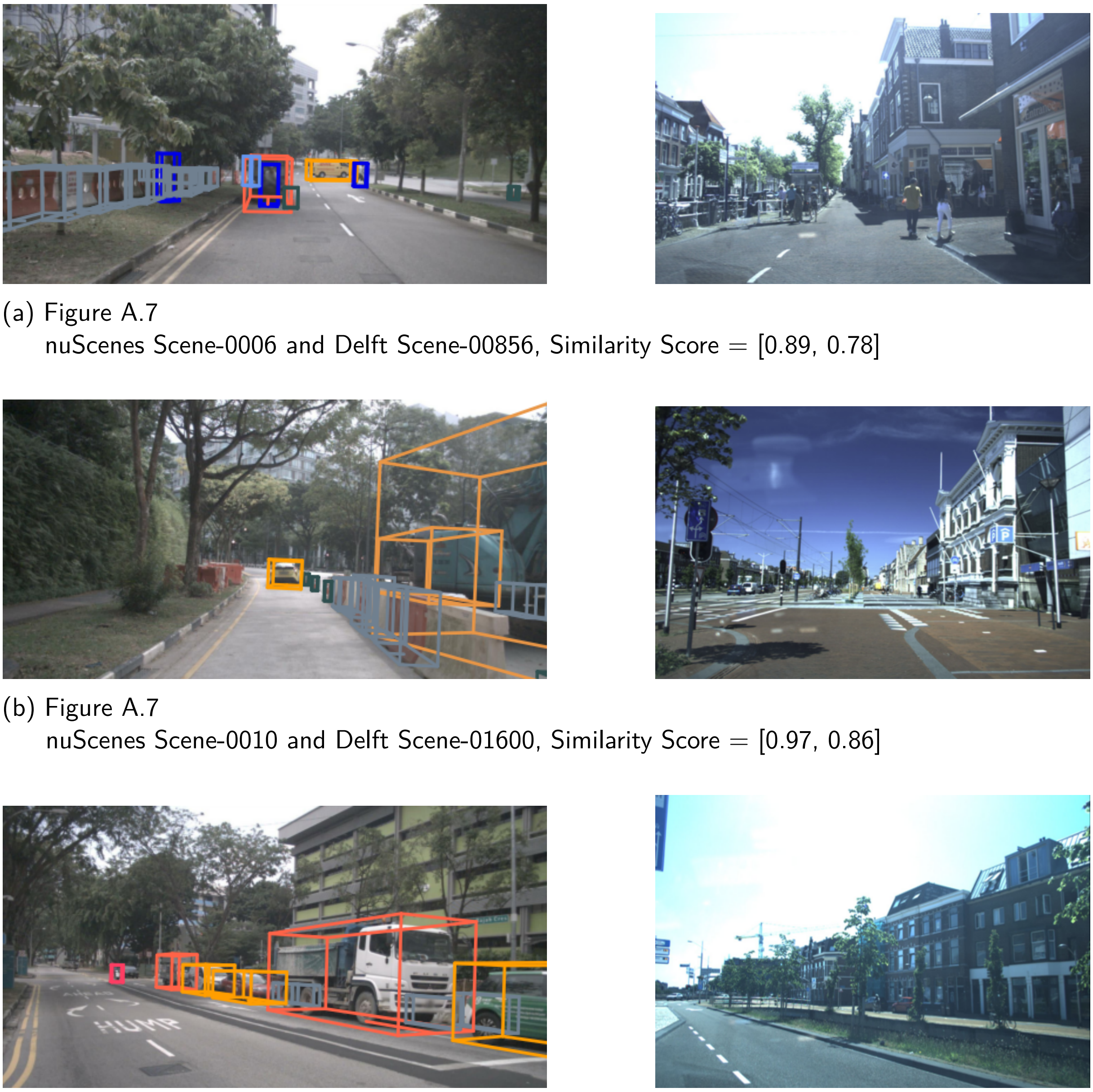
(c) nuScenes front vs full

(d) nuScenes front vs Delft

(e) nuScenes front vs full

(f) nuScenes front vs Delft

Figure A.6: Probability Distribution of Yaw Orientation [Car(a,b), Pedestrian(c,d) and Bike(e,f)]

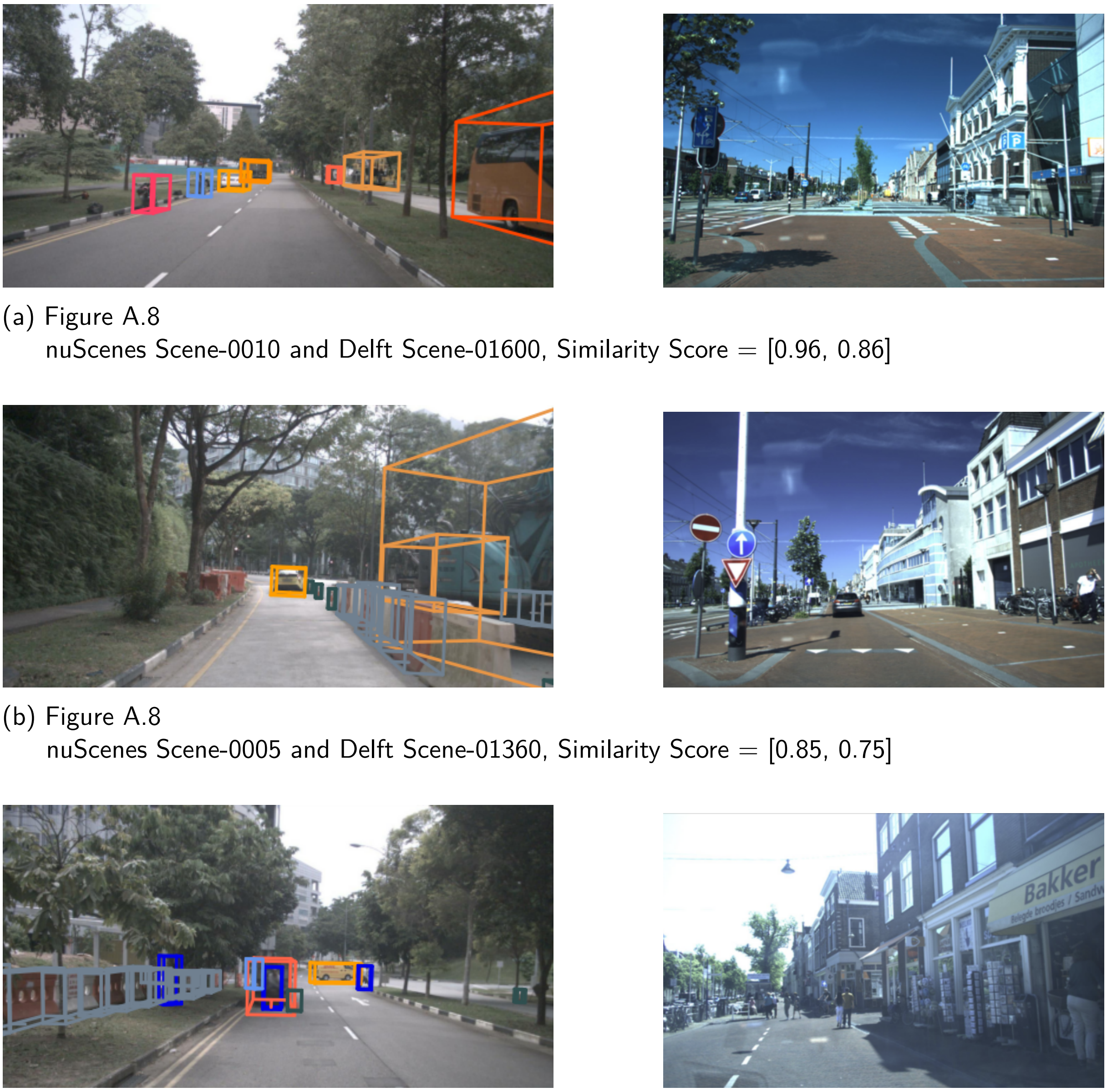


(a) Figure A.7
nuScenes Scene-0006 and Delft Scene-00856, Similarity Score = [0.89, 0.78]

(b) Figure A.7
nuScenes Scene-0010 and Delft Scene-01600, Similarity Score = [0.97, 0.86]

(c) Figure A.7
nuScenes Scene-0002 and Delft Scene-01768, Similarity Score = [0.78, 0.75]

Figure A.7: Similar scenes from Cosine Method

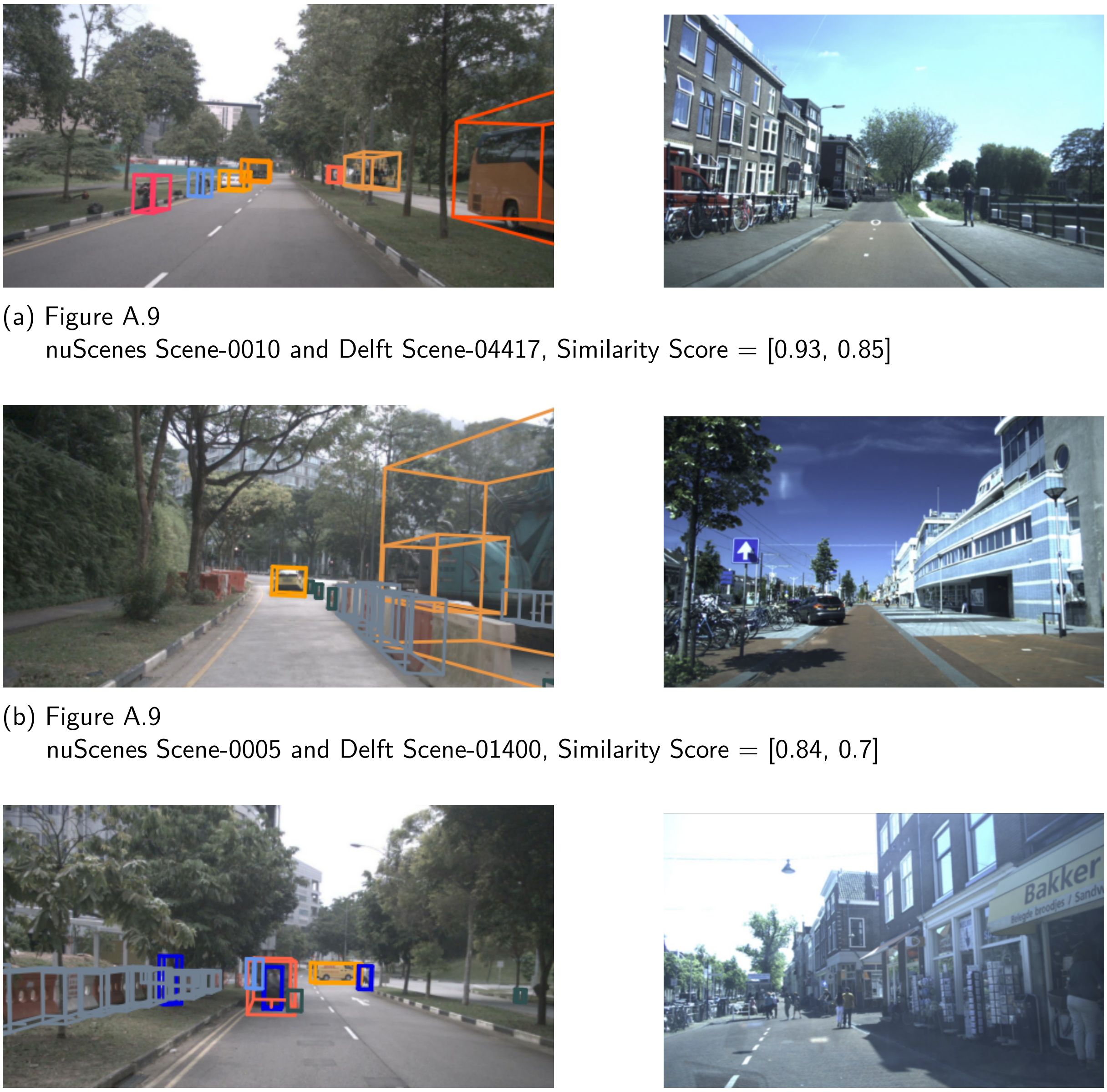


(a) Figure A.8
nuScenes Scene-0010 and Delft Scene-01600, Similarity Score = [0.96, 0.86]

(b) Figure A.8
nuScenes Scene-0005 and Delft Scene-01360, Similarity Score = [0.85, 0.75]

(c) Figure A.8
nuScenes Scene-0006 and Delft Scene-00828, Similarity Score = [0.79, 0.8]

Figure A.8: Similar scenes from Euclidean Method

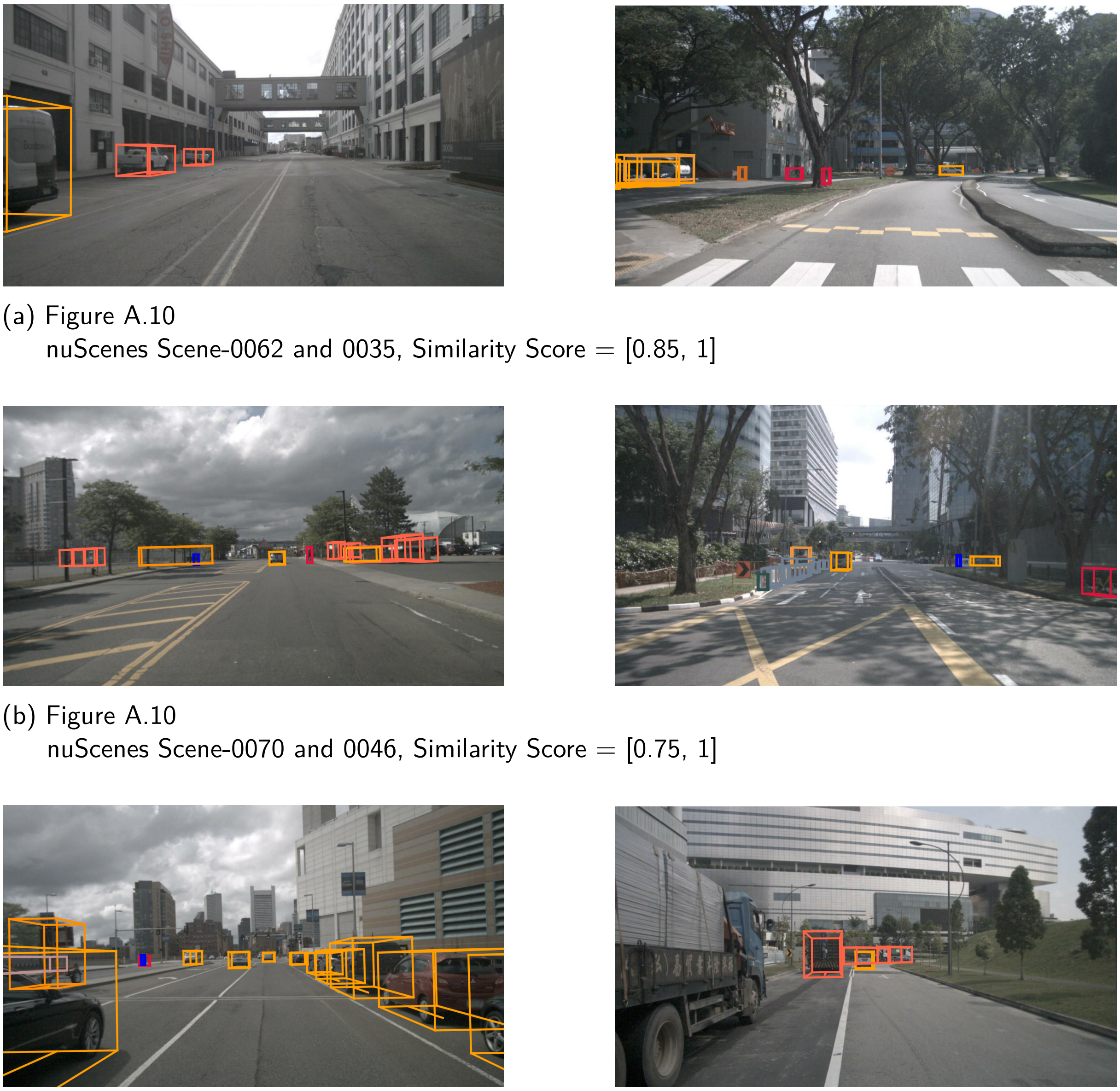


(a) Figure A.9
nuScenes Scene-0010 and Delft Scene-04417, Similarity Score = [0.93, 0.85]

(b) Figure A.9
nuScenes Scene-0005 and Delft Scene-01400, Similarity Score = [0.84, 0.7]

(c) Figure A.9
nuScenes Scene-0006 and Delft Scene-00828, Similarity Score = [0.78, 0.74]

Figure A.9: Similar scenes from Manhattan Method



(a) Figure A.10
nuScenes Scene-0062 and 0035, Similarity Score = [0.85, 1]

(b) Figure A.10
nuScenes Scene-0070 and 0046, Similarity Score = [0.75, 1]

(c) Figure A.10
nuScenes Scene-0098 and 0054, Similarity Score = [0.7, 1]

Figure A.10: Similar scenes from nuScenes Map

Bibliography

[1] Y. Li, Y. Zhong, Y. Zhang, and C. Shen, "Learning cross-modal deep embeddings for robust scene recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8795–8804.

[2] H. Zhang, T. Xu, H. Li, *et al.*, "Self-supervised data augmentation for improved diversity of data and training efficiency," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019, pp. 0-0.

[3] McKinsey and Company, "Autonomous vehicles: Technology, trends, and transformations," 2019. [Online]. Available: `https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/disruptive-trends-that-will-transform-the-auto-industry`.

[4] S. Chen, L. Wang, Z. Chen, and S. Chen, "A review of perception and control techniques for autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3828–3844, 2020. DOI: `10.1109/TITS.2019.2959710`.

[5] Society of Automotive Engineers, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE International, Tech. Rep., 2018.

[6] C. A. Smith and C. J. Brown, "Lidar sensors for self-driving cars: An overview," *IEEE Sensors Journal*, vol. 18, no. 11, pp. 4294–4301, 2018.

[7] Y. Kim and M. Kang, "Deep learning for autonomous vehicles: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 7, pp. 2895–2913, 2020.

[8] Y. Li, C. Zhang, and Y. Chen, "Multi-sensor fusion for perception system of autonomous vehicles: A review," *Information Fusion*, vol. 52, pp. 266–279, 2019.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[10] G. Egnal, S. Milz, and C. Becker, "High definition map-based localization for autonomous vehicles: A review," *IEEE Intelligent Transportation Systems Magazine*, vol. 12, no. 1, pp. 33–45, 2020.

[11] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, "Defining and substantiating the terms scene, situation, and scenario for automated driving," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015.

[12] Ò. Lorente, I. Riera, and A. Rana, "Scene understanding for autonomous driving," *CoRR*, vol. abs/2105.04905, 2021. arXiv: `2105.04905`.

[13] S. Zhang, G. Wu, J. P. Costeira, and J. M. F. Moura, “Understanding traffic density from large-scale web camera data,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4264–4273. DOI: 10.1109/CVPR.2017.454.

[14] J. D. Kim, J. A. Perrone, and R. B. Isler, “The effect of differences in day and night lighting distributions on drivers’ speed perception,” *Perception*, vol. 46, no. 6, pp. 728–744, 2017, PMID: 27923941. DOI: 10.1177/0301006616684236.

[15] R. Blin, S. Ainouz, S. Canu, and F. Meriaudeau, “Road scenes analysis in adverse weather conditions by polarization-encoded images and adapted deep learning,” in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 27–32. DOI: 10.1109/ITSC.2019.8916853.

[16] L. Tisljaric, S. Fernandes, T. Caric, and J. Gama, “Spatiotemporal road traffic anomaly detection: A tensor-based approach,” *Applied Sciences*, vol. 11, no. 24, p. 12 017, Dec. 2021. DOI: 10.3390/app112412017.

[17] A. A. Kashyap, S. Raviraj, A. Devarakonda, S. R. N. K, S. K. V, and S. J. Bhat, “Traffic flow prediction models – a review of deep learning techniques,” *Cogent Engineering*, vol. 9, no. 1, F. Galatioto, Ed., p. 2 010 510, 2022. DOI: 10.1080/23311916.2021.2010510.

[18] Federal Highway Administration (FHWA), *Roadway maintenance*, Accessed May 1, 2023. https://highways.dot.gov/safety/other/safety-and-roadway-maintenance-link, 2021.

[19] National Highway Traffic Safety Administration (NHTSA), *Traffic safety facts*, Accessed May 13, 2023. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813206, 2021.

[20] Federal Highway Administration (FHWA), *Manual on uniform traffic control devices for streets and highways*, Accessed June 5, 2023. https://mutcd.fhwa.dot.gov/, 2021.

[21] AAA Foundation for Traffic Safety, “2021 traffic safety culture index,” AAA Foundation for Traffic Safety, Washington D.C., Tech. Rep., 2022.

[22] M. Scholtes, L. Westhofen, L. R. Turner, *et al.*, *6-layer model for a structured description and categorization of urban traffic and environment*, 2021. arXiv: 2012.06319 [cs.OH].

[23] D. Cavaliere, S. Senatore, and V. Loia, “A multi-perspective aerial monitoring system for scenario detection,” in *2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS)*, 2018, pp. 1–6. DOI: 10.1109/EESMS.2018.8405820.

[24] M. A. Mohammad and R. S. Ioannis Kaloskampis Yulia Hicks, “Ontology-based framework for risk assessment in road scenes using videos,” *Procedia Computer Science*, vol. 60, pp. 1532–1541, 2015, Knowledge-Based and Intelligent Information and Engineering Systems 19th Annual Conference, KES-2015, Singapore, September 2015 Proceedings, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2015.08.300.

[25] J. Wang, C. Zhang, Y. Liu, and Q. Zhang, "Traffic sensory data classification by quantifying scenario complexity," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1543–1548. DOI: 10.1109/IVS.2018.8500669.

[26] W. Shi and L. Liu, "Dataset and benchmark," in *Computing Systems for Autonomous Driving*. Cham: Springer International Publishing, 2021, pp. 109–142. DOI: 10.1007/978-3-030-81564-6_5.

[27] H. Caesar, V. Bankiti, A. H. Lang, *et al.*, "Nuscenes: A multimodal dataset for autonomous driving," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 618–11 628. DOI: 10.1109/CVPR42600.2020.01164.

[28] A. Palffy, E. Pool, S. Baratam, J. F. P. Kooij, and D. M. Gavrila, "Multi-class road user detection with 3+1d radar in the view-of-delft dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4961–4968, 2022. DOI: 10.1109/LRA.2022.3147324.

[29] S. Ettinger, S. Cheng, B. Caine, *et al.*, *Large scale interactive motion forecasting for autonomous driving : The waymo open motion dataset*, 2021. arXiv: 2104.10133 [cs.CV].

[30] N. Gählert, N. Jourdan, M. Cordts, U. Franke, and J. Denzler, *Cityscapes 3d: Dataset and benchmark for 9 dof vehicle detection*, 2020. arXiv: 2006.07864 [cs.CV].

[31] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013. DOI: 10.1177/0278364913491297.

[32] J. Geyer, Y. Kassahun, M. Mahmudi, *et al.*, *A2d2: Audi autonomous driving dataset*, 2020. arXiv: 2004.06320 [cs.CV].

[33] B. Wilson, W. Qi, T. Agarwal, *et al.*, *Argoverse 2: Next generation datasets for self-driving perception and forecasting*, 2023. arXiv: 2301.00493 [cs.CV].

[34] O. Schumann, M. Hahn, N. Scheiner, *et al.*, *Radarscenes: A real-world radar point cloud data set for automotive applications*, 2021. arXiv: 2104.02493 [cs.LG].

[35] X. Weng, Y. Man, J. Park, Y. Yuan, M. O'Toole, and K. M. Kitani, *All-in-one drive: A comprehensive perception dataset with high-density long-range point clouds*, 2021.

[36] Y. Jia, J. Wu, and Y. Du, "Modeling and simulation of rainfall impacts on urban traffic flow: A case study in beijing," in *Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems*, L. Zhang, X. Song, and Y. Wu, Eds., Singapore: Springer Nature Singapore, 2016, pp. 475–484.

[37] A. M. Pérez-Marín and M. Guillen, "Semi-autonomous vehicles: Usage-based data evidences of what could be expected from eliminating speed limit violations," *Accident Analysis and Prevention*, vol. 123, pp. 99–106, 2019, ISSN: 0001-4575. DOI: https://doi.org/10.1016/j.aap.2018.11.005.

[38] H. Wang, T. Li, Y. Li, *et al.*, *Openlane-v2: A topology reasoning benchmark for unified 3d hd mapping*, 2023. arXiv: 2304.10440 [cs.CV].

[39] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517–6525. DOI: 10.1109/CVPR.2017.690.

[40] A. Paz and G. Arechavaleta, "Online optimization of humanoid walking trajectories for passing through a door," *Robotics and Autonomous Systems*, vol. 115, pp. 61–72, 2019, ISSN: 0921-8890. DOI: https://doi.org/10.1016/j.robot.2019.01.014.

[41] P. Papantoniou, E. Papadimitriou, and G. Yannis, "Review of driving performance parameters critical for distracted driving research," *Transportation Research Procedia*, vol. 25, pp. 1796–1805, 2017, World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016, ISSN: 2352-1465. DOI: https://doi.org/10.1016/j.trpro.2017.05.148.

[42] L. Riazuelo, J. Civera, and J. Montiel, "C2tam: A cloud framework for cooperative tracking and mapping," *Robotics and Autonomous Systems*, vol. 62, no. 4, pp. 401–413, 2014, ISSN: 0921-8890. DOI: https://doi.org/10.1016/j.robot.2013.11.007.

[43] M. Ding, Y. Huo, H. Yi, *et al.*, *Learning depth-guided convolutions for monocular 3d object detection*, 2019. arXiv: 1912.04799 [cs.CV].

[44] Z. Wang, W. Ren, and Q. Qiu, *Lanenet: Real-time lane detection networks for autonomous driving*, 2018. arXiv: 1807.01726 [cs.CV].

[45] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 690–703, 2017. DOI: 10.1109/TNNLS.2016.2522428.

[46] E. S. Ye, "Object detection in rgb-d indoor scenes," M.S. thesis, EECS Department, University of California, Berkeley, Jan. 2013.

[47] V. Milanés, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura, "Cooperative adaptive cruise control in real traffic situations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 296–305, 2014. DOI: 10.1109/TITS.2013.2278494.

[48] nuScenes by Motional), *Car setup*, Accessed Nov 20, 2023. https://www.nuscenes.org/nuscenes, 2019.

[49] TU Delft, *Sensors and data*, Accessed Nov 20, 2023. https://github.com/tudelft-iv/view-of-delft-dataset/blob/main/docs/SENSORS_AND_DATA.md, 2022.

[50] ISO 23150:2023, *Road vehicles — data communication between sensors and data fusion unit for automated driving functions — logical interface*, Accessed June 20, 2023. https://www.iso.org/obp/ui/en/iso:std:iso:23150:ed-2:v1:en, 2023.

[51] TU Delft, *Annotation information*, Accessed May 20, 2023. https://github.com/tudelft-iv/view-of-delft-dataset/blob/main/docs/ANNOTATION.md, 2022.

[52] Vipin Sharma, *Kitti coordinate transformations*, Accessed May 20, 2023. https://towardsdatascience.com/kitti-coordinate-transformations-125094cd42fb, 2021.

[53] Dimensions.com, *Toyota prius*, Accessed May 25, 2023. https://www.dimensions.com/element/toyota-prius-2016, 2016.

[54] nuScenes, *Can bus*, Accessed June 5, 2023. https://www.nuscenes.org/tutorials/can_bus_tutorial.html, 2020.

[55] TU Delft, *Ego vehicle velocity*, Accessed June 1, 2023. https://github.com/tudelft-iv/view-of-delft-dataset/issues/47, 2023.

[56] G. J. M. Rosa, "The elements of statistical learning: Data mining, inference, and prediction by hastie, t., tibshirani, r., and friedman, j.," *Biometrics*, vol. 66, no. 4, pp. 1315–1315, 2010. DOI: https://doi.org/10.1111/j.1541-0420.2010.01516.x.

[57] nuScenes, *Orientation angle of the bounding box*, Accessed June 1, 2023. https://github.com/nutonomy/nuscenes-devkit/issues/21, 2023.

[58] M. Triola, *Elementary Statistics*. Pearson, 2018, ISBN: 9780134462455.

[59] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis* (Applied Multivariate Statistical Analysis). Pearson Prentice Hall, 2007, ISBN: 9780131877153.

[60] S. Siegel and N. Castellan, *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill international editions statistics series). McGraw-Hill, 1988, ISBN: 9780070573574.

[61] Quizlet, *Monotonic relationship image*, Accessed June 28, 2023. https://quizlet.com/ca/629984576/psych-277-midterm-1-lo-flash-cards/, 2023.

[62] H. Caesar, V. Bankiti, A. H. Lang, *et al.*, "Nuscenes: A multimodal dataset for autonomous driving," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11 618–11 628, 2019.

[63] A. A. Goshtasby, "Similarity and dissimilarity measures," 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:120107178.

[64] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[65] Mohammed Terry-Jack, *Nlp: Everything about embeddings*, Accessed September 15, 2023. https://medium.com/@b.terryjack/nlp-everything-about-word-embeddings-9ea21f51ccfe, 2019.

[66] C. Aggarwal, *Data Mining: The Textbook*. Springer International Publishing, 2015, ISBN: 9783319141428. [Online]. Available: https://books.google.de/books?id=cfNICAAAQBAJ.

[67] Continental Engineering Services, *Ars 408*, Accessed September 15, 2023. `https://medium.com/@b.terryjack/nlp-everything-about-word-embeddings-9ea21f51ccfe`, 2019.