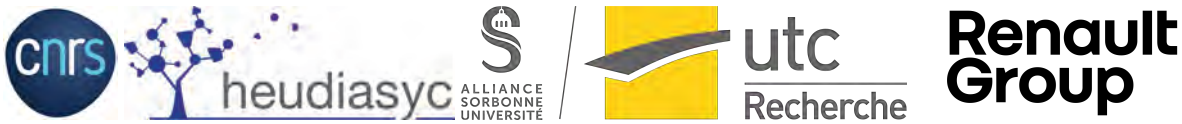


A World Model enabling Information Integrity for Autonomous Vehicles

PhD thesis defended by

Corentin Sanchez

Doctoral School **University of Technology of Compiègne**

Prepared at **Heudiasyc UMR CNRS – UTC 7253**

Academic Field **Computer Engineering, Robotics and Automation**

Defended on the 3rd of may 2022

Reviewers:	Samia Ainouz	INSA Rouen, France
	Jorge Villagra	Centre for Automation and Robotics, Spain
Examiners:	Romuald Aufrère	Univ. Clermont Auvergne, France
	Véronique Berge Cherfaoui	Univ. de Technologie de Compiègne, France
Industrial Advisor:	Alexandre Armand	Renault S.A., France
Supervisors:	Philippe Bonnifait	Univ. de Technologie de Compiègne, France
	Philippe Xu	Univ. de Technologie de Compiègne, France

Remerciements

Ces travaux ont été menés au laboratoire Heudiasyc de l'Université de Technologie de Compiègne. Cette thèse s'est déroulée dans le cadre du laboratoire commun SivaLab avec le groupe Renault. La plateforme véhicules autonomes du laboratoire Heudiasyc fut un réel atout, de nombreuses expérimentations ont pu être conduites et ont permis d'apporter une contribution significative.

Je tiens à remercier toutes les personnes qui m'ont permis de mener ces travaux de thèse à bien durant ces trois années.

Je souhaite tout d'abord remercier chaleureusement mes directeurs de thèse Philippe Xu et Philippe Bonnifait, pour leur expérience, leur implication et leur dévouement tout au long de ces trois ans. Vous m'avez permis de réaliser des travaux de recherche sur un sujet passionnant, et ce dans les meilleures conditions d'encadrement possibles.

Je remercie également Alexandre Armand de m'avoir accompagné dans mes travaux et d'avoir apporté son expertise industrielle acquise auprès du Groupe Renault. Tu m'as donné l'opportunité de pouvoir me plonger sur ce sujet des véhicules autonomes au coeur de la recherche en entreprise.

Je tiens également à remercier les membres de mon comité de thèse. Merci aux rapporteurs Jorge Villagra et Samia Ainouz pour leurs commentaires et l'intérêt qu'ils ont porté à mes travaux. Merci à Romuald Aufrère et Véronique Cherfaoui.

Merci à Lounis Adouane et Romuald qui ont suivi l'avancement de mes travaux et dont les discussions ont été constructives.

J'adresse également mes remerciements aux autres membres du laboratoire Heudiasyc qui ont accompagné mon quotidien, m'ont accueilli, soutenu et apporté leur aide. Merci à Alexis, Antoine, Joelle, Stéphane et Thierry pour leur implication.

Enfin, je tiens à remercier sans restriction mes proches et ma famille. Merci à mes supporters, mes grands parents, à mes frères et soeur, aux membres rapprochés de ma garde personnelle et bien sûr à mes parents qui ont contribué à cette réussite.

Résumé

Pour conduire dans des environnements urbains complexes, les véhicules autonomes doivent comprendre leur environnement de conduite. Cette tâche, également connue sous le nom de "connaissance de la situation", repose sur une représentation virtuelle interne du monde faite par le véhicule, appelée "World Model". Cette représentation est généralement construite à partir d'informations fournies par de multiples sources. Les cartes de navigation haute définition fournissent des informations préalables telles que la topologie du réseau routier, la description géométrique de la route et des informations sémantiques, en incluant le code de la route. Le système de perception fournit une description de l'espace et des usagers de la route évoluant dans l'environnement du véhicule. Conjointement, ils fournissent des représentations de l'environnement (statique et dynamique) et permettent de modéliser les interactions. Dans des situations complexes, un "World Model" fiable et non trompeur est nécessaire pour éviter des prises de décision inappropriées et assurer la sécurité. L'objectif de cette thèse est de proposer un nouveau formalisme sur le concept de "World Model" qui répond aux exigences de "connaissance de la situation" pour un véhicule autonome. Ce "World Model" intègre des connaissances préalables sur la topologie du réseau routier, une représentation basée grille au niveau des voies, sa prédiction dans le temps et surtout un mécanisme de contrôle et de surveillance de l'intégrité des informations.

Le concept de "World Model" est présent dans de nombreuses architectures de véhicules autonomes, mais il peut prendre des formes très diverses et parfois seulement implicites. Dans certains travaux, il fait partie du processus de perception alors que dans d'autres, il fait partie d'un processus de décision. La première contribution de cette thèse est une étude sur le concept de "World Model" pour la conduite autonome couvrant différents niveaux d'abstraction pour la représentation de l'information et le raisonnement.

Ensuite, une nouvelle représentation est proposée pour le "World Model" au niveau tactique combinant des objets dynamiques et des informations d'occupation spatiale. Tout d'abord, une approche descendante basée sur les graphes utilisant une carte haute définition est proposée pour extraire les zones d'intérêt par rapport à la situation du point de vue du véhicule. Elle est ensuite utilisée pour construire une "Lane Grid Map" (LGM), qui est une représentation intermédiaire de l'état de l'espace du point de vue du ego véhicule. Cette approche descendante est choisie pour évaluer et caractériser les informations pertinentes de la situation. En plus des états occupés et libres classiques, nous caractérisons davantage l'état inconnu par les notions de zones dites neutralisées et sûres qui fournissent un niveau plus profond de compréhension de la situation.

Une autre contribution au "World Model" est un mécanisme de gestion de l'intégrité qui est construit sur la représentation Lane Grid Map. Il consiste à gérer l'échantil-

lonnage spatial des cellules de la grille afin de prendre en compte les erreurs de localisation et de perception et d'éviter les informations trompeuses. Indépendamment de la confiance accordée aux informations de localisation et de perception, la LGM doit être capable de fournir des informations fiables au module de prise de décisions afin de ne pas prendre de décisions dangereuses.

La dernière partie de la stratégie de "connaissance de la situation" est la prédiction du "World Model" basée sur la représentation LGM. La principale contribution est de montrer comment une prédiction classique au niveau des objets s'adapte à cette représentation et que l'intégrité peut également être étendue au stade de la prédiction. Il est également décrit comment une zone neutralisée peut être utilisée dans l'étape de prédiction pour fournir une meilleure prédiction de la situation.

Le travail s'appuie sur des données expérimentales afin de démontrer une application réelle d'une représentation complexe de la "connaissance de la situation". L'approche est évaluée avec des données réelles obtenues grâce à plusieurs véhicules expérimentaux équipés de capteurs LiDAR et IMU avec corrections RTK dans la ville de Compiègne. Une carte haute définition a également été utilisée dans le cadre du laboratoire commun SIVALab entre Renault et Heudiasyc CNRS-UTC. Le module de "World Model" a été implémenté (avec le logiciel ROS) afin de répondre à une application temps réel et est fonctionnel sur les véhicules expérimentaux à des fins de démonstrations.

Abstract

To drive in complex urban environments, autonomous vehicles need to understand their driving context. This task, also known as the situation awareness, relies on an internal virtual representation of the world made by the vehicle, called world model. This representation is generally built from information provided by multiple sources. High definition navigation maps supply prior information such as road network topology, geometric description of the carriageway, and semantic information including traffic laws. The perception system provides a description of the space and of road users evolving in the vehicle surroundings. Conjointly, they provide representations of the environment (static and dynamic) and allow to model interactions. In complex situations, a reliable and non-misleading world model is mandatory to avoid inappropriate decision-making and to ensure safety. The goal of this PhD thesis is to propose a novel formalism on the concept of world model that fulfills the situation awareness requirements for an autonomous vehicle. This world model integrates prior knowledge on the road network topology, a lane-level grid representation, its prediction over time and more importantly a mechanism to control and monitor the integrity of information.

The concept of world model is present in many autonomous vehicle architectures but may take many various forms and sometimes only implicitly. In some work, it is part of the perception process when in some other it is part of a decision-making process. The first contribution of this thesis is a survey on the concept of world model for autonomous driving covering different levels of abstraction for information representation and reasoning.

Then, a novel representation is proposed for the world model at the tactical level combining dynamic objects and spatial occupancy information. First, a graph based top-down approach using a high-definition map is proposed to extract the areas of interests with respect to the situation from the vehicle's perspective. It is then used to build a Lane Grid Map (LGM), which is an intermediate space state representation from the ego-vehicle point of view. A top-down approach is chosen to assess and characterize the relevant information of the situation. Additionally to classical free-occupied states, the unknown state is further characterized by the notions of neutralized and safe areas that provide a deeper level of understanding of the situation.

Another contribution to the world model is an integrity management mechanism that is built upon the LGM representation. It consists in managing the spatial sampling of the grid cells in order to take into account localization and perception errors and to avoid misleading information. Regardless of the confidence on localization and perception information, the LGM is capable of providing reliable information to decision making in order not to take hazardous decisions.

The last part of the situation awareness strategy is the prediction of the world model based on the LGM representation. The main contribution is to show how a classical object-level prediction fits this representation and that the integrity can also be extended at the prediction stage. It is also depicted how a neutralized area can be used in the prediction stage to provide a better situation prediction.

The work relies on experimental data in order to demonstrate a real application of a complex situation awareness representation. The approach is evaluated with real data obtained thanks to several experimental vehicles equipped with LiDAR sensors and IMU with RTK corrections in the city of Compiègne. A high-definition map has also been used in the framework of the SIVALab joint laboratory between Renault and Heudiasyc CNRS-UTC. The world model module has been implemented (with ROS software) in order to fulfill real-time application and is functional on the experimental vehicles for live demonstrations.

Table Of Contents

Nomenclature v

General Introduction 1
Introduction . . . 1
Problem Statement and Manuscript Organization . . . 2

Chapter 1. World Model for Autonomous Vehicles 5
1.1 Introduction . . . 5
1.2 Key Concepts . . . 6
1.3 Challenges and Problem Statement . . . 8
1.4 Representation of Scene Elements . . . 9
1.4.1 Static information . . . 10
1.4.2 Dynamic information . . . 13
1.4.3 Relations between entities . . . 19
1.5 Architectures and Implementations . . . 20
1.5.1 Hierarchical cognitive approach . . . 20
1.5.2 World Model module . . . 25
1.6 Conclusion . . . 33

Chapter 2. An Architecture for a WM 35
2.1 Introduction . . . 35
2.2 Functional Architecture . . . 35
2.3 World Model Module . . . 37
2.3.1 WM abstraction levels . . . 37
2.3.2 WM functions . . . 39
2.3.3 WM layers . . . 40
2.3.4 WM integrity . . . 43
2.4 Modules Pipeline . . . 44
2.4.1 HD-Map . . . 44
2.4.2 Localization . . . 45
2.4.3 Perception . . . 47
2.4.4 Decision-making module . . . 47
2.4.5 WM module . . . 48
2.5 Conclusion . . . 48

Chapter 3. A Lane Level World Model **51**
3.1 Introduction . . . 51
3.2 Related Works . . . 52
3.3 Interaction Graph . . . 53
3.3.1 Problem statement . . . 53
3.3.2 Areas of interest . . . 54
3.3.3 Formalism . . . 56
3.3.4 Applications . . . 61
3.4 LGM: a Lane Level Representation . . . 61
3.4.1 Problem statement . . . 61
3.4.2 Lane level context . . . 63
3.4.3 Efficiency and scalability advantages . . . 64
3.4.4 Occupancy characterization . . . 66
3.4.5 Integrity evaluation metrics . . . 70
3.4.6 Experiments . . . 74
3.5 Enhanced Space Characterization . . . 78
3.5.1 Tactical representation . . . 79
3.5.2 Theory of evidence frames . . . 80
3.5.3 Characterization process . . . 81
3.5.4 Common frame of discernment . . . 87
3.5.5 Application and Implementation . . . 90
3.6 WM Spatial Information Overview . . . 93
3.7 Conclusion . . . 94

Chapter 4. Set-based LGM Prediction **95**
4.1 Introduction . . . 95
4.2 Related Works . . . 96
4.2.1 Operational level . . . 97
4.2.2 Tactical level . . . 98
4.3 LGM Prediction Formalism . . . 101
4.3.1 Problem statement . . . 101
4.3.2 Notations for prediction . . . 102
4.3.3 Reachability . . . 103
4.3.4 Occupancy . . . 107
4.3.5 Neutralization . . . 109
4.3.6 Integrity . . . 113
4.4 Experiments . . . 116
4.4.1 Case study . . . 116
4.4.2 Setup . . . 116
4.4.3 Results . . . 117
4.5 Conclusion . . . 133

General Conclusion **135**
Synthesis 135
Perspectives 137

Appendix A. Experimental Setup **145**

Appendix B. LGM Generation **151**

Appendix C. LGM Distance Metric **153**

Appendix D. Frames of Discernment Combination **155**

Appendix E. Prediction of a Neutralized Situation **159**

List of Figures **161**

List of Tables **165**

Bibliography **165**

Nomenclature

Acronyms

AOI	Area Of Interest
CA	Constant Acceleration model
CD	Constant Deceleration model
CV	Constant Velocity model
FOV	Field Of View
OL	Operational Level
SL	Strategic Level
TL	Tactical Level
DM	Decision Making module
KDB	Knowledge Database
LP	Localization and Perception modules
WM	World Model
NTI	Neutralized Time Interval
IG	Interaction Graph
LGM	Lane Grid Map
OG	Occupancy Grid
TIR	Target Integrity Risk

Symbols

$V_i; E_i$	Road user or entity
$LGM\ (t)$	Lane Grid Map at a given time t
c	LGM cell
M	HD map
l	Link of the map M
$s(t) = [\underline{s}(t); \bar{s}(t)]$	Curvilinear abscissa interval
$v(t)$	Speed
$a(t)$	Acceleration
L_{Ch}	Lane changing mode
L_{Cr}	Lane crossing mode
L_K	Lane keeping mode
L_{Me}	Lane merging mode
AOI	Area of interest
F	Free
H	Hidden
N	Neutralized
O	Occupied
S	Safe
U	Unknown
V	In field of view

General Introduction

Introduction 1
Problem Statement and Manuscript Organization 2

Introduction

Nowadays, the intelligent vehicles domain faces many challenges. There is a strong emulation that involves many actors such as car manufacturers, service companies, multimedia firms, research universities. Several works have carried out on the legislation and about the impact of such technology[1] that this new mobility would bring in road safety. It could provide more efficiency than human drivers and be less subject to faults. It should be known that human mistakes cause nearly 90 % of road accidents[2]. 83 % of accidents at intersection areas are in urban areas and 19 % of road fatalities happen at intersections. A complete situation understanding is necessary to drive in such hazardous environments. In addition to that, 10 % of fatal accidents are refusal of priority. An autonomous intelligent vehicle has to take into account and respect traffic laws but should also be capable of avoiding hazard coming from other users or hidden areas. Thus, modeling interactions is essential.

There are five levels of automation[3] going from Driver Only to Full Automation. Levels 1&2 encompass the known functions as Advanced Driver Assistant Systems (ADAS). The navigation tasks such as keeping lane, speed or suddenly break in case of emergency are basic functionalities that still require the monitoring of the driving task and the driving environment by the driver. In order to reach higher level of automation, the monitoring of the driving situation has to be performed by the autonomous vehicle.

Level 3 autonomous driving systems are able to handle very structured and clear situations like highways. A vehicle at level 4 or above should be capable of handling

[1] www.nhtsa.gov/technology-innovation/automated-vehicles-safety [Online]

[2] www.securite-routiere.gouv.fr/la-securite-routiere-en-chiffres-lobservatoire-national-interministeriel [Online]

[3] www.oica.net/wp-content/uploads//WP-29-162-20-OICA-automated-driving.pdf [Online]

every type of situation encompassed in its operational design domain and does not need the driver's intervention to get in safety. The vehicle has to be capable of driving among other vehicles, obeying the traffic rules and respect the interactions with the other road users and with the infrastructure. Several difficulties lie in respecting these interactions. An autonomous vehicle senses its surrounding environment through its sensors and continuously takes decisions. Several processes are involved before taking the final decision. First, the autonomous vehicle must localize itself in order to drive from point A to point B. Then, the data provided by perception sensors such as LiDARs or cameras is processed by perception algorithms in order to build and supply more significant information about the surroundings. It enables the vehicle to move in its environment. In addition to that, in order to drive safely in an open environment, a vehicle has to take into account prior knowledge as traffic laws. The road layout is structured such that a topological layer enables to better understand how to interact with vehicles sharing the common space. Semantic information is also important. Based on this information, a prior knowledge database is useful in order to take proper decisions. Finally, the situation must be understood so that the decision-making can plan the maneuver, the speed profile and the trajectory to adopt.

While many efforts have been deployed in the past decades to build up core functionalities such as obstacle detection, localization, mapping, trajectory planning or control, there are still challenges and bottlenecks. Operating in urban areas remains challenging to reach a level 4 autonomous driving system. Many sophisticated methods have been proposed to solve each of these issues. Moreover, one should keep in mind that all these tasks need to be run in real time. Even though one may suggest that time processing and memory shortage is only a current technological limitation, the relatively low cognitive charge needed by a human to drive a car may suggest that a more "intelligent" formalization is needed to combine all these basic functionalities. In complex situations such as those encountered in urban areas, the road user behaviors are governed by strong interactions with the others, and with the road network. In such situations, reliable situation awarenessis therefore mandatory to avoid inappropriate decisions.

Problem Statement and Manuscript Organization

Situation awareness raises several challenges such as representation of scene elements and comprehension of the current situation. They are key capabilities for autonomous driving that are addressed from many points of view in the literature. An internal representation of the driving situation, called World Model, has to provide a situation understanding for autonomous navigation. This concept of World Model is present in many architectures even if sometimes it is implicit. This representation has to manage static and dynamic information in order to represent the

scene, the interactions between road users and to anticipate hazardous situations. The World Model aggregates all the available information for decision making. As it is the last stage of information representation, it must not provide misleading information, even in the case of missing information. The core issue of this PhD thesis is therefore on the design and study of a World Model with internal mechanisms enabling information integrity.

This manuscript is organized as follows.

Chapter 1 provides a review of existing architectures with different structural compositions and internal mechanisms that deals with an internal world representation. This chapter aims at defining key concepts that must be part of situation awareness capabilities. Representations used in the literature to tackle this issue are described. A cognitive approach that may be used to handle the capability of abstraction of information and the use of prior knowledge is addressed. Finally, the concept of a World Model (WM) module that can provide a consensus is deepened.

Based on the literature review, chapter 2 aims at presenting the architecture considered in this PhD thesis. The place of a WM in the functional architecture of an autonomous vehicle and its functionalities are further detailed. The different modules of this architecture are also described. In order to lead experiments in real conditions and to show real-time applications of the concepts presented all along this manuscript, a section presents the different modules that have been implemented in order to provide the required inputs to the WM. The framework that enables to provide the different results of this manuscript is also detailed.

Once the framework of situation awareness has been settled, the question of the information provided to the decision-making module is raised. Based on the concept of abstraction levels that has been developed, the WM must not provide misleading information. It must also have the capability of being aware of missing information in its current driving situation. An autonomous vehicle has to understand as well as possible what is happening in its environment based on the information available to it. It has to be aware of each element it will interact with and to be aware of the hazard that may happen.

In chapter 3, a novel representation is proposed for the spatial layer of the WM combining dynamic objects and spatial occupancy information. It focuses on the spatial information representation that enables to focus the situation understanding on spatial interactions. A graph based top-down approach using a high-definition map is proposed to extract the areas of interests with respect to the situation from the vehicle's perspective. This so-called Interaction Graph representation (IG) is then used to build a Lane Grid Map (LGM), which is an intermediate space state representation from the ego-vehicle point of view. This representation aims at providing a deeper level of understanding of the situation where information is missing. The notion of focus of attention, which is an action to acquire additional information, is therefore a consequence of the knowledge that something is missing in order to

have a complete situation understanding. A top-down approach is chosen to assess and characterize the relevant information that are in areas of interest. Depending on the vehicle field of view, new notions are introduced with neutralized and safe areas. This work led to a publication at the IEEE Intelligent Vehicles Symposium [Sanchez et al. – 113] and two patent applications [Sanchez et al. – 114, 115].

One of the most critical aspects of situation awareness is reasoning on information while ensuring integrity of an enhanced situation representation. Integrity of the information managed by the WM for situation understanding is a critical issue. It must not provide misleading information that could lead to a hazardous decision-making and has to handle the unavoidable uncertainties. Thus, chapter 3 also shows how the LGM representation is a useful tool to handle integrity management. A mechanism to control and monitor the integrity of information is described. This work led to a publication at the IEEE Intelligent Vehicles Symposium [Sanchez et al. – 116].

Situation awareness does not only consist of the comprehension of a situation. It requires much more advanced understanding levels that enable reasoning such as the prediction of a situation.

Thanks to the enhanced situation understanding provided by the LGM representation through the neutralized constraint, chapter 4 shows a trustworthy application that enables to improve the decision-making process in such situation thanks to a Neutralized Time Interval indicator. A set-based prediction analysis is provided with the LGM representation. Another main contribution of this chapter is to show that the integrity can also be extended at the prediction stage. This work led to a submission at the 2022 IEEE Intelligent Vehicles Symposium.

The work and results presented in this PhD thesis rely on experiments that demonstrate a real application of a complex situation awareness representation. The approach has been evaluated with real data obtained thanks to several experimental vehicles (see figure 0.1) and a high-definition map of the city of Compiègne. Additionally to the WM module, many other related packages have also been implemented and are used within the SIVALab laboratory.

Figure 0.1: Experimental vehicles used within the SIVALab laboratory.

Chapter 1

World Model for Autonomous Vehicles

1.1 Introduction 5
1.2 Key Concepts 6
1.3 Challenges and Problem Statement 8
1.4 Representation of Scene Elements 9
1.5 Architectures and Implementations 20
1.6 Conclusion 33

1.1 Introduction

In robotic applications, decision-making processes are involved as soon as a mission to be achieved must be decomposed into several sub-tasks. Each of them may be decomposed into elementary tasks that are easily manageable to be converted into output signals to actuators. These processes differ accordingly to the level of difficulty of a given mission, e.g. navigate in a complex environment, and therefore may require different levels of understanding of a situation.

The granularity and accuracy of the acquired information relies heavily on perception and localization systems. Despite the level of difficulty of the goal to be achieved and the perception system that enables to supply observations on the environment, a key point is the ability to build an internal representation of the surroundings of the robot. It aims to understand the situation, reason and achieve an objective.

The representation of this information as well as the comprehension and reasoning capacities, depend on the level of requirement of the task to be accomplished. Driving in a closed and constrained environment with limited variation in environmental conditions, e.g. lighting condition, differs from driving in an open area with

different weather conditions and multiple interactions. In advanced robotics applications, e.g. autonomous driving applications, a cognitive approach inspired from human behaviors helps to understand mechanisms that are involved.

This chapter aims at providing a review of the literature of world modeling representations and architectures dealing with these problems, mainly in the robotic field and from a system point of view. As mentioned, the problem is addressed from a situation awareness point of view. Existing approaches more specifically related to autonomous driving systems are explored in order to extract several concepts that address the capabilities of situation awareness. All the figures in this chapter are illustrations that have been taken from the literature.

1.2 Key Concepts

Generally speaking, and from a cognitive point of view, the problem addressed is the situation awareness, a capability that a cognitive process from sensing to acting must have. [Albus – 3] first describes a point of view of intelligence mechanisms bases where a World Model (WM) representation stands between sensory processing and behavior generation. It corresponds to the internal "estimate of the state of the world". In [Endsley – 37], the author tackles the situation awareness theory in complex systems. As depicted in figure 1.1, three pillars of a situation awareness model are highlighted and may be applied to the field of autonomous vehicles:

1. Perception Of Elements In Current Situation: This corresponds to elements of the scene provided by the localization and perception modules, which constitute the basis of the information that the system has processed and extracted from an ego-vehicle point of view. One or several representations of information are possible.

2. Comprehension Of Current Situation: This corresponds to the situation understanding of the vehicle surroundings. At this step, the internal world model must provide tools that enable reasoning mechanisms in order to build additional knowledge.

3. Projection Of Future Status: This last step is about prediction. It represents the highest level of understanding. Once the system has grasped the ins and outs of a situation, it must be able to anticipate its evolution. Several scenarios may be inferred from a single situation. This is a hard task to perform, even for human drivers, as reasoning is different from one individual to another. This is what differentiates a predictive behavior from a reactive one.

Given these concepts, a module such as the WM is therefore part of a situational awareness concept. It aims at providing the basis of an internal scene representation,

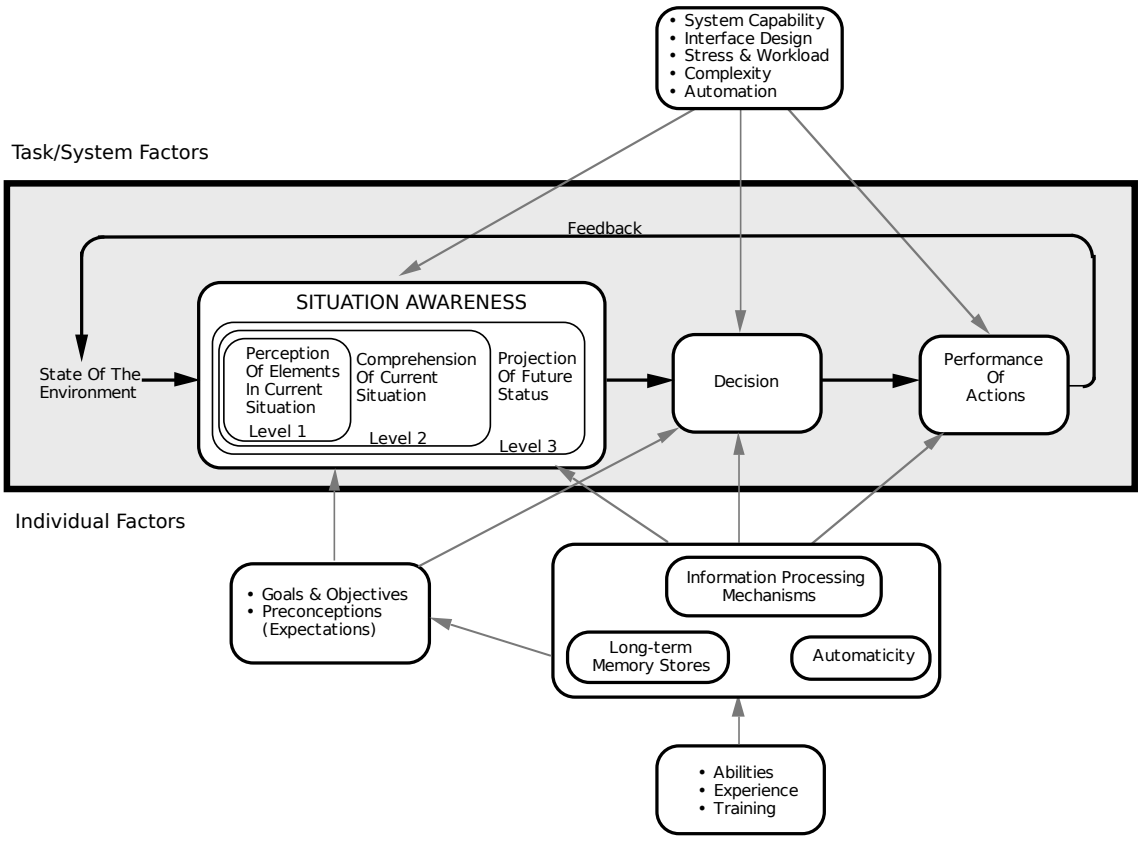


Figure 1.1: Situation Awareness model for decision-making from [Endsley – 37].

which, from an egocentric perspective, should allow reasoning on a given situation and may lead to many scenarios analysis.

Although the terms scene,situation and scenario are used for describing situation awareness, even in the specialized literature, it is common to find different definitions of these terms [Elrofai et al. – 36, Geyer et al. – 43]. Therefore, it is of importance to clarify their meaning and to define their scope. [Ulbrich et al. – 138] propose a definition of these terms that can be considered as a reference applied to the intelligent vehicle field. Thereafter, each of this term is related to situation awareness.

Scene representation For an intelligent vehicle, a scene consists of a collection of cohabiting road entities, as it is perceived by the stakeholders, including the static and dynamic road entities but also spatial occupancy of the space. The scene also includes the relationships among elements, i.e. the relations that may exist between them and that may have an impact on their behavior and situations. More concretely, the compilation of the perception data and of the prior information stored in a digital map can be considered as a manner to represent driving scenes. The relationship between entities is complex to sense, but such information can be provided by digital maps. However, as for a human being, it is not possible to be

aware of all elements that are part of a scene. Information is partial but does not prevent from taking decisions in order to achieve a goal. Especially for autonomous vehicles, it is necessary to distinguish the missing information from the necessary information among the scene elements, which leads to the definition of a situation.

Situation understanding A situation is deducted from a scene considering additional information like the behaviors of road users, intentions, decisions, etc. Moreover, only a selection of surrounding entities have an impact (direct or indirect) on the situation of a vehicle. Situation understanding relies on the capability to identify which perceived entities are the most relevant, i.e. must be considered by decision-making, by associating all perception and map data as well as likely interactions between entities. In the end, situation understanding consists of giving sense to scenes. This is a prerequisite for the elaboration of a projection of this situation into the future. The objective is to be able to anticipate by considering one or more scenarios.

Scenario prediction A scenario consists of a sequence of situations that are the consequence of actions or events applied to elements of a primary situation depending on their goals or evolution models. For an intelligent vehicle, the prediction of the likely scenarios that may happen in the future, i.e. anticipation, is necessary to ensure safe decisions and also comfortable behaviors.

Each step of situation awareness enhances the comprehension of the vehicle surroundings in order to provide as much information as possible to the decision-making. In the literature, the decision-making (DM) module may refer to many notions such as local planning, navigation (or global planner) or even maneuver planner. The term decision-making has the meaning of encompassing any type of these decisions, which are in fact hierarchical decisions.

1.3 Challenges and Problem Statement

There are many challenges that need to be tackled in order to fulfill situation awareness requirements in autonomous driving field. Scene description, situation understanding and scenario prediction describe well the tasks that should be performed by an intelligent vehicle before taking decisions, given the information from a perception system and prior information from a digital map. Situation awareness involves several mechanisms that may differ from one architecture to another. This raises several conception issues.

First, information representation needs to be addressed. In the literature, information representation depends on the task that needs to be fulfilled. In robotics

systems, there are many environment representations from static to dynamic representation including continuous or discrete representations. Thus, information representations as a whole and ensuring an overall consistency is mandatory in a world model module.

Such a representation must provide to decision-making, metrics that are required for decision processes. Information must be as complete as possible in order to be aware of all the elements of a situation that are compulsory for its understanding. As a consequence, the question of missing information arises. The autonomous system has to be aware of what information it may acquire and what information is truly available.

Uncertainties also need to be taken into account. Depending on the situation, they may lead to a misunderstanding of the driving context. Misleading information is a real issue as it may lead to hazardous decisions and thus situations.

Then, the understanding of the driving situation is important. Assuming that a perception system is able to provide relationships between entities, interactions still need to be inferred. The road layout provides the spatial relations between entities. However, to understand their interactions, one need a higher level of understanding using for instance traffic laws and traffic signs.

There is a need to reason over multiple types of information (spatial occupancy, road users, digital map information, traffic laws, etc.). What is relevant and necessary among this information needs to be identified and sorted. Indeed, missing information in a driving scene does not have the same meaning or importance depending on the situation that is considered. As a consequence, it is of importance to consider consistency and integrity of the inferred information. Reasoning must not lead to provide misleading information.

Finally, prediction of a situation must be tackled. Anticipating what is likely to happen in an immediate/middle/long term horizon is the highest capability of a system that is capable of a good situation understanding.

1.4 Representation of Scene Elements

In general, intelligent vehicles have at their disposal information that represents scene elements. As explained before, a scene is defined by a collection of static and dynamic entities, as well as the relations that may exist between them.

The representation of these types of information can take plenty of forms, depending on the techniques used to generate them, i.e. in general from a perception system, and those to exploit them, i.e. in general a decision-making system. The literature shows that certain data representation techniques allow representing a single type of information, while other allow to represent several ones. In [Badue et al. – 11],

information representation is surveyed according to sensor point of view and more specifically with respect to the perception or the localization module.

Furthermore, the level of detail can be highly different between different techniques, leading to different levels of abstraction [Ilievski et al. – 59]. This section aims to present a brief review of the main techniques of scene data representation available in the literature. They will be classified with respect to data sources and type of information represented, with comments about the level of abstraction.

1.4.1 Static information

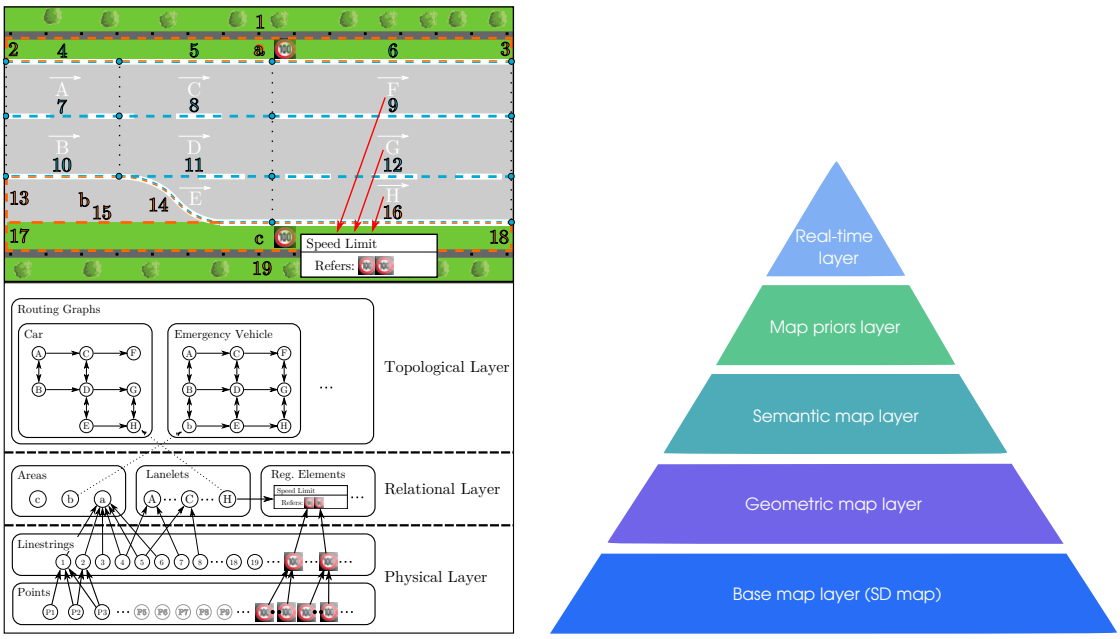
All road scenes contain features related to the road network. The advantage of these features is that they are very long to change, and except after road works or construction of new roads, they rarely change. They can therefore be easily stored in a database, and used as prior information by intelligent vehicles. The road network offers rich information about the context and greatly contribute to the situation understanding process. Whilst this section does not aim to be dedicated to digital map technologies, the following paragraph presents a summary of the literature dealing with this subject, and that seems relevant for the problem of scene representation.

Standard Definition Maps

The first function of digital maps within vehicles is for path planning and guidance of human-driven vehicles, through personal navigation devices. Such maps are topological. They describe the road network as a graph, including basic semantic information like the street names or road signs. Most of the time, such maps are not precise, and describe the road network at road level. This means that there is no direct representation of each navigation lanes. These maps are usually called Standard Definition (SD) maps. The topological layer enables to provide traffic laws information and to infer relationship between road entities for autonomous driving. However, in autonomous driving applications, enriched maps with finer representation, such as for lane geometry for instance, are widely used.

High Definition Maps

Today, in the domain of autonomous driving, maps are widely used in the objective to achieve level 4 and level 5 systems. Moreover, these maps have to offer a very detailed representation of the road network for handling complex maneuvers. In that way, most of autonomous driving actors are using High Definition (HD) maps: BMW [Aeberhard et al. – 1], Lyft (Toyota, with GeneralMotors and NuTonomy



(a) Lanelet2 map illustration from [Poggenhans et al. – 106]

(b) Structure of Lyft HD maps from [Chellapilla – 25].

Figure 1.2: Examples of HD maps structures for autonomous driving.

partnerships) [Chellapilla – 25], Uber or Waymo[1]. They consider maps as an additional source of information which works beyond standard perception sensors field of view and which allows predicting situations more easily.

The DARPA challenge organized in 2007 allowed for the development of the route network definition format (RNDF) for a HD representation of the road network [DARPA – 30]. This format specifies road segments describing navigation lanes, the position of road features such as the position of road signs and lane width. In that sense, these maps are topological maps with a limited number of additional attributes aiming at providing context information to autonomous vehicles. Several other map formats exist such as open street map (OSM) [Haklay and Weber – 49], openDrive[2] for simulation or even commercial map providers with HERE[3] or TomTom[4]. They can be decomposed in multiple ways as in [Bagschik et al. – 12, Ulbrich et al. – 137], although it is mostly the content of the map that is of importance, rather than the format properly said.

[Poggenhans et al. – 106] proposes the Lanelet2 framework including a new format for HD maps and associated tools, intended for autonomous vehicles purposes. As

[1] www.caranddriver.com/news/a29417731/self-driving-cars-city-mapping [Online]

[2] www.asam.net/standards/detail/opendrive [Online]

[3] www.here.com/platform/HD-live-map [Online]

[4] www.tomtom.com/products/hd-map [Online]

shown in figure 1.2a, three layers compose a Lanelet map: the topological layer that represents the road network, the relational layer that represents geometries like lanes and areas as well as the relations which may exist between them, and finally the physical layer that includes road features like the lines, road edges, points, etc. This representation is interesting because it allows representing the road with different abstraction levels under a graph based representation. Furthermore, it helps to understand relations between lanes, and therefore may facilitate reasoning about scenes, given traffic law rules that are included within the map.

Whilst the main actors of autonomous driving technologies do not communicate much about the HD maps they are using, Lyft published several articles online on this subject [Chellapilla – 25, Efland and Rapp – 33]. As shown in figure 1.2b, five layers compose the Lyft maps. The foundations of these maps are based on an SD map as described above. The lanes are described in the geometric layers, while the semantic layer includes features like traffic lights, road signs, or lane markings. This layer also includes information about the relationships, which define how lanes work together. The fourth layer is the map prior layer and contains information that is approximate and used as hint for the vehicle navigation. For instance, this layer may include information about zones where vehicles often park. It can be used by the vehicle to more rapidly take the decision to overtake parked vehicles. Finally, the top most layer concerns real time knowledge. This layer is regularly updated and contains real time information such as the traffic state or observed speeds.

Localization maps

Maps may also be used for localization purposes. A geo-referenced map is often used in absolute localization. Two sensors are mainly used for map-based localization: LiDARs and cameras.

Provided a high density of points given by a LiDAR, localization can be done with a stored point cloud map. The environment point cloud has to be stored in order to enable the map matching between sensor data and the stored data. In [Dubé et al. – 32], the point cloud representing the entire map is built from a dataset and segmented. Then a real-time point cloud is used for localization with the map. However, using point clouds is time consuming and costly in terms of memory.

A commonly used approach is localization with a feature map [Burki et al. – 24, Ghallabi et al. – 44]. Visual features processed from sensors are matched with features already stored in a map. A feature map encompasses all kinds of features such as landmarks, key points, road signs or road markings. Such representations are very well suited to represent lane markings and road edges. Plenty of solutions exist to fit models to the lane shapes such as straight lines, polylines, polynomials or splines as described in [Bar Hillel et al. – 15].

1.4.2 Dynamic information

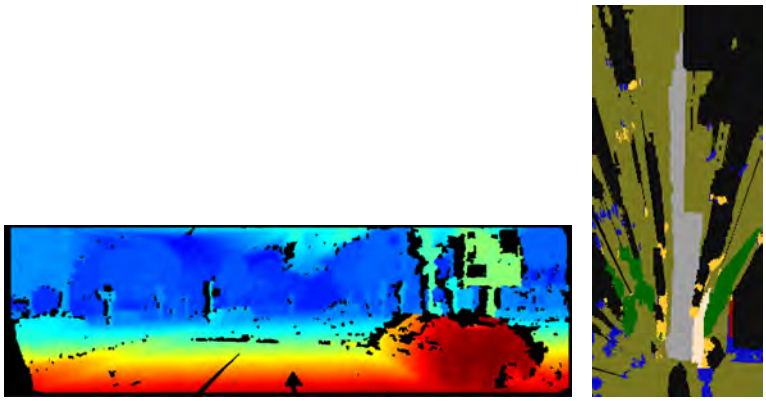
The sensors embedded on the vehicle (including V2X) allow perceiving entities evolving in the surroundings. This provides rich and dynamic information. The literature shows that several methods exist to represent the surrounding environment of autonomous vehicles. [Schreier – 121] surveyed different existing methods, highlighting several characteristics such as continuity, number of dimensions and compactness. This section provides a summary of these approaches, as well as a few others, with distinction between approaches offering parametric representations, and approaches offering discretized representations.

1.4.2.1 Feature based Representations

Objects maps The most compact representations of environments are usually achieved by parametric representations. Among the few parametric approaches that may exist in the literature, feature maps (also called object maps) probably represent the most intuitive approach. They represent environments through collections of geometric points and shapes that represent objects, lines, areas, etc. The local dynamic map presented in [Papp et al. – 102] represents such features as well as dynamic objects. For the representation of dynamic entities such as road users, sets of scene elements are often used [Schmidt et al. – 118]. A situation is defined by the combination of sets of scene elements put into relation with mapping functions. From a geometry point of view, scene elements are usually represented by parallelepipeds or polygons to approximate their shape and size, with associated attributes such as state, type, including uncertainties (figure 1.4a).

Whilst such representations are very compact and easy to manipulate, their main limitation is that they work under the closed world assumption as defined in formal systems of logic used for knowledge representation [Reiter – 110]. It means that this approach assumes that detected features exist and implies that the complementary space, i.e. where there is no description of a feature or object at a given place, is empty, i.e. free. This is done independently from the perception sensors set capabilities as its field of view or hidden zones for instance. This assumption is very strong and is a problem for safety-critical systems. The Closed World Assumption is opposite to the Open World Assumption. Indeed, in an Open World Assumption, an area with no detected object inside does not imply that this area is free from any object. This assumption is more robust to missed detection. Whilst these notions are initially used for knowledge representation, they are well appropriate to describe the feature map limitation.

Semantic overlays Vision systems with several cameras embedded in the vehicle are widely used. They are light-sensitive sensors, giving 2D images in the front plan. Used in stereo, they also allow to get relative distances and depth information which



(a) (b)

Figure 1.3: Examples of semantic overlay representations. (a) Disparity map shown in [Xu et al. – 148]. (b) Semantic overlay used with an occupancy grid from [Erkent et al. – 39].

can be represented in a front 2D disparity map (figure 1.3a). In the treatment of camera data, a great number of works have been done in segmentation area [Cheng et al. – 26]. Segmentation of images can enhance occupancy grid properties adding a semantic overlay [Erkent et al. – 39] if depth is known (figure 1.3b). LiDARs data can also be used with camera images giving a more precise depth [Zhu et al. – 154]. The point cloud can be labeled thanks to segmentation with images also [Mei et al. – 87]. Cameras have a large field of view and enable to recognize objects, road markings, etc. In the literature, it is common to associate scene-understanding problems with image segmentation [Xu et al. – 148]. However, these topics are rather related to perception processes than situation understanding as described in this chapter.

1.4.2.2 Continuous spatial information representations

The representation of spatial information is complementary to feature based representations as it models surroundings regardless of the delimitation of the scene elements.

Free space maps In order to overcome the limitations of object maps, free space maps have become a common solution. This approach does not give a representation of the environment through the presence of objects (static or dynamic), but on the contrary it represents the environment through the description of spaces where objects are absent. The free space is a 2D description (bird view) of the environment. One of the main advantages of this approach is its ability to represent the non-

presence of objects in a straightforward manner; it allows safe navigation is these zones.

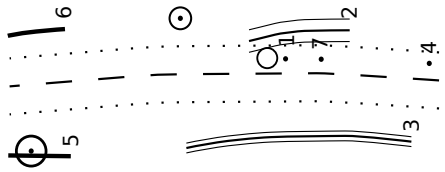
Many ADAS and AD smart camera providers (Mobileye[5] as in figure 1.4b, Horizontal Robotics, Stradvision, etc.) propose systems which return the description of the free space and/or navigable spaces. These systems usually estimate the free space after road segmentation, and return it in the form of polygons. Reliability of such information can be improved with multi-sensor fusion [Xu et al. – 148]. In [Mayr et al. – 86], an approach for generating training data used to label the drivable area with a self-supervised learning is proposed.

Parametric free space maps were recently proposed in the literature [Schreier et al. – 122]. These spaces are represented through the combination of "parametric curve and geometric primitives" (figure 1.4c), and encompass the free drivable space. Uncertainties on the system inputs can be considered, and result into probability distributions associated to the parameters. The advantage of this approach is the compactness of the generated data. Furthermore, it represents the boundaries with obstacles, but also boundaries with unknown environments (see orange boundaries). However, this approach, as it does not consider objects in a direct manner, makes prediction operations difficult. Moreover, the level of abstraction is too low to represent objects as well as interactions. As the same authors explain it, this approach is well suited to represent static parts of the environments, but it needs to be coupled with another approach, such as object maps for dynamic object representation.

Potential fields maps Another continuous approach to represent road environments is potential fields [Moreau et al. – 93]. In the domain of robotics, potential fields are used for representing the robot surroundings and to facilitate path planning. Contrary to feature maps and free space maps, potential field maps lose all notions of objects, features and areas. There is no semantic included within this sort of representation which is fully numerical. Objects and constrains applied to the vehicle are represented as attractive or, most of the time as repulsive forces. Whilst the initial versions of potential fields aimed at modeling the occupancy of perceived static objects or features, more recent versions allow to consider also dynamic objects by considering their speed components to represent risks [Pierson et al. – 104] (see figure 1.4d).

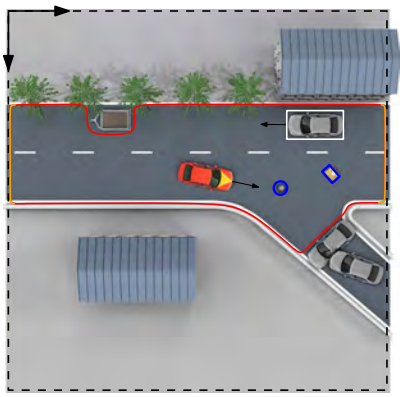
However, this representation totally disregards objects and semantics. This makes it difficult to perform prediction operations based on a potential field map. Nevertheless, similarly to free space maps, it could probably be coupled with object and feature maps to take the best of each approach.

[5] www.mobileye.com/our-technology [Online]

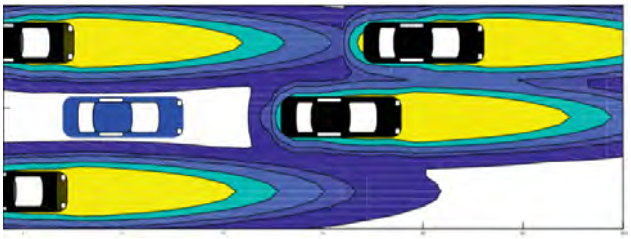


(a) Feature map from [Schreier – 121].

(b) Semantic free-space from Mobileye technology.



(c) Parametric free-space from [Schreier et al. – 122].



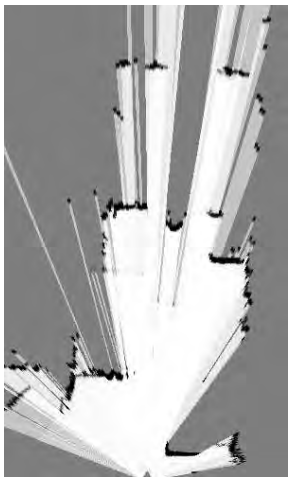
(d) Risk potential field from [Pierson et al. – 104].

Figure 1.4: Continuous environment representations.

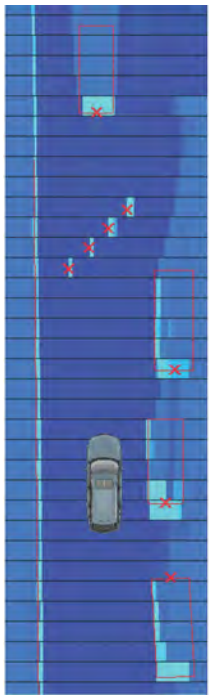
1.4.2.3 Discretized spatial information representations

Whereas continuous and parametric representations offer the advantage of compactness, they are not always the most convenient representations as inputs for decision-making and navigation systems. Furthermore, for parametric representations, it is complex to find models that perfectly fit to the road features. It may result into imprecise or erroneous representations. Discretized representations allow getting around this problem.

Occupancy grid maps Occupancy grid maps are probably the most popular discretized presentation of environments as they are used for localization, planning, tracking and navigation [Heide et al. – 51, Mouhagir et al. – 94]. This applies to many sensors such as LiDARs [Laugier – 73, Yang et al. – 149], sonars [Martin et al. – 84, Ribo and Pinz – 111] or radars [Slutsky and Dobkin – 128, Werber et al. – 145]. It consists in breaking down the vehicle surroundings into a finite number of cells (figure 1.5a). The range of the cells varies depending on the required precision, but also on the resolution of the sensors that are used to generate the grid. Grids

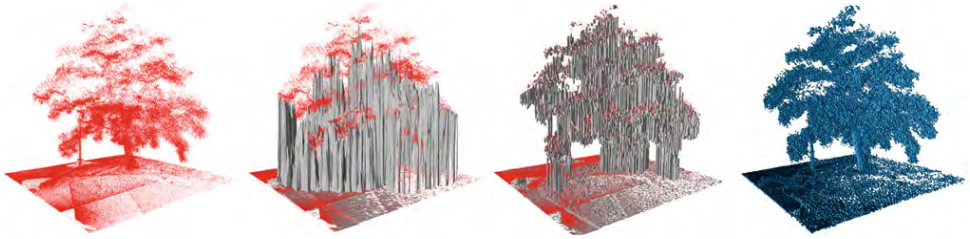


(a) Occupancy grid from [Negre et al. – 97].



(b) Interval map from [Weiherer et al. – 144].

(c) Stixel representation from [Pfeiffer and Franke – 103].



(d) 2.5 and 3D representations. From left to right: raw point cloud, elevation map, multi-level-surface map, voxel map from [Hornung et al. – 55].

Figure 1.5: Discretized environment representations

may be modeled with a Cartesian representation, quad-trees, or also with a polar representation as it is presented in [Badino et al. – 9].

In general, the state of the cells is defined by a distribution over the values it can take. Most of the time, it is either free or occupied [Elfes – 35]. For automotive applications, cells can be characterized as drivable and not drivable as it is proposed in [Michalke et al. – 89]. Probabilistic frameworks are the most popular to estimate the state of the cells considering uncertainties, however evidential frameworks offer the advantage to make the difference between unknown information and conflicting information as shown in [Moras et al. – 92]. This allows adding the "unknown" label to the list of possible states, which make the perception system more efficient in dynamic environments.

The first versions of occupancy grids assumed that the environment is static, which makes it difficult to be used in dynamic environments. Extensions allow the consideration of dynamic environments, like Bayesian Occupancy Filter (BOF) and its extension Hybrid-Sampling Bayesian Occupancy Filter (HSBOF) which allows to estimate the velocity of each cell through probability distributions [Laugier – 73, Negre et al. – 97]. Furthermore, other approaches propose to perform tracking within grids through particle filters [Danescu et al. – 29] or neural networks [Piewak et al. – 105]. Prior knowledge from maps can be used to improve the estimation of cell transitions, and thus to enhance long-term prediction of objects motion [Gindele et al. – 46]. Moreover, object maps can be fused with grids to make the most of the tracking, which is more convenient to be performed at the object level [Bouzouraa and Hofmann – 21].

Alternative grid maps Despite occupancy grids being widely used, several grid representations can also be found. A few alternative grid representations with different dimensions are proposed:

- Interval maps are 2D grid map which describes the vehicle environment only in the along-track direction [Weiherer et al. – 143] as shown in figure 1.5b. In following work [Weiherer et al. – 144], interval maps were extended to interval occupancy maps to describe free, occupied and unknown areas.
- A risk metric may be considered instead of a simple occupancy in grids with lambda fields [Laconte et al. – 71]. Physical constraints are used in order to compute a risk based on the kinetic energy of obstacles. As a consequence, collision avoidance is performed in a more efficient maneuver as it does not only consider occupancy but also both physical parameters of the ego-vehicle and the obstacle such as masses and velocities.
- Elevation maps are a 2.5D representation of the environment (figure 1.5d). It has been extensively used for ground robotics purposes, and then applied for intelligent vehicle-related research [Malartre et al. – 82, Nguyen et al. – 98].

- Multi Level Surface maps represent an extension of elevation maps where each cell of the grid does not store height information, but stores a list of surface patches (figure 1.5d). Each surface patch in a cell reflects the possibility of traversing the 3D environment according to the height [Triebel et al. – 136].

- Voxel grids provide a 3D representation of environments and allow to model occupancy with free and occupied states of cubic cells in the 3D space (figure 1.5d) [Schmid et al. – 117, Seo and Chou – 125]. Nevertheless their memory consumption is very high, which makes it difficult to be used in real time.

- An original approach is used with a stixel world that offers a lateral discretization of the environment on the vertical (camera) plan (figure 1.5c) [Badino et al. – 10, Pfeiffer and Franke – 103, Schneider et al. – 119].

Occupancy grids are very popular as they allow to consider all types of sensors, and to represent navigable spaces very explicitly. However, their implementation request high memory and bandwidth capacities.

1.4.3 Relations between entities

Another important source of information to model is the relations between road users and the autonomous vehicle. In a simplified way, two main categories can be mentioned: two scene elements may be related with geospatial or semantic relationships. Ontologies are a way for representing semantic interactions. For instance, the local dynamic map presented in [Papp et al. – 102] has been enriched with ontologies for this purpose in [Eiter et al. – 34]. Work has been done on this field in order to have reasoning about a given situation [Armand et al. – 6, Bagschik et al. – 12, Kohlhaas et al. – 64, Ulbrich et al. – 137].

Generally speaking, a commonly used strategy is to rely on geospatial relationship. The topological layer of a digital map is used to infer the potential relations between entities that are related to a spatially structured environment [Kuhnt et al. – 69, Schulz et al. – 124]. This work focuses on interaction-aware predictions based on intentions and expectations of road users. Dynamic Bayesian networks are used to achieve their goal. The situation can be interpreted as a graph. It should be noticed that interactions are considered between the scene participants from a list of objects. It can be noticed that interactions related to areas of the scene are not addressed.

Once information representation is addressed, situation understanding exploiting this information has to be deepened. The next section describes a literature review of concepts that are part of system awareness and how this problem has been tackled.

1.5 Architectures and Implementations

As already mentioned, situation awareness may involve many abstract concepts. Various works show several ways to achieve this goal. Through multiple architectures, two global concepts are highlighted.

First, architectures can be analyzed from a hierarchical cognitive perspective. Through the analogy with the human being, several mechanisms are presented. The second issue is whether these mechanisms must be part of the perception side, part of the decision-making side or part of a World Model, as no consensus exists.

1.5.1 Hierarchical cognitive approach

1.5.1.1 Levels of abstraction

Given the fact that an autonomous vehicle has to drive without the intervention of the driver, the vehicle has to behave more or less as a human. Some cognitive approaches aimed at modeling the human brain mechanisms and transposed them to a robot.

J. Rasmussen makes a distinction between skill-based behavior, rule-based behavior and knowledge-based behavior as depicted in figure 1.6 from [Rasmussen – 107]:

- Skill-based behavior is related to notions of reflexes or reaction to a stimulus. It is assimilated to the human ability to react to an action or stimuli.
- Rule-based behavior shows that rules are used in order to take decisions. It can be assimilated to information recognition and association to knowledge in a database where all rules that have to be applied are stored.
- Knowledge-based behavior can be compared to the strategy that has to be adopted given a known or new situation encountered. It aims at building additional knowledge in the knowledge database.

Those three grades enable to fit human behavior with the behavior a robot should have. In the case of an autonomous vehicle, human actions have to be replaced by an intelligent robot that has to fulfill the same tasks.

Other abstraction levels denominations are used in several published works dealing with autonomous vehicle architectures. The navigation, guidance and stabilization terms are used by [Donges – 31]. The stabilization level is the lowest level where the local trajectory is planned. If one goes up in the abstraction levels, there is the guidance level, responsible for the behavior planning. At the top, there is the navigation level which is in charge of the mission planning in a global context. E. Donges compares the behavior levels of J. Rasmussen as depicted in figure 1.7 and shows the different levels complementarity of them.

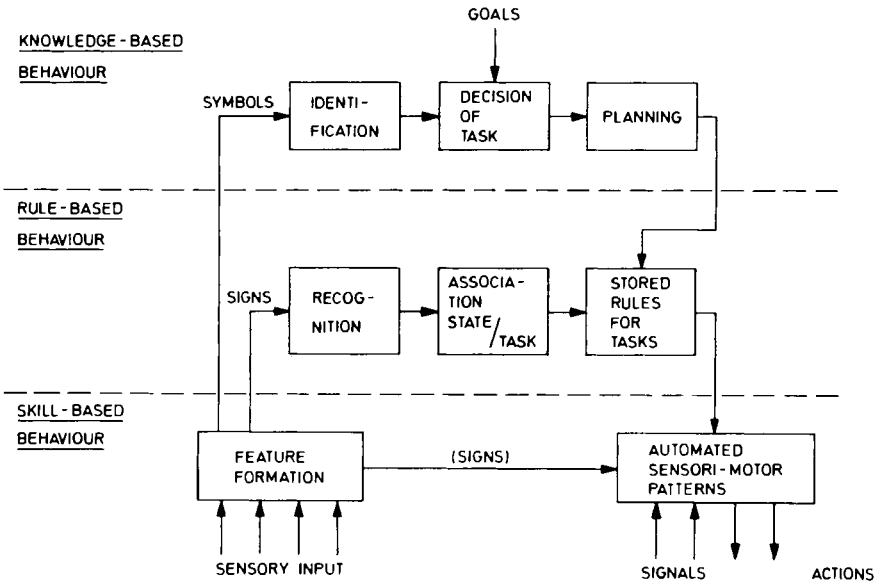


Figure 1.6: Levels of human behaviors presented by [Rasmussen – 107].

These terms of navigation, guidance and stabilization are similar to the terms presented by [Michon – 90] where the terms used are respectively: strategic at the navigation level, tactical at the maneuvering level and operational at the control level (figure 1.8).

The strategic level is used for the global planning process. In the DM module, it plans the itinerary. The tactical level is used for maneuver planning. It plans the maneuvers the vehicle has to take. The maneuver planner enables to take into account this context as it is shown in the "environment model" mentioned in [Orzechowski et al. – 101]. The operational level is dedicated to the control and to the planning of the local trajectory the vehicle has to follow.

In the following, the terms used to distinguish the levels of abstraction are the strategic level, the tactical level and the operational level. It highlights a hierarchical decomposition of internal mechanisms and of the world representation of situation awareness. The higher is the level, the more abstraction is gained. At the lowest level, the local planner focuses on the trajectory to plan within a short horizon of time and distance. At the intermediate level, the task gains in abstraction as the objective is to plan maneuvers. At the highest level, the mission planer focuses on a larger horizon of time and distance with coarser information needs.

These levels show a decomposition in at list two structural blocks that are part of an autonomous vehicle: the perception system (linked to inputs, i.e. sensors) and the decision-making system (linked to output, i.e. actuators). Several mechanisms enable to navigate from an abstraction level to another and are present in multiple autonomous vehicles.

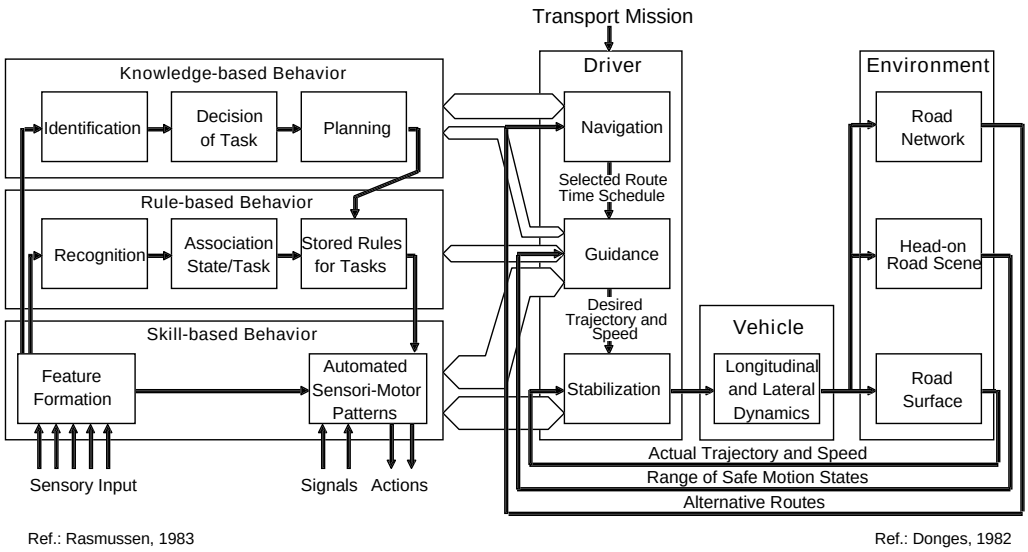


Figure 1.7: Architecture of the levels of abstraction presented by [Donges – 31].

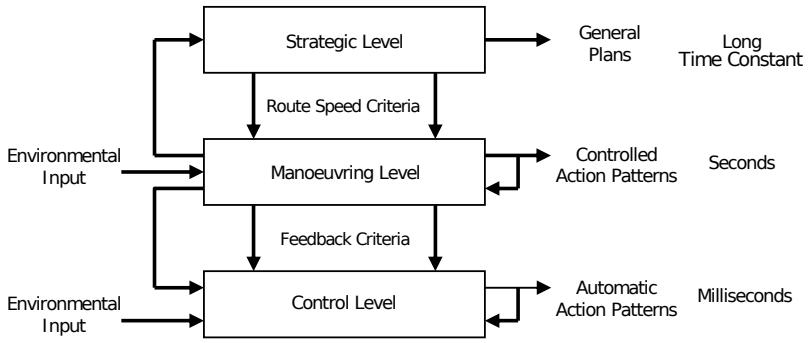


Figure 1.8: Architecture of the levels of abstraction presented by [Michon – 90].

1.5.1.2 Top-down and bottom-up mechanisms

Two mechanisms of a situation awareness capability are highlighted in order to go from a scene point of view to a situation understanding. First, the selection process enables to take into account a goal-oriented context in order to select relevant information related to the driving task. Then, the augmentation process enables to reason on information selection of the situation. Additional information such as risk analysis or safety metrics, with inter-distances for instance, enable to enhance the situation understanding. As shown with [Refaat et al. – 108], a convolutional neural network can be used to get an agent prioritization, which corresponds to the selection process. It orders road users from most important to least important.

The hierarchy in levels are reused in several works [Czarnecki – 28, Ulbrich et al. – 137]. A more detailed view of an intelligent autonomous vehicle architecture is presented in [Ulbrich et al. – 139] that has been formerly introduced with several articles [Matthaei and Maurer – 85, Ulbrich et al. – 137, 138]. The whole functional

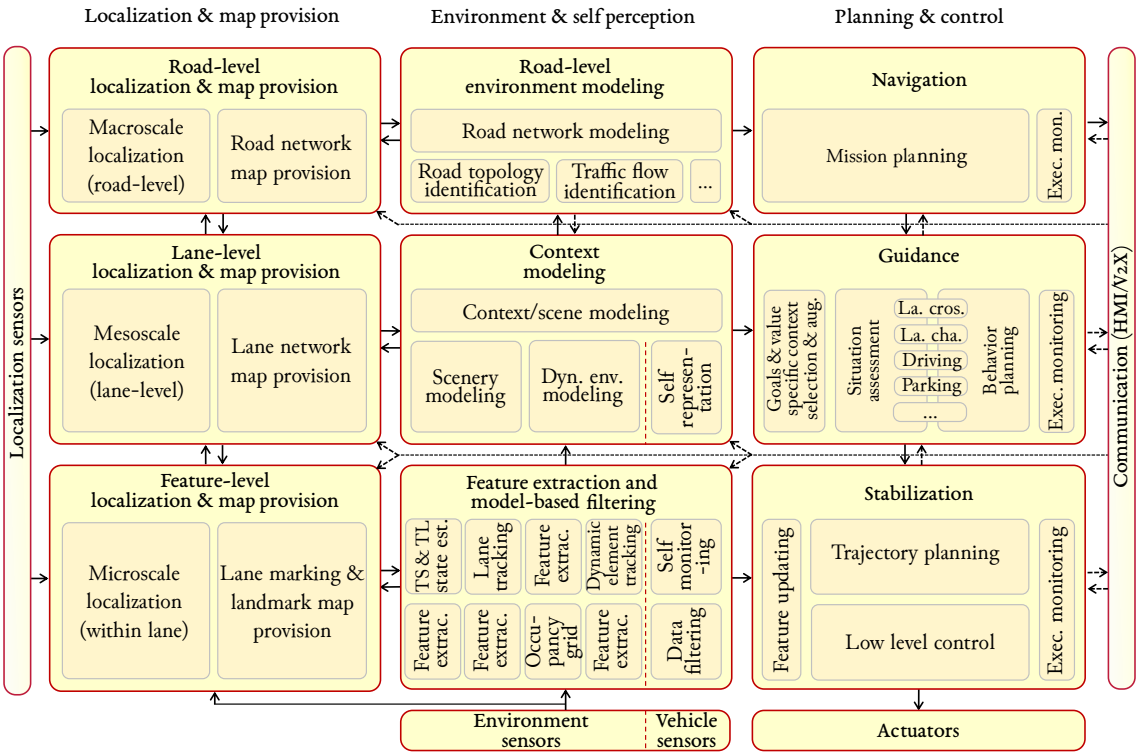


Figure 1.9: Functional system architecture for an intelligent autonomous vehicle presented in [Ulbrich et al. – 139].

architecture is presented in detail and illustrated with figure 1.9. The selection process and augmentation process to get a situation understanding happen in the guidance level of the "Planning & control" module. A few open issues are mentioned and solutions can be explored.

As it can be noticed, there is no feedback from the "Planning & control" module to the "Context modeling" and from this "Context modeling" to the "Feature extraction and model-based filtering". Depending on the situation context, a focalization process in order to improve perception mechanisms could be interesting.

In addition to that, it is mentioned that the "context modeling" brick of the "Environment & self perception" module only outputs a scene whereas the situation understanding happens in a different module, in the "guidance" block. Furthermore, the place of a prediction process is not clearly defined between these modules despite the authors mention that it could stand in different modules depending on each module needs.

[Stiller et al. – 130] use the terms knowledge, rule and skill and apply them to the automotive field. In [Ulbrich et al. – 139] (see figure 1.10), the authors show that skill-based, rule-based and knowledge-based behaviors are notions that are orthogonal to the three levels described above ("navigation", "guidance" and "stabilization").

		Processing level		
		Skill-based	Rule-based	Knowledge-based
Driving task	Navigation	Daily commute	Choice between familiar routes	Navigating in foreign town
	Guidance	Negotiating familiar junctions	Passing other car	Controlling a skid on icy roads
	Stabilization	Road following around corners	Driving an unfamiliar car	Learner on first lesson

Figure 1.10: Correspondence between the levels of abstraction explained in [Ulbrich et al. – 139].

Human mimicry has been used for a long time. J. Albus presented in [Albus – 3] an analogy between the building of intelligence by a human being and an architecture for autonomous robots who would have the same mechanisms of cognition. He detailed the human way to gain information, process it, analyze it, take decisions and act. He established a complete architecture that could be synthesized in a complete functional architecture (deepened in the next section).

Following the idea of human like capabilities, [Aufrère – 7] uses the human visual capabilities in order to extract concepts that can be applied to mobile robots, such as attention mechanisms. In this work, based on [Fernandez – 40], two cognitive concepts are used: exogenous mechanisms and endogenous mechanisms.

The exogenous mechanism is data driven based, in the sense that it gathers all types of information, filters and processes them so that at the end the brain is aware of what it has to take into account in order to take decisions.

The endogenous mechanism is at the opposite: it is model-based. The goal and the information needed in order to achieve this goal are known. The information that has to be looked for is prior information. That is why only this information will be searched and the brain will try to match its mental representation of this information to what its eyes really perceive.

These two mechanisms will be retained as they show well the bottom-up and top-down approaches that can be used in the autonomous vehicle field. Especially nowadays, the artificial intelligence field which tries to build up an end-to-end learning scheme where the input would only be the data and the output would be directly the actions on the actuators with no use of any prior information. The top-down approach is complementary and uses prior information such as the road layout or the traffic laws. It enables to be less sensitive to noises and improves performance. As prior information is known, some feedback can be made to the perception that could focus on specific tasks or areas.

1.5.2 World Model module

Work has been carried out with the aim of having a cognitive approach to the robotic field. In several works, the concept of a World Model (WM) has been introduced and conjugates some of the notions shown previously with the autonomous driving field.

1.5.2.1 Computer vision representations

With the scale-space theory approach addressed in computer vision [Tony – 134], several principles are similar and can be assimilated to the ones presented by Albus [Albus – 3]. While processing a signal, i.e. an image, it can be represented with a multi-scale representation. At first, the signal is stored as it is in its original state. Then, as the time goes, only the processed information is kept. An image made of pixels can be abstracted into several type of information, e.g. into geometric segmentation, where the fine-scale information is thus suppressed. This mechanism enable to simplify the data. Thus, the information has several levels of representation. As one goes up in the levels, the information will be coarser and more simplified. The same concept is present in [Latecki – 72], where four main categories of representations of an image can be made.

As surveyed in [Janai et al. – 61], there are many approaches to abstract information from images. Early, the grand objective of computer vision systems is to provide a complete semantic and behavioral interpretation of an input image by reasoning about a 2D or 3D representation. The authors address the notion of scene understanding in this field. It aims at considering physical constraints, temporal behaviors and relationship between traffic participants. However, these tasks were progressively oriented toward machine learning algorithms in order to overcome challenges such as expressiveness needed to autonomous driving tasks or a unified evaluation of scene understanding approaches.

The methods used to abstract information are also surveyed in [Janai et al. – 61]. They may be object oriented with objects detection and objects tracking. These methods aims at defining and localizing traffic participants in the road scene. They may provide a more or less abstracted information representation depending on the granularity of the information. Semantic segmentation gives semantic information on areas of the scene, e.g. it can separate all the pixels belonging to persons from the road. Semantic instance segmentation enables to further discriminate individual persons in the image. The panoptic segmentation combines both representations [Kirillov et al. – 63].

1.5.2.2 End to end representations

In end-to-end learning, the objective is to take inputs of a system to transform them directly into outputs. There is no intermediate representation enabling the understanding of the system decisions. Learned behaviors with neural networks are sensitive to under or over represented situations in their training database. Consequently, there is a lack of explainability [Zablocki et al. – 151] to ensure safety in driving situations. However, there are learning approaches that take into consideration intermediate representation such as Variational Auto-Encoder (VAE) [Ha and Schmidhuber – 47]. It is hard to provide information integrity on such a representation as it is not directly interpretable. Recent works propose to add intermediate interpretable representation in an end-to-end approach [Zeng et al. – 152]. Objects state with their prediction, i.e. a trajectory and a velocity profile, are provided to the motion planner.

1.5.2.3 Autonomous vehicles

Early in [Albus – 2], J. Albus presented a reference model architecture for intelligent unmanned ground vehicles. Previously presented in [Albus – 3], the architecture had the purpose to formalize the way the human brain performs. This architecture comprises four modules (figure 1.11) :

- The Sensory Processing module (SP)
- The Value Judgment module (VJ)
- The World Model module (WM)
- The Behavior Generation module (BG)

The SP module takes as input sensors data and processes them. The WM stores the world representation and makes prediction. A Knowledge Database is available for the WM and stores prior information and knowledge. The VJ evaluates plans given by the WM and situations encountered. At the end, the BG module takes the decisions and sends them to the actuators. This is a brief summary of the different modules. This architecture has been reused in several papers [Balakirsky and Lacaze – 14, Seward et al. – 126].

This architecture is composed of several layers of abstractions and there are four modules at each level. As one goes up in levels, the range in space and time increases while the resolution decreases. This means that information is very precise over a short distance at a lower level with a low level of abstraction. However, prediction can be made only on a short time span and up to a short range. Whereas at the top layer, prediction can be made on a longer distance but at a coarse level (road level and not precise trajectory level).

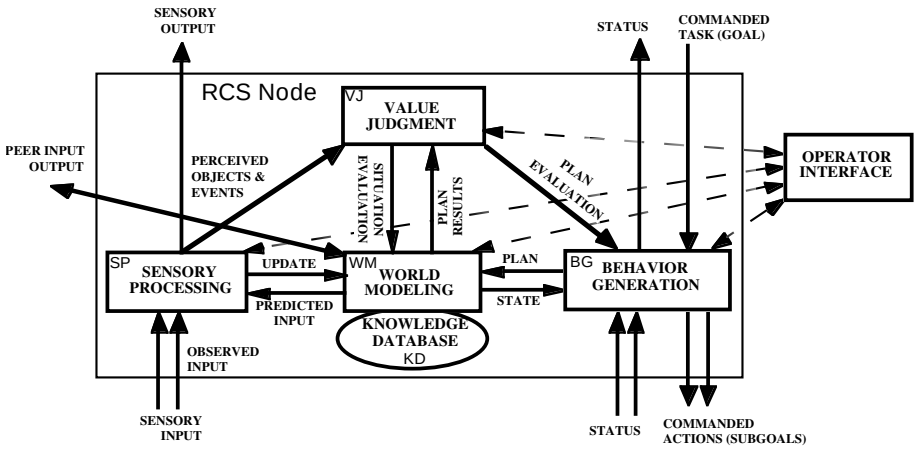


Figure 1.11: Model architecture for intelligent unmanned ground vehicles introduced by [Albus – 2].

In his journal publication [Albus – 3], J. Albus makes an analogy between the intelligent operation of a robot and the human intelligence. He builds up a whole architecture that highlights human capabilities like how the visual cortex works and how the sensed information leads to actions.

In [Ulbrich et al. – 139], the authors present in a detailed way the background that led to their proposed architecture but do not mention a WM module in their architecture. Context and scene modeling is rather part of the \ environment & self perception" whereas selection & augmentationprocesses and situation assessment are part of the \ planning & control".

In several publications, a module, which can be named differently but which represents what is called a WM is mentioned but not really described. In [Taş et al. – 133], an environment model is shown (figure 1.12). In this article, the scene understanding and the situation prediction, that could be part of the WM process, are rather situated in the understanding and planning layer (which corresponds more or less to the DM module) that enrich the environment model, showing a feedback between these modules. The environment model has the duty to provide confidence values.

Based on the functional system architecture diagram of the autonomous vehicle of ULM University [Kunz et al. – 70], a complete diagram is presented in [Tas et al. – 132]. In figure 1.13, there is an environment model module but located in the Perception & Scene Understandingmodule. Prediction and situation analysis processes are located in this module rather than in the Behaviour & Motion Planning. It can be noticed that the three main types of information representation given in this figure are:

- A digital map with ego position: this can be related to the representation of the ego-vehicle state.

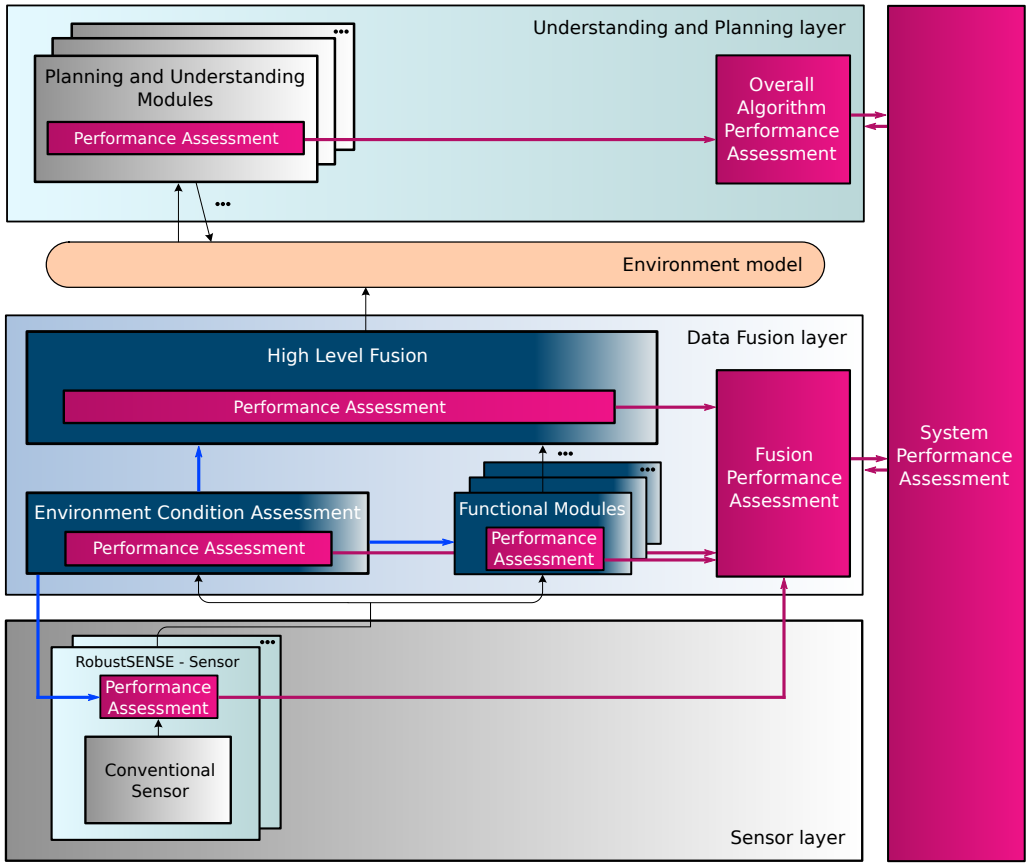


Figure 1.12: Automated vehicle system architecture presented in [Taş et al. – 133].

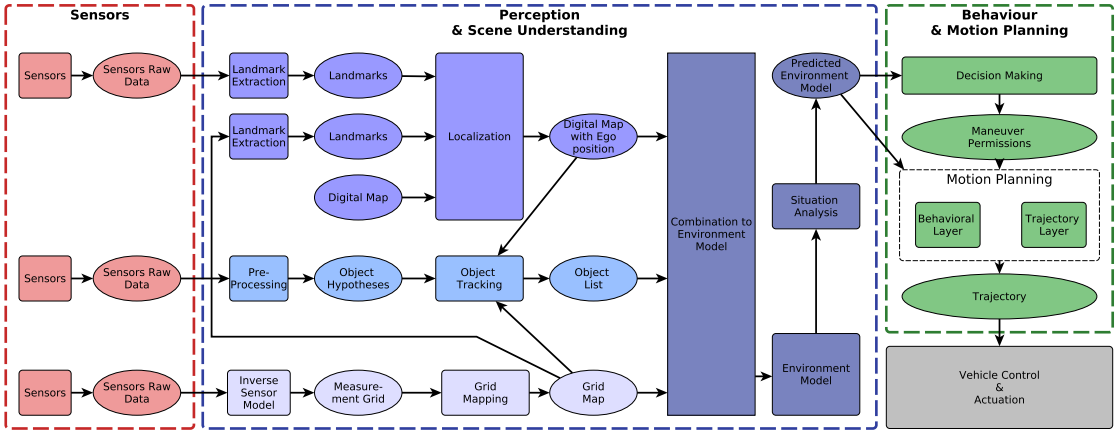


Figure 1.13: Functional system architecture diagram of the autonomous vehicle of ULM University [Kunz et al. – 70] presented in [Tas et al. – 132].

- An object list: this list is encompassed in an object map representation.
- A grid map: which could be one spatial representation of the space state.

These three information representations (described in section §1.4) are given as inputs to a situation awareness capability.

In [Furda and Vlacic – 41] (figure 1.14a), the WM purpose is to provide information input for the DM system. Several functionalities of the WM are described. It stores prior information but also perception and localization information from internal sensors but as well from V2X communication. It cyclically merges and updates this knowledge database, it infers interactions and relations and it provides to other modules its internal world representation but also notifies relevant events.

In [Regele – 109] (figure 1.14b), a description of a WM abstraction levels and their corresponding functions is described. Objects description and representation differ according to the abstraction level required.

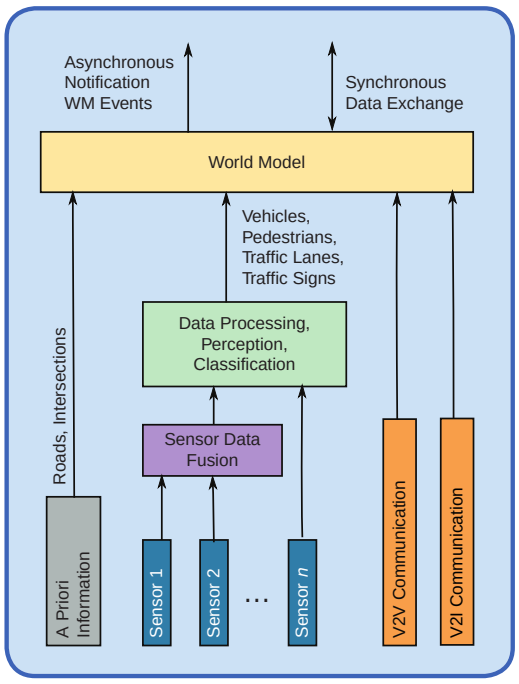
In [Papp et al. – 102], the local dynamic map corresponds to the WM representation they show in their architecture. It consists of a generic object database representation that contains entities of the world constructed with map data, sensory and communication inputs.

In [Gheţa et al. – 45] (figure 1.15), basic concepts of a WM representation are presented. It shows how the WM must use the prior knowledge database. The bottom-up process is illustrated with information acquisition from data that led to learning and increase the knowledge database. The top-down process is illustrated by prior knowledge that is used for situation understanding and that may lead to focus and request for specific information at low level.

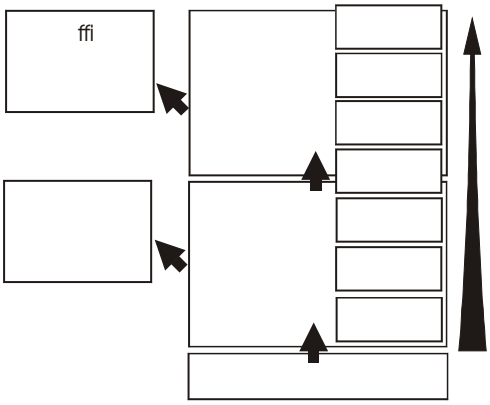
A WM can also provide risk evaluation and safety analysis [Shalev-Shwartz et al. – 127]. Based on the situation representation, decisions taken by an autonomous vehicle must fulfill safety evaluation requirements. It is of importance to have an intermediate representation as it allows formal verification. In order to monitor the behavior of an autonomous vehicle and determine the potential dysfunctions, the understanding of a driving situation has to be interpretable for a human being through its representation. This is one of the issues encountered by current state of the art of end-to-end approaches.

In many works as in [Trentin et al. – 135], the motion prediction module is separated from both the perception and decision-making ones. An additional representation is built from a map and perception data to compute interactions and intentions of road users. From a world model point of view, expectations are built with the top-down approach where intentions are inferred from a bottom-up approach.

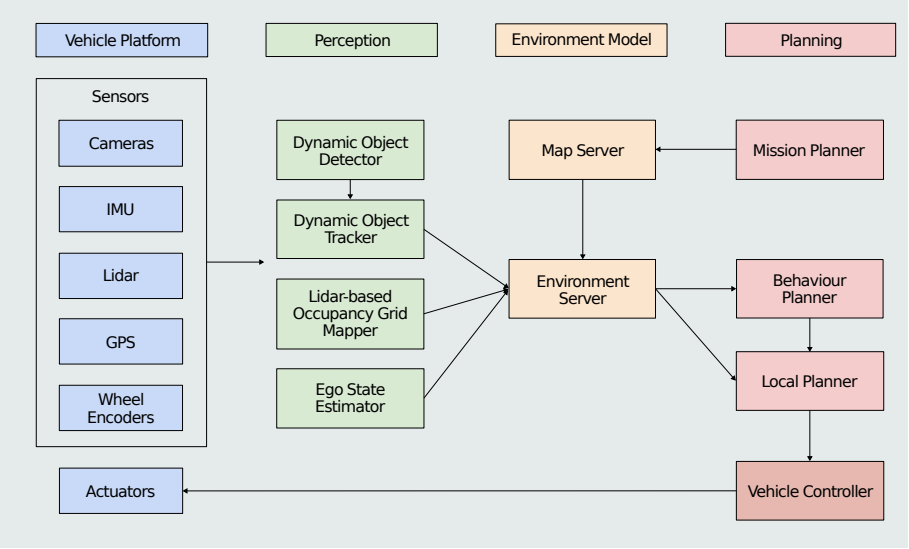
There are architectures where the WM is just conceptually described [Coutaz et al. – 27, Waslander – 142] (in figure 1.14c and figure 1.14d) and others that describe



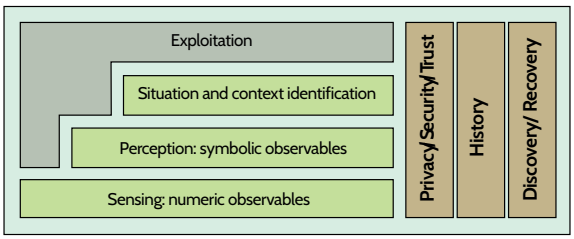
(a) Autonomous vehicle architecture presented in [Furda and Vlacic – 41].



(b) World model structure presented in [Regele – 109].



(c) Baseline autonomous driving architecture from Autonomoose project in [Waslander – 142].



(d) Levels of abstraction for context modeling from [Coutaz et al. – 27].

Figure 1.14: Different WM modules in several architectures.

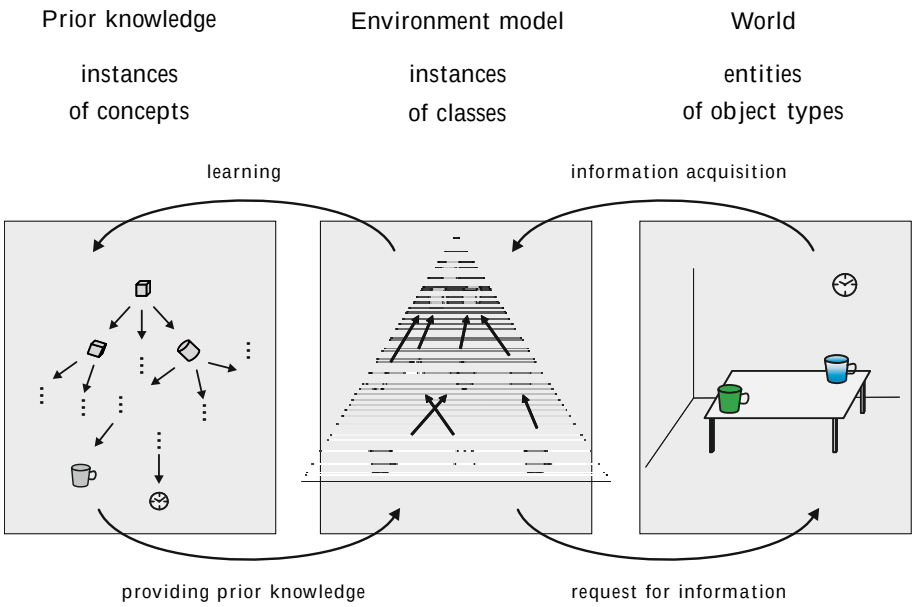


Figure 1.15: The three pillar information of a world model presented in [Gheţa et al. – 45].

abstraction levels from metric representation at low level to symbolic one at high level as in [Lefèvre et al. – 75, Menzel et al. – 88].

In [Behere and Törngren – 16], a WM is described with general concepts. This module may be passive, i.e. a simple information database, or reactive, i.e. reasoning over data is possible like the prediction process. However, it does not clearly present a settled definition. In figure 1.16, the WM is apart from the other modules whereas in the text description it may be part of the decision and control component. In addition to that, an additional semantic understanding bloc that is part of a perception component is represented. A focalization process, also mentioned as a feedback, is allowed from the semantic understanding block the sensing one.

In table 1.1, different elements and concepts addressed in the different architectures previously presented are compared. It summarizes briefly the notions that are mainly presented in the corresponding articles. It appears that there is no consensus about the presence of a WM module. The functionalities that are part of a WM appear in different modules depending on the architecture. For instance, scene understanding or prediction can be found in the DM module or in the perception module. A WM that encompasses in one module all the functionalities for decision-making and autonomous navigation is a key issue.

Table 1.1: Comparison of notions addressed in several architectures for autonomous systems. If there is not a detailed description of the notion then it is not taken into account. P=perception. DM=decision-making. L=localization. env.=environment

Architectures \ Notions	*World Model*	*Abstraction levels*	*World or env. representation*	*Scene or situation understanding*	*Prediction*	*Knowledge database*	*Vehicle architecture*
[Gheţa et al. – 45]	x	x	x	x		x	
[Taş et al. – 133] (from KIT)	x			x (DM)	x (DM)		x
[Tas et al. – 132] (inspired from ULM)			x (P)	x (P)	x (P)		x
[Ulbrich et al. – 139] (from TUB)		x	x (P)	x (DM)		x (L)	x
[Furda and Vlacic – 41] (from Griffith University)	x		x			x	x
[Regele – 109] (from Stuttgart University)	x	x	x			x	
[Papp et al. – 102] SAFESPOT (platform)	x		x			x	x
[Behere and Törngren – 16] (from KTH University)	x					x	x

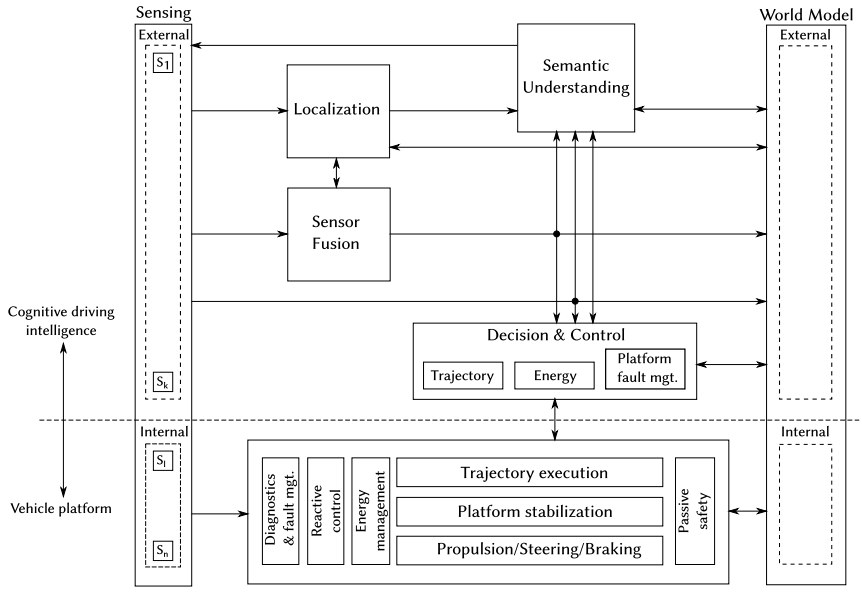


Figure 1.16: Functional architecture proposed by [Behere and Törngren – 16].

1.6 Conclusion

Situation understanding is a key capability in autonomous driving applications. In order to study how this problem has been addressed in the literature, a review of existing architectures with different structural compositions and internal mechanisms has been provided.

Based on the state of the art that has been proposed, it appears that a World Model (WM) is an efficient means to manage safety in autonomous navigation because it is a model-based approach. Moreover, it is a relevant way to deal with the explicability of decisions even if this notion is not addressed in this PhD thesis.

A WM is located between the perception and localization modules and the decision-making one. It fulfills the situation awareness concepts: scene representation, situation understanding, and prediction. Therefore, several representations to manage the information of the autonomous vehicle are necessary.

For the road environment, prior information is mainly provided by HD maps with geometrical, topological and semantic information. Regarding dynamic information, two main categories are distinguished. An approach based on objects representing the road users is not sufficient to model the vehicle surroundings and the agents that are in interaction. Spatial information is complementary whether it is continuous as a parametric free-space or discrete as occupancy grids. The dynamic road entities, which interact with the ego vehicle, and the space state information have to be managed simultaneously in real time and has to handle the unavoidable uncertainties.

Additionally, semantic information and graph based representations brings a more complete scene description. Indeed, it is of importance to model the relationships between road entities in order to take into account the interactions in the current driving situation.

Important concepts that should be part of a WM have been highlighted. Several abstraction levels are used in information representation. In this chapter, three abstraction levels are considered from the highest to the lowest: Strategic Level(SL), Tactical Level (TL), and Operational Level (OL). Information in a WM module is represented at each level of abstraction. This gives rise to mechanisms such as bottom-up and top-down processes that enable to go from one abstraction level to another. The bottom-up process enables to infer knowledge where the top-down one enhances processes such as selection with attention mechanisms (or focalization) and prediction. Situation understanding is also characterized by a selection process that selects relevant dynamic objects for the ego-vehicle point of view. Finally, prediction is the most complex task to achieve to complete situation awareness. All these mechanisms illustrate how reasoning can be managed within a WM.

In addition to that, the propagation of uncertainties coming from perception information needs to be handled properly. A WM also has to ensure information consistency and provide non-misleading information to the DM module to ensure safety. Therefore, integrity of the information managed by the WM for situation understanding is a critical issue.

A WM that fulfills these requirements is proposed in the following of this manuscript.

Chapter 2

An Architecture for a World Model

2.1 Introduction 35
2.2 Functional Architecture 35
2.3 World Model Module 37
2.4 Modules Pipeline 44
2.5 Conclusion 48

2.1 Introduction

Based on the literature review discussed in chapter 1, this chapter aims at presenting the architecture considered in this PhD thesis. Based on classical modules found in the literature, the place of the World Model (WM) in such an architecture and its functionalities are detailed. Finally, in order to lead experiments in real conditions, the last section presents the different modules that have been implemented in order to provide the required inputs to the WM. The framework that enables to provide the different results of this manuscript is also detailed.

2.2 Functional Architecture

The functional architecture proposed in this PhD thesis is depicted in figure 2.1. The modules of a standard architecture are represented along with the WM module. Each arrow indicates an input/output information. Focusing on the WM, red arrows model the inputs (dashed) and outputs (continuous) to the external modules.

In order to navigate in its environment, the ego-vehicle needs to localize itself and to perceive its surroundings. These tasks are performed respectively by the Localization and Perception modules (LP modules). They take as inputs data from the ego-vehicle sensors such as an IMU (Inertial Measurement Unit), cameras, LiDARs,

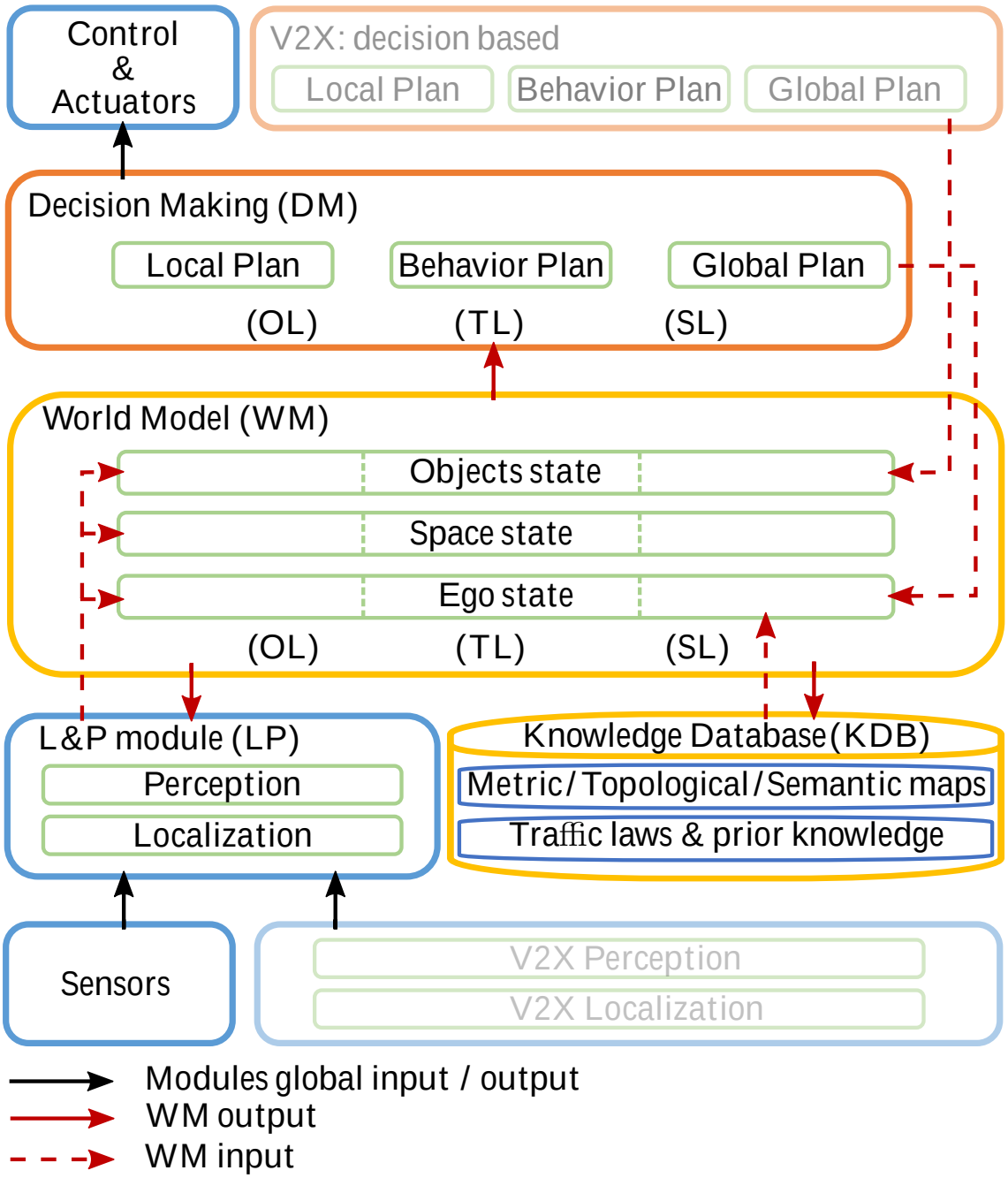


Figure 2.1: Proposed functional architecture for autonomous vehicles.

or odometers (non-exhaustive list). Processes such as data fusion, objects detection, object tracking and semantic segmentation enable to supply augmented and interpretable information representations to the WM with the description of the space and of road entities evolving in the vehicle surroundings.

The ego-vehicle is foreseen with a Decision-Making (DM) module. Three sub-modules, that are representative of the different levels of abstraction, are shown with: the local planner at the Operational level (OL), the behavior planner at the Tactical level (TL) and the global planner at the Strategic level (SL). In this architecture, the WM is independent from the DM module but the feedback of its information may improve the WM internal mechanisms and overall performance. It has a pruning effect and enhances selection processes.

For an autonomous vehicle, situation awareness [Endsley – 37] is a key capability towards safe and comfortable decision-making and navigation. The WM enables to build the internal environment representation of the autonomous vehicle. Three layers are distinguished: the ego-state layer, the space-state layer and the objects-state layer. The WM is located between the LP and the DM modules and has at its disposal a Knowledge Database (KDB). It corresponds to prior knowledge that can be stored such as maps (metric, topological, semantic) and traffic laws. The WM provides its internal representation to the DM module. In addition to that, it can update the Knowledge Database, or provide feedback to the LP module so that further processing may be applied thanks to focalization.

V2X modules are shaded as they were not studied in this thesis but are integrated in the functional architecture. It corresponds to information supply by the communication between the vehicle and another system such as a vehicle or the infrastructure. V2X information can be split into two parts. Localization and perception information are distinguished from decision-based one. Indeed, V2X information from a LP module may be fused with the ego-vehicle LP module (for example, another vehicle can share the objects it detects and transmit its global position). While in the other hand, decisions taken by road users are higher level information that may be supplied to the WM module (for example, a vehicle can share the path it wants to follow or indicate a lane change). It behaves in a similar way to the DM module.

The following section presents the WM internal representation and functionalities introduced by chapter 1.

2.3 World Model Module

2.3.1 WM abstraction levels

The WM has to model the driving situation in real time and to fulfill situation awareness capabilities, i.e. the requirements needed to solve the scene representation

and its corresponding situation understanding.

As seen previously, a key notion that is used in this architecture is the notion of levels. Each representation that is present in the WM owns a subpart at each level. In the proposed architecture, three levels are taken into account:

- Operational level (OL)

  This level is the lowest one. The OL concerns dense information updated at a high rate and is reactive to events that occur close in time and space, e.g. the ego-vehicle pose (position and heading) with its related speed. It offers a resolution of a few meters that provides a fine metric representation of the vehicle surroundings, which allows for short-term predictions.

- Tactical level (TL)

  This level is the intermediate one. It corresponds to lane level information. The prediction of a situation at this level is on a longer time horizon like maneuvers.

- Strategic Level (SL)

  This level is at the top and highlights information at its most abstracted view. This highest level with the highest abstracted knowledge is able to represent interactions. It offers a resolution of several hundred meters, and provides mainly a symbolic representation of the environment. For instance, the mission planer plans a global trajectory over a large horizon of time and space. This level has a longer view of the situation but with a coarse prediction (several seconds ahead). Information is updated at a low rate.

As going up in the levels, the information is more and more abstracted in terms of representation. Levels in the WM hierarchy are defined by temporal resolution of events and spatial resolution of maps. At each higher level, two phenomena can be observed:

- Density of information decreases about an order of magnitude at each higher level. Information is merged as the level increases. Spatial resolution and memory of information are getting coarser.

- Spatial visibility and time horizon increases about an order of magnitude at each higher level. As the level decreases, information is concentrated in a closer and closer area, which means that a focus is done on events that will occur in a shorter time span.

2.3.2 WM functions

As the World Model is at the intersection of both the LP and the DM modules, there is a double mechanism:

- Top-down approach

  The top-down approach is model-based. Prior information at the highest level is used in order to focalize the attention. For instance, by knowing the ego vehicle path, most relevant areas for the autonomous vehicle in the space state representation can be deduced. Then, this information can be propagated into the lower levels as areas that are expected to be observed. It helps to determine if information is missing and if additional focus of perception is needed. This approach enables to build expectations and to focus the attention towards relevant information.

- Bottom-up approach

  At the opposite, the bottom-up approach is data driven. Low-level information is processed in order to build additional abstracted information. The information provided by the LP module is processed in order to compute intentions of the dynamic objects, for example. Consequently, one can have at the top layer interactions and a situation understanding based on the prior information and the data.

The two approaches are in fact complementary and a well-performing system has to combine the two.

Several functions that are present in those two approaches can be highlighted:

1. Selection process

   As presented in [Ulbrich et al. – 138], a distinction has to be made between a scene and a situation. To go from the first to the second, a selection process is needed. Indeed, not all dynamic objects present in the scene are relevant for the situation in which the ego vehicle is. This is why, thanks to the context modeled at the strategic level taking into account potential interactions, most relevant objects in the most relevant areas can be selected for more computational process at lower levels like tracking. With this mechanism, only important vehicles will be tracked and less memory and time computing will be used. This enables to enhance the WM performance.

   Additionally, the focalization process is part of a feedback to the LP module. It must be mentioned that focalization may help to improve localization and perception algorithms.

2. Augmentation process

   In order to infer knowledge, the information given by the perception system is combined with the prior information stored in the geospatial database (i.e. a map) which contains traffic laws and a static representation of the world. This augmentation process enables to infer and build knowledge while processing information from a level to another.

3. Prediction process

   The WM has also a predictive role about what will happen in the future. Prediction can be done at each level of abstraction.

2.3.3 WM layers

There are three types of information to store and update in the WM. Each one is represented by a layer that owns the three levels of abstraction. As one goes up in abstraction levels, different understanding processes take place in order to build a complete situation understanding representation.

2.3.3.1 Ego vehicle state layer

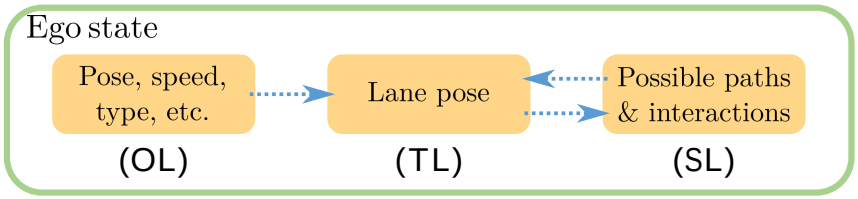


Figure 2.2: Ego-vehicle state layer representation. Each arrow shows a process that enables to enhance information representation.

The first layer to introduce is the ego vehicle state layer (figure 2.2). This layer permits to understand a scene from a single vehicle point of view. Scene elements are taken into account from the ego vehicle perspective.

- At the OL, the ego state is represented by metrical and low-level information: pose, speed, acceleration, etc. Information is updated in real time by the localization output. The layer records past poses on a short time interval and may infer possible future positions in a short time span.
- At the TL, the left arrow on figure 2.2 shows the augmentation process to get the lane pose representation. In this situation a map-matching, for instance, is applied. Thanks to the geometric and topological map, the corresponding lane pose is retrieved. It enables to infer the situation context with the actual ego vehicle position.

- At the SL, thanks to lane pose information, two mechanisms are highlighted with two arrows. From the TL to the SL, the process is an augmentation process. This process may infer through the topological and semantic map, road information as the possible paths. On this higher level of abstraction, time horizon is about several seconds up to few minutes. It models at first all the possible future paths of the ego vehicle whatever the mission planner decision is. The top-down process from the SL to the TL is a focalization process based on prior information. Indeed, possible paths are given to the lane pose representation in order to enhance the map-matching process. Instead of using the whole map, this process is improved by using a sub part of this map in order to be more efficient. Based on the current lane pose, the WM is also able to query the surrounding map information to the knowledge database on a large distance horizon.

Optionally, at this last level, this layer may take as input a decision-making feedback information as pictured on the proposed architecture. Consequently, the possible paths tree is pruned by the desired path provided by the mission planer. This mechanism enables the internal representation to be independent from any kind of decision and to be able to run over multiple possibilities of a single situation. The second advantage is to be able to be more efficient by focusing on a sub part of the situation thanks to a decision-making feedback.

2.3.3.2 Dynamic objects state layer

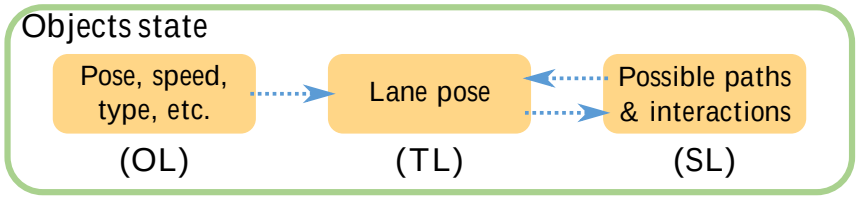


Figure 2.3: Objects state layer representation at each level of abstraction. Each arrow shows an augmentation process that enhances information representation.

This layer is similar to the ego state layer (figure 2.3). Instead of modeling a single vehicle, it represents local objects (road users). The internal mechanisms are identical to those presented in the ego state layer.

- At the OL, a list of dynamic objects is updated and provided by the perception module. It corresponds to all objects that have been seen in the vehicle surroundings.
- At the TL, a lane pose representation is associated to each object. For simplicity, objects are considered as road users in the rest of the manuscript. They are not limited to vehicles but no classification is done and used in this work.

- At the SL, possible paths representation is computed for each vehicle. It can be highlighted that, in a similar way to the ego state layer, the mission path could be provided by the mission planner of the corresponding vehicle through V2X communication, i.e. called V2X decision based in the architecture. As a consequence, the possible paths tree is pruned. At this level, interactions of road users are considered. Based on their possible paths, potential interactions between each entity of the situation can be inferred.

From the autonomous vehicle point of view, the mission path of each perceived road user is not necessarily supplied. That is why, as soon as the vehicle intention is known or detected, it can enhance the situation understanding by pruning the possible paths tree in order to better anticipate interactions.

2.3.3.3 Space state layer

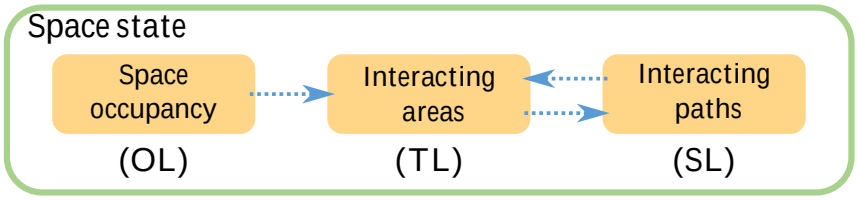


Figure 2.4: Space state layer representation at each level of abstraction. Each arrow shows an augmentation process that enable to enhance information representation.

This layer differs from the others by the fact that it handles the state of the space of interest instead of punctual information. If one is interested in the drivable space on the carriageway, this state can be for example free, occupied, hidden and even more refined as it will be presented later in the manuscript.

- At the OL, a space occupancy representation is supplied by the perception module with metrical and low-level information. As already mentioned, this representation may be multiple (polygons, occupancy grids, etc.).
- At the TL, a new representation called **Lane Grid Map** (LGM) is introduced in this work. It aims at supplying information at an intermediate level of abstraction to the DM module. This representation uses prior knowledge supplied by the upper SL representation in order to define the interacting areas the vehicle must pay attention. It corresponds to a selection process. Thanks to the lowest representation, information is merged in order to build a more compact information representation. In addition to that, the way reasoning can be applied to such a representation will be shown.
- At the SL, the interacting paths representation aims at describing potential interactions with respect to the road structure and not with potential road

users as commonly used. This representation is called **Interaction Graph** (IG). It aims at defining interacting paths on a spatial dimension in order to take into account the areas of interest. Indeed, as the possible paths of the ego vehicle are known, thanks to prior knowledge, i.e. the topological and semantic map, corresponding areas of interest can be extracted in order to focus the attention in relevant areas.

Processes that are transverse of these three main types of information enable to infer relations that exist between the ego vehicle, the space and the road users. They are also part of the selection process, which enables at the end to select which information is relevant and to go from a scene understanding to a complete situation understanding.

2.3.4 WM integrity

In order to manage the confidence of information and the risk to provide misleading information to the DM, the notion of integrity can be cited. This criterion has been developed and formalized in the aviation field. This field has been explored for urban environments as presented in the review of literature of [Zhu et al. – 155]. Integrity is an indicator of the reliability that can be placed in the provided information. This notion of integrity has also been applied for localization systems of intelligent vehicles for instance in [Hage et al. – 48]. Depending on the context, i.e. the navigation task, it reflects the ability of a system to remain in a confidence domain, which enables to decide to trust or not the information. Using an HD map, [Li et al. – 78] present integrity applied to lane-level map-matching.

In order to achieve this task, an accepted risk level must be defined by a Target Integrity Risk (TIR), which depends on the navigation context. For instance, a vehicle drives on a single lane. Its estimated position must not be misleading such as, despite uncertainty, the true position of the vehicle must be within the lane area. If not, integrity is lost and the navigation task cannot trust the vehicle pose anymore. Given a TIR, the system should not underestimate the uncertainty the system can have on its input values and manage the misleading rate to remain below the TIR.

Uncertainty of information provided to the WM has to be taken into account. Since the vehicle evolves in uncontrolled open environments, there may be strong disturbances, such as adverse weather conditions [Blin et al. – 19] that degrade its localization and perception capabilities. The WM has to be aware of the situation and to be as much confident as possible and, it has to not provide misleading information that may lead to hazardous decisions at the same time.

The WM is able to define what it is aware of and what information is missing. For instance, hazard coming from another road user due to an inappropriate behavior that the WM has analyzed is differentiated from hazard coming from uncertain or

missing information that will lead to a misleading inappropriate behavior of the ego vehicle. As a consequence, the WM is allowed to provide misleading information lower than a given TIR for the current Operational Design Domain. For example, the WM is allowed to mischaracterize a given TIR percentage of the navigation space actually occupied by dynamic objects, e.g. by indicating that the space is free instead of occupied which can be particularly hazardous.

2.4 Modules Pipeline

The functional architecture has been implemented in order to lead experiments in real conditions. It aims at providing real-time applications of this PhD work. The ROS system with its architecture of packages and nodes has been used to implement the different bricks of the autonomous vehicle software architecture. The final situation representation can be seen with the RViz plugin. Additionally to the WM module and for its needs, the following packages have also been implemented for this work. The software architecture illustration is provided in appendix A.

2.4.1 HD-Map

The knowledge database is represented by a HD map (referenced with centimeters level of accuracy) with metric, topological and semantic information. Here, the map of Compiègne is stored in a Spatialite database. It consists of an SQL database with useful features to retrieve geographic data.

The geometric layer is composed of points and polylines that represent the lane centers of the road and are called links. Useful attributes such as widths are also supplied.

The topological layer consists in providing next and previous links attributes along with an interaction table that put links in relation. However, the type of relation needs to be computationally determined, e.g. merging or crossing. It is a connected and oriented graph representation.

The semantic layer enables to provide useful information such as priorities, i.e. if there is a stop, a yield or a traffic light sign, but also speed limits.

The lanes in the map are represented by their center as a polyline with their nodes being geo-referenced with centimeter accuracy. The graph structure behind the map is used at both the topological and metrical levels. The borders of each lane are also encoded.

The map contains additional features, e.g. link borders, road signs, points of interest, useful for the localization and perception systems that are not detailed here.

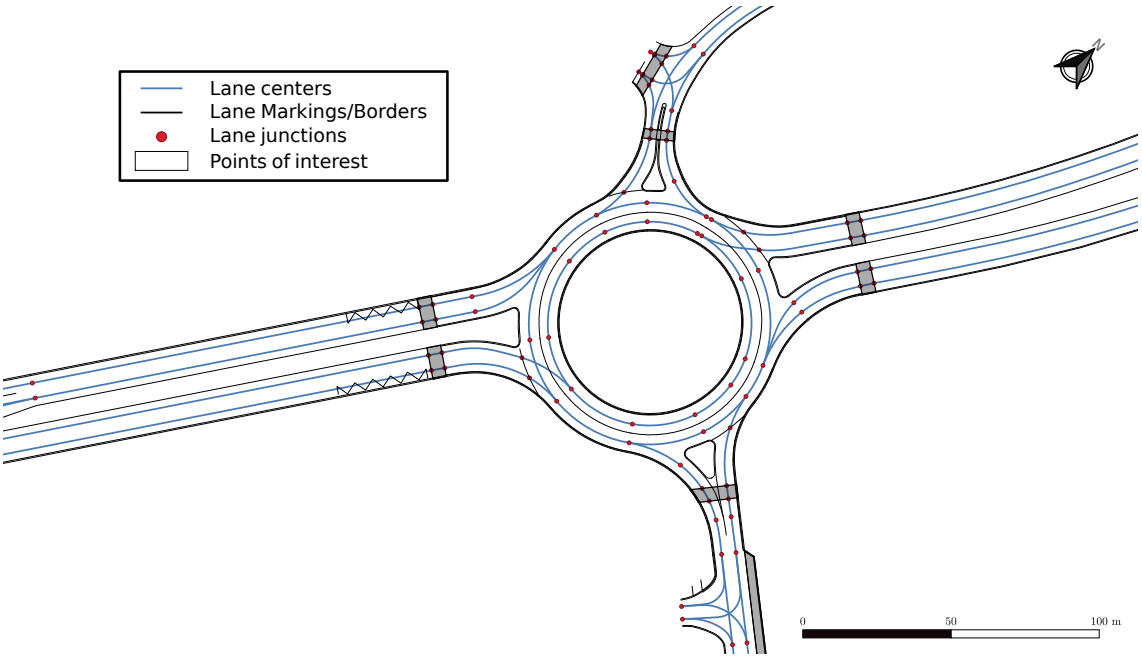


Figure 2.5: Visualization of a map tile from the HD map in the QGIS software.

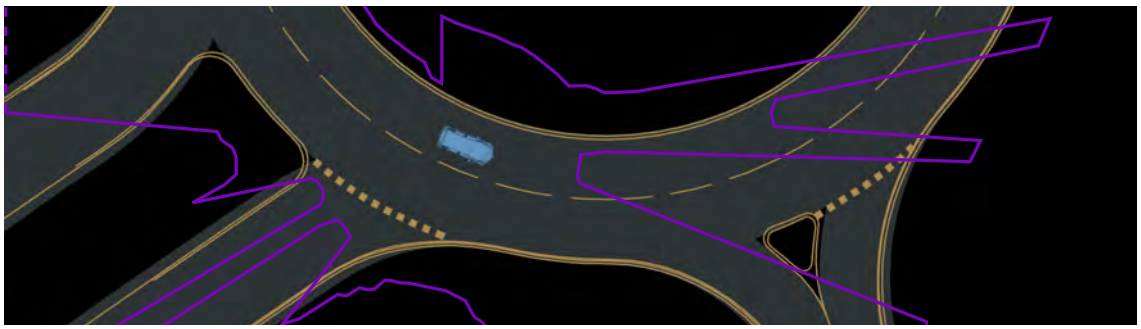
Figure 2.5 shows the road layout of a roundabout provided by the HD map. A module that handles the HD map (map_server) has been implemented in order to be able to load and to process the map under a ROS architecture. It enables the WM module to query a map tile around its position given a radius. Several additional functionalities have been developed in order to construct a functional road graph.

It must be noticed that the assumption of perfect map information is done. Mapping errors and not up-to-date map data are not addressed in this work.

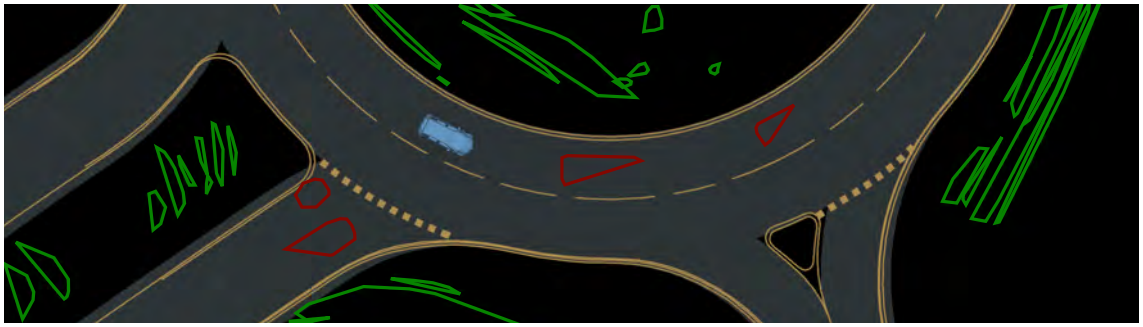
2.4.2 Localization

A highly accurate localization is supplied to the WM. Using a high-end GNSS/IMU solution with local RTK corrections, the autonomous vehicle will first be considered as being localized with centimeter accuracy. A simple module that supplies the pose of each experimental vehicle is used.

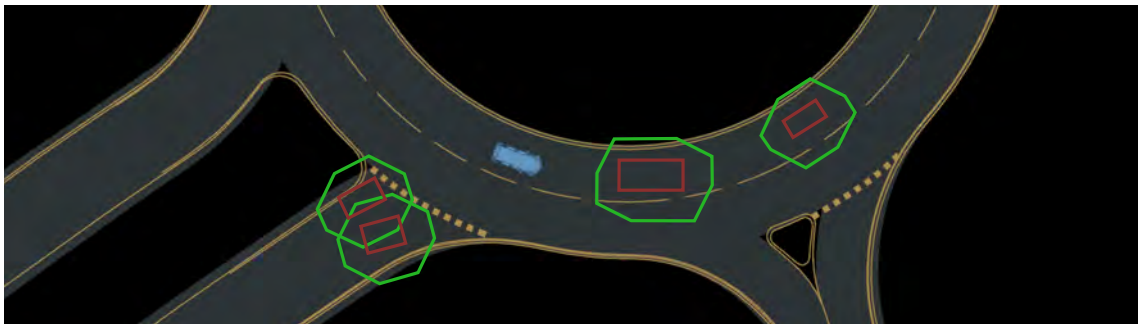
In order to simulate a less accurate localization, a simple module that applies a Gaussian noise on the vehicle pose based with a given standard deviation has been implemented. Thus, uncertainty of localization will be considered.



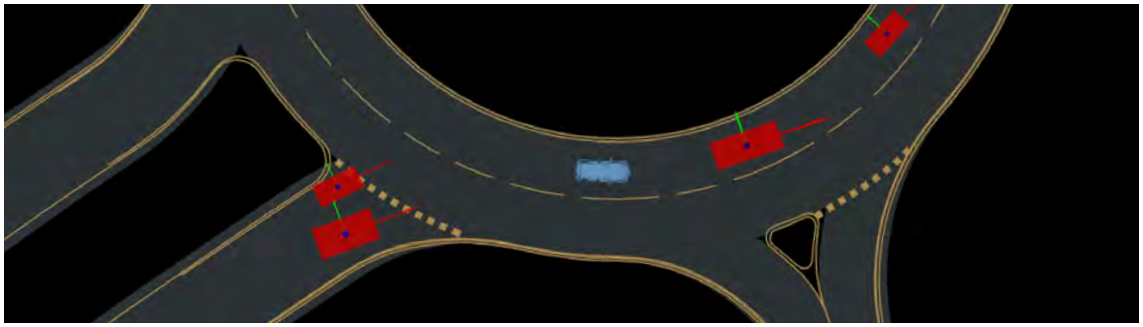
(a)



(b)



(c)



(d)

Figure 2.6: Perception visualization on RVIZ. (a) Contour of the free-space polygon. (b) Filtered polygons: intersecting the map lanes (red), outside (green). (c) Uncertainty of localization propagated on the road users hulls (green). (d) Bounding boxes of the road users.

2.4.3 Perception

Two types of information are supplied to the WM: space occupancy information and a list of objects. In the proposed approach, the polygon-based method has been implemented for its practicality at the expense of an occupancy grid. A simple geometric ground fitting based on Principal Component Analysis algorithm and a clustering algorithm was used on point-cloud data to extract objects clusters surrounding the ego vehicle [Zermas et al. – 153].

A free space polygon is generated over a radial grid with the minimum distance over the points of the clusters. One can compare it for instance to a parametric free space [Schreier et al. – 122]. The free space polygon considers that the area inside is free (see figure 2.6a).

To build the list of objects, the geometric layer of the HD map has been used to filter only clusters on or partially on the road (see figure 2.6b). It could be seen here as a focalization process from the WM to the perception module. The selected clusters are then tracked in order to estimate their heading and speed. An additional module combines this information with the convex hulls of the clusters that are provided to the WM as bounding boxes with a pose, speed and dimensions estimates (see figure 2.6d). Objects are considered road users.

The occupancy of each object is also given under a polygon considered as being occupied. These clusters are convex hulls and can take into account the localization uncertainty of the vehicle. The uncertainty propagation is further used in experimental results. Each polygon object observed is augmented by the localization uncertainty. The method presented in [Bernardi et al. – 18] has been implemented. Each vertex of the polygon generates several other vertices with a transformation that takes three times the standard deviation on the pose in each direction. Given all the resulting points, a new convex hull is obtained. In other words, polygons are enlarged by taking into account the error added on the ego-vehicle pose with a confidence value of 99:7 % (see figure 2.6c).

2.4.4 Decision-making module

The future path that must be followed by the ego-vehicle is supplied to the WM module. An A* algorithm has been implemented in order to build the mission path that is published at a rate of 0:25 Hz. This is a minimal implementation that is needed to build the information representation at the strategic level. The behavior of the DM module has not been implemented. It aims to take the WM representation output to take the proper decisions.

2.4.5 WM module

The World Model is implemented in C++. The WM owns three managers that process concurrently the three abstraction levels tasks of each representation, i.e. the ego state, the objects state and the space state. For the ego and objects state layers, the map-matching that have been implemented corresponds to the Lanelet method as described in [Hery et al. – 52]. It enables a fast implementation that ensure the continuity of the position and heading matching in real time.

The OL is updated by the perception and localization modules. The processes of the TL run at a rate of 10 Hz where those of the SL run at a rate of 0:25 Hz.

2.5 Conclusion

This chapter proposes a functional architecture for an autonomous vehicle with a WM module. Its internal structure and mechanisms have been detailed.

The requirements of the WM presented in the proposed architecture aim to achieve situation awareness. Each information layer enables to model a different part of a scene representation. These layers have different abstraction levels where bottom-up and top-down processes may enhance the situation understanding.

For this purpose, the WM has the duty to generate and maintain a best estimate of the state of the navigation context at each level. This best estimate describes the state and attributes of road users, classes, situations, and relationships taking into account uncertainties and integrity. Indeed, the WM must also check consistency of information in order not to provide misleading information to the DM module.

Unlike the perception module, it makes multi-informational fusion of information, i.e. it combines different types of information source: prior information such as maps and traffic laws with objects and spatial information. It gives rise to reasoning mechanisms such as relevant information filtering or prediction at different levels of abstraction.

Advantages of having a WM is to be more efficient in terms of memory consuming and time processing. Only pertinent information given the navigation context is processed for the driving task. Thus, the WM has to be able to have enough computational power in order to be efficient and to run in real time. Compromises have to be done between high-density information stored and the memory capacity in order to be as efficient as possible.

It should be highlighted that the WM is a whole system that could provide much more utility that only being a component of an autonomous vehicle. Indeed, as safety is a key point in this field, the WM could also be a tool for evaluating the system performances (risk evaluation or safety analysis for instance). As its purpose

is to model a virtual representation of the environment, the mechanisms that take place are easily understandable for a human. The WM may also be a powerful tool for simulation or even for driving assistance. It could enable to design the autonomous vehicle requirements based on its efficiency. However, as described in [Lotz – 80], even if qualitative metrics were found in the literature, it remains hard to evaluate such a system with quantitative metrics.

Chapter 3

A Lane Level World Model

3.1 Introduction 51
3.2 Related Works 52
3.3 Interaction Graph 53
3.4 LGM: a Lane Level Representation 61
3.5 Enhanced Space Characterization 78
3.6 WM Spatial Information Overview 93
3.7 Conclusion 94

3.1 Introduction

The second level of situation awareness is the comprehension of the current situation. Indeed, information representation must enable enhancement through augmentation processes in order to supply information at a higher level of understanding. It must be highlighted that the vehicle must be aware of the unexpected and its potential lack of information [Endsley – 38]. The contribution of this chapter is brought through the information representation of the spatial layer of the World Model (WM) module. This work aims at providing to decision-making an intermediate representation that enables to take decisions at the maneuver level.

While the operational level provides a scene understanding, the tactical level aims at providing a situation understanding as complete as possible. For this purpose, the ego-vehicle must first focus and analyze information that is in its vicinity. The selection process is thus applied to each information representation of the WM. The ego-vehicle must focalize on its position in the scene but also on the space where it is likely to evolve and on the scene elements with which it may interact. From a spatial layer point of view and at the strategic level, i.e the highest level, prior knowledge is used in order to infer areas that are of interest for the ego-vehicle.

The first part of this chapter aims at describing a so-called Interaction Graph (IG) representation that enables to build a spatial information representation at a strategic level. It permits to enhance situation understanding in the areas of interest of the ego-vehicle.

Once these areas have been defined, a lane level representation is built upon the IG that is used to focus on relevant information. The Lane Grid Map representation (LGM) is introduced and aims at providing enhanced information at the tactical level. This spatial representation handles occlusions and infers additional knowledge in hidden areas to facilitate safe decision-making.

One of the main purpose of the WM is to guarantee the integrity of the information so that the safety of a situation can be evaluated. Indeed, sources of uncertainty has to be taken into account but not only. Situation understanding is a complex task that requires access to a consistent and non-misleading representation of the vehicle surroundings. Spatial representation also enables to take into account the lack of information. The WM has to handle missing information in order not to lead to an erroneous decision-making. As a consequence, the final information representation must not provide misleading information as it could lead to a poor or inconsistent decision-making. Once the situation information representation is built, the management of the information integrity by the LGM is presented.

At operational level, information representation is supplied by the perception and localization modules, which may give an integrity confidence value. At the strategic level, prior information is provided by a knowledge database, e.g. a high definition map database. This map is assumed to be consistent and valid. As a consequence, map errors and its integrity must be deepened but are not part of the integrity evaluation.

3.2 Related Works

As mentioned in chapter 1, there is a need to combine spatial occupancy representations with objects representation, as it is presented in [Tas et al. – 132] architecture. Occlusions in specific situations need to be addressed and may need objects representation. Several decision-making approaches enable to take into account occluded areas. POMDP are used for handling occlusions for static obstacles [Bouton et al. – 20] as well as dynamic ones [Hubmann et al. – 58, Schorner et al. – 120]. The authors in [Hoermann et al. – 53] use several layers of information: a dynamic grid, an object list and a map of non-observable regions to assess likely collisions. Some articles handle directly polygons of occluded areas [Narksri et al. – 96], in some cases the center-line is used [Hubmann et al. – 58] and in other cases the segment border is used [Orzechowski et al. – 100].

In specific situations, additional information is inferred. A patented method [Burca – 23] shows the mechanism to determine a visibility distance by intersecting the path of a vehicle from a map with the dynamic field of view. This example shows a characterization of the following path of the vehicle by two states: the free space that is of interest and the hidden one. The Responsibility Sensitive Safety[1] (RSS) model presented in [Shalev-Shwartz et al. – 127], also highlights the spatial safety requirement with objects representation. The concept of a safety distance is introduced for several situations. For decision-making evaluation, distances computation based on the situation encountered by the ego-vehicle is often used as a metric. It can be expressed as a Time-To-Collision (TTC) or an Enhanced-Time-To-Collision (ETTC) [Minderhoud and Bovy – 91] or even a Time-To-Entrance (TTE) [Noh – 99]. These works highlight the use of several metrics used by decision-making modules.

Several works need a spatial representation handling occlusions and objects for tasks like decision-making but also risk analysis [Yu et al. – 150] or safety evaluation [Hoermann et al. – 53]. In [Orzechowski et al. – 100], reachable sets are used for predicting the path and intention of a potential road user that could be hidden in an occluded area. Particle filters are also used in order to determine the plausibility of a hidden road user [Narksri et al. – 96]. Then, as presented in [Hubmann et al. – 56], a global method for motion planning could use non-misleading information provided by such a context-modeling module.

When hidden areas appear, one needs to assess whether all of them are relevant for a safe decision-making. The notion of areas of interest presented handles this requirement as it enables to identify where relevant information is missing. As demonstrated in [Sun et al. – 131], the term Social Perceptionis defined in order to reason, i.e. infer information in occluded zones from the behavior of road users.

The following section presents the IG and the LGM representations that aim at handling most of these requirements. Additional information that can be inferred from occluded zones and from objects are integrated in a unified representation.

3.3 Interaction Graph

3.3.1 Problem statement

At the strategic level, information representation is more abstracted than at the tactical level but corresponds to a longer time and a further distance horizon. At this level, graphs and semantics representations can be found for instance in [Bagschik et al. – 12] or [Ulbrich et al. – 137]. The purpose of this higher representation is to provide interactions representation through the three different layers, i.e. the ego

[1]Latest concept is available on mobileye.com/responsibility-sensitive-safety

state, the space state and the objects state. The strategic level takes into account the ego-vehicle possible paths but also road users possible paths in order to put them into relation. Using prior information from a knowledge database, a focus on areas that are pertinent to the current situation is inferred. From a spatial occupancy layer perspective, these areas are called of interest and are defined according to their degree of relevance.

This representation highlights a top-down process because it allows a focus of information from a prior knowledge. It enables a focalization process. In the literature, mechanisms to focus mainly on objects in the scene can easily be found as in [Refaat et al. – 108] or [Leurent and Mercat – 77]. From an area of interest point of view, there is image analysis for example with [Han et al. – 50] where areas of interest are defined mainly in relation to the presence or not of objects in the scene. It is rarer to find focusing processes on a spatial representation of the situation.

3.3.2 Areas of interest

At the strategic level, lanes of interest are extracted from the road map given the ego-vehicle future paths. From these intended paths, one can define all the driving paths that may interact with the car during its cruise. Interacting paths are extracted from the topological layer of an HD map. For the sake of simplicity, it is assumed hereafter that a single future path is chosen for the ego-vehicle. The following would apply in a similar way for each possible future path.

Two types of interacting paths are differentiated in this work:

- The **primary order** paths are directly in interaction, i.e. shares a common driving space with the ego-vehicle intended path. These paths are the ones for which the vehicle needs to have information, typically whether they are occupied by another road user, for decision-making. It is especially true if the ego-vehicle has to yield the priority.
- The **second order** encompasses paths that have an indirect interaction with the autonomous vehicle, in the sense that they have a direct interaction with the primary order paths of the ego-vehicle itself. The vehicle does not necessarily need information about the second order paths to drive, but they are able to provide additional information about the primary ones.

Four types of interactions that widely cover common driving situations, called interaction modes,are considered: lane keeping, lane changing, lane merging and lane crossing. In figure 3.1 the interaction modes are defined based on the topological layer of the road. In order to simplify secondary order paths in the table, only one path of the primary order paths (surrounded in black) of a given interaction mode is analyzed.

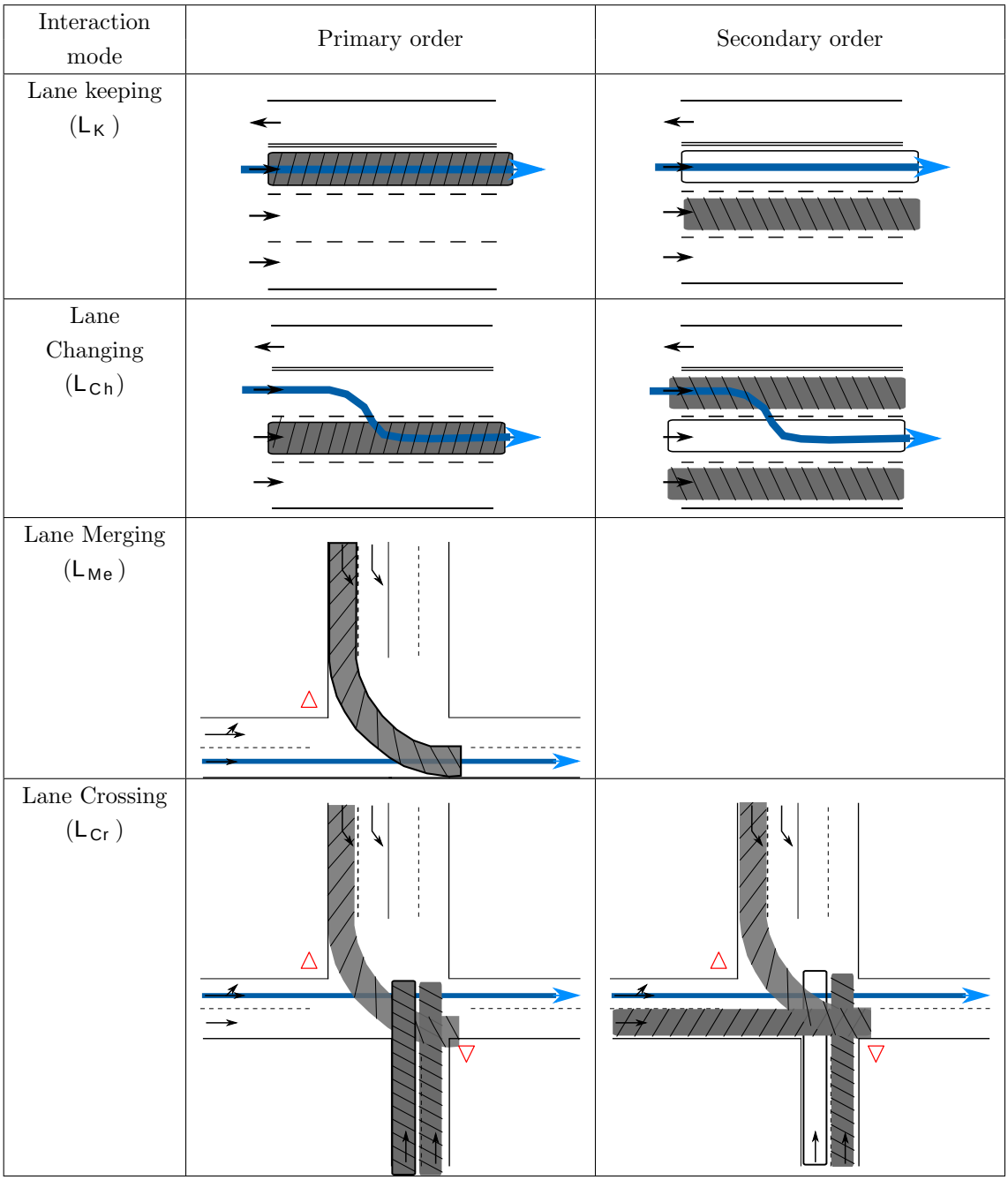


Figure 3.1: Lanes of interest (road users specific) represented in gray are based on their importance order and on the lateral interaction mode. The autonomous vehicle planned path is represented in blue. For visibility issues the secondary order shows the paths of interest corresponding to one lane of the primary order (surrounded in black).

In lane keeping mode for instance, the followed lane belongs to the primary order area of interest and is unique. The changing lane belongs to the secondary order, as it is the primary order path of the lane changing interaction mode.

In figure 3.2 the secondary order paths of the lane merging mode are described. Secondary order paths shown are primary order paths of the merging path represented by the black polygon. First, it must be noticed that lane keeping mode and lane merging mode of the merging path are encompassed by the path itself. Then, lane changing mode gives the path shown in figure 3.2a. Finally, lane crossing mode corresponds to the paths shown in figures 3.2b and 3.2c. The result is given by figure 3.2d.

3.3.3 Formalism

This section describes the formalism used to define the IG representation. First, the topological map representation used to build the nodes is presented. Then, the graph construction is addressed. At the Operational level (OL), the notation $[:]^{(OL)}$ will be used. $[:]^{(TL)}$ is used for the Tactical level (TL), and finally, $[:]^{(SL)}$, for the Strategic level (SL).

3.3.3.1 HD map representation

The HD map, denoted M, is a topological connected graph, which represents the road layout with very high accuracy. Each node of this graph is a portion of the road that is called a link l and which represents the center lane. This link can be represented at different levels of abstraction: as a unique ID, as a geometrical succession of cells or as a polyline made of points.

Definition 3.1 SL link

At the Strategic level, a link l is defined as a unique ID such that $\mathsf{l}^{(SL)} \in \mathsf{N}$.

In this graph, illustrated in figure 3.3, there are several types of connections. The first one is the child connection and is an oriented connection. A node can have one or several children or parents. The second connection is an adjacency connection. One node can have one or several adjacent nodes, i.e. lanes that are possible changing lanes. The third one is a crossing connection. If two links share a common crossing point then they are assumed to cross each other. The last connection is when a link merges with another. An illustration of these connections is provided with the lanes of a roundabout entrance (figure 3.3b).

As a consequence, each link stores a list of links that correspond to a given relation, i.e. $\mathsf{crossing_links}$, $\mathsf{merging_links}$ and $\mathsf{adjacent_links}$ attributes.

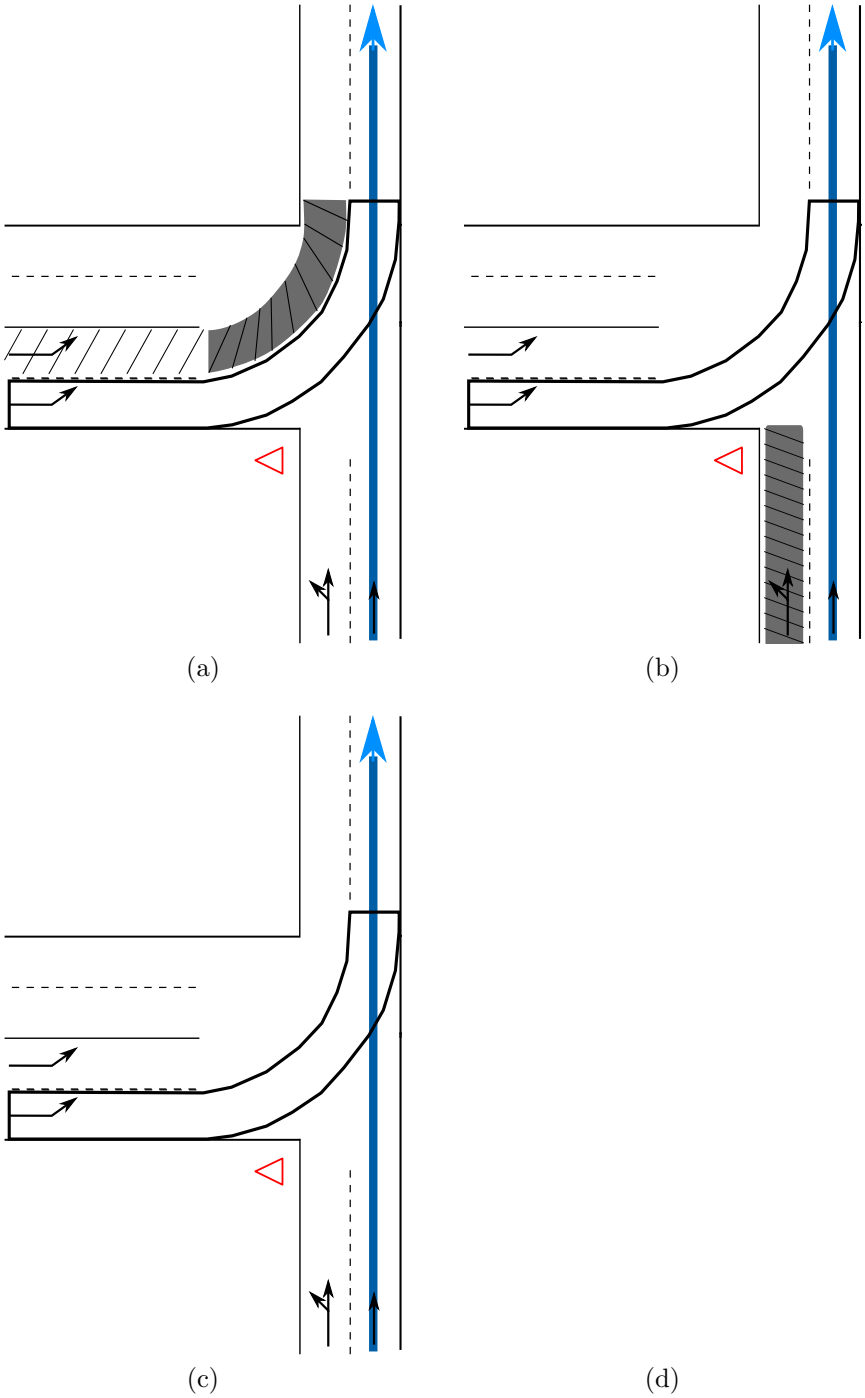
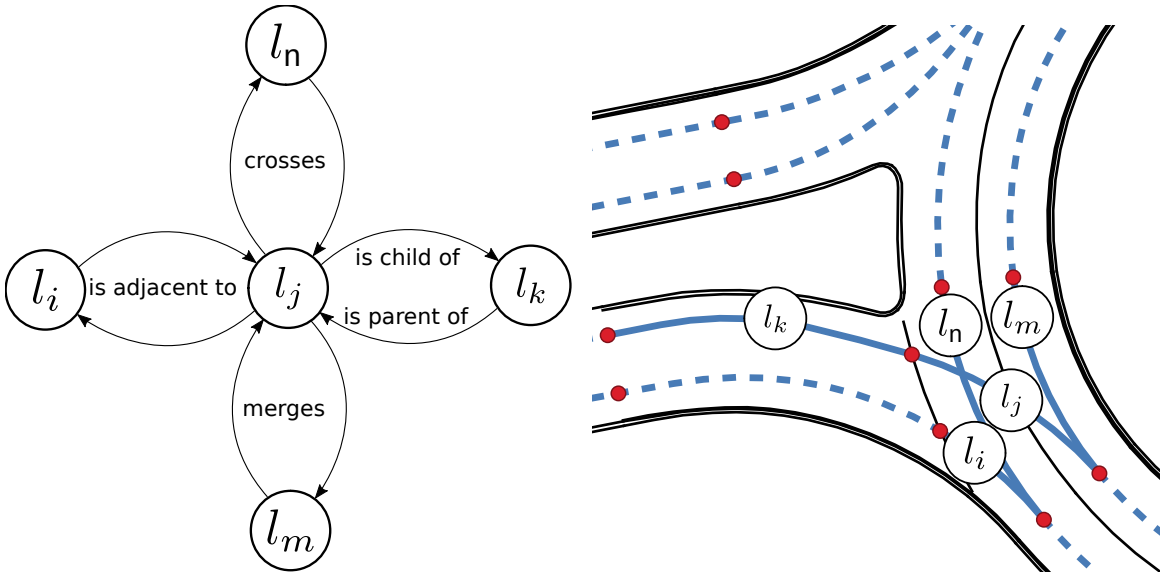


Figure 3.2: Secondary order paths of a lane merging mode. The primary order merging lane is circled in black. (a) It depicts the lane changing mode of the merging lane. (b) and (c) These figures depict the lane crossing modes. (d) This figure is the sum of all the secondary order lanes (which is presented in figure 3.1).



(a) (b)

Figure 3.3: Topological connections of the HD map. Each link l may have the following links to another link: crosses, merges, is-adjacent-to, is-parent-of or is-child-of. (a) Graph of connections. (b) HD map tile.

3.3.3.2 IG representation

The IG is built at the strategic level of the space state layer. Each node g of this graph is a collection of links from the HD map representation that needs to be defined. Each relation between the IG nodes also has to be specified.

First, as the HD map is a connected graph of links $l^{(SL)}$, each node g of the IG can be decomposed into rooted trees of links as defined in the graph theory. The following definitions are thus introduced.

Definition 3.2 Directed rooted tree

A directed rooted tree is a tree in which one vertex of the graph has been designated as the root and where the edges are assigned a natural orientation, i.e. away from or towards the root.

Thanks to this definition, a forward or backward path tree can now be defined. Each node of the interaction graph is either a forward or a backward path tree. In addition to that, the root element of a directed rooted tree is designated by root(< tree >).

Definition 3.3 Forward/Backward path tree

A forward path tree g is a directed rooted tree of the map M considering only the is-child-of relation. A backward path tree g is a directed rooted tree of M considering only the is-parent-of relation.

Then, three relation types that may exist between two path trees, i.e. two nodes of the IG, are defined.

Definition 3.4 Crossing relation

Let g_1 and g_2 be two path trees (forward or backward). The **crossing relation** is defined as:

$$g_1 \text{ crosses } g_2 \Leftrightarrow \exists\, l^{(SL)} \in g_2 ; \text{root}(g_1) \text{ crosses } l_{OL}$$

Definition 3.5 Merging relation

Let g_1 and g_2 be two path trees (forward or backward). The **merging relation** is defined as:

$$g_1 \text{ merges } g_2 \Leftrightarrow \exists\, l^{(SL)} \in g_2 ; \text{root}(g_1) \text{ merges } l^{(SL)}$$

Definition 3.6 Changing relation

Let g_1 and g_2 be two path trees (forward or backward). The **changing relation** is defined as:

$$g_1 \text{ is-changing-of } g_2 \Leftrightarrow \forall\, l_1^{(SL)} \in g_1 ; \exists l^{(SL)} \in g_2 ; l_1^{(SL)} \text{ is-adjacent-to } l_2^{(SL)}$$

The IG consists of a directed rooted tree which root corresponds to a forward path tree L_K, i.e. the lane keeping mode of the ego-vehicle with $L_K = \text{root}(IG)$. The other nodes of the IG are backward path trees with one of the defined relations among crossing, merging and changing. They correspond to the lane crossing L_{Cr}, lane merging L_{Me}, and lane changing L_{Ch} modes. The IG has a depth of order two. The first order nodes correspond to the primary nodes and the second order, to the secondary ones.

As the lane keeping mode of the ego-vehicle is generated by the ego state layer, interactions mode are explored on each link of this path. The purpose is to get primary interaction graph nodes. The algorithm 1 describes this process. Two functions related to the HD map are used. The map.get_max_backward_dir_path function enables to extract from the map, i.e. the links graph, the maximum spanning backward path tree given a minimum distance length. The second function, map.split_into_max_backward_dir_paths, enables to take a list of links and to output a minimum list of backward path tree thanks to the map also. In a second step, this process is similarly applied on the primary nodes instead of the root node in order to get secondary order nodes of the IG.

The figure 3.4 represents the interaction graph built at the strategic level. The possible paths of the ego-vehicle computed by the ego-state representation correspond to the lane keeping node. Primary and secondary nodes are then built. This definition of the IG assumes that the graph depth is of two but could be greater. Adding one or more depth levels could be an improvement if needed.

Algorithm 1: IG primary order nodes

```
Input  : root(IG)                          //Ego-vehicle future path
         d                  //Max distance horizon of an area of interest
Data: map
Output: IG_nodes                //Interaction Graph primary order nodes
L_Cr_nodes  [ ]                          //Nodes with a crossing relation
L_Me_nodes  [ ]                           //Nodes with a merging relation
L_Ch_nodes  [ ]                          //Nodes with a changing relation
adj_links   [ ]                                //List of adjacent links
foreach link in root(IG) do
    foreach crossing_link in link.crossing_links do
        L_Cr_nodes.add(map.getmax_backward_dir_path(crossing_link; d))
    end
    foreach merging_link in link.merging_links do
        L_Me_nodes.add(map.getmax_backward_dir_path(merging_link; d))
    end
    foreach adjacent_link in link.adjacent_links do
        adj_links.add(adjacent_link)
    end
end
L_Ch_nodes   map.split_into_max_backward_dir_paths(adj_links)
IG_nodes   [L_Cr_nodes; L_Me_nodes; L_Ch_nodes]
```

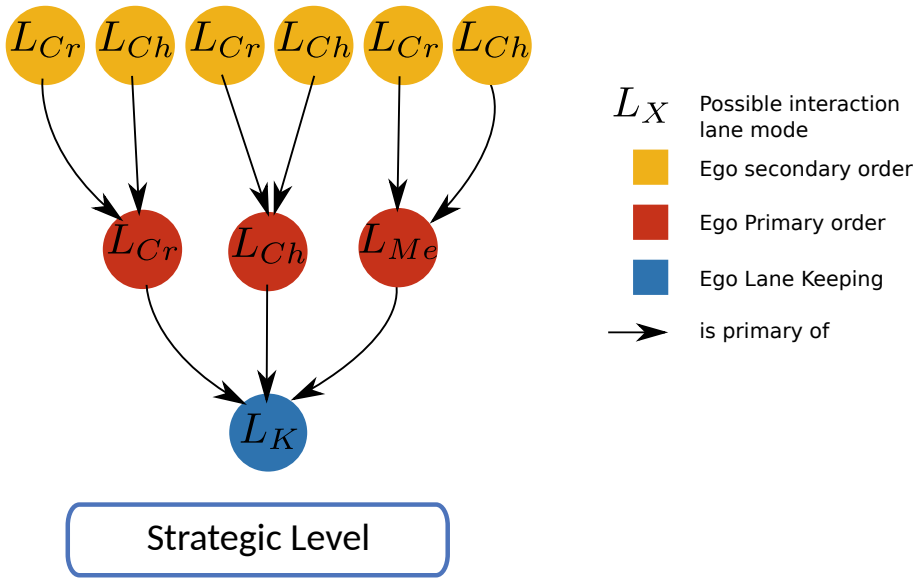


Figure 3.4: Schema of the Interaction Graph. The root node corresponds to the lane keeping mode (blue). Primary order nodes are shown in red with different possible interaction modes. Similarly, secondary order nodes are shown in yellow.

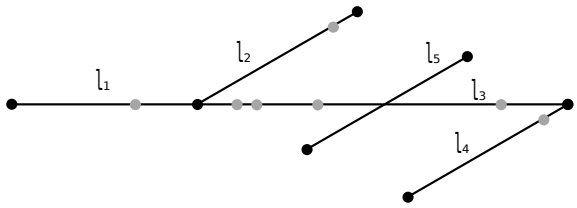


Figure 3.5: Sketch of a simplified link definition based on the topological layer of the HD map. l_i represents a simplified link composed of several segments.

3.3.4 Applications

In the HD map of Compiègne, a link is represented by an id that corresponds to a portion of the road lane depending on multiple parameters (new road sign, speed limit change, bumper, etc.). However, in order to compute the interaction graph as described, it has been defined a simplified link as a road lane where the end and start diverge or merge. The path between the start and the end is thus unique as shown in figure 3.5. Thereby, each link may be linked to a list of merging links, a list of crossing links and a list of adjacent links thanks to the topological layer.

A dataset has been recorded in the city of Compiègne. The figure 3.6 displays a real-time application of the IG computing. The lane keeping mode of the ego-vehicle (in blue) is represented by the blue line. Primary order lanes are in red and secondary order in orange. In bottom left, the ego-vehicle faces a merging lane. This merging lane encounters two crossings: the oncoming lane and the diverging lane.

It is also possible to compute possible paths of each road user of the objects state representation, i.e. their lane keeping mode. It enables to determine which of them are relevant in the current driving situation. Rather than displaying future paths of road users, figure 3.7 shows road users relevancy. Road users that are in relation to the IG are in red where others are in gray.

As a result, areas of interest of the ego-vehicle based on interaction modes and the vehicle situation are defined. The next section describes the tactical information representation that focuses on these areas.

3.4 Lane Grid Map: a Lane Level Representation

3.4.1 Problem statement

At the tactical level, the autonomous vehicle has to convert the topological information provided at the strategic level by the Interaction Graph (IG) into a metrical representation. Moreover, the vehicle has to focus on areas of interest that are in its vicinity in terms of time and distance. By defining a threshold in terms of interacting

Figure 3.6: Interaction Graph models in a real-time experiment led in Compiègne. The ego-vehicle is represented in blue. The primary order lane keeping paths is represented by the blue center-line. Other primary order paths are shown in red. Secondary order paths are in orange. For visual convenience, lane changing primary order paths are not displayed. Furthermore, some paths may be both primary and secondary but are represented with only one color. (The buildings are displayed in gray)

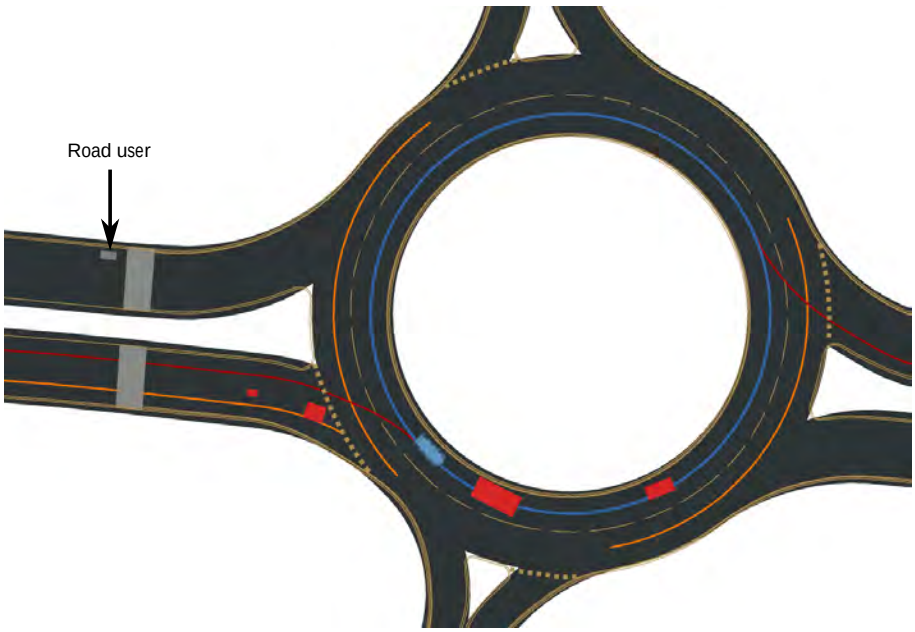


Figure 3.7: Visualization of road users relevancy on Rviz. Relevant road users are in red where others are in gray (only one is visible here).

distance, the interacting lanes are geometrically pruned and a spatial Area Of Interest (AOI) is constructed. The AOI is therefore composed of the primary one $\mathsf{AOI}^{(1)}$ and the secondary one $\mathsf{AOI}^{(2)}$, all the space outside of the $\mathsf{AOI} = \mathsf{AOI}^{(1)} \cup \mathsf{AOI}^{(2)}$, denoted as $\overline{\mathsf{AOI}}$, is considered out of interest as there will be no interaction with the ego-vehicle.

The proposed representation, called lane grid map (LGM), allows having a more detailed representation and aims to quantify the available and missing information in these areas. Such a representation must handle occlusions and may infer additional knowledge in order to facilitate safe decision-making. Indeed, some parts do not present any danger to the ego-vehicle. A formalism that allows their characterization and manipulation in real time is proposed. This will therefore firstly give rise to the characterization process of the AOI using perception information from exteroceptive sensors such as cameras or LiDAR. Then, a characterization process of the AOI using reasoning with prior map information has been developed and aims at improving situation understanding in unknown areas.

3.4.2 Lane level context

The proposed lane level grid representation, i.e. the LGM, is a spatial grid based on a discretization of the lanes within the AOI along their longitudinal direction. Based on the geometric layer of an HD map, the lanes are discretized in successive contiguous quadrangles, i.e. cells, that cover in a continuous way the space of the

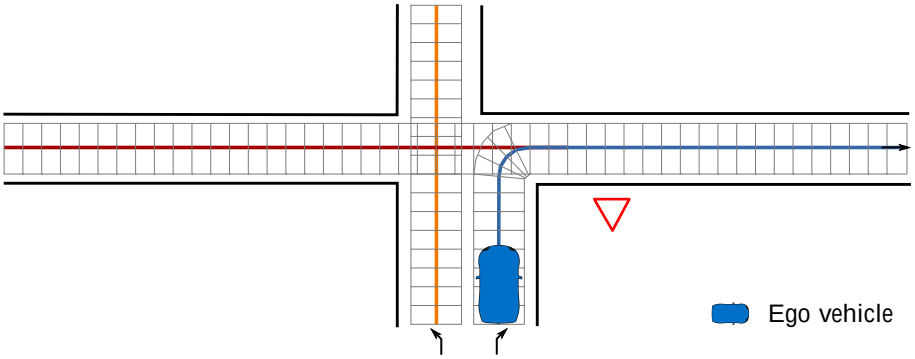


Figure 3.8: LGM example with a cross intersection. For the lane keeping mode (blue), the lane merging of first order and the lane crossing of second order of the ego-vehicle are shown in red and orange respectively.

AOI . A more detailed explanation of the LGM's cells computation is supplied by appendix B. Thus, the LGM has two parameters to be set: a lateral cell width and an along track cell length that is called sampling step parameter.

In figure 3.8, the LGM is represented at a cross intersection situation. The ego-vehicle is merging with the red lane, which is of first order, and the lane in orange is of second order.

The information supplied by the LGM is twofold. It enables to provide distances between entities with respect to the geometric topology of the road. Indeed, as each cell is defined by the sampling step parameter, in the along track direction of a road lane, distances are computed by aggregating the number of cells. A description of this process is detailed in appendix C. On the other hand, each cell encompasses spatial information delimited by its quadrangle shape. As a consequence, occupancy state information may be supplied conjointly with distance. For instance, the free distance to a road user or to a hidden area can easily be computed.

In order to supply a spatial information, information stored by each cell needs to be computed. This process is called characterization process It may take different spatial metrics such as an occupancy or a risk for instance. As occupancy is widely used and implemented in robotics applications, this information representation is used in the presented work.

3.4.3 Efficiency and scalability advantages

In classical occupancy grids (OG), all the space surrounding the vehicle is characterized regardless of its relevance. On the contrary, the LGM takes into account the situation in which the AV drives, i.e. in a given road layout, therefore it only stores information within the relevant lanes (AOI) provided by the IG at the strategic level. As a consequence, the LGM encodes the information of the space surrounding

the vehicle in a more compact and efficient way. Each cell of the LGM has two values: the width c_w, which corresponds to a lane level width and a sampling step length c_l. At the opposite, the OG is a spatial sampling over the entire space given square cells of width s_w. For an application example, one considers an area of 50 m over 50 m ($A = 2500\,\text{m}^2$), regardless of its geometric shape, that needs to be encoded in both representations.

The OG has a number of cells $C_{OG} = d\frac{A}{s_w^2}e$ and a visibility distance $d_{OG} = \sqrt{A}\,\text{m}$ as the area covers a 50 m square with the OG representation. The vehicle is located in the center of both representations and the example of a straight road lane is taken. The LGM has a number of cells $C_{LGM} = \frac{A}{c_w\, c_l}$ and a visibility distance $d_{LGM} = \frac{A}{c_w\, c_l} \quad c_l = \frac{A}{c_w}\,\text{m}$ as the area is covered by the LGM. For example:

$$\left.\begin{array}{c} c_w = 1\,\text{m} \\ c_l = s_w = 1\,\text{m} \end{array}\right) \qquad \begin{array}{c} d_{LGM} = d_{OG}^2 \\ \text{with } C_{LGM} = C_{OG} \end{array} \tag{3.1}$$

The increase of the visibility distance is easily explained by the fact that cells are located in a focused area for the autonomous driving. Computationally the number of lanes can be raised by decreasing the visibility distance that would be dispatched on each lane. In figure 3.9, an approximate number of cells in the LGM with respect to the sampling step parameter is plotted. Three different situations are taken into account in order to show the qualitative evolution of cells. The more complex the intersection, the higher the number of cells. The primary order lane distance horizon is around 100 m and around 50 m for secondary order lanes. The graph shows an inverse function. With a sampling step of 5 m, the number of cells decreases to an average of cells of [100 500]. As a consequence, the LGM efficiency and its memory consuming can be optimized depending on requirements. The memory usage by the LGM and characterization process speed depend on the sampling step parameter. It can be noticed that on experimental vehicles, with a frequency of 10 Hz, the computed LGM may reach up to a sampling step of 20 cm.

It should also be noticed that the LGM representation is easily scalable. Indeed, the sampling step may vary and easily enables to aggregate cells with low difficulty. On top of that, the fact that cells fit properly the road layout enable to have a good fidelity information representation over distances.

As cells are independent from the geometrical layer of the map, the delimitation of a number of cells to take into account for a portion of the road would not be easily defined. Distances over a grid are also not easy to fit, as it does not reason in the coordinate system of the road lane. These phenomena are two limitations of the OG that are used in the literature.

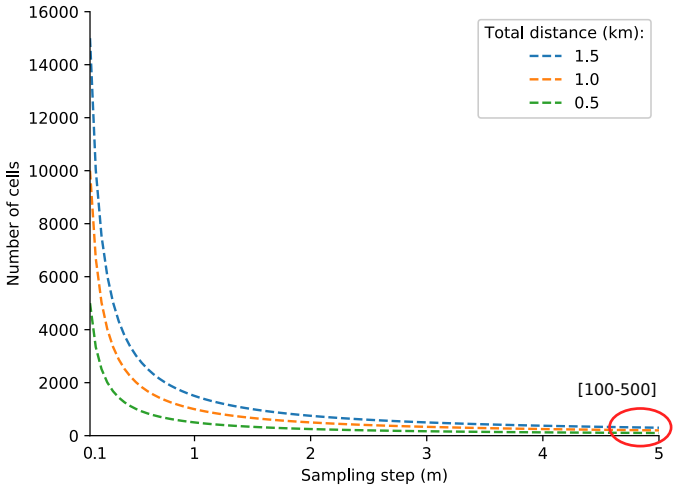


Figure 3.9: Number of cells in a LGM. The values presented are taken from a dataset run in Compiègne. The values are approximate and give a qualitative overview of the evolution of the number of cells in the LGM as a function of the sampling step parameter. The evolution of the number of cells is given for three distance amounts encoded by the LGM.

3.4.4 Occupancy characterization

In order to characterize the LGM, the formalism used to define the state of the cells is presented. The following section introduces the evidential theory that is used at the lane level context similarly to the evidential occupancy grids formalism as presented in [Moras et al. – 92].

3.4.4.1 Theory of evidence

In classic approaches, probabilities are used in occupancy grids to characterize the space surrounding the vehicle. It computes a belief of occupancy on each cell with a probability: from 1 for occupied (O) with a maximum confidence to 0 for free (F). However, this formalism does not enable do distinguish missing information from conflicting and uncertain information. In both cases, the cell may hold a probability of 0:5. In the case of total ignorance, e.g. a case where a cell is not perceived by the system, because there is no reason to prefer one state to the other, the probability 0:5 is taken, leading to $\mathsf{P}(\mathsf{O}) = \mathsf{P}(\mathsf{F}) = 0{:}5$. Now, supposing that two independent sensors supply conflicting information:

$$\begin{cases} \mathsf{P}_1(\mathsf{O}) = 0{:}75 \\ \mathsf{P}_1(\mathsf{F}) = 0{:}25 \end{cases} \quad \text{and} \quad \begin{cases} \mathsf{P}_2(\mathsf{O}) = 0{:}25 \\ \mathsf{P}_2(\mathsf{F}) = 0{:}75 \end{cases}$$

The combination leads to $\mathsf{P}(\mathsf{O}) = \mathsf{P}(\mathsf{F}) = 0{:}5$

Evidential theory handles in a better way missing and uncertain information. This approach enables to manipulate sets. In addition to that, probabilities can be seen as a particular case of the evidential theory. The theory of evidence combines sets and probabilities. The set of all possible classes is called the frame of discernment. For instance, the frame Ω encompasses what is free and what is occupied, its complementary.

$$\Omega = \{\text{Free}; \text{Occupied}\} = \{F; O\}$$

This frame of discernment has four sets : $\{F\}; \{O\}; \{F; O\}$ and an empty set $;$. Each set A has a probability mass function m such that:

$$m(;) = 0 \tag{3.2}$$

$$\sum_{A} m(A) = 1 \tag{3.3}$$

If $m(\Omega) = m(\{F; O\}) = 1$, it means that there is a total ignorance. Instead of a single probability, each set has a plausibility value Pl and a belief value Bel which represent, respectively, an upper and lower bound:

$$Pl(A) = \sum_{B \cap A \neq ;} m(B) \tag{3.4}$$

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{3.5}$$

Belief shows the confidence that is given to the set A. The difference between the plausibility and the belief shows the lack of information. A mass function can be transformed into a probability distribution through different methods (e.g. the pignistic function). $Bel(A) \quad P(A) \quad Pl(A)$ always holds true.

With two independent sources of information, the corresponding mass functions m_1 and m_2 can be combined with the Dempster's rule to obtain a new mass function denoted $m_1 \quad m_2$:

$$\begin{aligned} (m_1 \quad m_2)(;) &= 0 \\ (m_1 \quad m_2)(A) &= \frac{1}{1 - } \sum_{B \cap C = A} m_1(B) m_2(C) \\ \text{with} \quad &= \sum_{B \cap C = ;} m_1(B) m_2(C) \end{aligned} \tag{3.6}$$

In the case of total ignorance, $\Omega_1 = \{F; O\}$ leads to

$$m_1(;) = m_1(\{F\}) = m_1(\{O\}) = 0 \quad \text{and} \quad m_1(\{F; O\}) = 1$$

When combining several sources of information, ignorance is taken into account thanks to the Ω mass and the conflict is determined by . In addition to that, the

Dempster's rule is also used to combine masses of a common frame of discernment of several frames.

An evidential grid is chosen in this work instead of a probabilistic grid because it is able to manipulate explicitly ignorance in addition to uncertainties, but still enables to compute equivalent probabilities. Moreover, in order to take decisions, threshold values must be chosen to cells state in order to consider their occupancy. It may result in considering a binary value over each set of the frame of discernment. As a result, for simplicity, evidential theory easily handles the set theory by applying binary masses.

The first frame of discernment that is used is $\Omega_1 = \mathsf{fFree; Occupiedg} = \mathsf{fF; O\,g}$. Thus, a cell can be characterized as being free, occupied, or unknown ($\mathsf{fF; O\,g}$ state). The visible free space and occupied areas representations come from the perception module and gives to the WM a complete representation of the world space occupancy. At the operational level, the WM stores an up-to-date spatial representation. The space characterization may be done using an occupancy grid (OG) [Hoermann et al. – 53] or a parametric free space [Schreier et al. – 122]. These two representations will be used as examples.

3.4.4.2 Evidential grid input

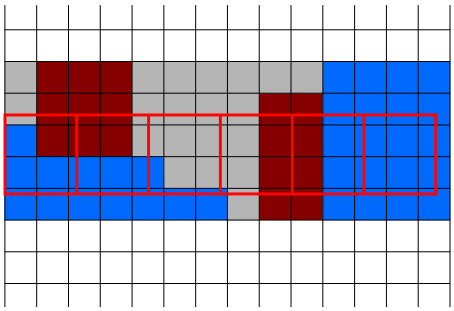
An evidential OG with the frame of discernment Ω_1 is given, for example, by the perception module as an input to the WM. Occupancy Grids are capable of handling uncertainties. As this grid is supposed to be supplied by the perception module, uncertainty of localization and perception are assumed to be already encoded in such representation.

The goal is to aggregate information of this grid in each cell of the LGM. The intersection between a cell of the LGM and the grid is computed such that one cell of the LGM encompasses $\mathsf{N_b}$ cells of the OG (figure 3.10b). Information from the $\mathsf{N_b}$ cells has to be aggregated into one cell. Table 3.10a presents the fusion rules. When merging two cells information, occupied state is more conservative and informative than the Ω state which is more cautious than the free state. Meaning that the aggregation of two cells is free if both of them are and occupied if at least one of them is. For an LGM cell, merging mass functions $\mathsf{m_1}$ and $\mathsf{m_2}$ from two OG cells can lead to the resulting mass function $\mathsf{m_{12}}$ with the assumption of independence between cells.

$$
\begin{aligned}
\mathsf{m_{12}(F)} &= \mathsf{m_1(F)m_2(F)} \\
\mathsf{m_{12}(O)} &= \mathsf{m_1(O)m_2(O) + m_1(O)m_2(F) + m_1(F)m_2(O) + m_1(O)m_2(\Omega) + m_1(\Omega)m_2(O)} \\
\mathsf{m_{12}}(\Omega) &= \mathsf{m_1}(\Omega)\mathsf{m_2}(\Omega) + \mathsf{m_1(F)m_2}(\Omega) + \mathsf{m_1}(\Omega)\mathsf{m_2(F)}
\end{aligned}
$$

$Cell_i$	$Cell_{i+1}$	$Cell_{res}$
F	F	F
F	O	O
F	Ω	Ω
O	O	O
O	Ω	O
Ω	Ω	Ω

(a) Table of aggregation rule of two cells of an OG. $Cell_{res}$ corresponds to the result of the fusion of the cells $Cell_i$ and $Cell_{i+1}$.



(b) LGM characterization with an OG. (free cells in blue, occupied cells in red and unknown in gray).

Figure 3.10: Characterization process of an evidential occupancy grid.

3.4.4.3 Polygons input

Instead of an evidential OG, a polygon that represents a visible free space Fs, and a list of polygons E_i that represents road users can be used (figure 3.11). The intersection between each cell c of the LGM and the polygons is computed. The area inside Fs is considered as free and each point is assumed to be part of an object E_i inducing an occupied area. If a cell is fully encompassed within the free-space Fs, it is considered as free. If it intersects a road user polygon, the cell is characterized as occupied. Finally, the other cells are considered with an unknown state. At the Tactical Level, without taking into account uncertainty of perception and localization, the mass function can be defined as follows:

$$\begin{cases} m(O) = 1 & \text{if } E_i \setminus c \neq \emptyset \\ m(F) = 1 & \text{else if } c \quad Fs \\ m(F) = 0 \\ m(O) = 0 & \text{else} \end{cases}$$

The characterized LGM in figure 3.10b is similar to LGM in figure 3.11. Occupied areas that intersect a LGM cell give an occupied state (red) to the cell. A whole cell can be classified as been occupied regardless of the object size. If a cell is within a free area, then the free state (green) is given, else the cell is defined as unknown (gray).

In order to simplify the LGM representation and decision-making, masses may be set to binary values depending on a threshold value on the fused masses in the case

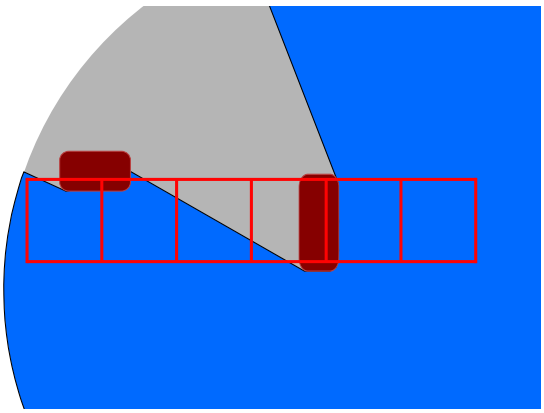


Figure 3.11: LGM characterization with polygons. Polygons representation: free space is represented by the blue polygon. Occupied polygons are shown in red and the gray area is unknown.

of an occupancy grid and depending on boolean results of intersections in the case of polygons.

It should be noticed that as well for OG as for polygon representation, objects and space representation are mutually linked. It could also be encoded in a map overlay.

3.4.5 Integrity evaluation metrics

The width of the LGM cells corresponds to the road lane width. The sampling step parameter of the cells enables to refine more or less the level of detail of the AOI . Bigger are the cells, coarser is the information. With the lane grid map representation, integrity is the ability to supply non-misleading information when characterizing the cell states. For example, characterizing a driving space as free when it is actually occupied may lead to a hazardous decision. The notion of integrity is used to guarantee a functional operating domain for the autonomous vehicle through the sampling method of the Lane Grid Map. Typically, an integrity loss of the LGM representation leads to a loss of availability of the system, as it does not guarantee to the decision-making that information is not misleading. The work presented shows how oversampling can lead to an integrity loss.

This section presents the evaluation method used to demonstrate the effectiveness of the Lane Grid Map. This methodology introduces the integrity criterion. The metrics presented aim at providing two types of information in a single representation. At maneuver level for the decision-making module, the LGM supplies both distances and occupancy integrity for the task to be fulfilled. In order to quantify misleading information from the characterization of a LGM, a ground truth has to be used.

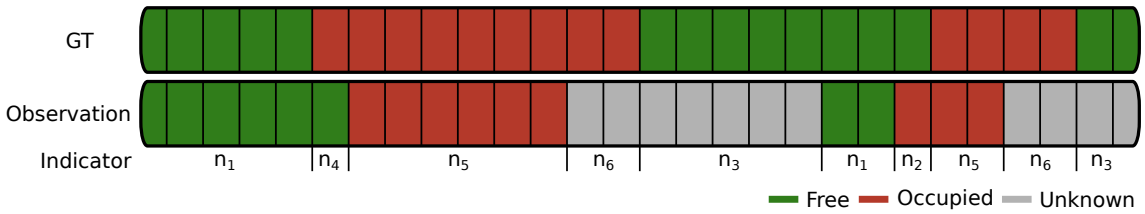


Figure 3.12: Correspondence between evaluation parameters and the LGM of both the observation and the ground truth (GT).

3.4.5.1 Ground truth generation

In order to evaluate the LGM model and efficiency, the characterization process is performed using the ground truth position of several vehicles. Given the accurate position of several vehicles on the road and their dimensions, the occupancy of these vehicles are approximated as rectangles. All the cells overlapped by these rectangles are considered as occupied on the ground truth LGM. Then, the rest of the cells are considered as being free. The cells of the ground truth LGM are either Free (F) or Occupied (O). The top row in figure 3.12 shows the ground truth of a LGM.

3.4.5.2 Observation generation

At the same time, the characterization process is applied with the on-board perception system. Using polygons approach, two types of information are supplied: the convex bounding polygons of the perceived road users and a free-space polygon. The perception system was described in section 2.4.3 on page 47. The bottom row in figure 3.12 shows the characterization by the ego-vehicle of the cells of the LGM into one of the three states: free (F), occupied (O) and unknown (U).

Over the ego-vehicle trajectory, the LGM (t) is generated at each time step t 2 [0; T]. At each time step, the LGM is characterized by the perception information and also by the ground truth of the objects. Combining the different configurations, for each cell of the LGM, there are six possible indicators N_i obtained from the combination of both characterizations as defined in table 3.1. The correspondence is shown as an example on the indicator row of figure 3.12. As a cell depicts the occupancy of a fixed step distance, these results are expressed not only in terms of cells but also in terms of distance and more precisely with a percentage. All the cells c^k of each LGM are associated with a distance n_j^k 2 fn $_1$; : : : ; n_6g. The N_i indicator corresponds to the total number of cells in state n_i and is written as follows:

$$N_i = \sum_{t2[0;T\,]c^k 2LGM\ (t)} \sum_{ij} n_j^k \quad \text{with} \quad {}_{ij} = \begin{array}{l} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{array} \tag{3.7}$$

Table 3.1: Parameters obtained with the combination of the ground truth and the observation.

	Observation		
Ground Truth	f	o	u
F	$F : f\ ?!\ n_1$	$F : o\ ?!\ n_2$	$F : u\ ?!\ n_3$
O	$O : f\ ?!\ n_4$	$O : o\ ?!\ n_5$	$O : u\ ?!\ n_6$

Two indicators are particularly relevant for the integrity analysis. N_2 corresponds to miss-classified free cells (n_2) considered as occupied. It is an overcautious indicator. N_4 is the most important indicator as it corresponds to the miss-classified occupied cells (n_4) considered as free. It is a highly miss-leading, i.e. hazardous, indicator that should be as low as possible. Indeed, if information of free space around the vehicle is considered whereas it is occupied, it can lead to a hazardous decision-making. At the opposite, if occupied information is given instead of a free one, it remains a cautious situation for the decision-making. The results section analyzes the rates of these two indicators. These rates must take into consideration only cells that have been characterized as free or occupied as it must not depend on the amount of unavailable information, i.e. unknown cells. The False Negative Rate (FNR) is the rate of N_4 over all occupied cells whereas the False Positive Rate (FPR) is the rate of N_2 over all free cells.

$$FNR = \frac{N_4}{N_4+N_5} \quad , \quad FPR = \frac{N_2}{N_1+N_2} \tag{3.8}$$

3.4.5.3 Localization and perception uncertainties

For a given integrity risk level, the characterization error represented by the FNR must be below the set risk level. A major contribution brought by the LGM is its ability to manage the integrity of the information thanks to its capacity to control the FNR via the sampling step, which has an impact on the overall LGM characterization performance.

In addition to that, integrity has to be evaluated with the impact of the localization errors. As HD maps are used to create the LGM in a global frame coordinates, perceived information in the ego-vehicle local frame is transformed into the global frame thanks to localization. As a consequence, when uncertainty of localization is taken into account, the characterization process has to provide, whatever the situation is, a confidence level of the information stored. In order to evaluate the robustness of the model, some Gaussian noises with different standard deviations have been added on the position of the vehicles ground truth. Thus, the impact of the localization uncertainty on the system can be studied.

At first, it is considered that there is uncertainty on the localization but it is not modeled in the perception outputs. It results in providing uncertain information

Figure 3.13: Aggregation and standardization of the LGM from initial step to another.

Table 3.2: Aggregation rule of two LGM cells. The information is condensed in one LGM cell.

	$Cell_{i+1}$		
$Cell_i$	F	O	U
F	F	O	U
O	O	O	O
U	U	O	U

without being aware of it. It aims to show that despite not handled uncertainties, the sampling step still enables to manage the characterization process.

Usually localization also quantifies its uncertainty, typically with a covariance matrix. This uncertainty can be propagated to the perception polygons by enlarging them. As localization errors on the ego-vehicle position are introduced, the uncertainty is propagated on the object hulls when they are transformed from the local coordinates of the car to the world coordinates.

In order to vary the sampling step, a methodology has been applied to replay the scenario for each step (figure 3.13). To proceed, a minimum step has been taken (step 1 of figure 3.13), which is typically used for the ground truth, and cells have been combined together to increase the step. For instance, with a sampling step of 0:2 m, information of two cells of 0:1 m has been combined by applying an aggregation rule (step i of figure 3.13). The aggregation rule is similar to the fusion rule of cells of an occupancy grid (table 3.2). Then, to be able to compare the ground truth with the initial step, the aggregated cells are split back to the standard minimal sampling step of the LGM used for the ground truth as shown in the last step of figure 3.13.

The aggregation rule enables to obtain in an efficient way the LGM with the desired sampling step. As the LGM is generated at the strategic level, it has a fixed sampling step that cannot vary dynamically to the tactical level requirements. As a consequence, assuming that the sampling step parameter may be set or requested dynamically over time, the LGM can be adapted accordingly.

3.4.6 Experiments

3.4.6.1 Setup

In order to validate the proposed method, experiments were carried out with three vehicles in real road conditions: two experimental Renault ZOE vehicles and one van (a Renault Master). A single dataset with sensors raw data has been recorded on an open road along a trajectory of 2:1 km in an urban area (see appendix A for the setup). The ego-vehicle was following the two others all along the trajectory. That is why, during this scenario, the polygons corresponding to the two perceived vehicles were used to compute the results.

When replaying data in real time, the LGM has been evaluated over 50 m in front of the ego-vehicle as the two others were driving just in front of it (figure 3.14) and generated at a rate of 10 Hz. Firstly, the LGM has been recorded when replaying the dataset with a minimum sampling step of 0:1 m. In a second time, data were replayed in real time in order to simulate the same scenario but applying a Gaussian noise on the ego-vehicle pose. It permits to add an error on the position with the same sensors and perception setup. Six replays are reported with standard deviation varying from 0 to 0:5 m.

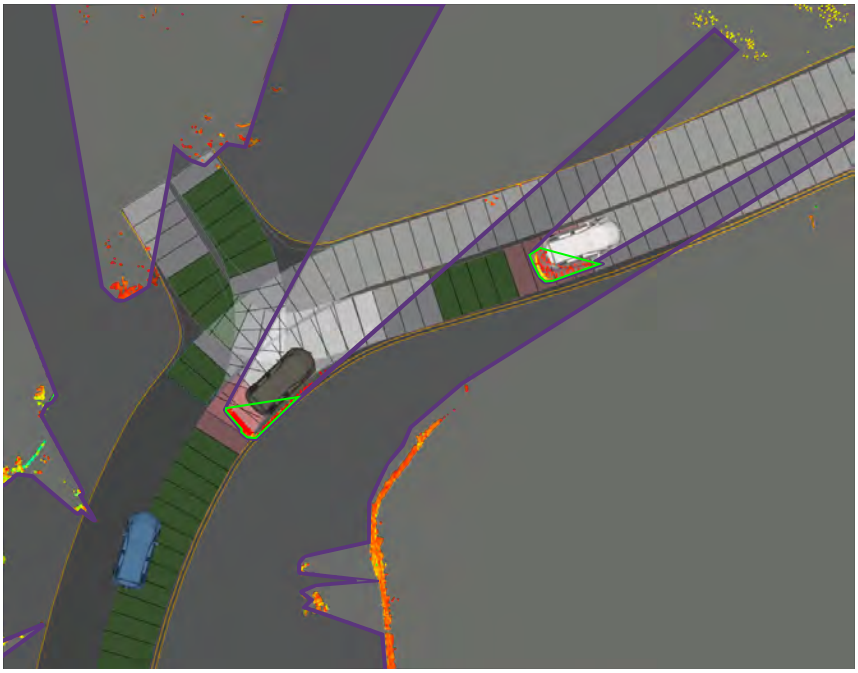


Figure 3.14: Lane Grid Map displayed in RViz with three vehicles. The ego-vehicle (blue) is following the two other ones (gray and white). Each cell is characterized as one of the following states: free (green), occupied (red), unknown (gray). The free space polygon is shown in purple and objects polygons are in green.

For each replay, there are two scenarios:

- The first one, with uncertainty propagation. Each observed polygon object is augmented taking into account the localization uncertainty (see 2.4.3 on page 47). The method presented in [Bernardi et al. – 18] has been implemented.
- The second one, without uncertainty propagation. The characterization process is performed with the polygons of the observed vehicles that do not take into account the noise applied on the localization. It shows a situation where the ego-pose is given to the world model module without information about potential errors.

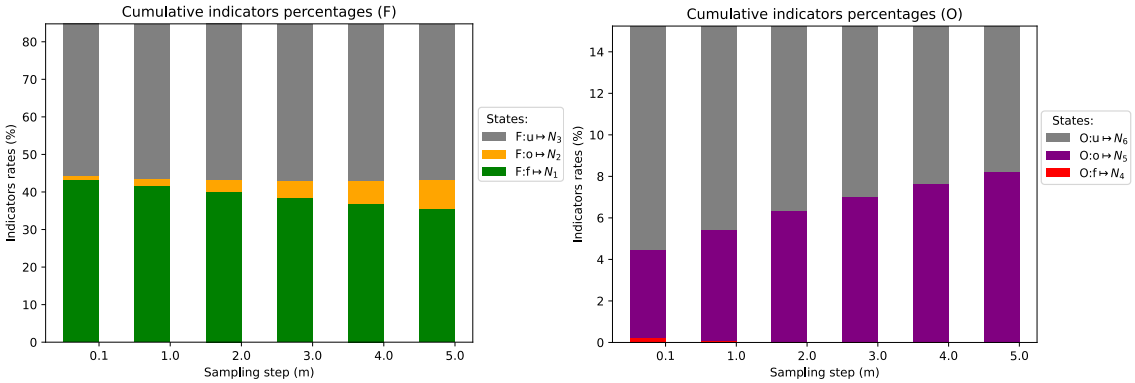
In addition to that, for each scenario a variation of the sampling step from 0:1 m to 5 m with a step of 0:1 m has been added. It gives the equivalent of 6 2 50 = 600 simulations to study the influence of the sampling step, the noise and the uncertainty propagation on the perception polygons.

3.4.6.2 Results

Figure 3.15 depicts the cumulative sum of each indicator N_i for several sampling steps of the LGM cells that are computed at the end of a scenario. Figure 3.15a, shows the indicators related to the truly free cells. As the sampling step increases, N_1 decreases and the misleading overcautious indicator N_2 increases. In figure 3.15b, shows the indicators related to the truly occupied cells. As the sampling step increases N_5 increases and the misleading hazardous indicator N_4 decreases.

Figure 3.16 depicts results obtained by the replay of data without propagating the uncertainty on the polygons of the detected vehicles. On the left of the figure, the FNR of the misleading occupation is plotted in function of the sampling step of the LGM. On the right, the FPR of the overcautious occupation is also plotted in function of the sampling step. Each curve shows the evolution of the rate value depending on the standard deviation of the noise added on the pose.

Figure 3.16a shows a decreasing level of the FNR when the sampling step raises. Indeed, the cells classified as occupied get bigger and, therefore, they cover a larger space that encompasses nearby cells that were incorrectly classified as free because of the localization error. As the standard deviation of the noise increases, the FNR decreases but at a higher level. This shows that it is possible to decrease the FNR under a given integrity risk by increasing the sampling step. For example, if a maximum integrity risk of 0:3 % is set, i.e. typically the risk associated with a 3 Gaussian confidence interval, for the FNR, the integrity requirement can be reached by setting the sampling step to the value depicted by the red dots in figure 3.16. To compute these red dots, the different curves are interpolated in a logarithmic scale (dashed curves) and intersected with the required target integrity risk (TIR). For



(a) (b)

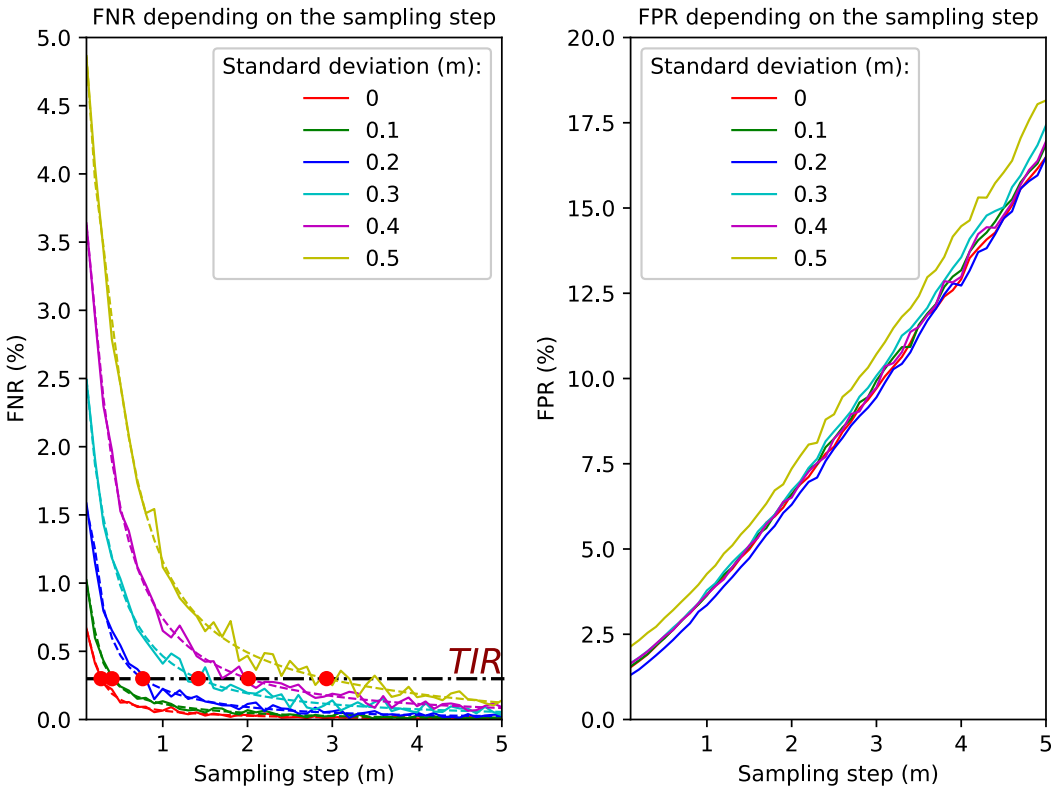
Figure 3.15: Cumulative percentages of the six $\mathsf{N_i}$ indicators. (a)

a localization noise with a standard deviation of 0:5 m, a sampling step of around 3:0 m is needed while for a standard deviation smaller than 0:2, a sampling step lower than 1 m is enough. The disadvantage of sub-sampling is that navigation quantities (such as distances to other vehicles, for example) may be too discrete to properly perform the current task. In other words, this graph shows that if the system not able to model the localization uncertainty, a correct sampling allows to keep the information safe as defined by the functional domain. For a given TIR, higher the uncertainty, higher the sampling step. The question can also be addressed the other way around: oversampling the LGM brings more misleading information since the curves increase as the sampling step gets closer to zero.

In figure 3.16b, one can observe that the evolution of the noise has a little impact on the FPR. As the aggregation rule is conservative on the occupied state, the FPR increases as the size of each occupied cell gets bigger. The close evolution of each curve can be explained by the fact that, whatever the noise is, the size of the occupied polygon is not modified which does not change the impact of the number of the cells that are observed occupied on the FPR.

Figure 3.17a shows the same curves as the previous figure with the difference that the uncertainty is now propagated onto the polygons which are enlarged. As a consequence, they cover a bigger space to ensure covering the correct occupied state. All the curves have the same shape with a much lower FNR. When the uncertainty is properly handled, integrity is kept at a given level for a shorter sampling step. For a localization noise with a standard deviation equivalent to 0:5 m or below, the maximum sampling step needed to remain under the TIR is close to 0:5 m as represented by the red dot in figure 3.17a.

As it can be observed on figure 3.17b, the consequence of making the polygons bigger is a higher variation on the FPR depending on the standard deviation of the noise. It is explained by the fact that the perception system is much more conservative

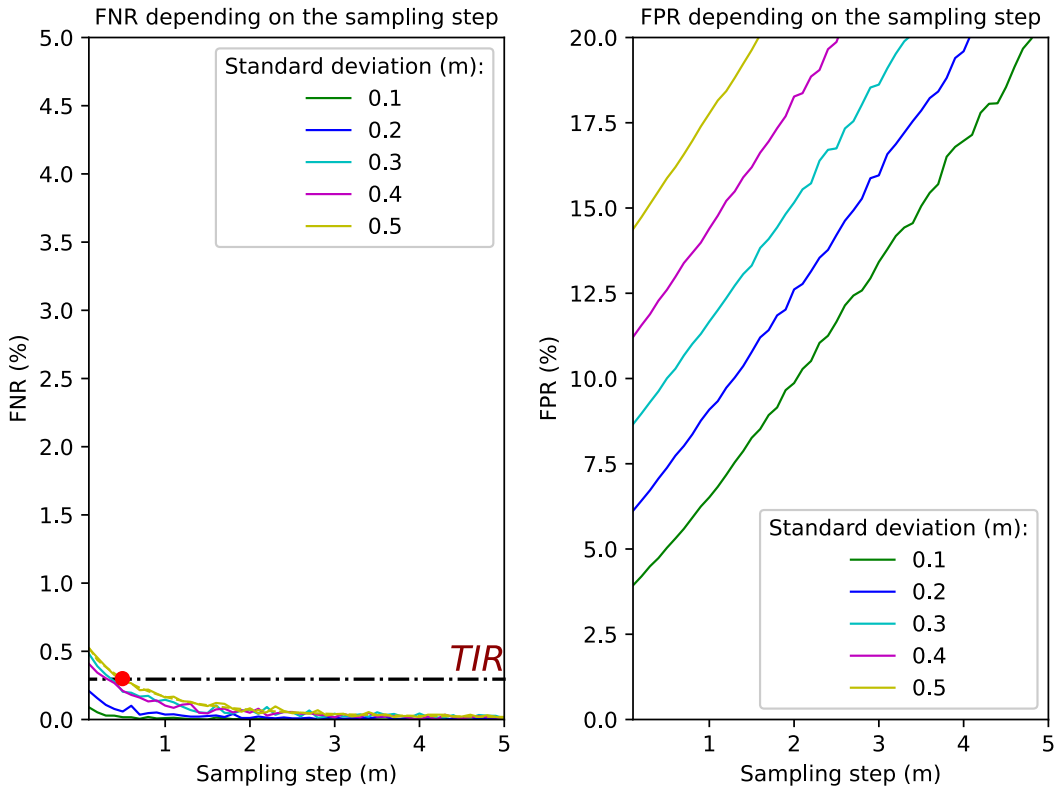


(a) (b)

Figure 3.16: FNR and FPR of the LGM without uncertainty propagation on the polygons of the detected vehicles. Random errors with different standard deviations are added to the localization estimates. The x-axis corresponds to the spatial sampling of the LGM.

as occupied area gets bigger as localization uncertainty increases. When the noise model is estimated and propagated, sub-sampling leads to a loss of efficiency as the FPR grows rapidly for no variation of the FNR. When uncertainty is taken into account and propagated to the polygons, it is possible to decrease the FNR while the FPR increases with the noise.

For a given TIR, with uncertainty propagation, it is therefore possible to lower the sampling step compared to no propagation. It brings the LGM closer to a finer representation. If localization uncertainty is not well defined, in order to keep a target integrity risk at 0:3 % for the FNR, the sampling step parameter should be equal or below 3:0 m. However, if uncertainty is well bounded, the sampling step parameter can be lowered to 0:5 m. Thus, information stored by the LGM may be refined. As a consequence, if the localization uncertainty is well defined, a lower sampling step is suitable in order not to supply misleading information. A low integrity risk level can be chosen. However, if the estimation error is not well bounded, a higher sampling step is needed to keep the same integrity risk level.



(a) (b)

Figure 3.17: FNR and FPR of the LGM with uncertainty propagation on objects hulls.

As a result, if a maximum allowed threshold of error for the misleading indicator, i.e. N_4 value, is chosen, the FNR is able to remain under this threshold thanks to the cells sampling step value. Thus, despite the uncertainty of the perception and the localization, a functional domain for the vehicle is guaranteed. According to the operational design domain, the sampling step parameter can be set accordingly. In practice, once the integrity requirement of the decision-making module has been set in terms of TIR, a data collection is required to monitor the integrity performance and pick the optimal sampling step satisfying this requirement. With a sufficient amount of driving recording, this parameter may be learned from data.

3.5 Enhanced Space Characterization

The following work aims to quantify the available and missing information in the areas of interest. A characterization process of the AOI using reasoning with prior map information has been developed and aims at improving situation understanding in unknown areas, e.g. hidden by other road users.

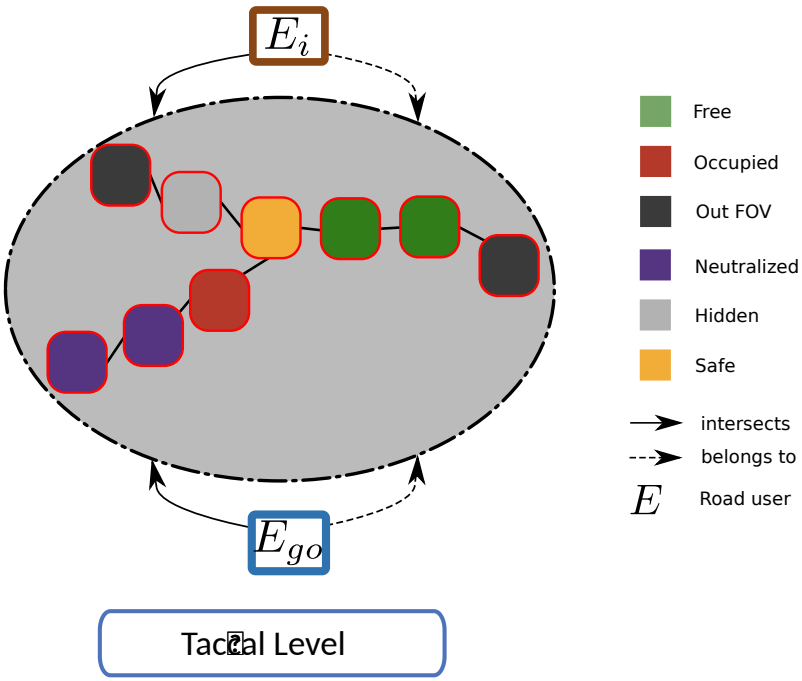


Figure 3.18: Schema representing the LGM part of an IG node.

3.5.1 Tactical representation

At the Tactical level, the lane grid map representation is a lower level representation of the AOI . As the IG is an oriented graph of the paths of interest, figure 3.18 shows an abstracted LGM representation of one path of interest, e.g. a primary order node. Each cell of the LGM represents a node that is characterized by one state. The Out FOV, Neutralized and Safe states shown are further described. As areas of interest are related to the ego-vehicle lane keeping mode, the LGM representation can be seen as a graph of successive states whose edges are represented by a distance to the ego-vehicle path.

Moreover, at this abstraction level, two relations may link a road user to the lane level representation. From an occupancy point of view, a road user may be on one or several lanes. This information is given thanks to the characterization process, provided that the road user is present in the object state representation. In addition to that, thanks to the map-matching process, i.e. a matched link, a road user may belong to one or several lanes. As a consequence, it is of importance to notice that this representation enables to determine if a road user only intersects or belongs to a primary or secondary order path (further details will be discussed when introducing the notion of neutralization).

The following section defines the characterization process of the cells of the LGM and the formalism used.

Table 3.3: Illustration of multi-class fusion.

Ω_1	F		O	
Ω_2	X		$\overline{X}$	
Ω_{12}	F X	F $\overline{X}$	O X	O $\overline{X}$

3.5.2 Theory of evidence frames

This section aims at providing a method to enhance the characterization process of the LGM. Free and occupied states are already handled. However, the objective is to better handle the unknown case. That is why, in order to add new states to better characterize this case, a refinement of the sets is proposed. Because mass functions are directly defined over sets of classes, refinement of information can be easily handled with the theory of evidence.

An example that fuses two frames of discernment is given as follows. Let $\Omega_1 = \{F; O\}$ and $\Omega_2 = \{X; \overline{X}\}$ be two frames of discernment. Table 3.3 shows the combination of the two frames of discernment where X corresponds to a new class with is complementary set $\overline{X}$. Semantically, both Ω_1 and Ω_2 represent the same space, only their respective decomposition differs. The first step is to express each mass function in the common frame. For instance, if $m^1(F) = 0{:}8$ then $m^{12}(\{F X; F \overline{X}\}) = 0{:}8$. The masses are converted such that:

$$
\begin{aligned}
m_1{}^1(F) &= m_1{}^{12}(\{F X; F \overline{X}\}) \\
m_1{}^1(O) &= m_1{}^{12}(\{O X; O \overline{X}\}) \\
m_1{}^1(\Omega_1) &= m_1{}^{12}(\Omega_{12}) \\
m_2{}^2(X) &= m_2{}^{12}(\{F X; O X\}) \\
m_2{}^2(\overline{X}) &= m_2{}^{12}(\{F \overline{X}; O \overline{X}\}) \\
m_2{}^2(\Omega_2) &= m_2{}^{12}(\Omega_{12})
\end{aligned}
$$

The second step is then to use the Dempster Shafer combination rule (see equation 3.6 on page 67) to compute all the masses of the sets of the common frame. The common frame of discernment is thus given by $\Omega_{12} = \{F X; F \overline{X}; O X; O \overline{X}\}$. In this example there are $2^{\mathrm{Card}(\Omega_{12})}$ masses. Two of them are given with:

$$
\begin{aligned}
m_{12}^{12}(\{F X\}) &= m_1{}^{12}(\{F X; F \overline{X}\}) \quad m_2{}^{12}(\{F X; O X\}) \\
m_{12}^{12}(\Omega_{12}) &= m_1{}^{12}(\Omega_{12}) \quad m_2{}^{12}(\Omega_{12})
\end{aligned}
$$

3.5.3 Characterization process

So far, the frame Ω has been decomposed in two different ways:

$$\Omega = \{AOI;\ \overline{AOI}\} = \{F; O\} \tag{3.9}$$

The common frame of discernment would be written as follows:

$$\Omega = \{F \setminus AOI; O \setminus AOI; F \setminus \overline{AOI}; O \setminus \overline{AOI}\} \tag{3.10}$$

As only the space inside the AOI is pertinent, the remaining sets of interests becomes:

$$\begin{aligned} F_{AOI} &= F \setminus AOI \\ O_{AOI} &= O \setminus AOI \end{aligned} \tag{3.11}$$

As the LGM is already a representation that focuses on the areas of interest, the first frame of discernment that is considered for the characterization processed is restricted to $\Omega_1 = \{F_{AOI}; O_{AOI}\}$. In the rest of this thesis, all information outside AOI is no longer considered and the notation $\Omega_1 = \{F; O\}$ is considered instead for simplicity.

In view of this, the next step is to characterize the AOI using perception information from exteroceptive sensors such as cameras or LiDAR. In figure 3.19, perception information is shown with occupancy information represented with polygons. The free space with the limit of the Field Of View (FOV) is displayed with a blue circle. Objects are represented in red inducing a hidden zone in gray.

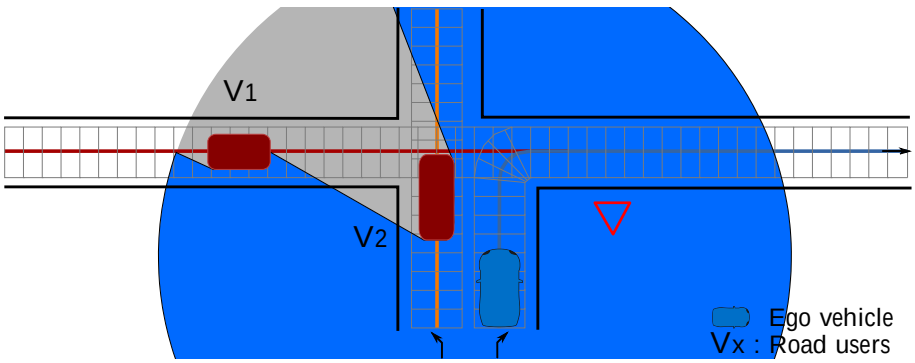


Figure 3.19: LGM example with perception occupancy information. Occupancy is represented by a polygon approach: free space (blue), objects (red).

Table 3.4: Characterization sets table of areas of interest.

	FOV			$\overline{FOV}$	
	Free	Occupied	Hidden		
			Unknown		
AOI	F	O	H	S : Safe N : Neutralized	U

The characterization process that uses the states present in the table 3.4 is presented in detail. The frame of discernment Ω, i.e. the space state, is decomposed into three categories: free (F), occupied (O) and unknown (U). One purpose of this is to handle in a better way missing information in areas of interest. The safe and neutralized concepts that enable to infer additional information where information is missing are presented. A distinction is also made between missing information that is within the FOV from outside the FOV.

Based on table 3.4, the hidden set $H = U \setminus FOV$, that represents the space within the FOV of the sensors but cannot be characterized because of occlusions, is distinguished from the rest of unknown space $U \setminus \overline{FOV}$ that are out of reach from the sensors range. The additional sets of interests proposed are:

$$\begin{array}{lcl} S & & \\ N & & \\ H & = & (U \setminus FOV) \cap (S \cup N) \\ U & = & (U \setminus \overline{FOV}) \cap (S \cup N) \end{array} \tag{3.12}$$

Similarly to the mass functions computed by the polygon based approach, binary masses are also considered on each discernment of frame presented in this section.

3.5.3.1 Field of view frame

The sensor setup of the vehicle can be characterized beforehand in order to determine its physical perception range. Prior to the perception itself, it is possible to define the spatial Field Of View (noted V) of the vehicle. Its complementary $\overline{V}$ corresponds to regions where the vehicle has no means to perceive anything using its own embedded sensors.

$\Omega_2 = V;\overline{V}$ is the second frame of discernment used. A field of view represented by a polygon (Fov) in the frame of discernment Ω_2 is considered (figure 3.20). The

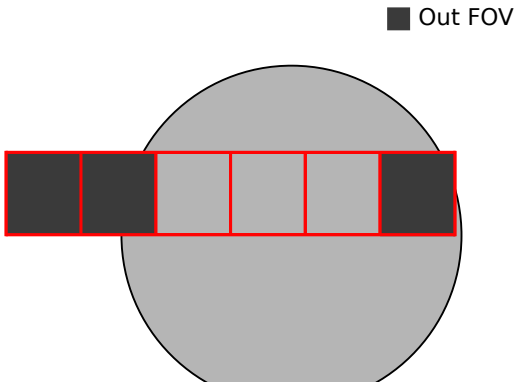


Figure 3.20: LGM characterized by a FOV (the circle represents a LiDAR FOV).

intersection between each cell c of the LGM and the Fov is computed. At the Tactical level, the mass function can be defined as follows:

$$\begin{cases} m(\overline{V}) = 0 \\ m(V) = 1 \end{cases} \quad \text{if } c \quad Fov$$
$$\begin{cases} m(\overline{V}) = 1 \\ m(V) = 0 \end{cases} \quad \text{else}$$

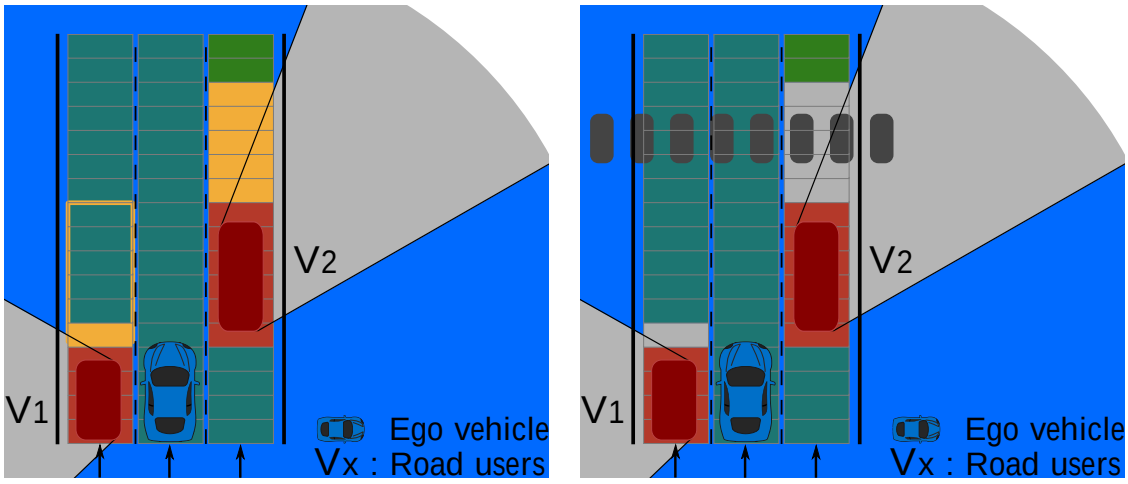
In figure 3.20, the LGM cells characterized as out of the Fov are in black. It should be noticed that in the case of an occupancy grid representation, the Fov could be considered as a map overlay.

3.5.3.2 Safe frame

When driving, the space in front of the road users needs to be obstacle-free for obvious safety reason. This space is defined by a safety distance that is an increasing function of the road user velocity. The faster a road user is driving the more space needs to be kept as free in front of it. This safety distance encompasses the driver's reaction time, the minimum braking distance and safety regulations. Therefore, when a road user is detected and its speed estimated, an area at its front denoted S is defined and is referred to as safe area thereafter.

The figure 3.21 presents a situation where the ego-vehicle intends to overtake a vehicle positioned on its right side. It is assumed here that vehicles constantly keep safe inter-distance with other vehicles so they can safely perform emergency braking to avoid collision in case of dangerous situations. It means that it is likely that the immediate space in front of any vehicle is free.

In figure 3.21a, road users travel at a certain speed. A safe area is defined in front of every detected object (yellow) by using its estimated speed. In figure 3.21b, the



(a) (b)

Figure 3.21: LGM with safe cells characterization on a straight road with only lane change interactions. Areas of interest are modeled and characterized: Free (green), Occupied (red), Hidden (gray), Safe (yellow). (a) and (b) show same perception information. (a) The road users travel at a certain speed. (b) The road users are stopped.

same situation with identical perception information is represented but with stopped vehicles before crossing a pedestrian crossing. In this case, there is no safe space. Therefore, the unknown area in front of it is characterized as hidden (gray). It must pay attention and take a cautious decision unlike the first case where despite the area is hidden (gray), the ego-vehicle may not pay as much attention. Overtaking a bus at a high speed on the highway does not involve the same attention mechanism as overtaking a bus stopped at a bus stop for instance.

In the case where this safe area is visible and characterized as free (as it should be) or occupied (it means that the theoretical safety distance is not respected), then the safe property can be ignored as it is less informative than the free or occupied knowledge. The safe property is therefore only useful to further characterize an unknown area which could be either hidden or out of the FOV.

In addition to the safety distance computation, there are two conditions so that a cell is classified as a safe cell. The first one is to belong to the likely path of its corresponding road user. The second one is that this cell belongs to a unique likely path. Indeed, if a road user may take two paths, both ways cannot correspond at a same time to the space needed for the object to stop.

The frame of discernment used is $\Omega_3 = \{S; \overline{S}\}$. For a sampling step parameter step, the safe area is defined by a number of cells N_b that are within a minimal distance d_{min} in front of a road user E_i such that $N_b = b\frac{d_{min}}{step}c$ with $b:c$ which corresponds to the inferior bound. The binary mass function is defined as follows:

$$\begin{cases} m(\overline{S}) = 0 \\ m(S) = 1 \end{cases} \quad \text{if } c \quad N_b$$
$$\begin{cases} m(\overline{S}) = m(S) \ = 0 \\ m(\Omega_3) = 1 \end{cases} \quad \text{else}$$

3.5.3.3 Neutralized frame

One important aspect of the interaction between road users is that they impose physical constraints among each other. Once a road user occupies a given space of the road, no other road users can cross this space without causing an accident.

In the proposed case study, objects belonging to the second order $AOI^{(2)}$ are used to further characterize the first order $AOI^{(1)}$. Because the lanes in $AOI^{(2)}$ have a direct interaction, e.g. crossing, with the ones in $AOI^{(1)}$, a road user in $AOI^{(2)}$ may obstruct the circulation in $AOI^{(1)}$. This will happen when the road user belongs to $AOI^{(2)}$ and occupies a spatial space that also belongs to a lane in $AOI^{(1)}$. All the space behind this road user along the lane in $AOI^{(1)}$ is neutralized in the sense that no other road users can go through the obstacle. This area is denoted N and is referred to as neutralized area thereafter.

In addition to that, there are two relations that may link any road user to an IG node as illustrated with figure 3.22.

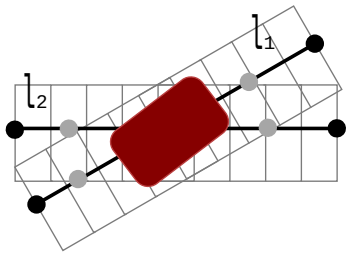


Figure 3.22: Schema of intersects and belongs torelations. The road users (red) belongs tol_1 and intersects l_2.

Definition 3.11 Belongs to relation

A road user E_i **belongs** (denoted) to a node g of the interaction graph IG if the result of the map-matching algorithm is a link $l_i^{(2)}$ that is part of this node:

$$E_i \quad g \ \emptyset \quad l_i^{(2)} \ 2 \ g$$

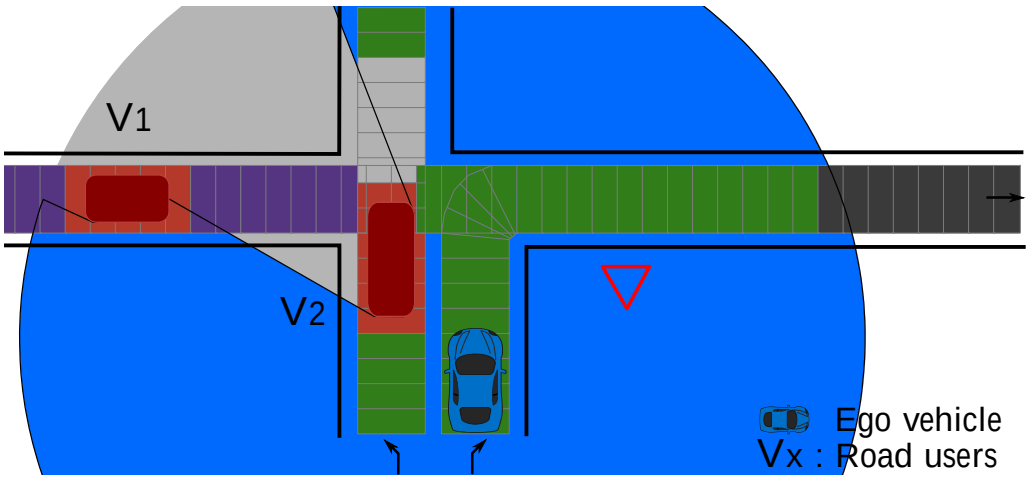


Figure 3.23: LGM with neutralized cells characterization on a T-intersection. Areas of interest are modeled and characterized: Free (green), Occupied (red), Out of range (black), Hidden (gray), Neutralized (purple).

Definition 3.12 Intersects relation

A road user E_i **intersects** (denoted #) a node g of the interaction graph IG if it is on a link $l_i^{(2)}$ that is part of this node but the map-matching did not match with this link:

$$E_i \# g \quad () \quad l_i^{(2)} \; 2 \; g \text{ and } :(E_i \quad g)$$

Property Relation unicity

If a road user E_i belongs to a path g of the interaction graph IG, then it does not intersect it. Similarly, if a road user intersects a path g of the interaction graph, then it does not belong to it.

$$E_i \quad g \Rightarrow : \quad (E_i \# g)$$
$$E_i \# g \Rightarrow : \quad (E_i \quad g)$$

The figure 3.23 presents a situation where a neutralized area is caused by a dynamic object. Within the visible area, the free cells (F) are shown in green. The cells occupied by the red cars are in red (O). The black cells are unknown (U) as they are out of the FOV. The gray cells are also unknown but, as they are inside the FOV and not observed because of occlusions caused by obstacles, they are categorized as hidden (H). Last, the purple cells highlight the neutralized (N) area. Like U and H, N is not observed, but its space is physically obstructed by the presence of other road users. It must be noted that the neutralized area may reach outside of the FOV. As it is neutralized by a perceived road user, the ego-vehicle has the possibility to take a decision under occlusion.

Like the safety case, the neutralized property is only relevant within the unknown category. In figure 3.23, the neutralized property can be ignored where a perceived road user stand as it is less informative than the occupied knowledge.

The frame of discernment used is $\Omega_4 = \{\text{Neutralized}; \overline{\text{Neutralized}}\} = \{N; \overline{N}\}$. For the neutralized characterization, a similar process to the safe characterization is used. A neutralized area occurs when a road user E_i **belongs** to a secondary order node $g^{(2)}$, $E_i \in g^{(2)}$, and when this road user also **intersects** it corresponding primary order node $g^{(1)}$, $E_i \# g^{(1)}$ with $g^{(2)}$ is primary of $g^{(1)}$. The neutralized area is composed of the $g^{(1)}$ cells denoted N_b that are ahead of the cells occupied by the road user. The binary mass function can be defined as follows:

$$\begin{cases} m(\overline{N}) = 0 \\ m(N) = 1 \end{cases} \quad \text{if } c \in N_{area}$$
$$\begin{cases} m(\overline{N}) = m(N) = 0 \\ m(\Omega_4) = 1 \end{cases} \quad \text{else}$$

3.5.4 Common frame of discernment

Once the characterization process for each frame of discernment is done, a common frame of discernment is built. Additionally, only the evidential grid at the operational level is fused over time thanks to the perception module. No fusion occurs at the tactical level between successive LGMs. The characterization process is done each time a LGM is generated at the tactical level.

In order to combine the four frames of discernment that characterize the AOI, the Dempster's combination rule (equation 3.6 on page 67) is applied to compute the different possible masses of the common frame and is further detailed in appendix D. Ultimately, the common frame of discernment is given by

$$\begin{aligned}\Omega_{1234} = \{&F \cap S \cap N \cap V; F \cap \overline{S} \cap N \cap V; O \cap S \cap N \cap V; O \cap \overline{S} \cap N \cap V; F \cap S \cap N \cap \overline{V}; F \cap \overline{S} \cap N \cap \overline{V}; O \cap S \cap N \cap \overline{V}; O \cap \overline{S} \cap N \cap \overline{V}; \\ &F \cap S \cap \overline{N} \cap V; F \cap \overline{S} \cap \overline{N} \cap V; O \cap S \cap \overline{N} \cap V; O \cap \overline{S} \cap \overline{N} \cap V; F \cap S \cap \overline{N} \cap \overline{V}; F \cap \overline{S} \cap \overline{N} \cap \overline{V}; O \cap S \cap \overline{N} \cap \overline{V}; O \cap \overline{S} \cap \overline{N} \cap \overline{V}\}\end{aligned}$$

where each focal element may represent a given cell's situation. However, several focal elements have a zero mass and are not actually observed. Assuming that the vehicle does not receive V2X information, focal elements with non-visible state have a zero mass. Thus there are only eight possible focal elements that may be used as depicted in table 3.5. Each focal element is associated to its corresponding situation representation (circled in red).

Given the classes O, F, S, N, H (see table 3.4 on page 82), table 3.6 shows the correspondence between characterized sets that are relevant and the initial sets. It is reminded that everything outside of the AOI is ignored. Looking at the O class, what is of interest is the belief and plausibility of the space that is both occupied

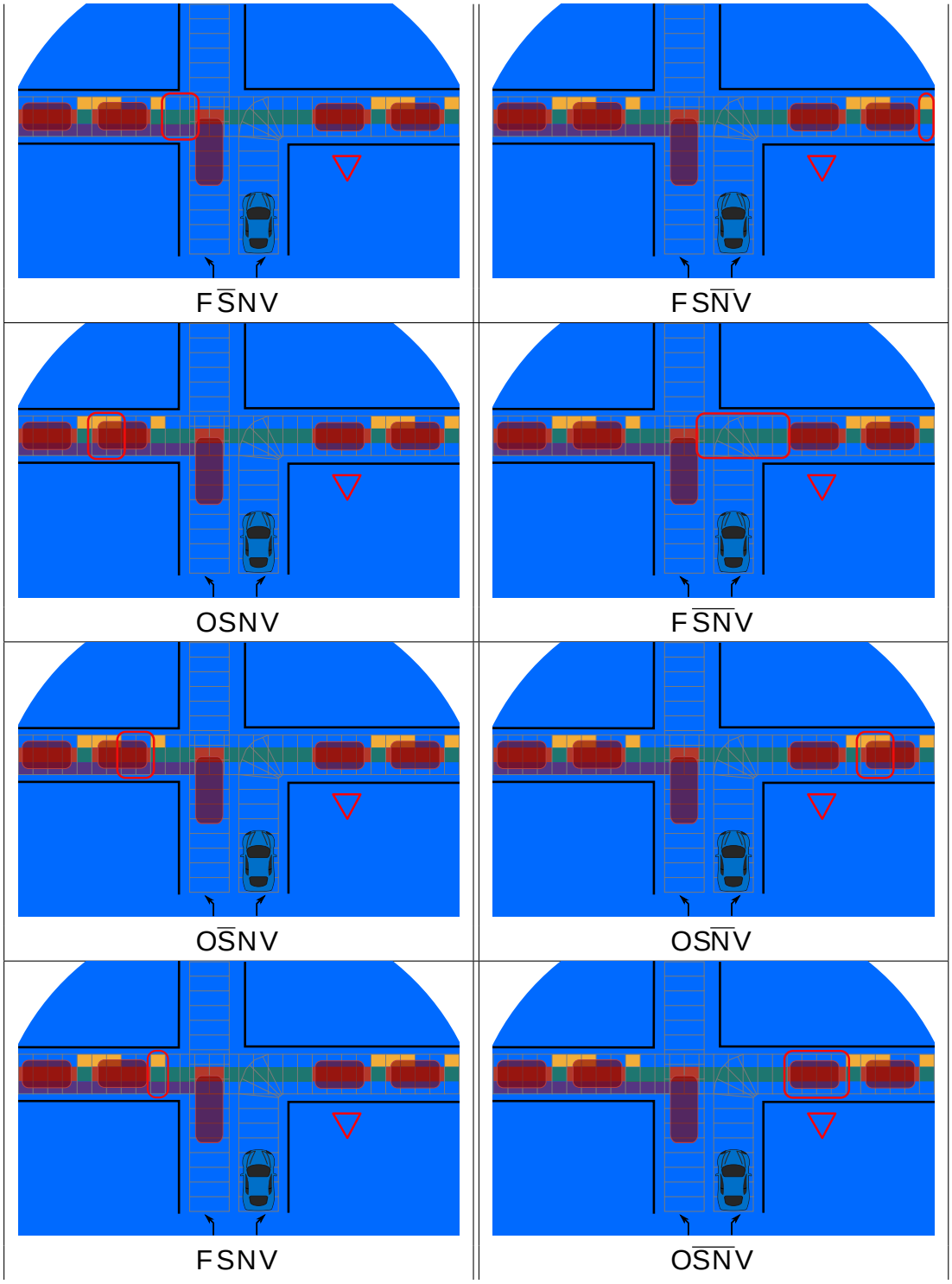


Table 3.5: Situation representation of each focal element of the common frame of discernment within the FOV.

and visible. In order to determine the most relevant state, all the belief functions in the table are computed. Based on these values, a decision-making can define rules or thresholds to determine the final state of the LGM cells.

Figure 3.24 depicts the set diagram of the characterized areas of interest. The free, occupied and hidden states are respectively in green, red and gray. The safe and neutralized states (yellow and purple) reduce the area of the hidden and out FOV areas.

Table 3.6: Evidential sets correspondence to the common frame of discernment.

Class	Notation	Initial sets	Description
Occupied	O	O \ V	Occupied and Visible
Free	F	F \ V	Free and Visible
Safe	S	S ∩ (O ∪ F)	Only Safe and unknown
Neutralized	N	N ∩ (O ∪ F)	Only neutralized and unknown
Unknown	U	V̄ ∩ (S ∪ N)	U is what is out FOV except if it's safe or neutralized
Hidden	H	V ∩ (F ∪ O ∪ S ∪ N)	Hidden is similar to unknown but correspond to what is in FOV

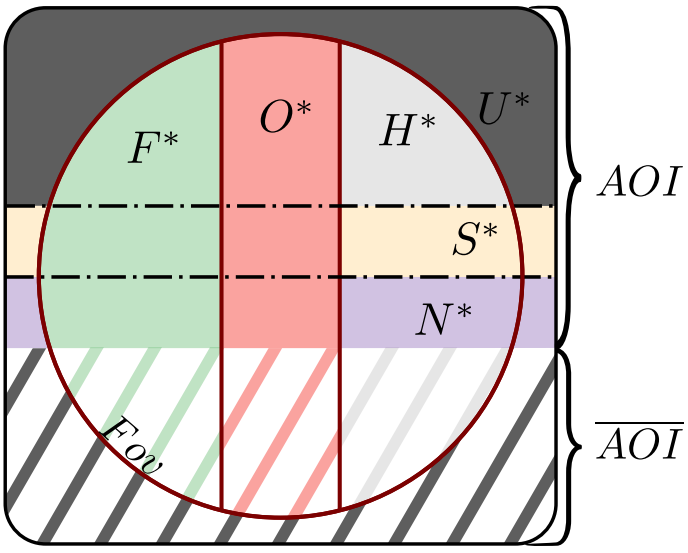


Figure 3.24: Set diagram of areas of interest. Colors show the characterization of the areas of interest: free (green), occupied (red), hidden (gray), safe (yellow), neutralized (purple) and out Fov (dark).

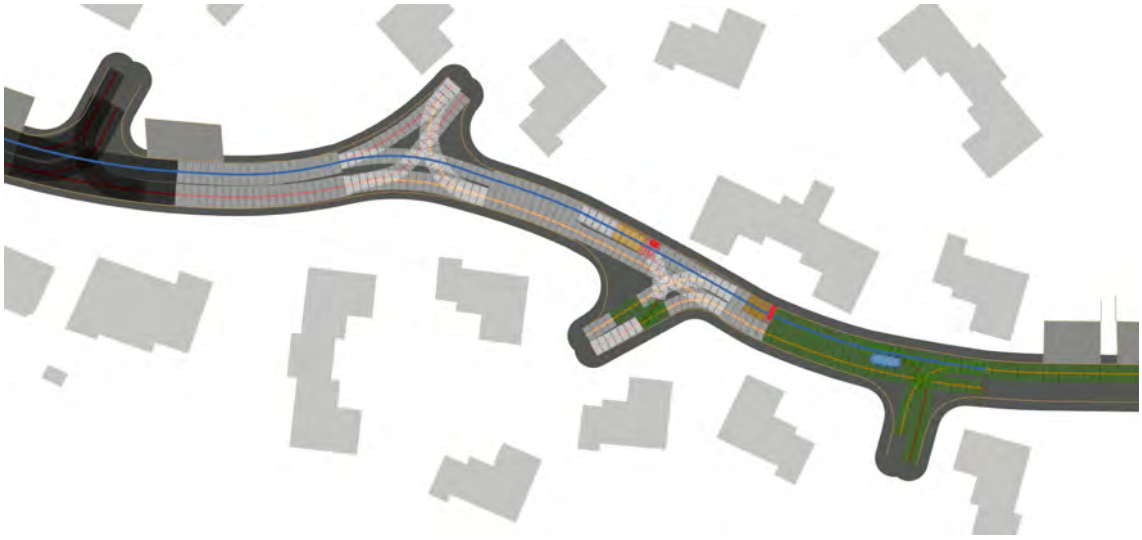
3.5.5 Application and Implementation

The LGM generation and its characterization process have been implemented on a Renault ZOE experimental vehicle of the Heudiasyc laboratory. For the computation of the AOI , a distance of interest of 100 m has been chosen and for the LGM, an arbitrary cell discretization step of 1 m has been set.

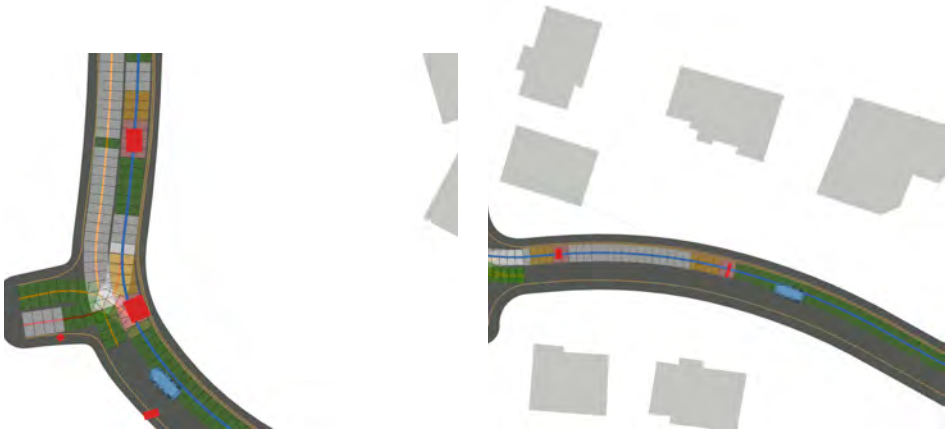
For the computation of the safe space, an emergency braking model with a deceleration of $a = 6\,\mathrm{m\ s^{-2}}$ is used. The safety distance associated to a vehicle driving at a speed v is computed as $d_{safe} = v^2/(2a)$:

The ego-vehicle paths of interest are extracted from the HD map. Figure 3.25 and figure 3.26 show the results of several LGM from real data. In the first one, the ego-vehicle drives on a simple road and follows two other vehicles. It shows safe areas in front of each vehicle as the corresponding LGM cells are hidden. In figure 3.26c, a neutralized area is shown in a roundabout entrance situation. Indeed, the ego-vehicle intends to enter the roundabout by merging with the outer lane. This lane is thus a primary order lane. The lane on the left side of the ego path is crossing the outer lane and is therefore considered to be of second order. As the vehicle on the left belongs to this secondary order lane and is on the corresponding primary order lane, the area behind is therefore characterized as neutralized.

It must be noticed that an advantage of the LGM representation is that it characterizes the surrounding space with respect to the situation context. Indeed, as it can be seen in both figure 3.25 and figure 3.26, there are overlapping cells that are characterized with a safe or neutralized state if they belong to a given link but the same area is considered hidden if the considered cells belong to an adjacent link. Information interpretation depends on the context. In the neutralized case of figure 3.26, cells of the merging link that are occupied by a vehicle that intersect it are different from cells of the secondary order lane that are occupied because the vehicle belongs to it. As the vehicle does not belong to the merging lane, the occupied area does not need to be taken into account in the decision-making process. In other words, it does not correspond to a vehicle that is likely to merge with the ego-vehicle.

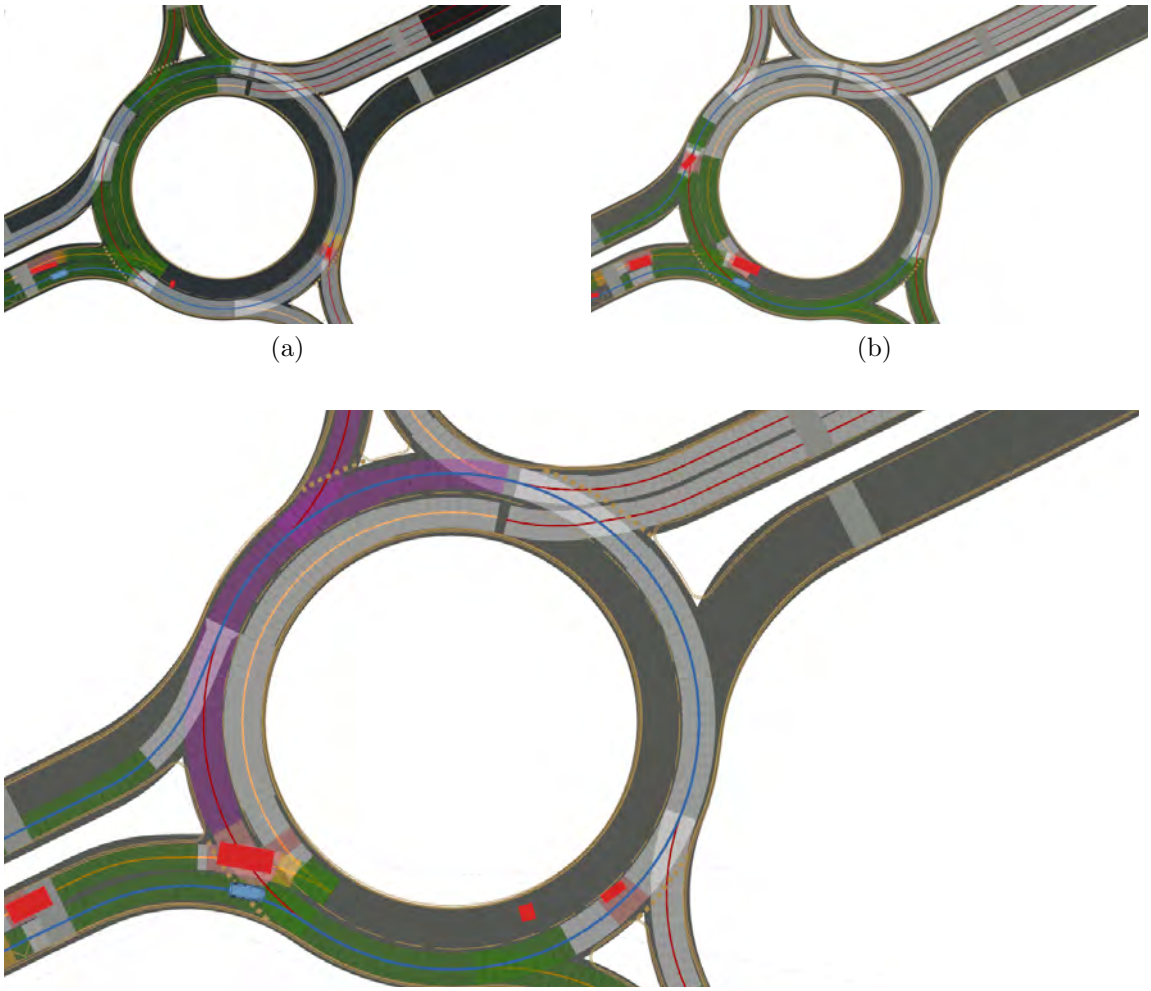


(a)



(b) (c)

Figure 3.25: LGM representations in a real-time experiment led in Compiègne. The ego-vehicle (blue) follows two vehicles (red). The free cells (green) are within the free space. Occupied cells (red) intersect the vehicle polygons. Safe cells (yellow) are in front of occupied cells. Other cells (gray) are partially or completely hidden.



(a) (b)

(c)

Figure 3.26: LGM representations at a roundabout in a real-time experiment led in Compiègne. The ego-vehicle (blue) enters the roundabout with a vehicle on its left side (red). The top figures show the situation before and after the neutralized area appears. The merging lane in the roundabout has free cells (green) and hidden cells (gray). At the bottom, the vehicle located on the secondary order lane causes an occlusion. As the vehicle is already engaged, the primary order lane is neutralized (purple).

3.6 WM Spatial Information Overview

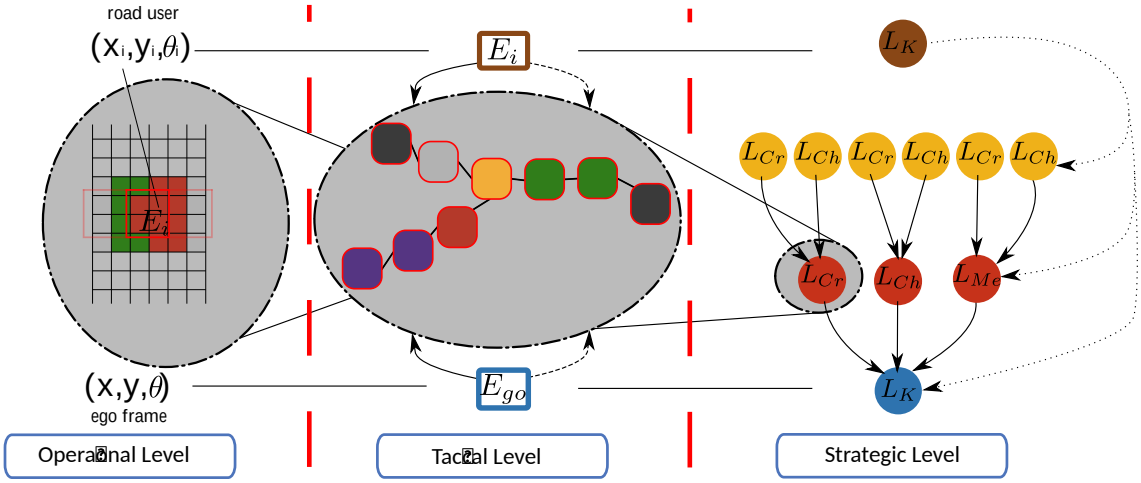


Figure 3.27: Spatial information representation at each level of abstraction. From the left to the right are represented respectively an occupancy grid, a LGM and the IG. Each node of the IG is decomposed in cells that belongs to the LGM and which are characterized thanks to an OG.

This section gives an overall view of the WM space state layer. Figure 3.27 presents the space state representation at each abstraction level as depicted in this chapter. Relations to other layers are represented in a simplified way. It can be observed that the lane keeping mode of road users is computed and can be related to the IG in order to better understand interactions.

At the strategic level, the interaction graph (IG) depicts lanes of interest for the ego-vehicle over several hundreds of meters and over a long time horizon. They are extracted from the topological layer of a map and define the interaction modes encountered by the autonomous vehicle. These areas may be of primary or secondary order and enable the autonomous vehicle to focus on relevant space parts and to look for the corresponding potential interactions. As described, despite the IG takes one future path of the ego-vehicle, it may also apply considering several futures. In addition to that, only the road network has been considered. It must be noticed that the notion of areas of interest can be extended to vulnerable road users spaces. The computation of the areas of interest are limited but could be enhanced: oncoming lanes, diverging lanes, etc.

At the tactical level, an intermediate lane level information representation approach has been proposed with a Lane Grid Map (LGM). This lane level grid is a discretization of the areas of interest and each cell is classified from most informative state: free, occupied, neutralized, safe, to less informative: hidden, out of field of view. This characterization process handles occlusions and enables to improve confidence on information. The safe and neutralized characterizations must enable to take

better decision under occlusion. Indeed, the aim is to build additional knowledge from the situation encountered by the autonomous vehicle. If an area is occluded but is considered as neutralized or safe for the autonomous vehicle, then a possible maneuver can be attempted.

3.7 Conclusion

In this chapter, a lane level representation called Lane Grid Map has been presented. On the bases of areas of interest provided by an Interaction Graph, the LGM aims at providing non-misleading information focusing in these areas. The characterization process enables to classify the LGM cells with one of the following states: free, occupied, safe, neutralized, hidden and out field of view.

It can be noticed that depending on the lack of information, a mechanism of attention and focusing on these zones can be set up. It can also be noticed that the areas of interest identified at the strategic level also allow a feedback to the perception in order to filter data for instance with respect to the map.

The LGM has many advantages. It is a scalable representation that enables to optimize time computing as a selection of information is performed. It also combines both spatial information and distance metrics. Once the characterization process occurred, the distance that separate each cell state to the vehicle can easily be retrieved.

The sampling step allows refining more or less the level of detail of the information stored in the LGM. Since all the cells encode the same amount of information, the more cells there are, the more information is contained in the LGM. The fewer cells there are, the coarser the representation of the situation becomes. In addition, it reduces the risk of providing misleading information to the decision-making module. The sampling step is a key stone to handle the integrity of the generated information. It has been shown how a proper sampling enables to fulfill integrity requirements. By increasing the sampling step, misleading information can be lowered under a Target Integrity Risk.

In the presented situation, the ego-vehicle entered the roundabout despite an occluded area in the areas of interest. For the neutralized case, the decision of the ego-vehicle to enter the roundabout relies on the fact that the neutralized area remained neutralized during the whole duration of the insertion maneuver. This representation space should therefore also be used for prediction purposes where the predicted positions of the tracked objects are used instead.

Chapter 4

Set-based Lane Grid Map Prediction

4.1 Introduction 95
4.2 Related Works 96
4.3 LGM Prediction Formalism 101
4.4 Experiments 116
4.5 Conclusion 133

4.1 Introduction

Once the autonomous vehicle has its own situation representation, i.e. a Lane Grid Map (LGM) in this work, and has multiple comprehension mechanisms that enable to understand the world in which it evolves, it still needs a last and the highest level of understanding which is the prediction of the situation. This mechanism is the highest capability in achieving a full situation awareness as mentioned in [Endsley – 37]. Therefore, the challenge for an autonomous vehicle is to be able to predict a situation in order to anticipate the intentions of the other road users and predict their behaviors. It has directly an impact on the decision-making process as it participates in the goal of taking the proper maneuver and planning the best trajectory without harming other traffic participants and ensuring safety. Prediction enables to plan actions by the decision-making module over a given time horizon. It corresponds to the most likely situations that the vehicle is about to encounter. Thus, through the prediction representation of a situation, many different scenarios as defined in [Ulbrich et al. – 138] can be explored.

Considering the levels of abstraction and the information layers of the WM system presented in section §2.3, prediction is present at each layer (i.e. ego state, space state, objects state) and level combination. As a consequence, the objectives of such a topic, which is a fairly large field of study, first need to be defined.

Indeed, prediction can be achieved at low levels, as the operational level, with only perception information. It concerns the field of trajectory prediction over a short time period. At the tactical level, prediction focuses on a larger time prediction than the operational level and may take into account map information as the topological layer. Prediction at operational level is not addressed in this work. The main interest of this work is to focus on prediction at the maneuver and lane level. It enables to model a situation over several seconds. Once the LGM representation has been settled, the subject of prediction of a situation is considered.

An objective of this work is to study a prediction representation at the tactical level, with the formalism of the LGM from the spatial layer. In intersections areas, interactions between road users may be considered in order to improve prediction. On the basis of some work extracted from the literature [Althoff and Magdici – 4, Koschi and Althoff – 65], the understanding of the current driving situation is exploited in order to improve the prediction mechanisms. A classic set-based prediction is improved by taking into account a new constraint based on a LGM augmentation process. This improvement shows how a better decision is made in specific situations. It also aims at providing to the decision-making a situation prediction with integrity considerations. Misleading information must not be provided in order not to lead to bad decision-makings. Another challenge of the prediction is to provide the most precise information over the longest possible time horizon. This must be done without compromising safety, as behaviors may be unexpected as the time goes by. Thus, a contribution on the integrity of the prediction through the LGM formalism is presented.

This chapter is divided into several parts. First, different prediction methods used in the literature are briefly presented according to the information representation type and abstraction level used. Then, the prediction formalism used here is presented. A section on integrity in a predicted LGM is given. Finally, the experiments section presents some results carried out with real data.

4.2 Related Works

As a reminder, the LGM is an intermediate information representation in the levels of abstraction. As illustrated in [Vendrell et al. – 140], several abstraction levels have to be considered for robot decision planning and prediction. Thus, a brief state of the art of several prediction methods at different representation layers and abstraction levels is presented. As prediction is part of a deeper situation understanding, there are several survey articles on motion prediction such as [Lefèvre et al. – 75] or [Rudenko et al. – 112] where prediction methods may be classified in several categories that are related to the operational and tactical levels (see figure 4.1). These methods may or not take into consideration interactions of the strategic level. Therefore, the methods are presented according to their abstraction level.

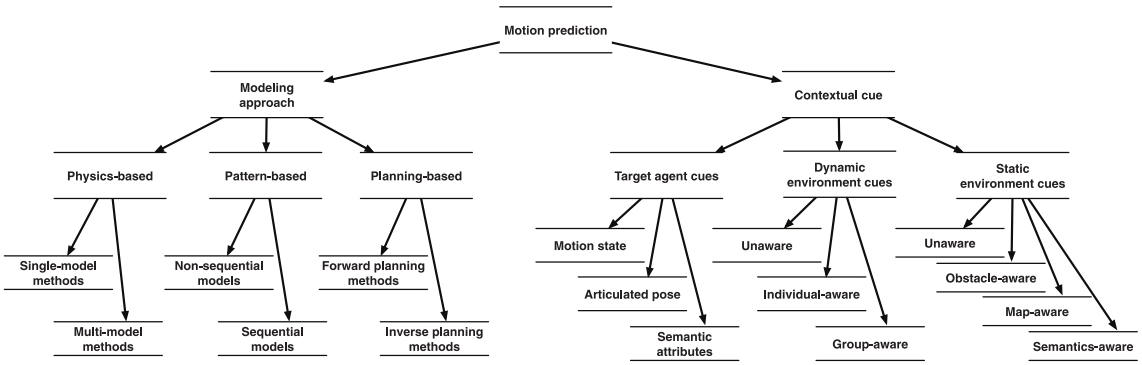


Figure 4.1: Overview of the taxonomy presented in [Rudenko et al. – 112].

4.2.1 Operational level

Two information representations can be distinguished: the one based on objects and the one based on spatial occupancy. At this level of abstraction, prediction of information concerns a situation evolution over a short time and space horizon and barely represents interactions.

4.2.1.1 Objects state layer

There are several works that have been carried out for predicting the trajectory of road users. Motion prediction models can be classified into three main categories: physics-based, maneuver-based and interaction-aware models [Lefèvre et al. – 75]. At this level, prediction aims at providing the closest trajectory that a road user is likely to take over the next few seconds and mainly corresponds to physics-based models. The goal is to predict as accurately as possible the position of road users without "complex" reasoning.

Several approaches can be highlighted. In commonly used strategies, there are numerous model-based approaches such as physical models [Schubert et al. – 123] or, for instance, the intelligent driver model presented in [Liebner et al. – 79]. There is no reaction to other traffic participants and the road topology layout is not taken into account. However, ignoring other information, including map information or interactions with other vehicles, limits the relevance of the prediction, which is why these methods are usually only effective in the short term. In addition, with these methods, it is impossible to anticipate changes in vehicle motions that could be caused by a change in the road context, a reactive behavior or the execution of a new maneuver. Furthermore, a review of human driver behavior prediction is presented in [Brown et al. – 22].

4.2.1.2 Space state layer

At the operational level, some work can be found on space prediction. Dynamic Bayesian Occupancy Grids [Baig et al. – 13] can be used in order to predict in a short time horizon the occupancy of the vehicle surroundings. In [Lee and Kum – 74] a predictive occupancy map is used in order to compute the risk and determine the safest trajectory with the lowest risk level. This prediction is used to compute an emergency trajectory on a short time horizon that corresponds to the operational level. In [Gindele et al. – 46] a prior map is used to improve motion prediction of dynamic cells. However, there are some disadvantages. The computation complexity can become intractable as the prediction computation is related to the grid resolution and the dynamic cells to track. It also has a direct impact on memory consumption. In addition to that, time prediction horizon is limited to the operational level and does not take into account interactions at this level.

4.2.2 Tactical level

This level accounts for lane level information. The prediction time horizon may reach several seconds that encompass maneuvers time execution. The vehicle is thus aware of topological map information. Prediction at this level has the possibility to incorporate interactions and semantic information.

4.2.2.1 Objects state layer

At this level of abstraction, the maneuver-based and interaction-aware models presented in [Lefèvre et al. – 75] are the most commonly used. Indeed, they are both more elaborated and are able to take into account the road topology and semantic information and even the situation and interactions with interaction-aware methods. Figure 4.2 depicts the overview of prediction methods with respect to the prediction level. A maneuver prediction enables to predict at a lane level the trajectory taken by the vehicle.

At this level, more elaborated models are used to predict path intentions or maneuvers of road users. Two main methods in this field may be distinguished: motion patterns [Augustin et al. – 8], and maneuvers estimation [Xie et al. – 146]. However, interactions are not taken into account in these approaches.

In interaction aware models, the assumption that drivers always try to avoid collisions as much as possible is made. They are more complex to implement since it is necessary to model driver reasoning. In [Trentin et al. – 135] and [Schulz et al. – 124], trajectories and maneuvers are inferred with Bayesian networks approaches.

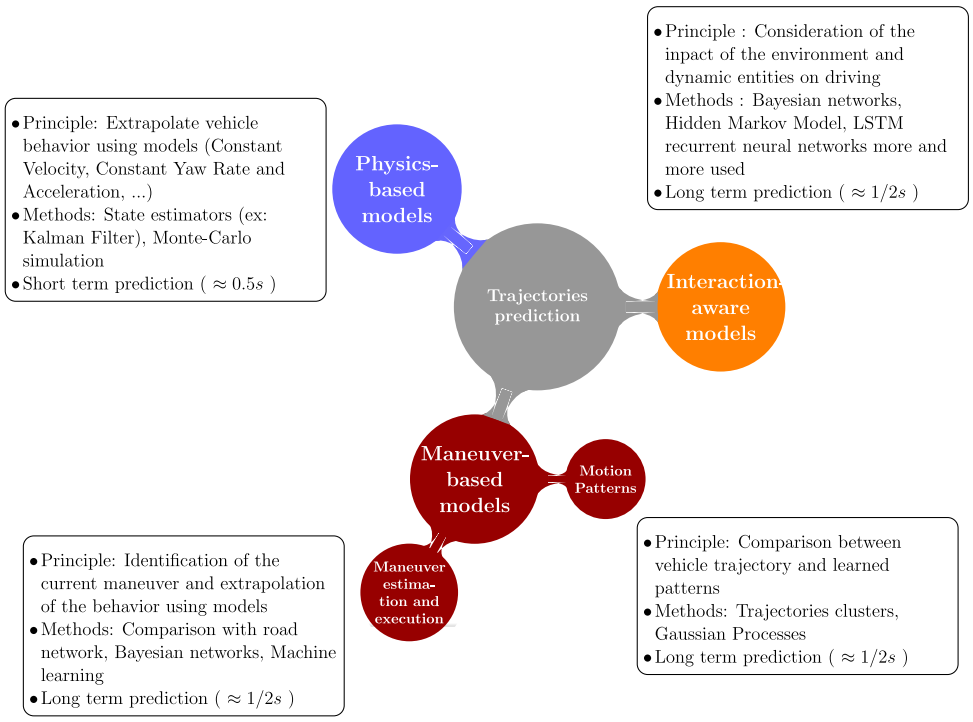


Figure 4.2: Overview of objects trajectory prediction methods.

In a more global aspect, neural network approaches tend to encompass as much as possible every aspect of a situation like prediction, as it does not explicitly model it. For instance, Deep Neural Networks (DNN) [Lenz et al. – 76] or Long Short Term Memory (LSTM) [Jeong and Yi – 62] are more and more used with an example of application in the Apollo Platform [Xu et al. – 147]. Plenty of methods have been studied: Multi-Relational Graph Convolutional Networks (MRGCN) in [Mylavarapu et al. – 95], Convolutional Neural Networks (CNN) that are present in Zoox applications [Hong et al. – 54], conditional Variational AutoEncoder (VAE) with [Ivanovic et al. – 60] but also new approaches with Graph Neural Network (GNN) used in the vector net world representation of Waymo [Gao et al. – 42]. Neural networks will greatly simplify the consideration of interactions since they will avoid any reasoning but have significant disadvantages, such as the large amount of data required for learning and the dependence of the result on these data. No focus is done on these approaches as they are not interpretable representations. In addition to that, these approaches are object-oriented and the will here is to address the spatial occupancy prediction of LGM that corresponds to the space state layer.

4.2.2.2 Space state layer

In [Althoff and Magdici – 4], reachable areas of independent vehicles, called reachable sets, are computed. Taking into account individual constraints, a reachable

space is computed for each vehicle. The solution thus returns, for a given time horizon, the space that can be occupied, i.e. likely to be reached, by a road user (typically a vehicle) during this interval.

In order to infer prediction of road users, a stochastic approach can be used as in [Hubmann et al. – 57]. Given an evolution model and a probabilistic density function, the behavior of a road user can be predicted on its corresponding lane trajectory. Uncertainties are managed in the along track direction. Reachable sets with a stochastic approach is presented in [Althoff et al. – 5] but following works moved to reachable sets based only on a set-based prediction.

There are several articles based on reachable sets that have been deepen [Althoff and Magdici – 4, Koschi et al. – 67, Magdici et al. – 81, Sontges et al. – 129] including a recent one with decision-making applications [Manzinger et al. – 83]. The main work is presented in [Koschi and Althoff – 66] within the SPOT project. In order to take into account interactions between road users, an intersection of different reachable sets is performed. This area is bounded by the road layout and follows some constraints. For instance, a negative speed is not allowed as vehicles always go forward. Reachable sets are computed and propagated in the road lanes directions thanks to topological information. Methodologies are presented to compute inferior and superior borders of the reachable sets depending on several parameters (speed limit, engine torque, maximum acceleration, etc.). In some other work [Koschi and Althoff – 65], some interactions are taken into account by adding constraints in order to refine the reachability analysis strategy. When a vehicle cannot overtake a vehicle on a single lane, its prediction must remain consistent. Their order on the lane is kept as they cannot overtake each other on a same lane. In another article [Koschi and Althoff – 66], a safety distance rule is added. When changing lane, safety distance must be kept, which results in adding new constraints to the prediction model. It should be noticed that a recent work has been done to add reinforcement learning with a safety layer based on reachable sets [Krasowski et al. – 68].

Reachable sets can easily be adapted to the LGM structure and represented in a predicted LGM. Indeed, reachable sets can be seen as a list of polygons. It should be remembered that the characterization process of a LGM works very well with polygon occupancy. As a consequence, the same process is applied with the reachable sets polygons. However, it will over bound the reachable space as shown in figure 4.3. The higher the sampling, the finer the representation. As a consequence, reachable sets can be seen as a finer representation of the LGM. Indeed, if a lateral sampling step and a longitudinal sampling step parameters are used, when they tend toward zero, the LGM representation tends to a continuous representation of the space and thus represents the continuous representation of the reachable sets. As parameters value increase, approximation is made and the discrete representation encompasses the continuous one.

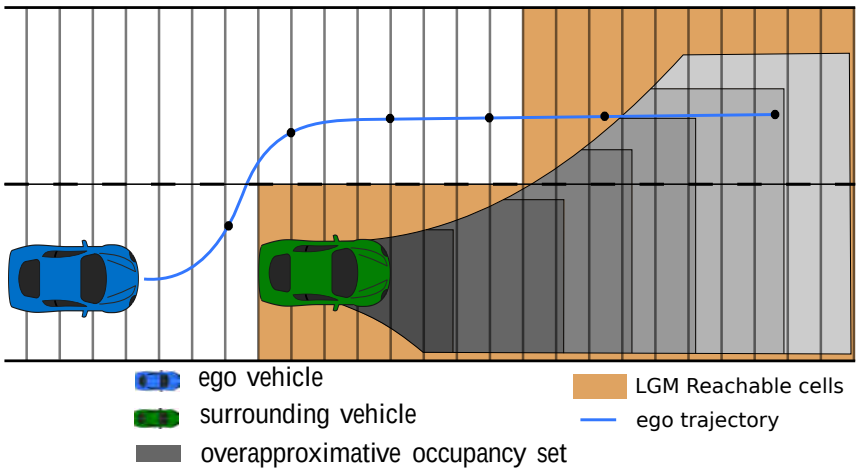


Figure 4.3: Reachable sets overlay on a LGM. Adaptation of the image from [Magdici et al. – 81].

4.3 LGM Prediction Formalism

4.3.1 Problem statement

This section shows how the prediction with the reachable sets approach can be easily used with the LGM representation. A strategy to enhance the prediction strategy is presented.

As discussed, this work aims at enhancing the reachability analysis. In particular, neutralized areas generated by road users are taken into account in order to add an additional constraint in the prediction process. When an area is characterized as being neutralized, road users or potential ones in this area cannot cross and go through. Figure 4.4 shows a good example of a situation where a neutralized area is detected. The ego vehicle wants to enter a roundabout but has no visibility due to the presence of a vehicle on its left side. As this vehicle is entering the roundabout, a neutralized area is characterized as described in the characterization process (see 3.5.3 on page 81). Thanks to the information stored in the LGM, interaction constraints are added in order to refine and improve the reachable sets without introducing misleading information.

Integrity of information provided by a predicted LGM is then addressed. The second main interest of the prediction work is not to explore a new method for predicting object behaviors but to bring additional constraints to increase the availability of the system under an integrity risk perspective. The goal to achieve is to remain consistent without introducing misleading information in this situation. It is explained how integrity of a predicted LGM is ensured with basic set-based prediction strategies. The ability to take decision even under occlusions without compromising safety has to be highlighted.

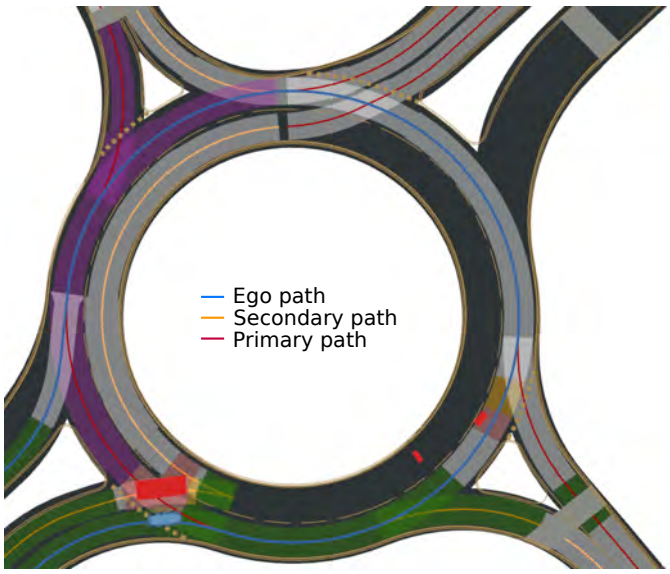


Figure 4.4: Use case of the prediction evaluation. Snapshot of a real situation that happened in the city of Compiègne. The ego vehicle (blue) follows its ego path (also in blue). A vehicle (red) located in the area of interest enters the roundabout at the same time, generating an occluded area. This area is characterized as being neutralized (purple). Free cells are in green and hidden cells are in gray.

This work focuses on the particular situation shown in figure 4.4 to explore prediction with the notion of neutralized case. The ego-vehicle (blue) enters the roundabout with a perceived vehicle (red) at its left side that is causing a neutralized area (purple). It aims to show that a safe decision can be taken and improved compared to classical approaches: the ego vehicle may enter the roundabout despite an occluded area. If the decisions taken are based only on perceived road users, a vehicle that is hidden in an occluded area cannot be taken into account and may lead to a bad decision making. This situation cannot be handled as the vehicle would enter the intersection despite the occlusion. When occluded areas are taken into account, what usually happens is overcautious decisions. Here with the proposed approach, availability can be increased without compromising safety. As a consequence, the prediction study will only be done in this chapter when a situation that has a neutralized case is encountered at a given time noted t_0. In this situation, new constraints are added while keeping consistency in the LGM prediction.

4.3.2 Notations for prediction

A LGM is generated at a given time t_0 and used to predict future states of the cells of this LGM. For each predicted time t in a time horizon $[t_0; t_0+\Delta t_N]$, the result gives a new LGM called predicted LGM denoted $\mathsf{LGM}\ (\mathsf{tjt}_0)$. In this chapter, it is assumed

that each cell c of the LGM, at a given time t denoted LGM (t), is characterized by a single state value among free, occupied and unknown (which encompasses the hidden, safety, neutralized and unknown states). Similarly, each cell of the predicted LGM has a state value among the states: F for free, O for occupied, U for unknown. As a consequence, a cell defined as reachable is a cell that is occupied (O state) or may be reached (U state) in the future. Free cells (still in the future) correspond to unreachable cells. An illustration of the prediction states notions is provided with table 4.1. The initial LGM cells that have a state different from free need to be predicted in order to determine cells that have a reachable state in the predicted LGM. The remaining cells belongs to its complementary, i.e. non-reachable cells (free).

Table 4.1: Prediction states and their relationship with reachable sets.

Non reachable	Reachable	
Free (F)	Occupied(O)	Unknown (U)

The LGM is therefore defined as a set of cells LGM $= \{c_i | \forall i \in [0; N]\}$ where each cell c_i is represented as a bounded interval of curvilinear abscissa $[\underline{c_i}; \overline{c_i}]$ along the center polyline.

As a neutralized area has been defined in a LGM, it can also be characterized in a predicted LGM. For the following prediction work, the assumption of an initial neutralized situation at time t_0 is made. As a consequence, during the prediction, a key point is to estimate how long the situation will remain with a neutralized lane, as it allows the ego vehicle to navigate while being protected. The duration of this neutralization is called Neutralized Time Interval (NTI).

The conversion of a LGM to a predicted LGM with the corresponding state values is subsequently described. The successive prediction steps are summarized in figure 4.5. First, the reachability notion is addressed. It describes the process that aims at defining reachable cells. Then, the occupation function is introduced and enable to characterize occupied cells. Finally, the last step is to take into account the neutralization constraint to update the final LGM.

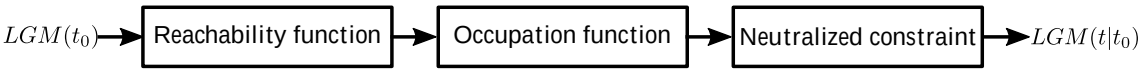


Figure 4.5: Prediction steps.

4.3.3 Reachability

For a given prediction time horizon Δt such that $t_0 + \Delta t \in [t_0; t_0 + \Delta t_{max}]$ a predicted LGM is created. The prediction stage first initializes all cells of the predicted

$\mathsf{LGM}\,(\mathsf{tjt}_0)$ as free (F). The reachability of the road users and hidden space is first addressed before the predicted LGM update. Then the occupancy analysis and the neutralization constraint are considered afterwards.

4.3.3.1 Road users

In order to predict cells that may be occupied by road users, their bounds are first predicted along the center line of the road lane as an along-track curvilinear abscissa interval $\mathbf{s}(t) = [\underline{s}(t); \overline{s}(t)]$, which can be seen as a simplified reachable set. Therefore, each distance, i.e. the upper and lower bounds, can be seen as a curvilinear abscissa that fits well the LGM representation. As road users also comply with the traffic laws, a reasoning over a distance along the curvilinear abscissa is made when computing reachable areas. A given prediction module, called an object predictor, that provides these bounds at a given time is used. The focus is on how to use it and how to manage its unavoidable imperfection. A common model with unknown acceleration has been implemented in the object state representation in order to obtain road users predicted bounds. The curvilinear abscissa is then given as follows:

$$\begin{array}{l} \underline{s}(t_0 + \Delta t) = \underline{s}(t_0) + \underline{v}(t_0) \quad \Delta t + \frac{1}{2}\underline{a}(t_0) \quad \Delta t^2 \\ \overline{s}(t_0 + \Delta t) = \overline{s}(t_0) + \overline{v}(t_0) \quad \Delta t + \frac{1}{2}\overline{a}(t_0) \quad \Delta t^2 \end{array} \tag{4.1}$$

where s is the curvilinear abscissa, v the velocity and a the acceleration under the hypothesis that the acceleration remains constant in the time interval $[t_0; t_0+\Delta t]$ and that the velocity is positive and bounded by v_{lim}, the maximum allowed velocity.

At the intermediate level representation, the lateral prediction of a road user is handled by its taken path. It is assumed that a road user prediction may extend on several paths as long as it belongs to the LGM. The capability of a vehicle to drive on a path outside the LGM is no more relevant for the ego vehicle as it implies that there is no further interaction with this vehicle.

4.3.3.2 Hidden space

For handling the hidden space, the same principles are applied. In the hidden space, there may be road users. If there are hidden vehicles, their lengths are unknown. Therefore, in the worst case, each cell can contain a potential road user whose speed is unknown. For each hidden cell, it is considered that there may be a virtual road user in the interval $\mathbf{s}(t_0)$ which corresponds to the bounds of a hidden cell. The bounds are predicted as the worst case scenarios, corresponding to a constant velocity equals to zero or the maximum allowed velocity v_{lim}:

$$\begin{array}{rcll} \underline{s}(t_0 + \Delta t) & = & \underline{s}(t_0) & \\ \overline{s}(t_0 + \Delta t) & = & \overline{s}(t_0) + v_{lim} & \Delta t \end{array} \tag{4.2}$$

Hidden cells prediction model is derived from equation (4.1). Non-observable parameters are assumed for speed and acceleration.

4.3.3.3 LGM update

Once bounds of road users and hidden cells have been computed. The predicted LGM (tjt_0) is updated with this information. Thus, the reachability function is defined as follows.

Definition 4.1 Reachability function

The reachability function R defines the cells c of the predicted LGM (tjt_0) with a reachable state U. For each predicted entity (real or virtual road user) lying at time t within an interval $\mathbf{s}(t) = [\underline{s}(t); \overline{s}(t)]$, the set of reachable cells $\mathsf{R}(\mathbf{s}(t))$ in the predicted LGM (tjt_0) is defined as all the cells c that intersect $\mathbf{s}(t)$.

$$\mathsf{R}(\mathbf{s}(t)) = \mathsf{fc\ 2\ LGM\ (tjt}_0\mathsf{)jc} \setminus \ \mathbf{s}(t)\ 6 = \mathsf{g} \tag{4.3}$$

All the reachable cells are updated at this stage with the label Unknown.

Figure 4.6 shows the prediction stages with the LGM update and the reachability function applied to the two road users for a time horizon Δt. Figure 4.6a shows a LGM (t_0) that has cells with free and occupied states at time t_0. Two vehicles V_1 and V_2 are considered (the ego vehicle is not represented). In figure 4.6b, the predicted bounds of the road users V_1 are represented with dotted lines. The cells that encompass the prediction bounds of each vehicle are retrieved. In figure 4.6c, LGM (tjt_0) is updated by defining cells with an unknown state (orange). It must be noticed that the overlapping distance of V_2 , i.e. the distance that corresponds to the overlapping area of the road user footprints, is shorter than the sampling step in this figure.$\overline{s}_1(t_0)$

Now, the same situation is analyzed but taking into account a hidden area. Figure 4.7 shows the prediction stages with the LGM update and the reachability function applied to a virtual road user for a time horizon Δt. In figure 4.7a a LGM with a hidden area is presented, i.e. cells characterized with a hidden state in the LGM (t_0). A potential hidden vehicle is represented by a virtual road user whose length L is unknown. In figure 4.7b, the predicted bounds of the virtual road users $V_{virtual}$ are represented with dotted lines. The cells that encompass the prediction bounds of each vehicle are also retrieved. In figure 4.7c, LGM (tjt_0) is updated by defining cells with an unknown state (orange). In this case, there is no longer hidden state as it becomes reachable. Indeed, in a cautious strategy, hidden areas may be reachable as their lower bounds assume a null speed. As a consequence, in the predicted LGM, cells with a hidden state generate predicted cells with an unknown state U.

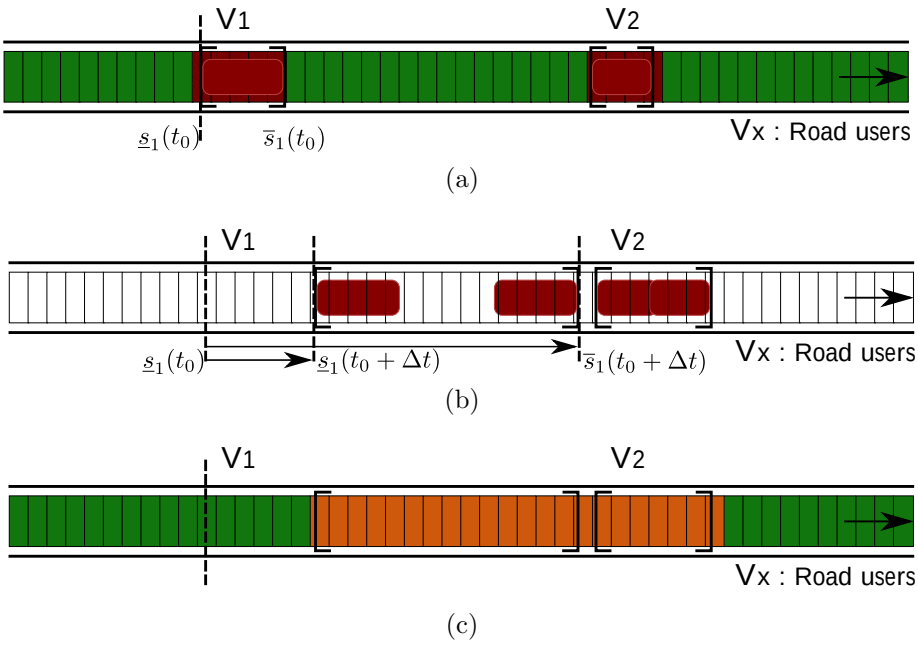


Figure 4.6: Illustration of prediction stages of LGM ($t_0 + \Delta t$jt) with road users prediction at time horizon Δt. (a) LGM (t_0) example with free (green), and occupied (red) cells. The rear boundary of the vehicle V_1 is represented by a dotted line. (b) The upper and lower bounds of the road users model based prediction are displayed. The footprint of the vehicle is displayed at each end of its reachable set. (c) The predicted LGM is updated with possibly occupied (orange, unknown) states. The overlapping distance of V_2 is shorter than the sampling step.

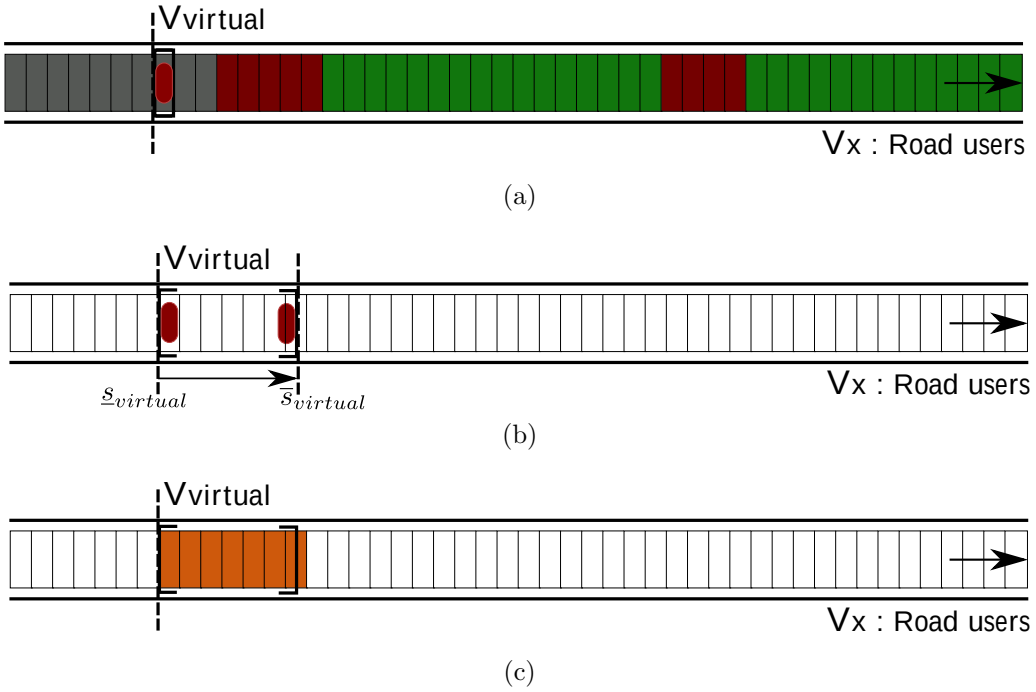


Figure 4.7: Illustration of prediction stages of LGM $(t_0 + \Delta tjt)$ with a hidden area prediction at time horizon Δt. (a) LGM (t_0) example with free (green), hidden (gray) and occupied (red) states. Each cell of the hidden area may be considered as a virtual road user of length equal to the sampling step parameter. (b) A hidden cell is considered as a virtual road user. Its model-based prediction is represented with its upper and lower bounds. Considering a minimum speed of 0, $\underline{s}_{virtual} = 0$. $\overline{s}_{virtual}$ is dependent to v_{lim} . (c) The predicted LGM is updated with possibly occupied (orange, unknown) states.

In this prediction step, independence between cells is made. This prediction is similar to the objects prediction with the assumption that the virtual road user length is of the size of the sampling step.

4.3.4 Occupancy

At time $t = t_0 + \Delta t$, a cell is occupied (with a state O) if it is completely overlapped by the footprint of a road user put at each bound of $s(t)$ (see figure 4.8a). Indeed, only a cell that is fully occupied can generate a neutralized area, i.e. whatever the position of the road user prediction. As a consequence, a partially occupied cell is only considered as reachable, as it is not actually fully occupied.

Road user lengths L are supplied by the objects tracker to the objects state layer. Thus, their footprint occupancy is defined as follows.

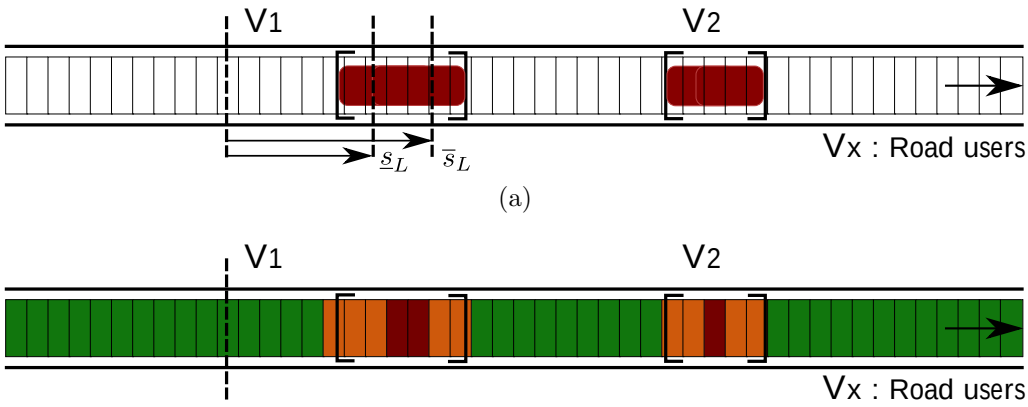


Figure 4.8: Illustration of prediction stages of LGM $(t_0 + \Delta tjt)$ with occupancy prediction at time horizon Δt. (a) Objects model based prediction with upper and lower bounds and with an overlap. The footprint of the vehicle is displayed at each end of their reachable set. (b) Final predicted LGM with an additional occupied (red) state, with the free (green, non-reachable), and possibly occupied (orange, unknown) states.

Definition 4.2 Footprint occupancy

For a given road user in the interval $s(t)$, its footprint occupancy $s_L(t)$ is defined as follows:

$$s_L(t) = \begin{cases} [\overline{s}(t) \quad L; \underline{s}(t) + L] & \text{if } \underline{s}(t) + L > \overline{s}(t) \quad L \\ ; & \text{otherwise} \end{cases} \tag{4.4}$$

In particular, if the predicted interval $s(t)$ is twice as long as the road user length L, then there is no occupancy.

The predicted LGM (tjt_0) is updated with this information. Thus, the occupation function is defined as follows.

Definition 4.3 Occupation function

The occupation function O of road users defines cells of the predicted LGM with an occupied state O. The set of predicted occupied cells $O(s_L(t))$ is defined as all the cells c in the predicted LGM (tjt_0) that are included in $s_L(t)$:

$$O(s_L(t)) = \{c \in \text{LGM}(tjt_0) \mid c \in s_L(t)\} \tag{4.5}$$

All the predicted occupied cells are updated with the label Occupied. Figure 4.8 illustrates a predicted LGM update with the occupation function given two road users V_1 and V_2 for a time horizon Δt. Every road user is predicted as an interval that generates occupied cells.

In figure 4.8a, the predicted bounds of the road users are represented with their footprint put at each ends of their bounds. The cells that are within the overlapping bounds of each vehicle are retrieved. In figure 4.8b, LGM (tjt_0) is updated by defining cells with an occupied state (red). The distance that corresponds to the overlapping area of each road user footprints is greater than the sampling step in this figure.

It must be noticed that hidden cells cannot generate occupancy but only reachability as prediction is done only with virtual vehicles. Indeed, there is no guarantee that they correspond to a real physical road users.

4.3.5 Neutralization

This section focuses on the prediction of a LGM with a lane characterized as neutralized. A particular attention is paid in the case where this lane remains neutralized for a defined time interval. This is the last step of the prediction phase.

As applying a constraint on a vehicle prediction is dependent of the prediction of another road user in this case study, the question of prediction order of road users arises. Indeed, the prediction of cells located in a neutralized area are constrained if and only if this area remains neutralized. An area can only remain neutralized if the prediction of another road user generating this situation has been done. It seems complicated to make predictions of some road users dependent from some others as their number can increase drastically in a driving situation. That is why, in the previous stages, all the road users have been predicted independently w.r.t. each other without any constraint.

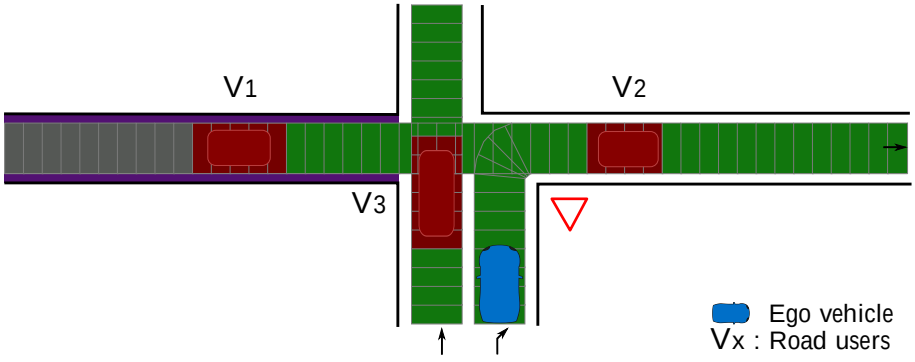


Figure 4.9: LGM at time t_0 with free (green), hidden (gray), and occupied (red) states. In this situation, a neutralized situation is represented (purple). The ego vehicle is in blue and intends to turn right without visibility on its left side.

In the final stage, if the neutralized area remains neutralized, the physical interaction induced by neutralization needs to be back-propagated on the reachable cells of the

road users and potential hidden ones located in the neutralized areas. It prevents prediction of some cells from reaching non-reachable areas.

Figure 4.9 shows the LGM with a neutralized case at t_0. V_3 is generating a neutralized situation as the ego vehicle intends to turn right. N denotes the set of neutralized cells (purple). The objective is to include the neutralized state into the LGM prediction.

For a predicted $LGM\ (tjt_0)$, there are two possible scenarios:

1. The neutralized area remains neutralized in the predicted LGM at time $t_0+\Delta t$ (see figure 4.10). As a consequence, a constraint is applied on reachable cells which were generated by road users or potential ones in the neutralized area. This backward constraint propagation is defined as follows.

   Definition 4.4 Backward constraint propagation

   Let the set $fs_1(t_0); s_2(t_0); : : :g$ be the set of entities (real or virtual road users) that are in the neutralized area N at time t_0. At a given time $t = t_0 + \Delta t$, let $O(t) = R(fs_1(t); s_2(t); : : :g)$ be the reachable cells of all these entities. Assuming that N is still neutralized in the predicted $LGM\ (tjt_0)$, all the cells in $O(t)$ outside of the neutralized area N defined as

   $$C = O(t) n N \tag{4.6}$$

   are actually not reachable by the entities in N. The backward constraint propagation consists in canceling all the updates that have been applied to the cells in C by any entity s_i.

   Figure 4.10 shows the prediction of the LGM at time t. In figure 4.10a, an occupied cell (in red) is still generating the neutralized area (purple area). As a result, in figure 4.10b, the backward constraint propagation is applied on the cells that are located outside the neutralized area. The difference can be seen in figure 4.10b where cells are set to free instead of unknown.

   In figure 4.11, it can be highlighted that the backward propagation applies except for cells that are defined as reachable by several road users at a time, i.e. common reachable cells of V_1 and V_4. Cells that are reachable by a road user that is not involved by the backward propagation (see V_4) are still reachable. In other words, a cell that is no more reachable by a vehicle can still be reachable by another road user (figure 4.11b).

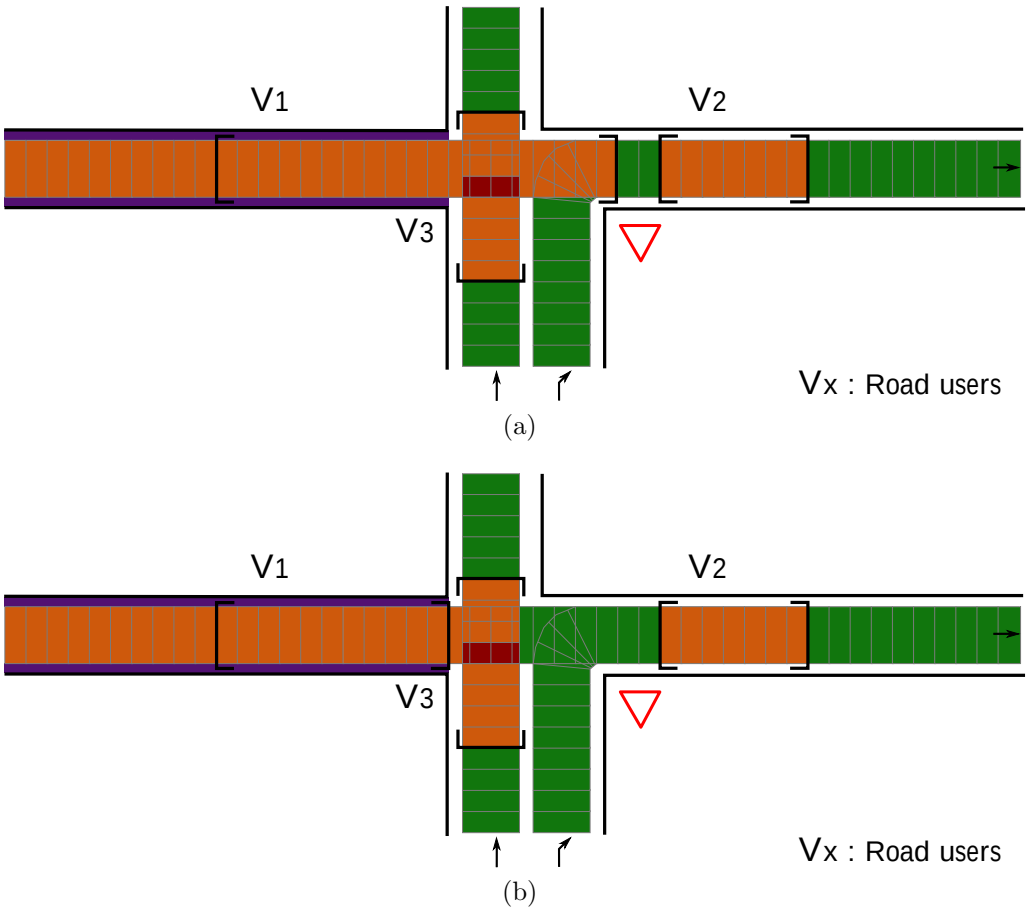


Figure 4.10: Predicted LGM ($t_0 + \Delta tjt_0$) with a neutralized situation that remains neutralized (purple) with free (green), reachable (orange), and occupied (red) states. (a) The predicted LGM is first built without any constraint. (b) The backward constraint propagation is applied on the reachable cells generated outside the neutralized area that are predicted from cells within the neutralized area.

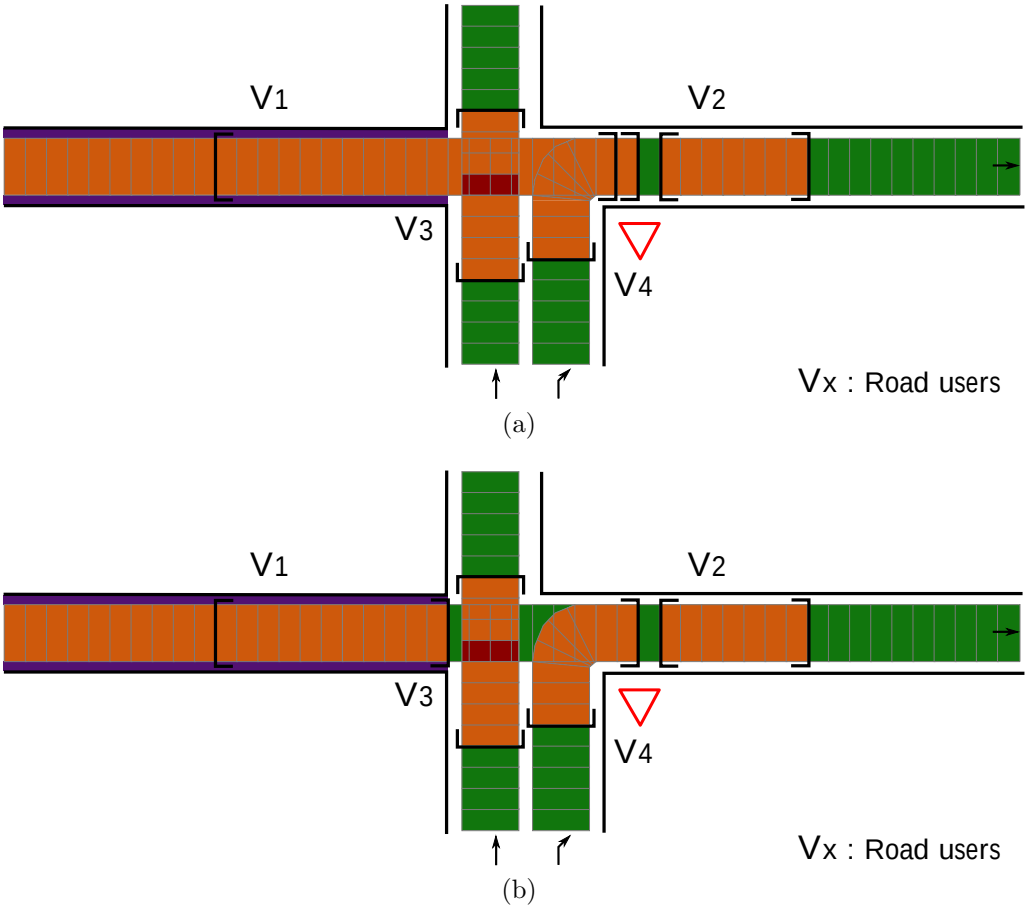


Figure 4.11: Predicted LGM (tjt_0) with a neutralized situation (purple) that remains neutralized with free (green), unknown (orange), and occupied (red) states and with multiple reachable sets overlap. (a) The predicted LGM is built without any constraint. There are reachable cells that are common to V_4 and V_1. (b) The backward constraint propagation is applied on the reachable cells generated outside the neutralized area that are predicted from cells within the neutralized area. However, due to V_4 prediction, some cells remain reachable.

2. The neutralized area is no longer neutralized in the predicted LGM at time $t_0 + \Delta t$ (see figure 4.12). No backward propagation is applied.

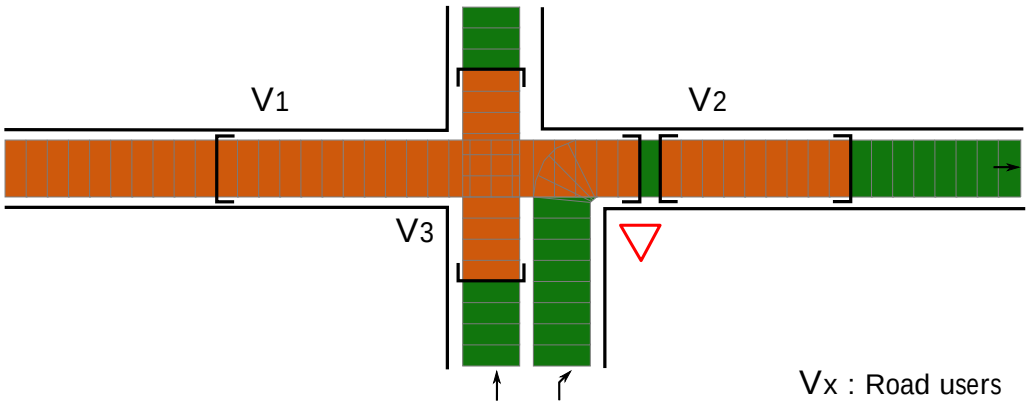


Figure 4.12: Predicted LGM ($t_0 + \Delta t|t_0$) where there is no more neutralized area. No backward propagation is applied on prediction of road users in the neutralized area.

4.3.6 Integrity

This section tackles the integrity management of the information provided by the predicted LGM. One call integrity, the ability of the predicted LGM to supply non-misleading information. To quantify the quality of the prediction, the predicted states of the cells of LGM ($t|t_0$) are compared to their actual states in LGM (t). Multiple situations can occur as detailed in table 4.2. There are two situations that may lead to hazardous situation (in red in table 4.2). The main objective is to provide evaluation metrics of these situations in order to manage misleading information.

Table 4.2: Correspondence between the predicted states of LGM ($t|t_0$) with the true ones of LGM (t).

True LGM (t)	Predicted LGM ($t|t_0$)			
	Occupied	Reach.	Non-reachable	
Occupied	Non-Misleading (1)		Hazardous (2)	Non-Hazardous (neutralized)
Free	Hazardous (neutralization)	Non-Hazardous (3)	Non-Misleading (4)	

First, when the true state is occupied but it was predicted as non-reachable, the ego-vehicle may decide to drive on an occupied area, which leads to a hazardous case (see (2) in table 4.2). Note that if this cell happens to be inside a neutralized area, then this miss-classification becomes non-hazardous as this information is of

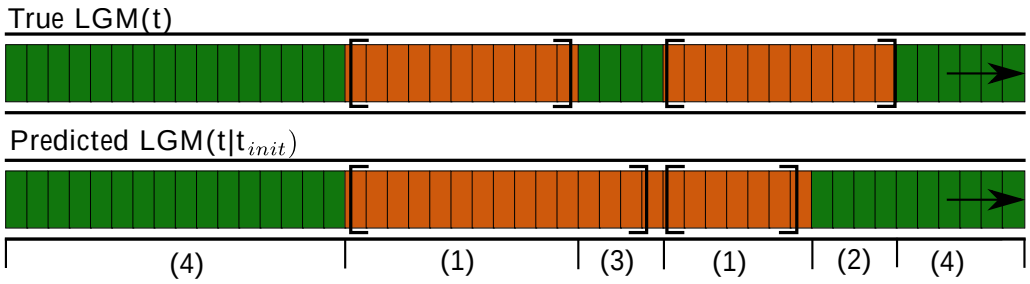


Figure 4.13: Illustration of table 4.2 (see numbers) with comparison of a true $\mathsf{LGM}\,(\mathsf{t})$ with its corresponding predicted $\mathsf{LGM}\,(\mathsf{tjt}_0)$.

lesser importance. This situation is quantified by the False Negative Rate (FNR) of the state prediction. At the opposite, when a free cell is predicted as occupied but does not generate a neutralized area, then the information is still misleading but not hazardous (see (3)). Indeed, considering wider reachable or occupied areas may leads to an over-cautious decision. As a result, the FNR is the most highly misleading and relevant indicator that is used to manage integrity of these hazardous situations. It corresponds to the rate of the number of cells that are non-reachable among the cells that are occupied with the true LGM. Figure 4.13 shows an illustration of table 4.2 without neutralized area consideration.

The second case is when the true state is free but is predicted as occupied and generates a neutralized area. This area may lead the ego vehicle to decide to cross when it is actually not protected from any road user driving from this area, which leads to a hazardous case (see (neutralized) in table 4.2). As a consequence, if the ego vehicle mistakenly considers a neutralized situation, i.e. considers occupied cells instead of free cells, it could lead to a poor decision-making. It can be noticed that, when a lane is truly neutralized, misleading information in the neutralized area is not considered as hazardous. As a second metric indicator, the prediction of the duration of the neutralization constraints called NTI (Neutralized Time Interval) is computed. This performance metrics corresponds to the predicted time $\mathsf{t}_{\mathsf{NTI}} = \mathsf{t}_0 + \mathsf{NTI}$ where there is a neutralized area for the last time.

Computing the NTI when a neutralized area occurs means finding the closest time for which there is no longer any cell fully occupied in the crossing area. The implementation consists in predicting the LGM over a time horizon with a given sampling time t and simply finding the first time when the neutralization ends. It corresponds to a lower bound approximation of the actual NTI.

The goal is to compute the NTI as close as possible to the real one but not overestimating it. Indeed, overestimating the NTI may lead to hazardous situation, e.g. the ego-vehicle estimates that it has the time to cross while it actually does not. Conversely, underestimating the NTI leads to overcautious decisions that guarantees the integrity of the navigation information but decreases the availability of the navigation function.

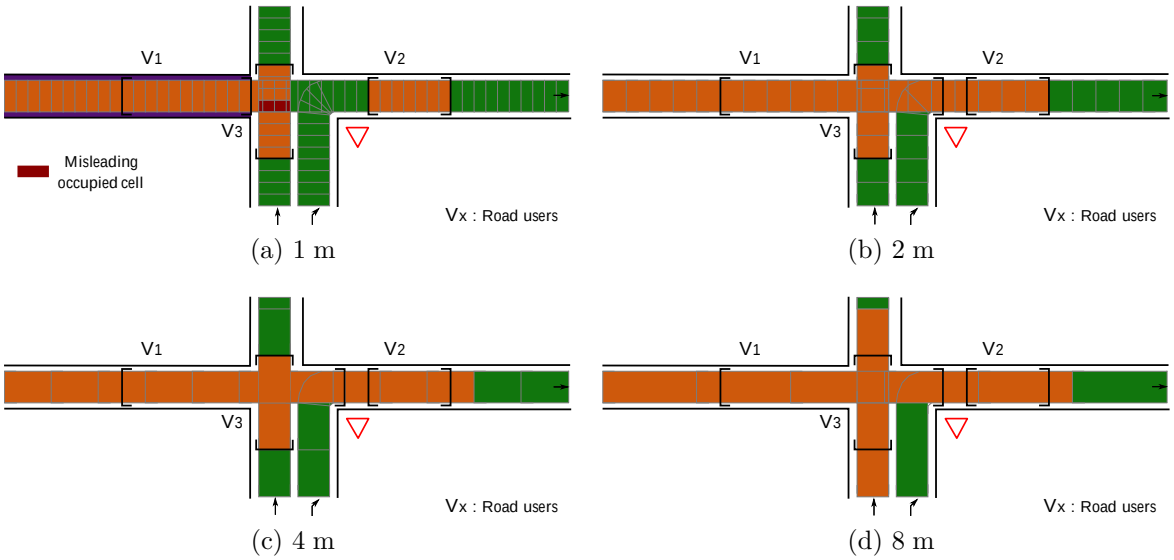


Figure 4.14: Several sampling steps for a same predicted LGM at a given Δt horizon are shown. As the sampling step increases, the neutralized area disappears. The reachable area of V_3 increases. Thanks to the sampling step parameter, the footprint occupancy of the vehicle does no longer fully encompass a cell, which leads to remove the occupied cell. As a consequence the situation is not considered neutralized anymore. The last figure shows that with a high sampling step, integrity of the situation can be reached but at the cost of a reduced decision-making space.

The quality of the estimated NTI depends on the prediction of the road users. For example, in figure 4.14, a mistakenly neutralized area is assumed with a sampling step of 1 m. The upper bound of the road user V_3 has been underestimated. In figure 4.14a, at some time t in the future, the LGM predicts that a cell will be fully occupied when it is not. This generates a neutralized area that may lead the ego vehicle to decide to cross. This situation leads to an overestimation of the NTI.

An effective way to mitigate hazardous situation (due to model errors, localization or perception errors in a LGM) is to enlarge the cell sampling step (see figure 4.14b). As the sampling step increases, for a same prediction time, the occupied cell get an unknown state. This sampling step should not be overestimated otherwise there will be a reduction in the availability of the navigation function, i.e there is a reduction of the free-space as seen in figure 4.14d. The NTI and the FNR metrics quantify the integrity of the information of a predicted LGM. Both metrics can be managed to be under a given application requirement, called Target Integrity Risk (TIR), by changing the spatial sampling step of the LGM.

4.4 Experiments

4.4.1 Case study

A dataset where neutralized situations occurred has been recorded in the city of Compiègne. The vehicles involved in this experiment was a Renault ZOE for the ego-vehicle and one Renault Master for the road user at its left. The situation used for the results is shown in figure 4.15. The ego vehicle (blue) is about to enter a roundabout. At this time, the merging lane is occluded by the van at its left side (red), which starts to generate a neutralized situation (this instant is called t_0). The same situation was reproduced three times (see appendix A for the setup).

The objective is not to provide a prediction over a Δt horizon as precise as possible, but to show that even if an inaccurate prediction model is used, it is still possible to generate a reliable prediction (in terms of non-misleading information) and to improve the decision-making process.

To evaluate the LGM in practice, the predicted $\mathsf{LGM}\,(t|t_0)$ is compared to the $\mathsf{LGM}\,(t)$ observed at the same time t by replaying the data offline with ROS.

Figure 4.15: Real-time situation used at initialization t_0 for prediction (Rviz). The ego vehicle (blue) is about to enter the roundabout. The merging lane (outer ring) is occluded by a vehicle (red) on its left lane.

4.4.2 Setup

4.4.2.1 Predicted objects bounds models

As mentioned before, the predicted bounds of each entity have to be computed at first at the object layer representation. To compute the bounds for a given

prediction Δt horizon, the equation 4.1 on page 104 is used with different intervals for the acceleration. The following three models, which are named according to their upper bound on the acceleration, are considered in order to compare and evaluate the impact on the results:

1. Constant acceleration (CA): $a(t) \in [-3.5; 4.0]\ m.s^{-2}$
2. Constant velocity (CV): $a(t) \in [-3.5; 0.0]\ m.s^{-2}$
3. Constant deceleration (CD): $a(t) \in [-3.5; -1.5]\ m.s^{-2}$

For each model, the velocity is also constrained within the interval $[0; v_{lim}]$, i.e. the vehicle cannot drive backward nor drive over the speed limit (where $v_{lim} = 50.0\ km.s^{-1}$ in the roundabout case). The CA model is the most pessimist one in terms of prediction, while the CD model is the most likely to be wrong in the long term for the studied situation. They all share the same lower acceleration bound.

The model used for hidden cells is based on equation 4.2 on page 104 where $v_{lim} = 50.0\ km.s^{-1}$. It results in a constant velocity model in this case.

It must be noticed that no other constraint is applied as the focus is on the neutralized case. For instance, the prediction of a vehicle following a leader vehicle is not constrained by the prediction of the leader vehicle.

4.4.2.2 Implementation

In order to store the predicted LGM, each cell of the LGM has a vector of its predicted states. As soon as a neutralized area is encountered as described in the case study, the prediction process is launched. Given a maximum prediction time horizon $\Delta t_{max} = 2$ s and a time sampling step of 0.1 s, each predicted $LGM(t|t_0)$ is computed based on the initial $LGM(t_0)$ and the neutralized constraint is applied depending on the situation. Indeed, the neutralized constraint is backward propagated if and only if the neutralized area remains neutralized. It enables to get the predicted NTI. As soon as the constraint is no longer applied, the predicted NTI is used for the reachability analysis of the neutralized cells. Indeed, they start to reach cells outside the neutralized area as soon as the NTI is over.

In order to show how integrity is handled through the prediction process, several sampling steps have been used by aggregating cells of the predicted LGM.

4.4.3 Results

The goal is to provide non-hazardous misleading information. The results that are presented in this section are twofold. In a first part, the purpose is to demonstrate the advantage of the neutralized constraint. It is possible to make a prediction that

is still consistent and enables to improve the reachability analysis, and thus the decision-making process, through the Neutralized Time Interval (NTI). The second part shows how the predicted LGM supplies non-misleading information thanks to its integrity management when performing the prediction process. In order to evaluate both integrity of a predicted LGM (with FNR) and the neutralized constraint (with NTI), the same situation has been exploited. It is of importance to show that this prediction representation does not bring misleading information that is handled by a good tuning of the sampling step parameter.

4.4.3.1 Reachability analysis

As previously described, the predicted LGM ($\hat{tjt}_0$) computed with different prediction models is compared to its corresponding LGM (t) observed at the same time. The first results present the impact of the model used to compute the NTI.

Neutralized Time Interval

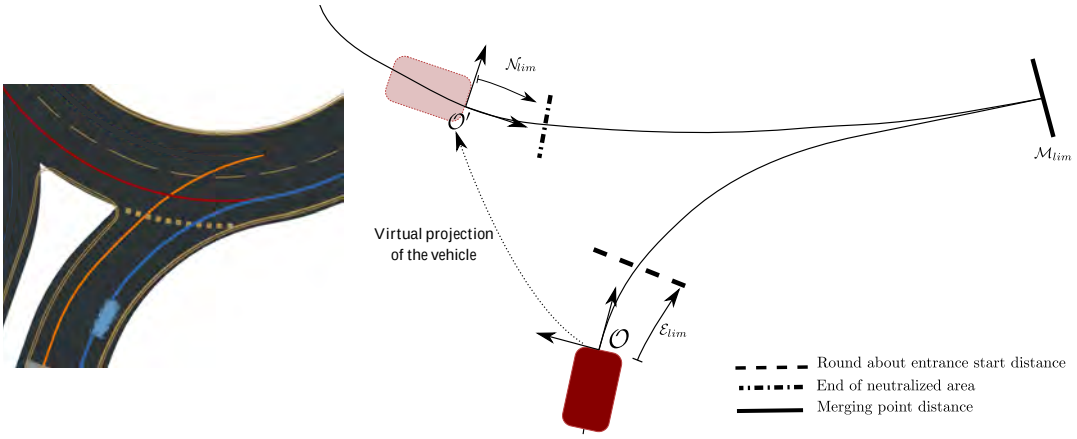


Figure 4.16: Sketch of the road layout at time t_0. This figure shows how curvilinear distances are computed. The origin reference is O and some remarkable distances are displayed. E_{lim} corresponds to the distance to the give way line. N_{lim} corresponds to the distance to reach the end limit of the neutralized area. M_{lim} corresponds to the distance to the merging point of the two lanes.

Figure 4.16 is a sketch of the situation (displayed at its left side). It shows the origin of the curvilinear abscissa $s(t)$ that is displayed on graph results and extracted from the dataset. As the vehicle is supposed to enter the intersection only if its corresponding merging lane (red) is free, the goal is to show how much distance of this merging lane is non-reachable, i.e. that remains free. As a consequence, the curvilinear abscissa of the ego-vehicle, i.e. its displacement along its trajectory, must be greater than the curvilinear abscissa of the end of the reachable area, i.e.

the ego-vehicle position must be above the reachable curvilinear abscissa on the graphs. In other words, the projection of the ego-vehicle on the merging lane must be within a non-reachable area. The distance traveled by the ego vehicle along the prediction process is computed from the origin O (front position of the vehicle at t_0). O^0 is the projection of O on the merging lane such that the merging point M_{lim} is equidistant.

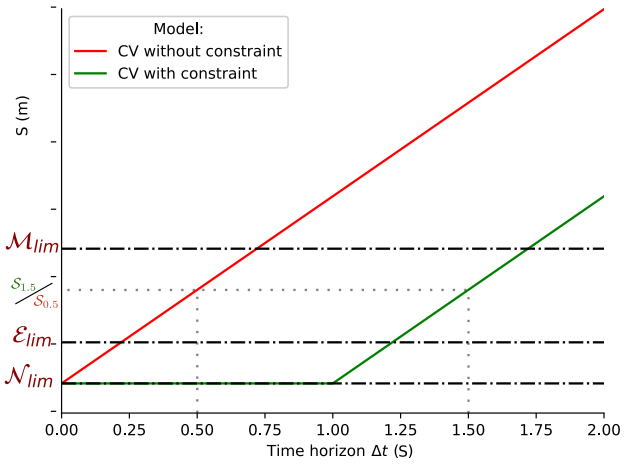
In a first step, theoretical results are presented in order to show the impact of the NTI. The merging lane with the neutralized area is composed of hidden cells. In figure 4.17, with theoretical results, the reachable distance prediction from O^0 is displayed as a function of time. The red curve draws the reachable distance evolution without considering the neutralized constraint. Here, the slope is directly dependent on the model used for hidden areas (CV model). The green curve has the same evolution but delayed, as it considers the neutralized constraint for a given NTI of 1:0 s. According to this NTI, the curve will be shifted by the predicted duration.

For a given Δt horizon, the area under the curve has to be considered as reachable whereas the area above is considered non-reachable (free). For $\Delta t \ 2 \ [0{:}0; 1{:}0]$, the reachable area in green is constrained by the end of the neutralized area limit whereas the reachable area in red shows a classical approach. Figure 4.17b illustrates the merging lane reachability of figure 4.17a at four different time horizons Δt.

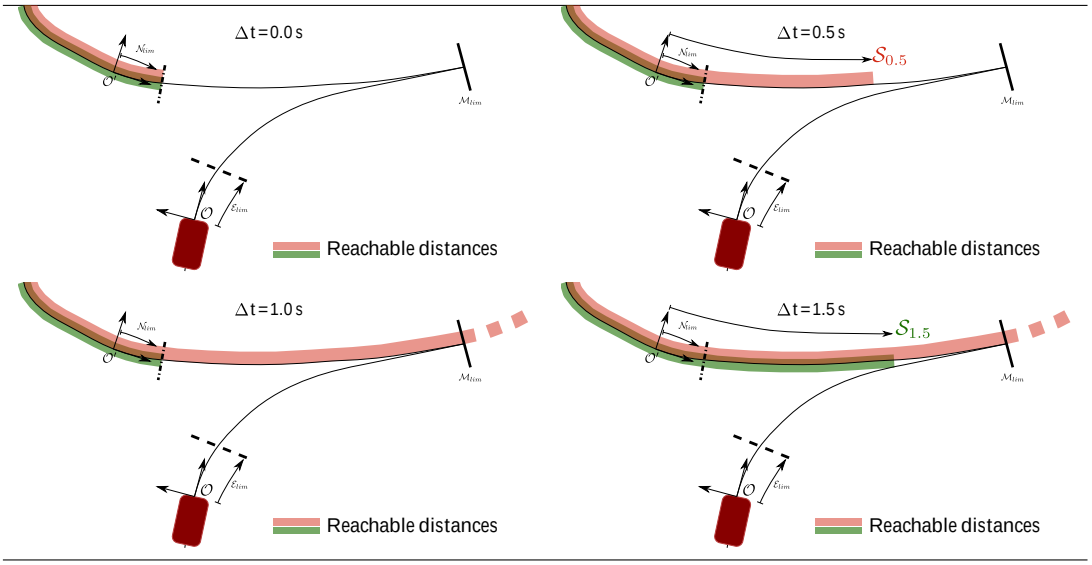
As a consequence, while the neutralized constraint is absent with the red curves, the green shows a different situation taking into account a prediction with the neutralized constraint. For instance, if the size of a road user generating the neutralized area is considered, the duration of the predicted neutralized interval may vary. Indeed, the green curve can model a road user with a certain length. Thus, as shown with the N_{lim} limit, the reachable distance remains the same for 1:0 s before evolving as soon as the neutralized area is not neutralized anymore. A small road user as a motorbike driving at high speed would imply a shorter predicted NTI.

The prediction model used for hidden cells is a matter of performance, as it does not affect the situation interpretation. That is why a simple maximum constant velocity model is used in the experiments for hidden cells of the LGM.

Three situations have been used in order to get experimental results. The same situation was repeated and they represent the same case at three different times in the recorded dataset with no model or parameter variation over time. Table 4.3 shows the NTI values the three models used for road users. Thanks to the corresponding observed $LGM\ (t)$ generated at each predicted time with a small sampling step of 0:2 m, the observed duration this area is neutralized is extracted. A comparison is made with the predicted NTI given by the prediction. It must be noticed that the values are similar in the three situations. It shows that the behavior of the van that is neutralizing the lane was similar when entering the roundabout.



(a)



(b)

Figure 4.17: The reachable curvilinear abscissa of any potential road user on the interacting lane from the O^0 origin is displayed w.r.t. the Δt horizon. (a) With a constant velocity model, the red curve shows an evolution of the reachable distance without taking into account the neutralized constraint. As shown with the green curve with the constrained applied, the reachability distance stays at the N_{lim} distance for 1:0 s before growing in a similar way as the red curve. At $\Delta t = 1{:}0$ s the vehicle that was neutralizing the interacting lane is not neutralizing it anymore. (b) Illustrations of the spatial reachable area for both situations for $\Delta t \in [0{:}0; 1{:}5]$: In these figures, the red curve shows the reachable area evolution without any constraint. The green curve shows how the neutralized constraint is applied until $\Delta t = 1{:}0$ s. The last figure shows that the green reachable area reaches M_{lim} much later than the red reachable area.

Table 4.3: Comparison of the NTI (s) in three similar situations with respect to the CA, CV and CD models. The sampling of the prediction at 100 ms makes the numbers have only one decimal value.

NTI (s)		Situation 1	Situation 2	Situation 3
Prediction	CA	1.2	1.2	1.2
	CV	1.7	1.6	1.8
	CD	2.0	2.0	2.0
Observed in real time		1.8	1.8	1.8

In figure 4.18 with experimental results, the reachable distances prediction is presented. Similarly to figure 4.17, the red line corresponds to the reachable distance with CV model without taking into account the neutralized constraint of the merging lane. The green line corresponds to the reachable distance with the observed neutralized time duration obtained with the observed LGM (t) generated in real time. As the situation has a neutralized area at t_0, as soon as the first LGM (t) generated does not have the neutralized area anymore with LGM (t_{NTI}), the NTI is retrieved from $NTI = t_{NTI} \quad t_0$. The intermediate blue line corresponds to the reachable distance with the corresponding predicted LGM (tjt_0) taking into account the neutralized constraint. Similarly, thanks to LGM ($t_{NTI}\, jt_0$), the NTI is computed. The better the prediction, the closer to the green curve. In this case, the CA model is used for the prediction of road users.

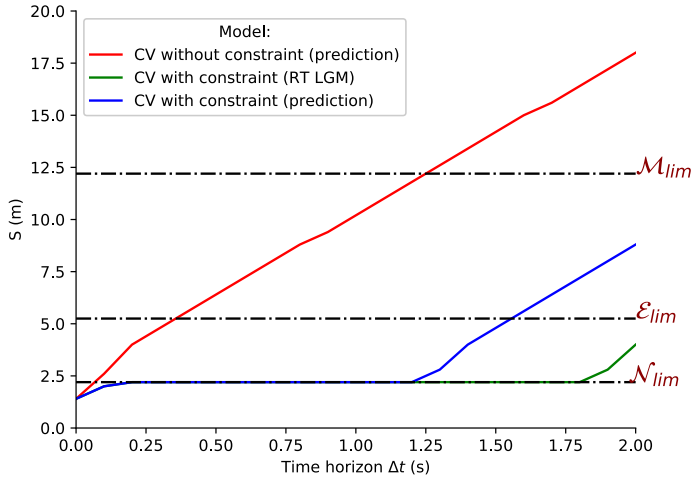


Figure 4.18: Experimental values of the reachable distance abscissa (s) w.r.t the time horizon (Δt) of a neutralized area. The CA model is used for the prediction of road users. The red curve is similar to figure 4.17a. The blue curve shows the reachable distance evolution taking into account the backward constraint propagation with the predicted NTI. Green curve shows the results with the NTI found with the observed LGM in real time (RT).

Decision models

In order to show how decision-making can be improved thanks to the constraint application, different possible decisions to take at t_0 are presented and compared to the real taken trajectory. Indeed, the goal is to show an example where, considering or not the NTI, a stop or entering the roundabout decision is possible while guaranteeing safety.

In figure 4.19, the neutralized constraint is not taken into account, which implies to consider the reachable distance of the red curve. A stop profile based on the initial ego-vehicle position s_0 and speed v_0 at t_0 is planned and drawn. The two pink lines represent the interval occupancy of the ego-vehicle planned trajectory, i.e. the front and back of the predicted vehicle position projected on the merging lane. In order to enter the roundabout, the vehicle must find a path that is in the non-reachable space. It means that its position must be above the reachable area shown in red. This area means that the merging lane is considered reachable until the merging point. As a trajectory is not feasible, the vehicle must be able to stop before the roundabout entrance located at a distance E_{lim} from the origin. $v_0 = 5{:}25\,\mathrm{m{:}s^{-1}}$ and the deceleration parameter used for the planned trajectory of the ego-vehicle is the same as the lower bound of the models ($3{:}5\,\mathrm{m\ {:}s^{2}}$).

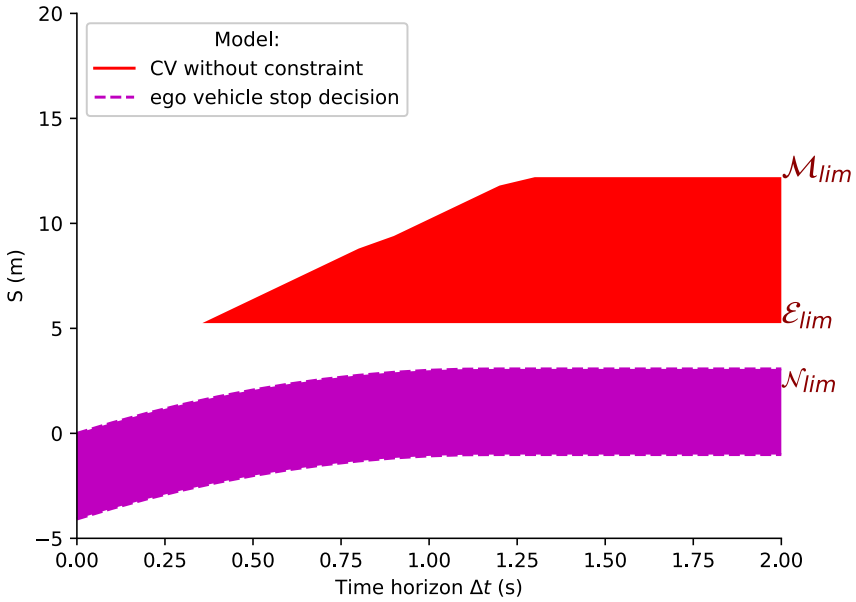


Figure 4.19: The neutralized constraint is not taken into account here. It shows a predicted stop trajectory to be taken by the ego-vehicle in order to stop before entering the roundabout. Given the ego speed at t_0, the pink curve shows the occupancy interval of the ego vehicle able to stop before it reaches the E_{lim} distance. The red area shows the relevant area of the merging lane that is defined as reachable.

In figure 4.20, the neutralized constraint is taken into account (blue curve). With a constant acceleration of $4{:}0\,\mathrm{m{:}s}^{2}$ and a maximum constant velocity of $8{:}3\,\mathrm{m{:}s}^{1}$, the projected model decision is able to plan a path that does not intersect the reachable area. In other words, there is a trajectory that enables the ego vehicle to go through the intersection while the lane is still free from any potential road user prediction. In this case, the goal is not to find the best model to plan the ego vehicle trajectory but to take a plausible evolution model that fits the situation.

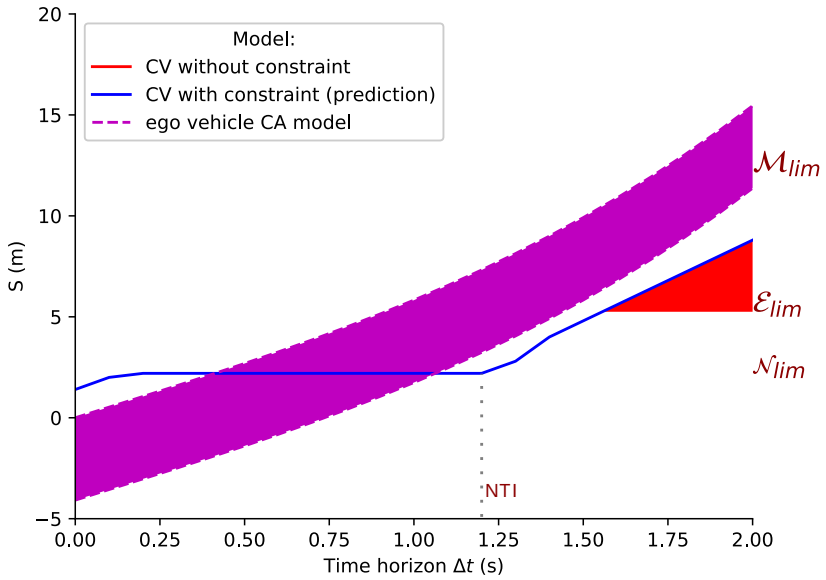


Figure 4.20: The neutralized constraint is taken into account. The neutralized time interval (NTI) has been predicted (blue curve). A predicted trajectory that follows a CA model to take by the ego-vehicle in order to pass the roundabout is shown (pink area).

In order to compare to what really happened in this situation, figure 4.21 shows the true position evolution of the ego vehicle with the corresponding real time observed NTI. It can be noticed that in this situation, the vehicle faced a sufficient NTI to enter the roundabout in the meantime.

In the situation presented, a CV model for potential road users and hidden cells prediction was used. It must also be noticed that depending on the prediction model used for road users present in the LGM, the predicted NTI may vary. Indeed, the larger the predicted bounds for a given Δt prediction horizon, the shorter the predicted neutralized time. Thus, based on several prediction models, integrity of prediction is explored.

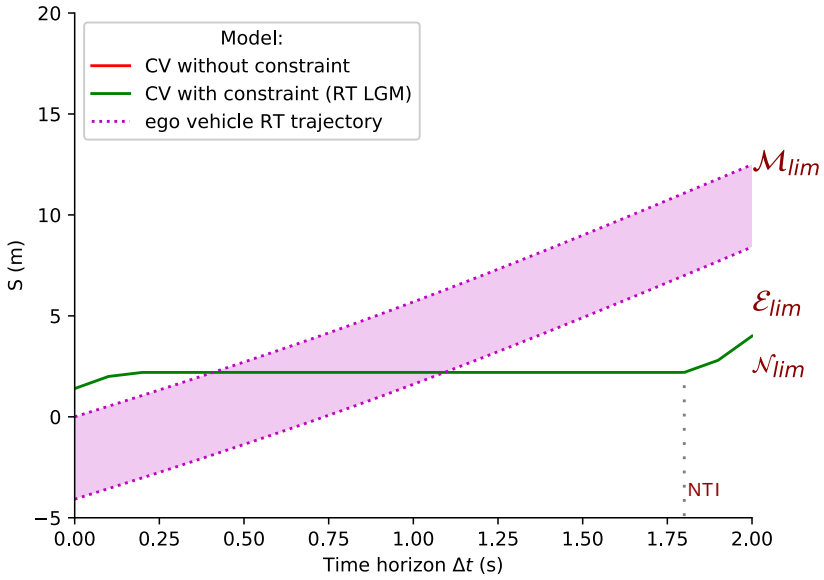


Figure 4.21: The pink curves display the true decision taken by the ego-vehicle. The neutralized time interval (NTI) is obtained by the real time LGM (t). It shows the real trajectory taken by the vehicle in order to pass the roundabout.

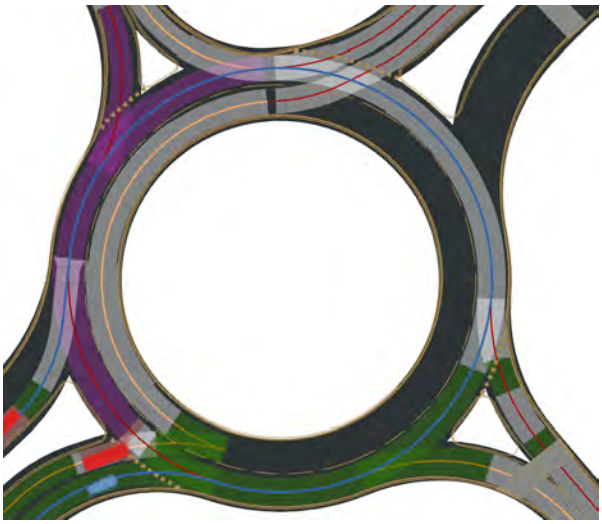
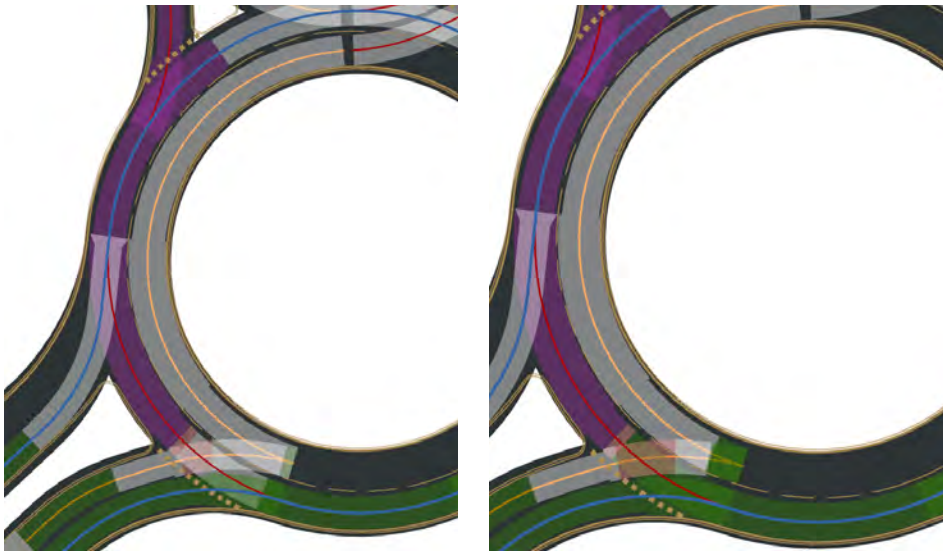


Figure 4.22: Initial LGM at t_0 displayed on Rviz. LGM (t_0) is displayed with a neutralized area (purple) caused by a vehicle (red) on the left side of the ego-vehicle (blue).

The initial situation, i.e. LGM (t_0), is presented in figure 4.22. Figure 4.23a depicts a predicted situation without taking into account the constraint of the neutralized area. The lane where the ego vehicle is likely to merge is shown as being reachable by merging traffic. When looking at figure 4.23b, a predicted situation with the



(a) (b)

Figure 4.23: Predicted LGM (tjt_0) on Rviz. (a) LGM prediction without neutralized constraints ($\mathsf{t} = 1{:}1\,\mathrm{s}$). (b) LGM prediction with neutralized constraints ($\mathsf{t} = 0{:}8\,\mathrm{s}$). (green=non reachable, red=reached, white=unknown, purple=neutralized area).

backward constraint propagation application is observed. It can be noticed that the lane where the ego vehicle is likely to merge is still free before entering the roundabout.

4.4.3.2 Integrity: NTI evaluation

The impact of the sampling step parameter on the predicted neutralized time interval is shown in figure 4.24. Each curve represents the evolution of the NTI using different prediction models for the road users: CD (blue), CV (green), CA (red). For each model, it can be observed that higher is the sampling step, lower is the predicted NTI.

The NTI obtained with the LGM (t) observed in replay is drawn with an horizontal gray line. A misleading situation occurs where the estimated NTI is larger than the true one as it can be seen with the CD model. By increasing the sampling step, the NTI is no longer misleading even with an approximate model. On their part, the CV model enables to maximize the NTI and the CA model is much more conservative but also more robust. With a more pessimistic model, it becomes harder to predict a long NTI and it brings a more conservative situation understanding. Results show how the sampling step enables to keep integrity by ensuring lower estimation of the predicted neutralized time interval. The prediction model CD shows the loss in

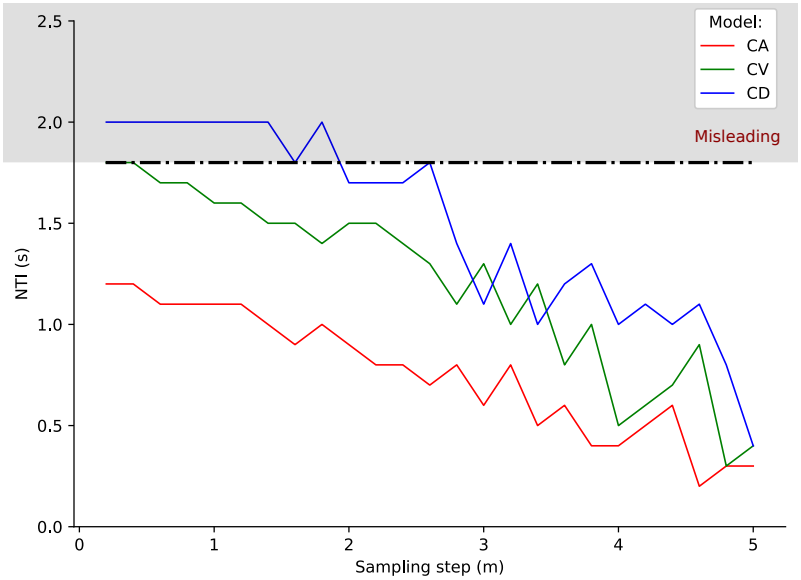


Figure 4.24: Evolution of the predicted NTI obtained as a function of the sampling step parameter. The three models CA, CV and CD are represented respectively by the red, green and blue curves. The NTI computed with the LGM (t) is shown with the dotted line. Values above this line are considered as misleading.

terms of integrity when a model does not bound the real behavior of road users. In order to fulfill the same requirement, the sampling step has to be increased up to two meters at least.

It must be noticed that the maximum NTI that can be obtained is bounded by the maximum predicted time horizon $\Delta t = 2\,s$.

As the sampling step grows, noise is observed on the NTI values due to a sampling step side effect (information becomes coarser). Figure 4.25 depicts how the NTI may vary from one step to another. On the left, as the sampling step increases, occupation may decrease accordingly. However, on the right, an example is given where occupation may disappear temporarily. Depending on the cell boundary, it can be observed that there is an intermediate sampling step where no cell is characterized as being occupied as no cell is within the occupancy bounds $\mathbf{s}_L$.

4.4.3.3 Integrity: FNR evaluation

In a second time, integrity of prediction is addressed. The second prediction integrity focus is on the LGM cells prediction. The same situations are used. However, the prediction of the road user that is next to the ego vehicle, and which is generating a neutralized area in the LGM, is now analyzed. The purpose is to show, with the three prediction models, how integrity of information is ensured regardless of the

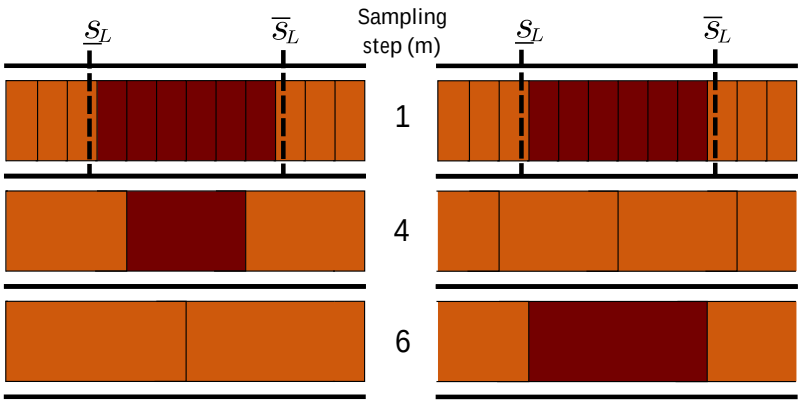


Figure 4.25: Sampling side effect on the NTI computation. On the left, occupation decreases as the sampling step increases. On the right, there may exist a sampling where there is no occupied cell. As cells are not fully occupied, they are set reachable with an unknown state. The sampling may impact the NTI. (occupied=red, reachable=orange)

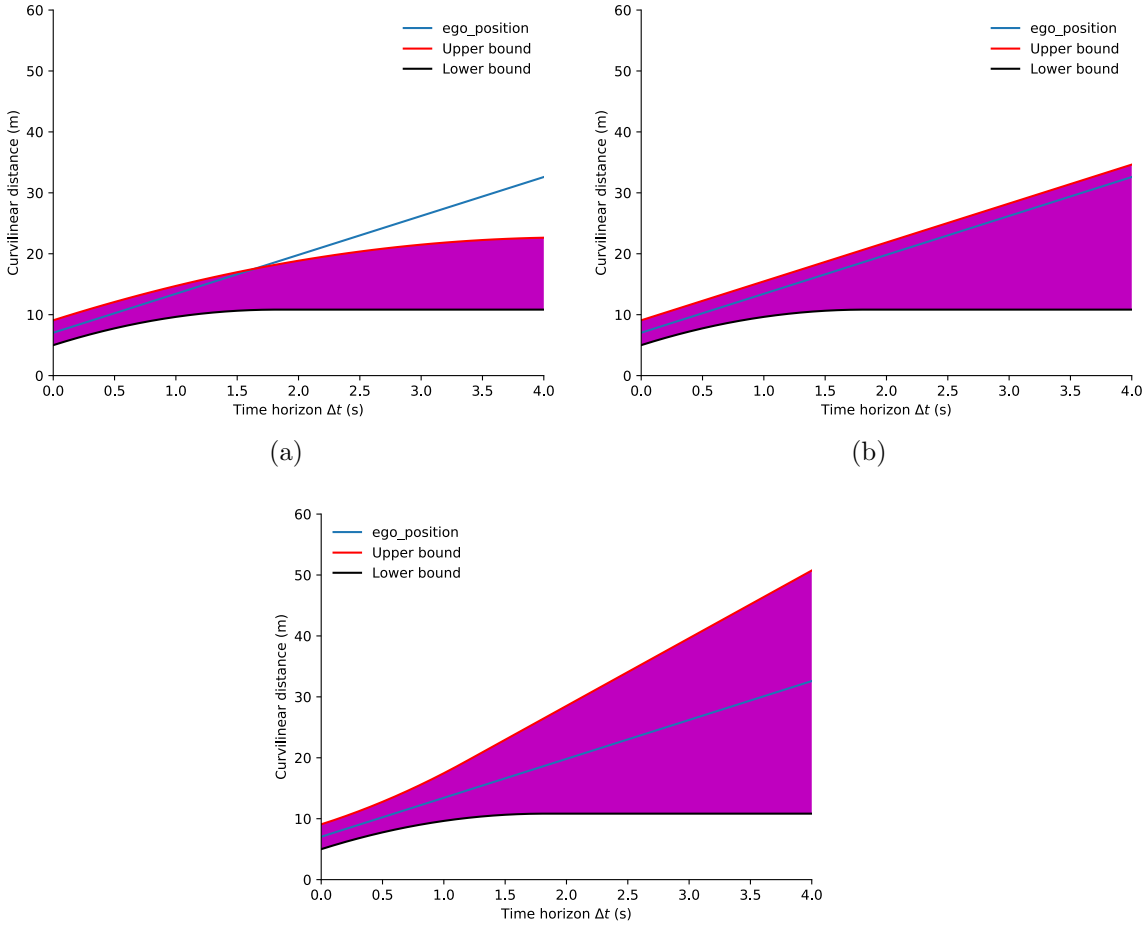
prediction model performance. Analysis of the results are based on one road user prediction on a single lane.

Simulated results In a first part, to facilitate the understanding of experimental results, some simulations were conducted. Given the LGM of a straight lane, a single road user, called ego-vehicle, driving with a CV model is considered. At a given time t_0, the predicted LGM (tjt_0) is generated over 4:0 s by assuming different models. This simulation enables to compare directly the true evolution of a road user (that is thus known) with an assumed evolution.

Figure 4.26 shows simulated results of the vehicle driving on the straight lane at constant speed (ego_position) with the three prediction models. In figures 4.26a, 4.26b, 4.26c are drawn respectively the predicted bounds of the CD, CV and CA models. Depending on the model pertinence, it shows how predicted bounds encompass, or not, the position of the ego vehicle at each prediction time. Its position (the center) is modeled with an upper and lower bound (rear and front).

These graphs represent the predicted bounds in a continuous space dimension. It can be observed that the reachable space of the CA model encompasses the vehicle position. The CV model is close also to the vehicle position whereas the CD one does not encompass it as the time horizon increases. The goal is now to show how misleading information is managed with the different models. In experimental results, the model of the ego vehicle is unknown.

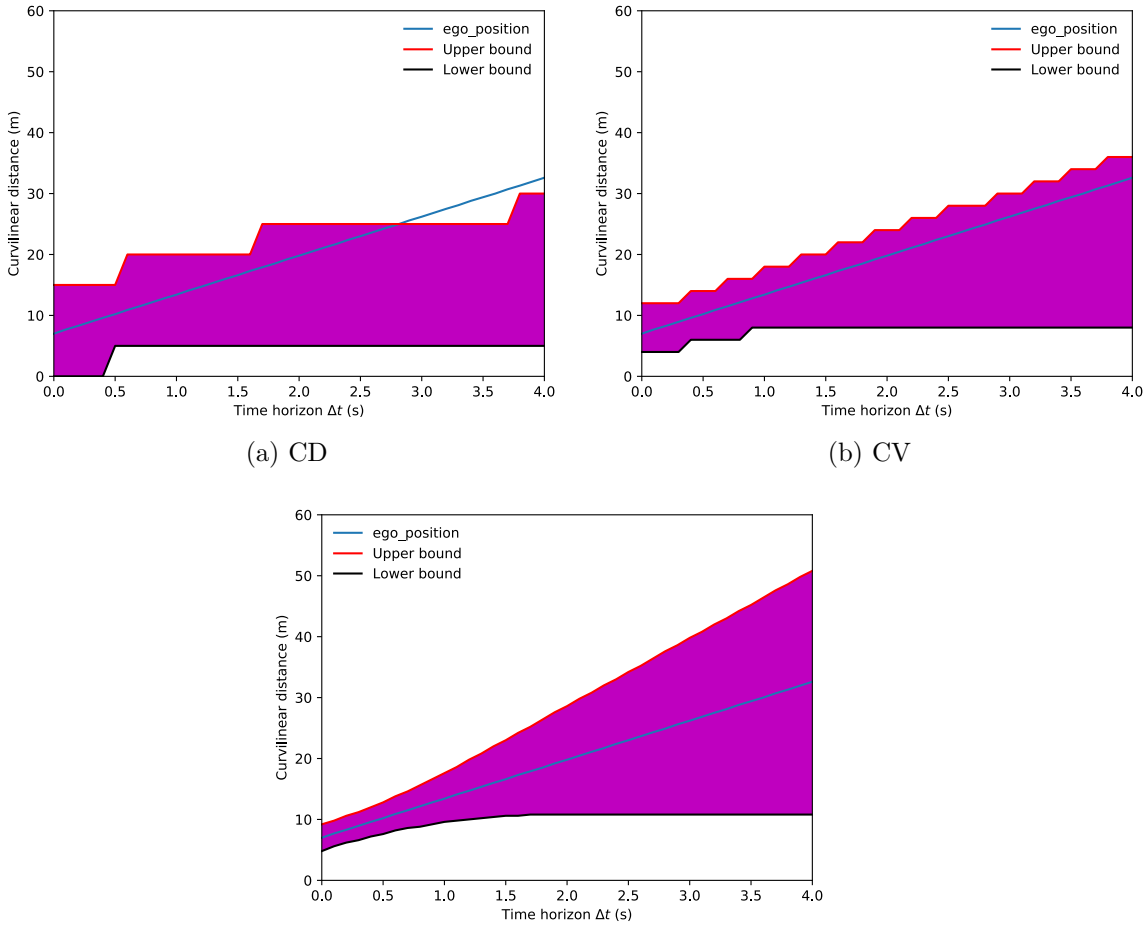
In figure 4.27, the minimum sampling step that enables the predicted reachable bounds to encompass the ego position for at least 2:0 s of prediction is displayed for each model. In figure 4.27a, as the CD model does not bound the ego position,



(a) (b)

(c)

Figure 4.26: The ego-vehicle position (blue) is modeled with an upper and lower bound (rear and front). For each prediction model, the reachable distance is filled by the purple area. The lower bound of each model corresponds to the black curve and the upper bound to the red curve. (a) CD model. (b) CV model. (c) CA model.



(a) CD

(b) CV

(c) CA

Figure 4.27: For each prediction model, the reachable distance is filled by the purple area. The lower bound of each model corresponds to the black curve and the upper bound to the red curve. (a) The figure shows the reachable area with CD model with a sampling step of 5:0 m. (b) It corresponds to the reachable area with CV model with a sampling step of 2:0 m. (c) The reachable area with CA model with a sampling step of 0:2 m.

a sampling step of 5:0 m is chosen in order to overestimate the reachable area. Compared to figure 4.26a, and according to the target integrity risk, integrity is kept for 2:0 s of prediction by rising the sampling step. In a similar way, to keep integrity over a larger Δt horizon, the sampling step can be increased. In figure 4.27b, a shorter sampling step of 2:0 m is set. As the CV model corresponds to a CV model like the evolution of the ego vehicle, it fits better the curve. Thus, a lower sampling step enables to keep integrity. In figure 4.27c, as the CA model over-bounds the ego trajectory, a very small sampling step of 0:2 m is sufficient.

If the prediction model is reliable for a short prediction time and then drifts, it is thus possible to set a sampling step that is dependent on the predicted time. As a consequence, even with extreme cases (as the CD model), a very high sampling-step enables to define the whole lane as reachable. Despite supplying information with a high integrity level, this clearly shows that it will strongly impact the decision-making process towards a very cautious decision as there is no more free cells. This induces a loss of availability of the system.

As mentioned before, the goal here is to provide non-hazardous misleading information. To do so, the integrity of this system is defined as its capability to keep a statistical FNR under a given threshold called target integrity risk (TIR).

In figure 4.28, the FNR is shown on simulations with the CD model for different sampling steps of a single vehicle prediction in order to compare with the experimental results. The FNR is high with the CD model as it corresponds to the worst one compared to the CV and CA models.

Figure 4.28a shows the FNR for different sampling steps obtained with a single situation. As the CD model is an inaccurate model, the FNR grows rapidly as the prediction time horizon increases. It can also be observed that for any given time horizon, increasing the sampling step enables to decrease the FNR. The jumps in the curves are a side effect of the sampling step. This effect is explained by the fact that cells defined as reachable rely on t_0. At this time, if the road user is close to the start or to the end of a cell, it has an impact on the update function.

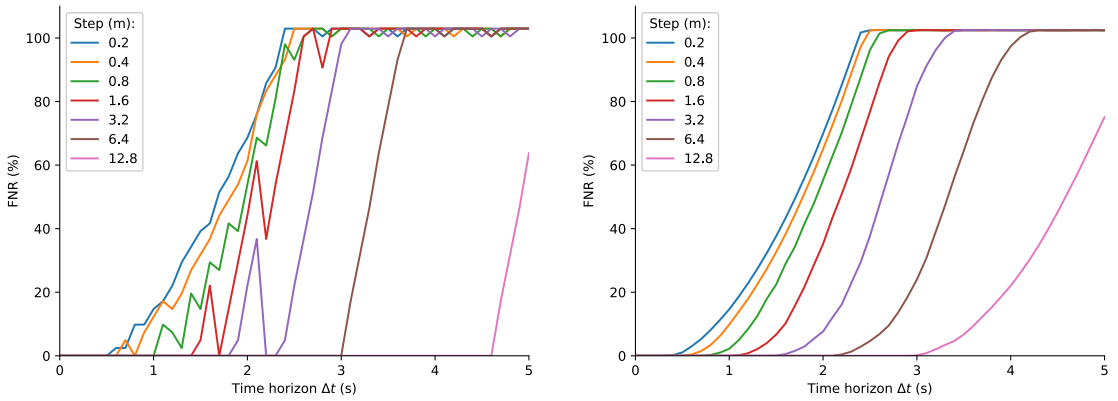


Figure 4.28: Evolution of the FNR w.r.t. the time horizon for different sampling step using simulated data (CD model). (a) Single simulation. (b) Average over more than one thousand simulations.

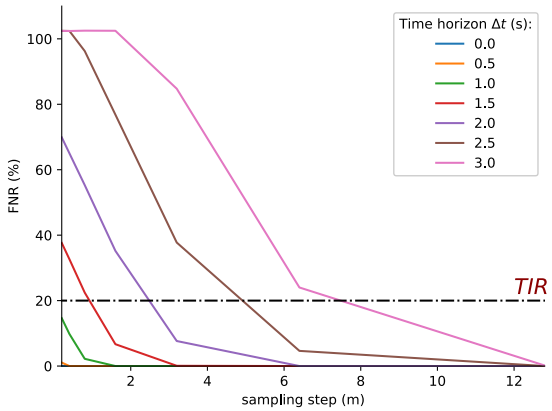
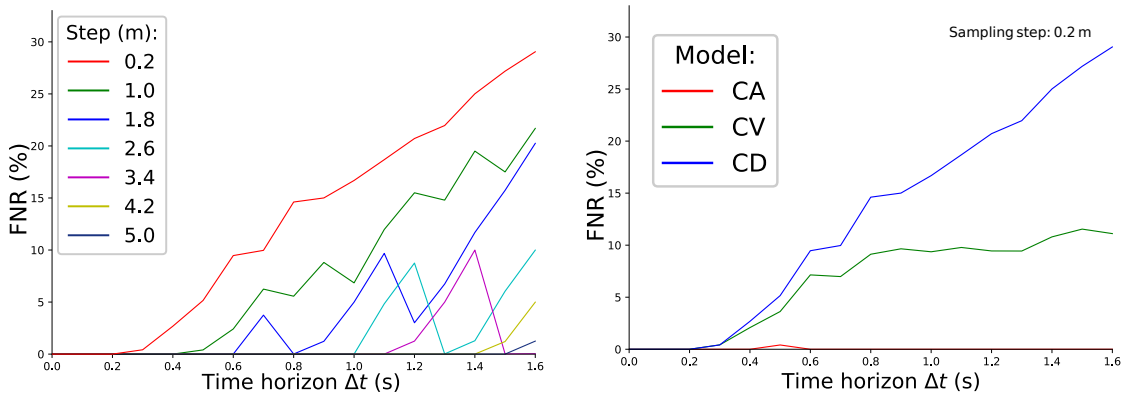


Figure 4.29: FNR evolution based on the sampling step parameter. (TIR=Target Integrity Risk) The further away from the origin, the more the curve corresponds to a distant predicted time horizon.

In order to get the true statistical FNR, figure 4.28b shows the true statistical FNR by averaging more than one thousand simulations using a different initial time t_0 each time. It can be observed that for any given time horizon, by increasing the sampling step, the FNR decreases.

Based on figure 4.28b, figure 4.29 shows the evolution of the FNR for a given predicted time according to the sampling step. Given a fixed TIR, this figure enables to highlights the integrity property of the sampling step in a same way as in the previous chapter with the integrity of a static LGM. Indeed, it shows that the further



(a) (b)

Figure 4.30: Evolution of the FNR of prediction with experimental results. (a) shows experimental results obtained with the CD model. Each curve displays the evolution for a given sampling step. (b) shows experimental results obtained with the three models for a sampling step of 0:2 m: CD (blue), CV (green) and CA (red).

the predicted time is, the higher the sampling step needs to be in order to remains under the TIR.

Experimental results In the experiments, three sequences have been recorded in the dataset. Figure 4.30 shows results that are representative of figure 4.28a with a CD model. The actual evolution model of the human road user is unknown, but the curves suggest that the CD model is likely to be inaccurate.

In figure 4.30a, the evolution of the FNR with the CD model is depicted for different sampling steps from 0:2 to 5 m. For any required prediction time horizon and TIR, it is possible to make the FNR below the TIR by increasing the sampling step. When the prediction time horizon increases, the bounds computed by the CD model are further away from the road user true position and requires a much larger sampling step for the correct cell to still catch the road user. On the contrary, if the evolution model is pessimistic, as it is the case for the CA model, the bounds computed by the model will likely contain the road user true position. Therefore, the LGM can be used with a much small sampling step.

Figure 4.30b illustrates the results obtained with a small sampling step of 0:2 m using the three models: CD, CV and CA. With the CA model, the FNR is close to zero whatever the time horizon is.

It should be highlighted that the prediction model performance is not evaluated here. If the prediction model is accurate enough to catch the road user position within the prediction bounds, then the requirement in terms of sampling step will not be

stringent. Conversely, if the prediction model is far from the reality, the sampling step will act as a parameter to manage the integrity of the whole system.

In practice, once an object predictor has been set, along with the integrity requirement of the decision-making module in terms of prediction time horizon and TIR, a data collection is required to monitor the integrity performance and pick the optimal sampling step satisfying the integrity requirement while maximizing the availability via the NTI.

4.5 Conclusion

In this chapter, set-based prediction analysis has been addressed based on the LGM representation at a tactical level. This work aimed at improving this process by using the neutralized characteristic that is supplied by an augmentation process of the LGM. It has been shown how the backward propagation of a neutralized constraint, i.e. a specific interaction, may enhance the decision-making process through the computation of the Neutralized Time Interval (NTI), which is a key metric for a decision making algorithm when a neutralized situation occurs. The results have shown how a better decision can be made using the NTI.

The results have also shown that the sampling step parameter enables to provide non-hazardous misleading information through prediction integrity management. Simulations and experiments led in Compiègne showed how to use the sampling step parameter in order to fulfill a target integrity risk requirement. For a desired prediction time horizon, a proper sampling step must be defined. Depending on prediction models, the higher the sampling step, the further the prediction is reliable. However, it may induce a loss of availability of the system as decision-making may be overcautious. With a sufficient amount of driving recording, this parameter may be learned from data. Depending on the operational design domain, the sampling step parameter can be set accordingly.

From a situation analysis perspective, prediction has been focused on a neutralized case at time t_0. Cells characterized with a safe state in the initial LGM have not been taken into account. It appears unnecessary to use this information in a predicted LGM. Indeed, a safe area models the minimal distance for a road user to stop, it can be encompassed or added to the prediction model of the road users. In addition to that, this implies that the prediction of a neutralized situation that is not neutralized at a given prediction start time t_0 is not addressed. The prediction of such a situation does not bring any additional information that could lead to a better decision (see appendix E for details).

It must be highlighted that cells prediction may apply on several types of cells as it only relies on an evolution model. For instance, from a spatial occupancy point of view, prediction of occupied cells that does not belong to any object of the object

layer only need an evolution model to be predicted in a similar way of hidden cells. For instance, if an object such as a scooter has been missed by an object detector but has been taken into account in a geometrical occupancy representation, prediction at the LGM cell level can still be performed.

In the presented approach, there are some concepts that enable computation efficiency of prediction at lane level. First, as the LGM is scalable, the sampling step parameter may be chosen in a convenient way. In addition to that, as the LGM focuses on areas of interest thanks to the interaction graph, only road lanes that are converging towards the ego lane are taken into account. If a road user is in the LGM, only paths that lead to the ego vehicle path are taken into account. Prediction of road users is thus considered only within the LGM as outside is not relevant. As a consequence, the computation of the prediction is focused only on the relevant part of the situation.

General Conclusion

Synthesis 135
Perspectives 137

Synthesis

This PhD thesis has been devoted to the study and design of a World Model module within the functional architecture of an autonomous navigation system. On the bases of situation awarenessrequirements for safe navigation and decision-making, the key issues of situation information representation, situation understanding and situation prediction have been addressed. The goal is to build up a unified world representation from the vehicle point of view that provides a situation understanding with integrity management such that the information provided to decision-making is not misleading. This is particularly crucial in complex situations such as intersections or roundabouts to avoid inappropriate decisions and to ensure safety.

In order to take decisions, information representation is a key stone. Perception systems enable to acquire, process and abstract sensed data from the surroundings of the autonomous vehicle. However, prior knowledge such as traffic laws and maps, that contain several abstraction levels of information, have to be taken into account in order to understand interactions with the other road users. It has been shown that there is a need to process different levels and sources of information, whether they are maps, road users or information on the free or occupied space.

Based on the state of the art that has been done, existing architectures with different structural compositions have different mechanisms and concepts to tackle situation understanding. An architecture with a World Model (WM) has been proposed to answer these different needs. Three abstraction levels of the information are considered from the highest to the lowest: Strategic Level(SL), Tactical Level (TL), and Operational level (OL). The OL provides information to the control module. The TL enables decisions such as maneuvers and the SL is long-term planning oriented. Mechanisms such as bottom-up and top-down processes enable to go from an abstraction level to another. At the highest level, prior information such

as a topological map leads to a selection process at lower levels with attention mechanisms.

A WM is an efficient means to manage safety, explicability and risk analysis of the situation the vehicle encounters. It has the duty to generate and maintain a best estimate of the state of the navigation context at each level. This best estimate describes the state and attributes of road users, classes, situations, and relationships taking into account uncertainties and integrity issues. It gives rise to reasoning mechanisms such as relevant information filtering or prediction at different levels of abstraction.

One of the most critical aspects of situation awareness is reasoning on information while ensuring integrity of an enhanced situation representation. In this PhD thesis, at the top level of abstraction, the Interaction Graph representation has been proposed to extract the areas of interests with respect to the situation from the ego-vehicle's perspective. At the intermediate abstraction level, a lane level representation, called Lane Grid Map (LGM) that combines detection of road users and spatial occupancy information, is built upon areas of interest. It aims at providing non-misleading information in these areas. A characterization process of the cells based on low-level information such as an occupancy grid has been presented. Additionally to classical free-occupied states, the unknown state is further characterized by the notions of out field of view, neutralized and safe areas that provide a deeper level of understanding of the situation. As the LGM enables to be aware of relevant missing information, it aims at providing additional information in these areas.

In classical decision-making processes, occupancy information or distances between road users are often used for local planning or for computing safety metrics like time to collision. The LGM has the advantage to combine both spatial information and distance metrics. Moreover, it is a scalable representation that enables to optimize computation. The sampling step allows the refinement of the level of detail of the information stored in the LGM.

Integrity of the information managed by the WM for situation understanding is a critical issue. It must not provide misleading information that could lead to a hazardous decision-making and has to handle the unavoidable uncertainties. The sampling step is a key stone to handle the integrity of the generated information. Another major contribution to the WM is an integrity management mechanism that consists in managing the spatial sampling of the grid cells in order to take into account localization and perception errors and to avoid misleading information. It has been shown how a proper sampling enables to fulfill integrity requirements. By increasing the sampling step, the misleading information rate can be kept under a Target Integrity Risk (TIR) specified by the navigation task to perform. This TIR can be seen as the maximum allowed level of misleading information. It aims at bounding errors within a given operational design domain.

The third pillar of situation awareness is the prediction capacity of the current situation. Indeed, a situation can be predicted at several levels of abstraction whether it is for the trajectory computation of a road user, for an intended maneuver or for potential interactions. In the literature, set-based prediction relies on reachable sets formalism that handles spatial information. This is well in accordance with the LGM representation. Constraints such as no overtaking or keeping safe distances between road users on a same lane show the need to take into account interactions between road users. The prediction work aimed at improving this process by using the neutralized characteristic that is supplied by the augmentation process of the LGM. Results show how a better decision can be made using the Neutralized Time Interval (NTI) indicator, which corresponds to the duration that a situation remains neutralized.

In addition to that, it has been shown how integrity can also be extended at the prediction stage. Integrity management is still performed thanks to the sampling step parameter. For the prediction case, the sampling step parameter is dependent on the prediction models. For a given prediction model, the objective, for instance, is to provide non-misleading information over a given prediction time horizon.

The work and results in this thesis rely on experiments that demonstrate a real application of a complex situation awareness representation. The approach has been evaluated with real data obtained thanks to several experimental vehicles and a high-definition map of the city of Compiègne. The proposed WM has been implemented (with ROS software) in order to fulfill real-time constraints and is functional on the experimental vehicles of the laboratory.

Perspectives

Although this PhD thesis has proposed a World Model module to tackle situation awareness and its related integrity, in particular with the LGM concept and its associated prediction process, there are still many points to explore. Perspectives and improvements are presented in relation to several contributions.

Interaction Graph related

IG expension In order to build up areas that are of interest for the current driving situation of the ego-vehicle with the Interaction Graph (IG) representation, the topological layer of the road network has been used. It uses prior information to infer relations that exist between different paths, e.g. merging or crossing relations. However, only road users were considered in this work within a road network. A similar approach could be extended towards vulnerable road users. A map that also

encompasses the cycle paths, sidewalks or pedestrian crossing would enable to define additional interaction modes and to expand the IG.

That said and globally speaking, additional interaction modes that was not addressed can be explored, e.g. diverging lanes or even oncoming lanes. The IG can also be extended in terms of depth level in addition to primary and secondary order lanes.

IG requirements The interaction distances that corresponds to the distance of the directed path trees of the IG have been set arbitrarily, sufficiently high to lead real time experiments. However, the greater the distance, the more the areas of interest. It has a direct impact on the performance of the LGM computation. At the opposite, taking low distances could lead to not meet decision-making requirements (not enough visibility information at intersections for instance). A compromise between computation efficiency and navigation requirements is to be considered and studies have to be done on this.

Lane Grid Map related

Characterization process In order to characterize the cells of the LGM, different frames of discernment have been used in order to determine if a cell has one or several states among free, occupied, neutralized, safe, hidden and unknown. In must be highlighted that the evidence theory has be used which means that the LGM process can be easily extended to additional semantic characterization. For instance, objects may generate occlusions that lead to define an LGM cell with a hidden state. It may be interesting to characterize and to know if a cell is hidden by a road user or by a building for instance. Thus, a state information of hidden cells could be decomposed in several sub-frames. Similarly to reachability in prediction, an improvement of the characterization process of a static LGM would also be to differentiate the truly occupied cells using the road users dimensions to occupied cells by their uncertainty when it has been propagated on the perception information.

In the presented work, a neutralized situation occurs when a road user that belongs to a secondary order lane prevents other road users from a primary order lane to reach an area of interest that is in the future path of the ego-vehicle. Only the occupancy of the road user is used and it must only intersect a cell of the primary order lane. These conditions of neutralization could be improved. Indeed, if a vehicle slightly intersects or is quasi-static, the neutralized case can be ambiguous. As a consequence, one could consider that the cell must not only intersect the road user occupancy but be within and the safety area could also be taken into account to determine if there is a neutralized situation (which could also impact the NTI value).

The characterization process that has been implemented, through the polygons approach, consists in computing intersections between these polygons and the cells of the LGM. This process has been parallelized as it is time consuming and it strongly depends on the number of cells. The shorter the sampling of the LGM, the higher the CPU load. As this process deals exclusively with geometric shapes, an implementation based on GPU acceleration could be a significant performance improvement.

Integrity evaluation Results on integrity have shown the important role of the sampling step parameter on misleading information. A proper sampling step must be set for the real-time LGM and for the predicted ones. To achieve this goal, this parameter can be learned from a large amount of data that must be representative of a desired operational design domain. Experiments with large dataset acquisition still need to be conducted. It will enable to study the sampling step effects varying multiple parameters. For instance, as uncertainty may increase with the speed of the ego-vehicle, it is of interest to be able to modulate the sampling step accordingly.

The along track sampling step has been addressed. However, a lateral sampling could also be explored. It mostly relies on the decision-making requirements granularity. In the experiments, the lateral width of the LGM cells corresponds to the width of the lane to which they belong. It corresponds to the simplest case where the ego-vehicle can drive in its corridor. If a lane change is required, in this case, the adjacent lane must be free from road users. If a lateral sampling step is provided, the ego-vehicle can better handle obstacles avoidance.

Sampling strategy The sampling step of the LGM is defined when the areas of interest of the IG have been computed. Thus, the LGM structure is updated at a low rate (strategic level), but its internal composition is updated at a higher rate (tactical level). In order to enable the LGM to provide a sampling step upon request, a good sampling strategy would be to first define the shortest sampling step as possible (depending on the hardware performance), and then to aggregate cells on the fly depending on the requested sampling.

In this perspective, the aggregation strategy also needs to be deepened. The results presented in the integrity section used an association-based aggregation rule to increase the sampling step but this process can still be improved. Indeed, like the resolution of an image, one could imagine that increasing the sampling step of the LGM leads to aggregate information of the cells that are within bigger cells but also its neighbor cells. It behaves as a smoothing mechanism and may reduce the sampling step side effect.

Prediction related

Models improvement The hypothesis made in the experiments can easily be extended to more complex assumptions. The models used for the prediction of road users can be improved. Indeed, as mentioned, predicted bounds of road users can be provided by a better predictor. Extension to include vulnerable road users can also be done based on the same principles. The evolution model should be set according to the object classification for instance.

Additional constraints The neutralized constraint has been explored in the set-based prediction work. It must be highlighted that additional constraints can be used and explored. For instance, no overtaking or keeping safe distances between road users on a same lane could be added in order to add consistency in prediction and also reduce pessimism with a better situation prediction analysis.

Sampling evolution The results presented have shown that the sampling step also enables to manage integrity of a predicted LGM. Similarly to the static LGM generation, a fixed sampling step has been used for the whole prediction horizon. However, as a prediction model may become imprecise as time goes by, one could also imagine that the sampling step may evolve and increase depending on the predicted time horizon.

Virtual road user enhancement As presented in [Wang et al. – 141], the concept of virtual road user may be used in hidden spaces. Each virtual road user is defined by a two-dimensional space with a position (in the hidden space) and a speed. In other words, each virtual road user is defined by the bounds of its possible speed and the bounds of its possible position in the hidden space. It gives a discretized 2D space of speed and position combination. Thanks to the ego-vehicle displacement, possible combinations of speed and position are no longer valid and the position and speed bounds become more accurate. As a consequence, it shows a strategy to improve the bounds computing of the reachable position of these virtual road users and to improve the overall reachability analysis from hidden areas. The worst-case scenario has been considered with virtual road users. Like with the neutralization, additional constraints enable to increase the availability of the system.

Global perspectives

Perception free-space In experiments and results, the polygon approach has been implemented in order to supply spatial occupancy information to the WM module. However, the free space generated by the perception is a snapshot observation and may be thus noisy. In the proposed architecture, ideally, the perception module

rather supplies estimates of the road users poses, e.g. using tracking, but also an estimate of the spatial occupancy. The perception module can fuse occupancy information over time and thus provide filtered estimates. This principle is well handled with occupancy grids. The free space used in the implementation of this work can be therefore replaced by a more reliable occupancy grid that integrates information over time and more sophisticated perception algorithms, e.g. deep learning-based. The same applies for the detection of road users and tracking.

Output LGM Once the LGM has been built by the WM module, a decision-making process that takes decisions based on this representation needs to be addressed. Thus, a comparison between maneuvers taken thanks to distances and occupancy information provided by an LGM and maneuvers based only on the detection of road users could be done. A decision-making module that takes into account an augmented situation representation, i.e. with the safety and neutralized notions, should exploit this advantage.

In addition to that, as integrity is managed by the LGM through the sampling step parameter, one could imagine that this parameter changes over time depending on the needs of the task to perform. Indeed, an LGM can be queried in real time with this parameter.

Knowledge database update A task that the WM can do is to update the knowledge database. The map maintenance is a well-known issue. However, thanks to the situation understanding of the ego-vehicle surroundings, this process can be improved. Typically, when a feature is detected by the perception system and is not present in the database, an addition can be made. At the opposite, when the feature is in the map but not perceived, its confidence may be lowered. However, in this case, it can be of importance to be able to take into account the situation context. For instance, a feature that is actually not visible from the vehicle point of view, e.g. due to occlusions that may be caused by static obstacles such as buildings or a dynamic road user, has to be considered.

WM evaluation metrics As mentioned, the WM can be used to design the autonomous vehicle requirements based on its efficiency through metrics such as risk evaluation and safety analysis. It is part of the functional architecture proposed and it appears that there are few means to compare this approach to other autonomous vehicle architectures. It still remains an open issue to evaluate directly a global architecture with quantitative metrics and to provide comparisons between architectures. Further study is needed to adapt existing qualitative metrics found in the literature [Lotz – 80] and proposed to evaluate the design of a system architecture.

Appendices

Appendix A

Experimental Setup

Platform

All experiments and results presented in this thesis have been tested using experimental data. The Heudiasyc laboratory has several experimental vehicles that can be used to test autonomous driving features and to acquire data. Several dataset series were recorded in the city of Compiègne. The experiments were carried out with three vehicles of the Heudiasyc laboratory: two experimental Renault ZOE vehicles (figure A.1a) and a Renault master van (figure A.1b).

(a)

(b)

Figure A.1: Experimental vehicles used to lead the experiments.

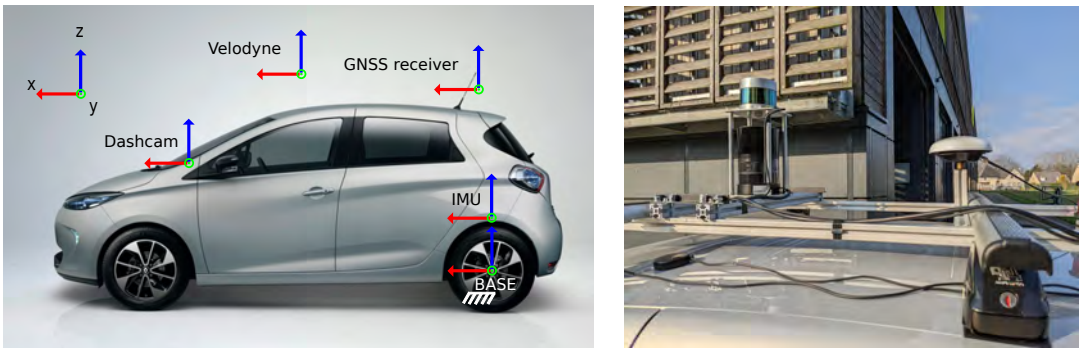
The laboratory purchased two Renault ZOE vehicles in 2013 and 2015 using funds from the Equipex ROBOTEX (ANR-10-EQPX-44-01). Those vehicles are first generation Renault ZOEs which have been modified to enable autonomous control.

In 2019 a third-generation Renault ZOE, was also purchased. This vehicle is not modified for autonomous use. It is therefore only dedicated to sensor acquisitions.

The third vehicle that has been used is a Renault master van. It has essentially been used for assistant purposes. In the experiments, this is the vehicle that had the duty to generate occluded areas for the ego-vehicle (blue one) as it has a length of 6:23 m and a height of 2:55 m. Thus, it enabled to generate neutralized situations.

Sensors

The setup of the sensors is depicted in figure A.2.



(a) (b)

Figure A.2: Sensors setup on the ego-vehicle. (a) Sensors positioning. The origin frame corresponds to the base symbol. (b) GNSS receiver and VLP-32C sensors on the roof of a vehicle.

Ground Truth system

All vehicles are equipped with a Novatel SPAN-CPT GNSS receiver. This system uses a GNSS receiver with RTK capability loosely coupled with a high-accuracy IMU. Post-processed GNSS PPK corrections, with "Inertial Explorer" from the Novatel company (this is called Post-Processed Kinematics - PPK), was used to have a centimeter-level localization of each vehicle. This system provides the ground truth state used for the localization. RTK positioning relies on a static receiver at a known position, on top of one of the university buildings, to estimate the errors affecting satellite signals. These errors can then be compensated by the vehicle

receiver to obtain an accurate positioning. The pose of each vehicle was supply at a rate of 50 Hz.

Velodyne VLP-32C

Only the Renault ZOE vehicles are equipped with Velodyne VLP-32C lidar sensors.It was used for the perception processes. This sensor has a 360 field of view with a theoretical range of 100 meters and runs at 10 Hz. It enables to build several types of spatial occupancy representation.

CAN Bus

Only the Renault ZOE vehicles provide access to their CAN Bus. The vehicle internal sensors are therefore accessible. The vehicle speed can be obtained through this bus.

Extrinsic calibration

Calibration of the sensors is essential to correctly model the link between the observations and the vehicle state. The perception sensors and the GNSS antenna position need to be identified.

To obtain the lidar and the Novatel SPAN-CPT extrinsic calibration, a FARO Vantage laser tracker was used. It enables to measure the position of points of the vehicle with an accuracy below the millimeter. The position of the center of the rear axle can be found by measuring points on the rear wheels. The position of the screw with which the sensor is attached to the vehicle was also measured. The measurement system can then provide the transformation from the base frame to the sensor frame.

High Definition Map

Experiments have been conducted in the city of Compiègne, France, where an HD map has been constructed (figure A.3). The map contains 57 km of lanes (a road with two lanes is counted twice).

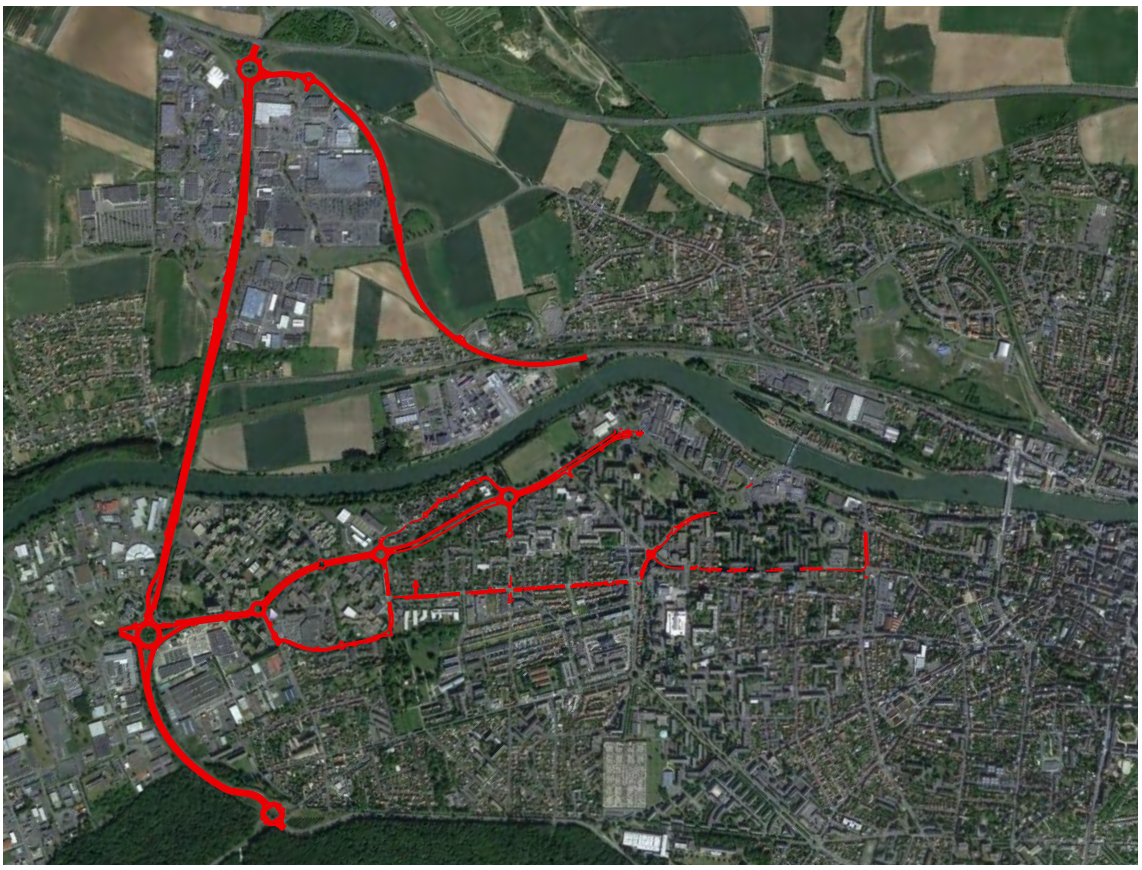


Figure A.3: HD map of Compiègne where experiments were conducted.

Datasets

The experiments and results presented in this manuscript were provided thanks to several datasets that have been recorded. The datasets have been recorded with the help of Stéphane Bonnet, Antoine Lima, Stefano Masi, Thierry Monglon and Anthony Welte.

Software

The Robot Operating System (ROS) middle-ware was used for the implementations and the Boost Geometry library for the geometric operations between the polygons of the areas of interest and the cells of the LGM. The Jsk library was used to manage the road users bounding boxes.

Figure A.4 shows the software architecture of this work. The modules that have been fully developed and implemented are shown in red. They enabled to compute the necessary inputs of the WM: