Urban environment perception and navigation using robotic vision : conception and implementation applied to automous vehicle

Giovani Bernardes Vitor

Par **Giovani BERNARDES VITOR**

Urban environment perception and navigation using robotic vision : conception and implementation applied to autonomous vehicle

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC

Soutenue le 26 septembre 2014
Spécialité : Technologie de l'Information et des Systèmes

D2155

Perception de l'environnement urbain et navigation s'appuyant sur la vision robotique : La conception et la mise en oeuvre appliquées au véhicule autonome

Giovani BERNARDES VITOR

Thèse soutenue le 26 Septembre devant le jury composé de :

Président:

Cairo Lúcio NASCIMENTO JUNIOR
Professeur
Instituto Tecnológico de Aeronáutica - ITA

Rapporteurs:

Patrick RIVES
Directeur de Recherche INRIA
Univ de Nice-Sophia Antipolis

Samuel BUENO SIQUEIRA
Senior Researcher
Centro de Tecnologia da Informação Renato Archer

Examinateurs:

Denis Fernando WOLF
Professeur
Universidade de São Paulo - ICMC/USP

Clésio Luis TOZZI
Professeur
Fac. Eng. Elétrica e de Computação UNICAMP

Jérôme DE MIRAS
Maître de Conférences HDR
Université de Technologie de Compiègne

Directeurs de Thèse:

Janito VAQUEIRO FERREIRA
Professeur
Univ. Estadual de Campinas

Alessandro CORRÊA VICTORINO
Maître de conferences - HDR
Univ. de Technologie de Compiègne

Invité:

Claudio GARCIA
Professeur
Escola Politécnica USP

Université de Technologie de Compiègne

Laboratoire Heudiasyc UMR CNRS 7253

26 Septembre 2014

Giovani Bernardes Vitor

URBAN ENVIRONMENT PERCEPTION AND NAVIGATION USING ROBOTIC VISION:

Conception and implementation applied to autonomous vehicle

PERCEPÇÃO DO AMBIENTE URBANO E NAVEGAÇÃO USANDO VISÃO ROBÓTICA:

Concepção e implementação aplicado à veículo autônomo

92/2014

CAMPINAS
2014

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA

Giovani Bernardes Vitor

URBAN ENVIRONMENT PERCEPTION AND NAVIGATION USING ROBOTIC VISION: Conception and implementation applied to autonomous vehicle

PERCEPÇÃO DO AMBIENTE URBANO E NAVEGAÇÃO USANDO VISÃO ROBÓTICA: Concepção e implementação aplicado à veículo autônomo

Doctoral Thesis presented to the School of Mechanical Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Mechanical Engineering, in the Area of Solid Mechanics and Mechanical Design.

Tese apresentada ao Curso de Doutorado da Faculdade de Engenharia Mecânica da Universidade Estadual de Campinas, como requisito para a obtenção do título de Doutor em Engenharia Mecânica.

Orientador: Prof. Dr. Janito Vaqueiro Ferreira (UNICAMP/FEM)
Orientador: Prof. Dr. Alessandro Corrêa Victorino (UTC/HEUDIASYC)

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO GIOVANI BERNARDES VITOR E ORIENTADA PELO PROF. DR. JANITO VAQUEIRO FERREIRA E PROF. DR. ALESSANDRO CORRÊA VICTORINO

ASSINATURA DOS ORIENTADORES

CAMPINAS
2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

V833u

Vitor, Giovani Bernardes, 1985-
Urban environment and navigation using robotic vision : conception and implementation applied to autonomous vehicle / Giovani Bernardes Vitor. – Campinas, SP : [s.n.], 2014.

Orientadores: Janito Vaqueiro Ferreira e Alessandro Corrêa Victorino.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica.

1. Robôs moveis. 2. Veiculos autônomos. 3. Visão por computador. 4. Inteligência artificial. 5. Sistemas inteligentes de controle. I. Ferreira, Janito Vaqueiro,1961-. II. Victorino, Alessandro Corrêa III. Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Percepção do ambiente urbano e navegação usando visão robótica: : concepção e implementação aplicado à veículo autônomo
Palavras-chave em inglês:
Mobile Robots
Autonomous vehicle
Computer Vision
Artificial Intelligence
Intelligent control systems
Área de concentração: Mecânica dos Sólidos e Projeto Mecânico
Titulação: Doutor em Engenharia Mecânica
Banca examinadora:
Janito Vaqueiro Ferreira e Alessandro Corrêa Victorino [Orientadores]
Cairo Lúcio Nascimento Junior
Samuel Bueno Siqueira
Clésio Luis Tozzi
Denis Fernando Wolf
Data de defesa: 26-09-2014
Programa de Pós-Graduação: Engenharia Mecânica

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
COMISSÃO DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
DEPARTAMENTO DE MECÂNICA COMPUTACIONAL

TESE DE DOUTORADO

URBAN ENVIRONMENT PERCEPTION AND NAVIGATION USING ROBOTIC VISION: Conception and implementation applied to autonomous vehicle

PERCEPÇÃO DO AMBIENTE URBANO E NAVEGAÇÃO USANDO VISÃO ROBÓTICA: Concepção e implementação aplicado à veículo autônomo

Autor: Giovani Bernardes Vitor
Orientador: Prof. Dr. Janito Vaqueiro Ferreira
Orientador: Prof. Dr. Alessandro Corrêa Victorino

A Banca Examinadora composta pelos membros abaixo aprovou esta Tese:

Prof. Dr. Janito Vaqueiro Ferreira
Universidade Estadual de Campinas
(FEM/UNICAMP)

Prof. Dr. Alessandro Corrêa Victorino
L`Université de Technologie de Compiègne
(Heudiasyc/UTC) - França

Prof. Dr. Samuel Bueno Siqueira
Centro de Tecnologia da Informação Renato Archer (CTI)

Prof. Dr. Clésio Luis Tozzi
Universidade Estadual de Campinas
(FEEC/UNICAMP)

Prof. Dr. Denis Fernando Wolf
Universidade de São Paulo (ICMC/USP)

Prof. Dr. Claudio Garcia
Universidade de São Paulo (EP/USP)

Prof. Dr. Cairo Lúcio Nascimento Junior
Instituto Tecnológico de Aeronáutica (ITA)

Firstly to God.
Then to my family and especially
To my lovely, loving and lovable wife...

"Don't limit yourself. Many people limit themselves to what they think they can do. You can go as far as your mind lets you. What you believe, remember, you can achieve."

Mary Kay Ash

Abstract

The development of autonomous vehicles capable of getting around on urban roads can provide important benefits in reducing accidents, in increasing life comfort and also in providing cost savings. Intelligent vehicles for example often base their decisions on observations obtained from various sensors such as LIDAR, GPS and Cameras. Actually, camera sensors have been receiving large attention due to they are cheap, easy to employ and provide rich data information. Inner-city environments represent an interesting but also very challenging scenario in this context, where the road layout may be very complex, the presence of objects such as trees, bicycles, cars might generate partial observations and also these observations are often noisy or even missing due to heavy occlusions. Thus, perception process by nature needs to be able to deal with uncertainties in the knowledge of the world around the car. While highway navigation and autonomous driving using a prior knowledge of the environment have been demonstrating successfully, understanding and navigating general inner-city scenarios with little prior knowledge remains an unsolved problem. In this thesis, this perception problem is analyzed for driving in the inner-city environments associated with the capacity to perform a safe displacement based on decision-making process in autonomous navigation. It is designed a perception system that allows robotic-cars to drive autonomously on roads, without the need to adapt the infrastructure, without requiring previous knowledge of the environment and considering the presence of dynamic objects such as cars. It is proposed a novel method based on machine learning to extract the semantic context using a pair of stereo images, which is merged in an evidential grid to model the uncertainties of an unknown urban environment, applying the Dempster-Shafer theory. To make decisions in path-planning, it is applied the virtual tentacle approach to generate possible paths starting from ego-referenced car and based on it, two news strategies are proposed. First one, a new strategy to select the correct path to better avoid obstacles and to follow the local task in the context of hybrid navigation, and second, a new closed loop control based on visual odometry and virtual tentacle is modeled to path-following execution. Finally, a complete automotive system integrating the perception, path-planning and control modules are implemented and experimentally validated in real situations using an experimental autonomous car, where the results show that the developed approach successfully performs a safe local navigation based on camera sensors.

Resumo

O desenvolvimento de veículos autônomos capazes de se locomover em ruas urbanas pode proporcionar importantes benefícios na redução de acidentes, no aumentando da qualidade de vida e também na redução de custos. Veículos inteligentes, por exemplo, frequentemente baseiam suas decisões em observações obtidas a partir de vários sensores tais como LIDAR, GPS e câmeras. Atualmente, sensores de câmera têm recebido grande atenção pelo motivo de que eles são de baixo custo, fáceis de utilizar e fornecem dados com rica informação. Ambientes urbanos representam um interessante mas também desafiador cenário neste contexto, onde o traçado das ruas podem ser muito complexos, a presença de objetos tais como árvores, bicicletas, veículos podem gerar observações parciais e também estas observações são muitas vezes ruidosas ou ainda perdidas devido a completas oclusões. Portanto, o processo de percepção por natureza precisa ser capaz de lidar com a incerteza no conhecimento do mundo em torno do veículo. Nesta tese, este problema de percepção é analisado para a condução nos ambientes urbanos associado com a capacidade de realizar um deslocamento seguro baseado no processo de tomada de decisão em navegação autônoma. Projeta-se um sistema de percepção que permita veículos robóticos a trafegar autonomamente nas ruas, sem a necessidade de adaptar a infraestrutura, sem o conhecimento prévio do ambiente e considerando a presença de objetos dinâmicos tais como veículos. Propõe-se um novo método baseado em aprendizado de máquina para extrair o contexto semântico usando um par de imagens estéreo, a qual é vinculada a uma grade de ocupação evidencial que modela as incertezas de um ambiente urbano desconhecido, aplicando a teoria de Dempster-Shafer. Para a tomada de decisão no planejamento do caminho, aplica-se a abordagem dos tentáculos virtuais para gerar possíveis caminhos a partir do centro de referencia do veículo e com base nisto, duas novas estratégias são propostas. Em primeiro, uma nova estratégia para escolher o caminho correto para melhor evitar obstáculos e seguir a tarefa local no contexto da navegação hibrida e, em segundo, um novo controle de malha fechada baseado na odometria visual e o tentáculo virtual é modelado para execução do seguimento de caminho. Finalmente, um completo sistema automotivo integrando os modelos de percepção, planejamento e controle são implementados e validados experimentalmente em condições reais usando um veículo autônomo experimental, onde os resultados mostram que a abordagem desenvolvida realiza com sucesso uma segura navegação local com base em sensores de câmera.

Résumé

Le développement de véhicules autonomes capables de se déplacer sur les routes urbaines peuvent fournir des avantages importants en matière de réduction des accidents, en augmentant le confort et aussi, permettant des réductions de coûts. Les véhicules Intelligents par exemple fondent souvent leurs décisions sur les observations obtenues à partir de différents capteurs tels que les LIDAR, les GPS et les Caméras. En fait, les capteurs de la caméra ont reçu grande attention en raison du fait de qu'ils ne sont pas cher, facile à utiliser et fournissent des données avec de riches informations. Les environnements urbains représentent des scénarios intéressant mais aussi très difficile dans ce contexte, où le tracé de la route peut être très complexe, la présence d'objets tels que des arbres, des vélos, des voitures peuvent générer des observations partielles et aussi ces observations sont souvent bruyants ou même manquant en raison de occlusions complètes. Donc, le processus de perception par nature doit être capable de traiter des incertitudes dans la connaissance du monde autour de la voiture. Tandis que la navigation routière et la conduite autonome en utilisant une connaissance préalable de l'environnement ont démontré avec succès, la compréhension et la navigation des scénarios généraux du environnement urbaine avec peu de connaissances reste un problème non résolu. Dans cette thèse, on analyse ce problème de perception pour la conduite dans les milieux urbains basée sur la conaissance de l'environnement pour aussi prendre des décisions dans la navigation autonome. Il est conçu un système de perception robotique, qui permettre aux voitures de se conduire sur les routes, sans la nécessité d'adapter l'infrastructure, sans exiger l'apprentissage précédente de l'environnement, et en tenant en compte la présence d'objets dynamiques tels que les voitures. On propose un nouveau procédé basé sur l'apprentissage par la machine pour extraire le contexte sémantique en utilisant une paire d'images stéréo qui est fusionnée dans une grille d'occupation évidentielle pour modéliser les incertitudes d'un environnement urbain inconnu, en utilisant la théorie de Dempster-Shafer. Pour prendre des décisions dans la planification des chemin, il est appliqué l'approche de tentacule virtuel pour générer les possibles chemins à partir du centre de référence de la voiture et sur cette base, deux nouvelles stratégies sont proposées. Première, une nouvelle stratégie pour sélectionner le chemin correct pour mieux éviter les obstacles et de suivre la tâche locale dans le contexte de la navigation hybride, et seconde, un nouveau contrôle en boucle fermée basé sur l'odométrie visuelle et tentacule virtuel est modélisée pour l'exécution du suivi de chemin. Finalement, un système complet automobile intégrant

les modules de perception, de planification et de contrôle sont mis en place et validé expérimentalement dans des situations réelles en utilisant une voiture autonome expérimentale, où les résultats montrent que l'approche développée effectue avec succès une navigation locale fiable basé sur des capteurs de la caméra.

List of Figures

List of Tables

List of Algorithms

Summary

CHAPTER 1

General Introduction

Learn from yesterday, live for today, hope for tomorrow.
The important thing is not to stop questioning.

(Albert Einstein)

1.1 Context and Motivation

This thesis concerns a doctorate in "cotutelle", under the scientific cooperation between the Department of Computational Mechanics of the Mechanical Engineering Faculty from Universidade Estadual de Campinas (DMC/FEM/UNICAMP) in São Paulo - Brazil and the Heudiasyc Laboratory, UMR7253 CNRS/UTC at the Université de Technologie de Compiègne (Heudiasyc/CNRS/UTC) in Compiègne - France. Justifying the scientific cooperation between the institutions over several years, UNICAMP has been working jointly with UTC in the exchange program between engineering students, called BRAFITEC (GELAS et al., 2007). Among one of the visits made by the program coordinator of the UTC at UNICAMP in 2008, professor Dr. Alessandro CORREA VICTORINO proposed a scientific cooperation more specifically inclined to Research and Technological Development between both institutions. After this visit, UTC received the first doctoral student in the scheme of "cotutelle", Arthur de Miranda Neto, to work in the Heudiasyc Laboratory. As a determining factor in this collaboration, the Heudiasyc laboratory has currently two labels, the LABEX MS2T and the ROBOTEX, which classify the laboratory as being a center of excellence in research and also in equipments inside and outside France. A consequence, this thesis is the second of more three "cotutelles"

that have been in progress, consolidating this cooperation toward the sharing and transference research of Technology and Innovation.

In this sense, this thesis is linked to the VERVE project[1] (Novel Vehicle Dynamics Control Technique for Enhancing Active Safety and Range Extension of Intelligent Electric Vehicles) from Heudiasyc - UTC and also, it is linked with the Autonomous Mobility Laboratory (LMA) team consolidated by the Ministerial Order N° 429 MCT as part of the National Research in Brazil.

This research received a scholarship from Coordination Improvement of higher Education Personnel (CAPES) under the process BEX N°9129/12-0 and National Council for Scientific and Technological Development (CNPq) under the process Nº209656/2013-1 in Brazil. It is important to mention that the experimental part of this work was carried out and funded in the framework of the Equipex ROBOTEX (Reference ANR-10-EQPX-44-01). It was supported by the French Government, through the program "Investments for the future" managed by the National Agency for Research in France.

Right behind almost every research work, no matter the subject, there is always a social context which motivates it. Considering the causes of death around the world, According to Bhalla et al. (2014), injuries and air pollution generated by motorized road transport were associated with six of the top 10 causes of death, including ischemic heart disease, stroke, lower respiratory infections, etc., in 2010 (BHALLA et al., 2014). The statistics indicates that the total number of road traffic accidents in the world remains unacceptably high at 1.24 million of deaths per year (WHO, 2013). While the total number of fatal traffic accidents has been slightly decreasing over the last couple of years, still more than 36.499 fatalities have been reported in Brazil and more than 3.992 cases have been registered in France in 2013 (WHO, 2013). Aside from the accidents in road transports, there is the undesirable side effect of significant source of energy consumption (KOJIMA; RYAN, 2010).

It is expected that the development of robots capable of self-driving is a viable solution to the problems mentioned. A robot is free of distractions and capable of making objective decisions at every moment in time, ensuring a higher degree of safety. This advanced technology to address the development of autonomous cars would not only potentially reduce traffic

[1] <http://www.hds.utc.fr/heudiasyc/catalogue-131/projets-133/projets-aser/article/verve/> accessed on: 05/08/14

accidents, but also considerably improve the global energy consumption while offering better driving comfort.

1.2 Autonomous Vehicles

For several years the design of autonomous vehicles is an area of particularly active research in the field of mobile robotics. In 1939, General Motors participated in an exhibition at the New York's World Fair with the philosophy of "Building the World of Tomorrow", which had the most popular exhibit entitled "Futurama". The industrial designer Bel Geddes had proposed exhibit a world 20 years into the future featuring automated highways as a solution to traffic congestion of the day which cars were powered by circuits embedded in the roadway and controlled by radio. According to Geiger (2013), much of these projections are like modern production lines working today. In the 1970s the first mobile robot was created, such as the robot "Shakey" by SRI (NILSSON, 1984). The first driverless car was demonstrated in 1977 by the Tsukuba Mechanical Engineering Lab, that ran at 30 km/h on a dedicated track at Japan (BENENSON, 2008). In 1986, a team of engineers leaded by professor Ernst Dickmanns equipped a Mercedes-Benz van with cameras and other sensors. Their experiments which were based on well-marked streets without traffic, successfully demonstrated a driverless car capable of driving in highway (DICKMANNS; MYSLIWETZ, 1992).

Afterwards, the European Commission began funding major research initiatives such as Eureka PROMETHEUS project, CyberCars and CyberMove, as well as Bodega, MobiVIP in France. They all have the objective of studying the problem of autonomous navigation in urban environments. In 1995, a semi-autonomous car driving up to 175 km/h in highway from Munich in Germany to Odense in Denmark, was demonstrated as one of the results of the Eureka project, where this semi-autonomous car had about 5% of human intervention to accomplishment of the complete experiment. Considerable results can be also seen by the research group of Alberto Broggi in 1996 and 2010 (BROGGI et al., 2012), amongst others. With the Grand Cooperative Driving Challenge in 2011, the AnnieWay team from Karlsruher Institut für Technologie (KIT), leaded by Christoph Stiller, has reached the state-of-art in autonomous platooning systems (GEIGER et al., 2012).

The USA, the most notable developments came firstly from the Carnegie Mellon University (CMU) Navlab project, which achieved 98% autonomous driving with manual longitudinal control using the RALPH (Rapidly Adapting Lateral Position Handler) software (POMERLEAU, 1995). Later, contributions in three widely known Defense Advanced Research Projects Agency, DARPA-sponsored competitions: Two "Grand Challenges" in 2004 and 2005 (BUEHLER et al., 2007), and a "Urban Challenge" in 2007 (BUEHLER et al., 2010). These competitions have enlarged the scope and boosted the production of new researches in perception, control and motion planning. These areas cover probabilistic localization methods, mapping techniques, tracking strategies, global and local planning and the decision management through behavioral and deliberative hierarchies. The DARPA challenges have demonstrated that embedded robotic system can completely operate a passenger car traveling over significant distances and manage complex situations arising from quasi-urban real conditions. In august 2012, the Google enterprise announced that its google driverless car, a Toyota Prius equipped with self-driving capabilities (WIKIPEDIA, 2012) which has completed over 300.000 miles without accident, under the guidance of Sebastian Thrun and the current team lead Chris Urmson.

While one might think the problem is solved, some of the aforementioned projects are targeted at tasks like highways driving, lane keeping/following, contrasting with this thesis that deals with the more challenging task of understanding urban scene situations. And others, that use more constraints like digital maps, or then, using expensive laser sensors to interpret the environment. Therefore, there is indeed an important work, particularly on improving the robustness and reliability of the techniques used.

1.3 Conception and challenges of autonomous vehicle in urban environment

After two decades of research progress, autonomous car has demonstrated enough technology to deal in indoor environments as well as in controlled outdoor environment. However, designing an autonomous car to drive in a dynamic urban environment is still a challenge. Even when the feasibility of the robotic task has been demonstrated with a prototype, defining the

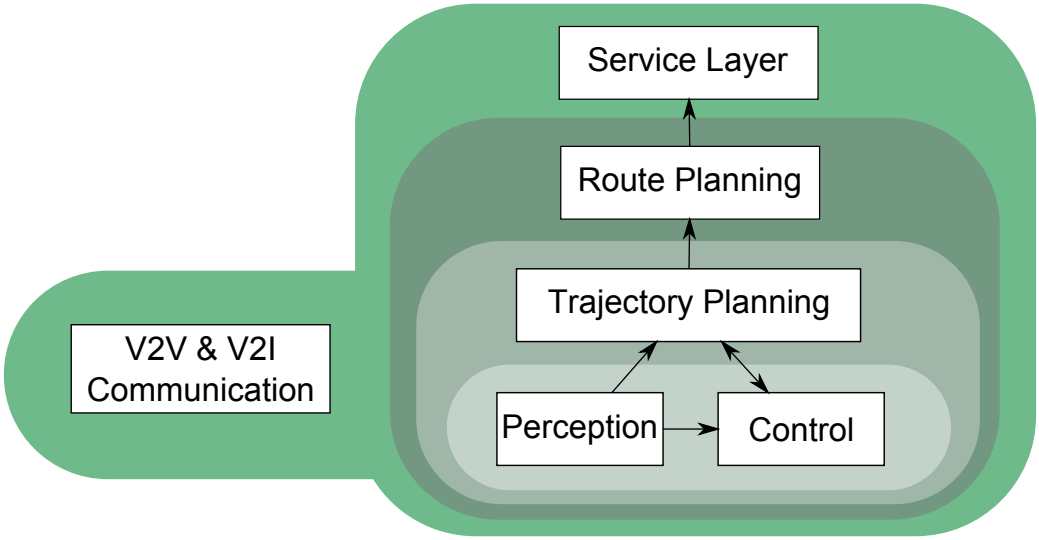


Figure 1.1 - Main areas that model a type of the automatic car concept. Adapted Source: (BENENSON, 2008)

best methods to accomplish this task, with the lowest constraints and the lowest costs, remains an open issues.

In this context, this thesis focuses on the scenario of autonomous car in inner-city scenarios, since it is the most challenging and the one with higher potential impact on daily life. In his work, Benenson (2008) explains that the development of autonomous car can be decomposed in several areas of research and development. These areas are roughly grouped in six fields of research that interact among them. As can be seen in Figure 1.1, developing such a robot presents multiple theoretical and practical challenges at different levels. A brief descriptions of the mentioned fields of research are given here, highlighting their conceptions and challenges.

First of all, the mechanics of the car should be automated. All vehicle controls, such as throttle, braking and steering can be controlled by electronic devices that can be coupled, or even replaced by some specific mechanical components. Nowadays the automotive industry has produced its commercial cars including several functionalities, starting from the automation of actuators, called by “drive-by-wire”, until sophisticated onboard computers that manage a complete automatic car system. In this sense, it is acceptable to consider that the mechanical and automation issues are in an advanced stage, and today, the biggest challenging is centralized at the signal and information processing layer.

The Service layer covers a wide area of research. It can include Human-Machine Interface (HMI) that explores new ideas focused on innovative interaction and visualization strategies to

keep increasing amount of information in vehicles easily accessible (SPIES et al., 2009). It can consider the new concept that autonomous vehicles will provide large savings by allowing travelers to rely on shared, self-driving taxis instead of personal vehicles, reducing ownership and parking costs (FAGNANT; KOCKELMAN, 2013; LITMAN, 2014). Consider for example the self-driving taxis. A business service model could be employed to process a request of a customer that is using a mobile device. This leads to the development and design of the hardware infrastructure, the software layer and the communication system required to accomplish this task. According to Benenson (2008), to complete a request for a self-driving car it is required to specify the HMI encounter location:

"Where is the user calling from?"

Trying to solve this question, the system could specify its location using GPS, WiFi localization and/or explicitly constructing an application to interact with the user demanding his information. Then, defined the two points, where the customer and the self-driving car are located, it is possible to generate a navigation plan to be executed by the autonomous taxis.

Navigation can be roughly described as the process of determining a suitable and safe path between a starting and a goal point for a robot traveling between them (LATOMBE, 1991; CHOSET et al., 2005). Among existing ways to structure a navigation, the concept applied in this work considers an abstraction in multilevel organization. It means that there are two distinct layers to make the navigation procedure (BENENSON, 2008), defined as global navigation or route planning, and the local navigation or trajectory planning, as illustrated in Figure 1.1. At the route planning, challenges are concerned to solve multiple related optimization problems to find the best solution from all feasible paths between these two points in a given map. This field of research tries to estimate the solution depending in what situation the problem is conditioned. Taking the last cited example, some questions arise (BENENSON, 2008):

"Which is the best path to reach point B from A?"
"How to predict the traffic flow?"
"Which vehicle should serve a request?"

These logistic aspects impose a complex challenge to be solved and suboptimal solutions can be demonstrated (AWASTHI, 2004; AZI et al., 2007; POLIMENI; VITETTA, 2012). An application example in this field is given by Google Maps, as shown in Figure 1.2.

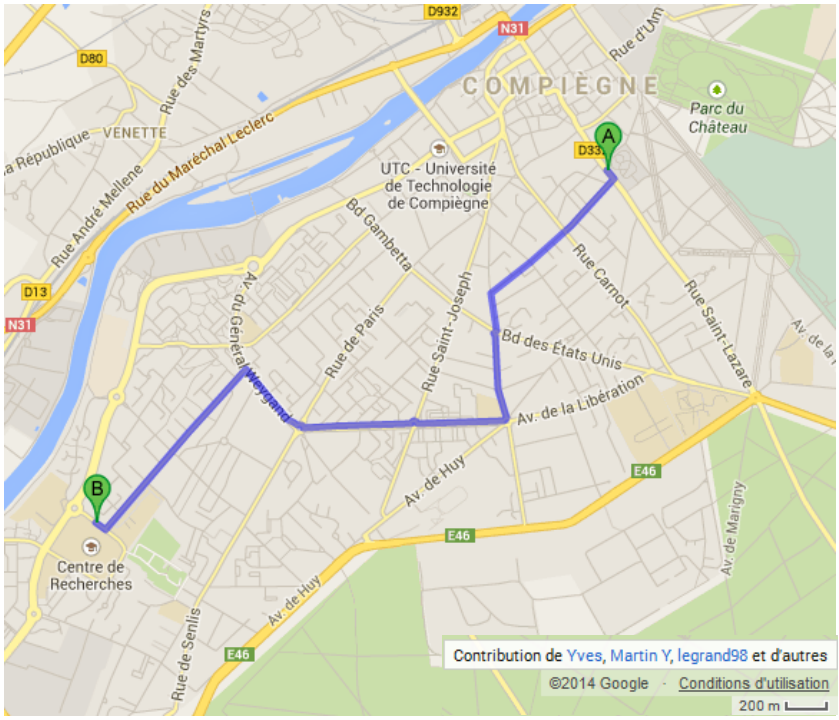


Figure 1.2 - A example of Route Planning given by Google Maps

At the trajectory planning, a safe trajectory has to be molded as a sequence of states in time conditioned to some aspects that should be taken into account, the accomplishment of the objective defined at the route planning, the obstacle avoidance and geometric restrictions of the car. Therefore, safe trajectory planning of nonholonomic robot in uncertain dynamic environment is an interesting challenge to be solved and recently it has been the focus of several researches with different strategies. In this case, the questions that arise are:

"How to reach a goal from where I am?" (Localization and planning)
"How to find a safe trajectory in an uncertain dynamic environment?" (Perception and planning)

Figure 1.3 shows a typical scenario of a local environment where the trajectory planning is conditioned. As can be seen, to perform the trajectory planning, the robot should be able to perceive the environment around it and then control its movements to follow the estimated trajectory.

The perception module is responsible to build a consistent representation of the environment, transforming data acquired from sensors in useful information. Depending on how robust and rich this representation is modeled, it directly impacts on the performance of the trajectory and control procedures. In the work of Levitt and Lawton (1990), some questions in the scope of

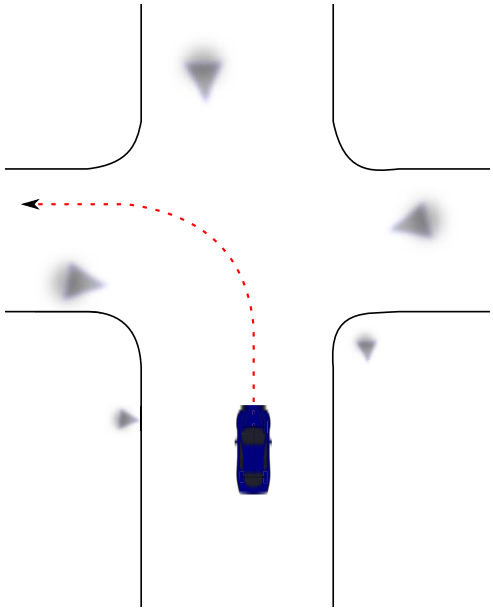


Figure 1.3 - An example of Trajectory Planning made at local environment.

Figure 1.4 - An example of urban scenario.

navigation, intrinsically related to the perception of the robot, are globally defined, as explained by Benenson (2008):

"Where I am?" (localization)

"What is around me?" (exoperception, obstacles detection)

"Are there indications on the road?" (application of specific perception)

The actual research in autonomous navigation addresses several fields of the perception that includes scene understanding, map representation and localization. In the next section the key elements to perform a safe navigation are discussed, considering these mentioned aspects, using strategies to lead with images acquired from urban scenario such as this one illustrated by Figure 1.4.

The Control module is responsible to ensure that the physical trajectory defined by the

Figure 1.5 - Vehicle to Vehicle Communication example [2].

planner is followed, having a closed control loop along its displacement. It is deeply coupled to the perception and trajectory modules. Poor environment model leads to divergence between reality and estimated observation. It forces the trajectory to perform abrupt paths, making hard its execution. Considering the geometric restrictions of the car, and its dynamics, the model that governs a nonholonomic car becomes a problem not trivial. Different models using linear and non-linear methods are found in literature (LAUMOND, 1998). However, non-linear control of nonholonomic robots is still an open problem for the general case (BENENSON, 2008).

To finish the explanation of Figure 1.1, the last module concerns the communication among vehicles. According to Santa et al. (2013), communication architectures integrating Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications will be the key of success for the next generation of cars. They aim at improvements in transportation, greater energy efficiency, less road construction, reducing collisions and safety of vehicle occupants, as illustrated in Figure 1.5.

In V2I, the infrastructure plays a coordination role by gathering global or local information on traffic and road conditions and then suggesting or imposing certain behaviors on a group of vehicles. In the case of V2V, it is more difficult due to its decentralized structure, it aims at organizing the interaction among vehicles and possibly develop collaborations among

[2]Source: <http://www.safercar.gov/v2v/index.html> Accessed on: 22/07/2014

them (WEIß, 2011). Actually, this field of research has been centralized at defining patterns of communication and also studying security assessment of Vehicle Services, since many applications will be offered to the drivers and smaller hand-held devices like Android and iPhones will be seamlessly integrated into the vehicles' networks. Thus, some questions arise to be answered:

"What is the best communication technology to be used for vehicular communication architecture?"
"How to design security systems able to void possible attacks?"
"How to manage decentralized structure among vehicles?"

Large discussion about challenges in this field can be seen in an Intermediate Report on the Security of the Connected Car (2012). Such a communication system is likely to interact at multiple levels of the self-driving cars.

1.4 Key elements for autonomous navigation

Based on the conceptions explained before, it is possible to derive some capabilities that are essential to the autonomous navigation and used in this study. Among all tasks, the autonomous navigation should be able to accomplish the following tasks:

To achieve these tasks, the vehicle needs to know its position, characterize its environment, plan its trajectory and follow it in a controlled way. Therefore, the autonomous vehicle is facing with the problems of localization, perception and planning.

1.4.1 Localization

The location task in urban environment is not trivial and it serves to answer the essential question mentioned before, *"Where I am?"*. Two approaches can be used:
It can be located in a relative way, with respect to its previous position (generally considering

its initial position). To do this, it must know its movement and, in this case, it is denoted by odometry. This technique consists of estimating the movement of the vehicle on an short instant of time using proprioceptive sensors. Thus, the displacement of the vehicle can be obtained by:

- Encoders placed on the wheels of the vehicle: speed measurement.
- Inertial units: measuring acceleration (accelerometers) and angular velocity (gyroscopes).
- Or exteroceptive sensors: Camera, LIDAR, sonar.

Furthermore, the vehicle may also be located in an absolute manner. In this case, the principal method used by excellence is the global positioning by satellites (GPS, GLONASS, GALILEO, etc). It provides a georeferenced position of receiver by measuring the time of reception of signals transmitted by satellites, enabling its localization. It should be mentioned that both approaches can be fused as well.

1.4.2 Environment perception

As mentioned, the environment perception can be defined as a complex process to model a consistent representation of the environment, transforming the data acquired from sensors in useful information. Then, the perception is employed to solve three kinds of problems:

Ego-Localization: As it was already considered in section 1.4.1, the perception can be used to self-localization of the environment by answering the question *"where am I in my local environment?"*. This problem is mainly based on the detection and correspondence of natural features of the environment in the referential of the vehicle, where the absolute position is known or estimated. In the case where the position of the detected elements are unknown, but they are static and incremental, it is possible to model the perception to build a map in which these elements are positioned relative to the vehicle, and then, based on this map, it is possible to locate itself. This problem is known as Simultaneous Localization and Mapping (SLAM) and had its formalism firstly defined by Smith and Cheeseman (1986)

Object detection, tracking and recognition: The main role of perception is to detect and characterize the obstacles in the environment of the vehicle. This action can be broken into

two basic tasks: the detection and localization of objects in the environment.

The detection process answers the following question: *"What is around me?"*. If there are elements present in the scene, this process is frequently complemented by the location task, which answers the subsequent question: *"Where are these elements located?"*.

These two tasks are the basis of any perception system. They require the use of exteroceptive sensors to gather data about the vehicle's environment. Once defined these tasks, it is possible to classify these elements using a machine learning process to provide a better understanding of the scene. Further, the tracking of mobile objects might be applied to establish a temporal relationship between detections of the sensor and also to increase the robustness of the detection front to measurement errors and occlusions.

Determining the free area of the road: The perception is also useful for determining the free areas, which means, the area where the vehicle can move without collision. In several works, the free area is defined using the dual of the obstacle detection. This area is determined implicitly, derived from the detection and tracking of obstacles. However, this is not always true, considering the urban environment. In this type of scenario, the free space is composed by the street and infrastructure elements such as sidewalk. Thereby, the navigable area must be explicitly determined, preventing the autonomous vehicle to drive over a sidewalk for example.

In the field of driverless cars, there are many approaches including this one, that use the explicit representation of the navigable space to perform the trajectory planning (URMSON et al., 2007; MONTEMERLO et al., 2008; CHERUBINI; CHAUMETTE, 2011).

1.5 Problematic of the thesis

In this thesis, the main difficulties associated to the perception in urban environment are analyzed. This problem is closely related to the problematic of autonomous navigation. Therefore, it is proposed a perception-planning method able to deal with an heterogeneous environment populated by static and moving objects. This scenario poses an interesting problem and is challenging due to the heavy occlusions caused by objects in the scene, difficulties caused

by the sensor noise leading to inaccurate, sporadic measurements and noisy depth estimates as well as limited camera field of view. Other, even if these problems are clarified in the perception, another two interesting issues arise in the planning and navigation process. The first one is associated with the fact that trajectory planning tries to reach a precise point on a local map, while current route planning provides information within a few meters of error (taking into account any route planning generated by a common commercial navigation product). In this sense, the low precision way points have to be matched to the centimetric precision map used for trajectory computation. The second one is related with the capability that the vehicle passes from one region to another, defined by the route planning (for example switching between two streets). Since the trajectory planning method tries to reach a goal defined by a specific vehicle state (a specific point on the road), its extension to consider the notion of "goal region" is non trivial.

The comprehension of this thesis tries to mold the work around the following statement:

"Perception in inner-city scenarios can be done without the need for adapting the infrastructure to ensure integrity on the local navigation process."

In particular, the proposed method tries to answer the following questions:

- *"Where are the streets?"* (Different from free space that does not distinguish street and sidewalk)
- *"Where are the vehicles located?"* (Differentiation of the various obstacles classes)
- *"Which car is moving?"*
- *"What is the current traffic situation?"*
- *"How to find a safe trajectory to reach a defined goal in an uncertain environment?"*
- *"How to bind the route planning in the trajectory planning?"* (Association of global and local navigation)

Under these circumstances, it is tried to answer the aforementioned questions using visual measurements alone, which provide substantial information of the environment such as texture and deep at low-cost, never get outdated (as map do) and mimic the human perception process. Further, cameras are already onboarded on standard vehicles differently from laser scans, that are still more expensive. It should be mentioned that cameras still have some restrictions such as weather and nighttime environments.

This thesis, as part of the field of visual perception for car-like robot, or the so called intelligent vehicles, is associated to the field of mobile robotics and computer vision. It deals, in context of SLAM, with the challenge of semantic context association in the local dynamic evidential grid. Considering the semantic context, it also addresses the problem of segmentation process and recognition of road detection and urban scene understanding. In the field of navigation, this work seeks to solve the last two concerned questions. Following these multidisciplinary domains that are merged to reach those objectives, this work proposes a solution system depicted in Figure 1.6. As can be seen, a set of tasks have been designed for accomplishment of an appropriate perception scheme which allows its validation into a safe displacement based on decision-making process in autonomous navigation. The solution approach is divided into four main tasks defined as (I) Semantic Context, (II) Dynamic Evidential Grid, (III) Local Navigation and (IV) Global Navigation.

The (I) Semantic Context task aims to urban road scene understanding. In this layer, it is developed four modules that produce meta-knowledge from a pair of stereo images. The first module, the Image Segmentation, is used to obtain the super-pixel segments. The Texton Map and Dispton Map modules are employed to compose a set of features to represent each super-pixel segment into the Machine Learning module. This last module receives these computed information and then it constructs feature descriptors for the whole super-pixel segments, addressing these vectors to be classified in a proposed classifier. Using this structure, the principle to get the meta-knowledge is applied for two complementary contexts. The first one concerns only Road Detection. In autonomous navigation, the basic and principal capability of such system is to recognize precisely the road in front of the vehicle, where any driving maneuver or vehicle control should be performed. The second context is characterized as an extension of road detection, considering more elements to be classified and in this case, performing the semantic urban road scene understanding. The output result of the (I) Semantic Context task is illustrated in Figure 1.6(a). Early road detection approaches, specifically related to vision-based principle, seek to use a large diversity of features ranging from spatial ray features, color, texture or then applying techniques in 3D data information such as V-disparity map. These features alone are not sufficient to deal in the context of inner-city scenarios yet. As a contribution, this thesis aims to combine 2D and 3D data, enhancing the detection and reducing the drawback of each approach running separately. In this sense, one of the proposed classifiers has become a cutting-edge approach in the field of road detection, where it will be detailed in Chapter 3.

The (II) Dynamic Evidential Grid task aims to perform a local perception mapping and characterization of static and moving obstacles using the output responses leveraged by the layer (I). The (II) Dynamic Evidential Grid layer comprises three modules responsible to model a dynamic local occupancy grid, applying the Dempster-Shafer Theory. The Structure from Motion (SfM) module is applied to get the rigid transformation between two successive pair of images. The Sensor Grid module builds a novel inverse sensor model that projects 3D point obtained from the disparity map into a metric grid, taking into account the noise in the stereo measurements and the uncertainty linked with stereo geometry reconstruction, where exponential error is observed increasing the distance. Furthermore, the meta-knowledge extracted from urban road scene understanding is associated into this proposed inverse sensor model, where it provides a better and reliable representativeness of navigable, infrastructure and obstacles areas. After that, the Perception Grid module performs the temporal fusion and mobile cells detection. There are several approaches in the field of occupancy grids that propose an inverse sensor model using laser sensors and also incorporating meta-knowledge information extracted from digital maps, which requires in this case a precise global localization to associate these information. Other approaches, taking advantage of the stereo vision to propose an inverse sensor model, are based on epipolar geometry techniques together with V-disparity methods. They extract 3D points and distinguish these points in obstacles and non-obstacles to reproject in the grid. Differently, this work proposes a method that takes into account the semantic context information with epipolar geometry to generate the inverse sensor model. Compared with studied approaches, it does not require neither some prior digital map information, nor a precise global pose estimation. The proposed Dynamic local Perception grid contributes to the grid-based approaches and allows to account semantic, dynamic and uncertainty aspects in the representation directly. The output result of this layer (II) can be seen in Figure 1.6(b) and it will be detailed in Chapter 4.

The proposed navigation adopts an hybrid approach. This hybrid approach mimics the conception like humans may drive in urban environments. The (IV) Global Navigation layer has the Route Planning module that uses a topological representation to generate the complete route planning. It is usually run as a planning phase before the robot-like car begins its journey. The actual approach considers only a local navigation task at the moment although it uses this theoretical concept to model a framework that binds the route planning in the context of path-planning, performing in this way a deliberative-reactive navigation. In this case, a metric representation principle is employed to local navigation task, which is more appropriate to

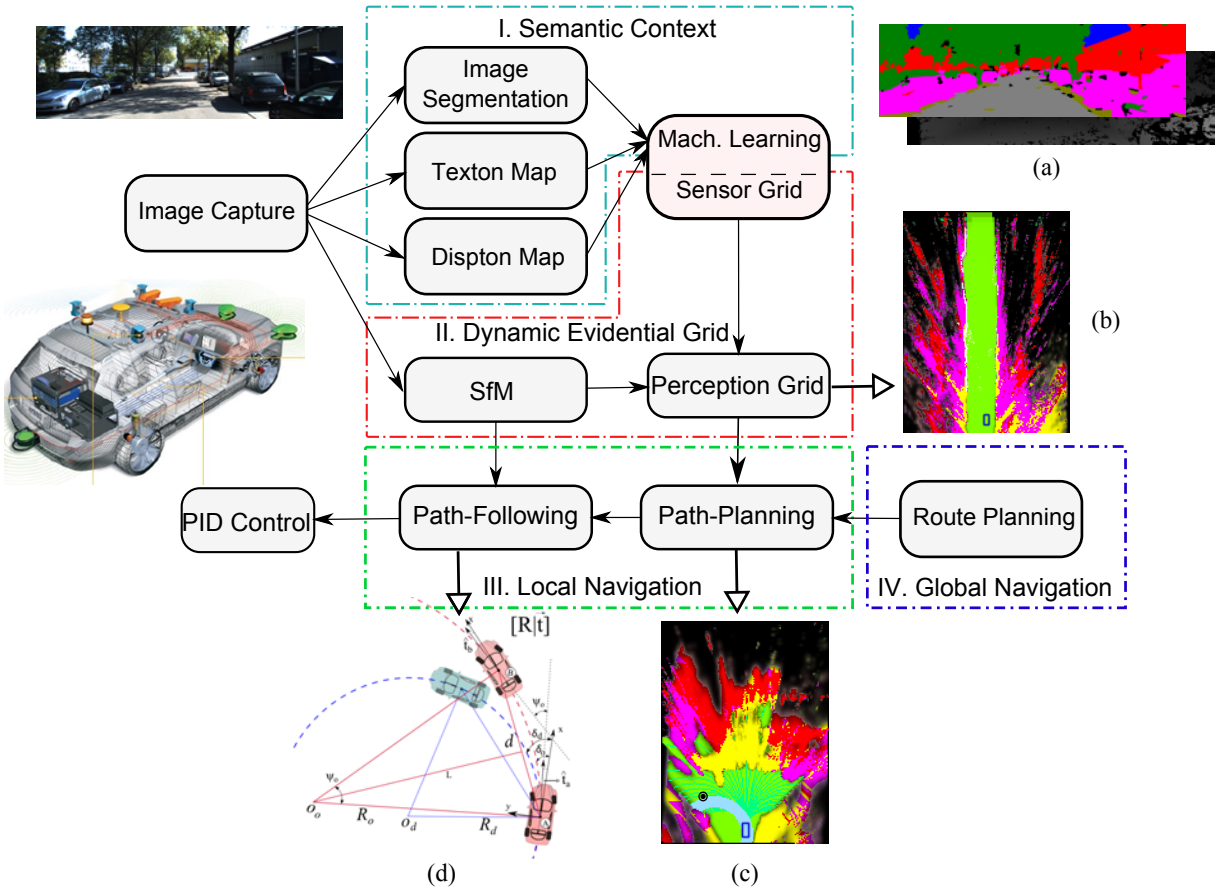


Figure 1.6 - The proposed solution for autonomous navigation using only a pair of stereo camera sensor.

reactive issues and is more robust to localization errors. Therefore, the (III) Local Navigation layer has two modules to accomplishment of this task. The first one, Path-Planning module, is responsible to generate the correct path by where the vehicle should go, applying a new extension of the virtual tentacles approach. The Path-Following module uses a simple proportional control to ensure the real execution of the desired commands based on a new closed loop control scheme. While the virtual tentacles approach is able to deal in the context of global navigation using the trajectory planning by precise GPS information, its original version is not modeled to deal with hybrid navigation strategy, becoming necessary some improvements to cope with it. As contributions in this field, a new strategy to select the correct path to better avoid obstacles and to follow the local task in the context of hybrid navigation is proposed, introducing the notion of *mutant point*, as can be seen into the output result of the Path-Planning module in Figure 1.6(c). Further, at the best of us knowledge this work is the first one to apply a specific path-following scheme, it means, a new closed loop control based on visual odometry and virtual tentacle is modeled to path-following execution, which is illustrated in Figure 1.6(d). These concepts and developments will be better explained in Chapter 5.

1.6 Contributions

Besides the aforementioned contributions, others works have been published along the period of doctorate, where several research lines were studied:

- VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. A histogram-based joint boosting classification for determining urban road (extended abstract). In: **International IEEE Conference on Intelligent Transportation Systems (ITSC)**. [S.l.: s.n.], 2014.

- VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. Comprehensive performance analysis of road detection algorithms using the common urban kitti-road benchmark. In: **Workshop on Benchmarking Road Terrain and Lane Detection Algorithms for In-Vehicle Application on IEEE Intelligent Vehicles Symposium (IV)**. [S.l.: s.n.], 2014.

- VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. A probabilistic distribution approach for the classification of urban roads in complex environments. In: **Workshop on Modelling, Estimation, Perception and Control of All Terrain Mobile Robots on IEEE International Conference on Robotics and Automation (ICRA)**. [S.l.: s.n.], 2014.

- LIMA, D. A.; VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. A disparity map refinement to enhance weakly-textured urban environment data. In: **International Conference on Advanced Robotics (ICAR), 2013 IEEE**. [S.l.: s.n.], 2013.

- VITOR, G. B.; LIMA, D. A.; VICTORINO, A. C.; FERREIRA, J. V. A 2d/3d vision based approach applied to road detection in urban environments. In: **Intelligent Vehicles Symposium (IV), 2013 IEEE**. [S.l.: s.n.], 2013. p. 952–957.

- BELENO, R. D. H.; VITOR, G. B.; FERREIRA, J. V.; MEIRELLES, P. S. Planeacion y seguimiento de trayectorias de un vehiculo terrestre com base en el control de dirección en un ambiente real. In: **Primer Congreso Inter-**

nacional sobre Tecnologias Avanzadas de Mecatronica, Diseño y Manofactura (AMDM). [S.l.: s.n.], 2013.

- BELENO, R.; VITOR, G. B.; FERREIRA, J. V.; MEIRELLES, P. Proposta de uma plataforma de testes para o desenvolvimento de veículos autônomos. In: **XIX Congresso Brasileiro de Automática (CBA)**. [S.l.: s.n.], 2012. v. 1, p. 2735–2742. ISBN 978-85-8001-069-5.

- VITOR, G. B.; BELENO, R. D. H.; FERREIRA, J. V. Fuzzy application for mobile robot navigation using computer vision in real-time. In: **In proceedings on Congreso Internacional de Innovación y Tecnología: Sistemas Mecatrónicos.** [S.l.: s.n.], 2011.

- KORBES, A.; VITOR, G.; LOTUFO, R. de A.; FERREIRA, J. Advances on watershed processing on gpu architecture. In: SOILLE, P.; PESARESI, M.; OUZOUNIS, G. (Ed.). **Mathematical Morphology and Its Applications to Image and Signal Processing**. [S.l.]: Springer Berlin / Heidelberg, 2011, (Lecture Notes in Computer Science, v. 6671). p. 260–271. ISBN 978-3-642-21568-1.

- VITOR, G. B.; KöRBES, A.; LOTUFO, R. D.; FERREIRA, J. V. Analysis of a step-based watershed algorithm using cuda. **International Journal of Natural Computing Research (IJNCR)**, IGI Global Disseminator of Knowledge, v. 1, p. 16–28, 2010. Available from: <<http://www.igi-global.com/article/analysis-step-based-watershed-algorithm/52612>>.

1.7 Outline of the manuscript

This thesis contains a detailed description of the methods, theoretical concepts, experimental observations, results and conclusions of this research. Every chapter constitutes a small step towards the implementation of a complete system for safe autonomous navigation in urban environment. Following the proposed system, the manuscript is divided as follows:

- In Chapter 2, it is presented the related works in the principal areas that concern the scope of this thesis, considering the key tasks presented in section 1.4 and positioning the proposed approach with respect to previous works.
- In Chapter 3, it is addressed the Urban Road Scene Understanding to autonomous driving (introduced by the I- Semantic Context in Figure 1.6). In this case, two phases are proposed to accomplish this task, that are the road detection and then the Urban Scene Understanding. A meaningful performance analysis considering three proposed algorithms are produced, highlighting the pros and cons of each one. Based on this fair performance comparison, the best algorithm is chosen to compose the process explained in the next chapter.
- In Chapter 4, it is described the formalism of occupancy grid based on Dempster-Shafer theory to manage uncertainties come from the dynamic urban environment (emphasized by the II - Dynamic Evidential grid in Figure 1.6). Additionally, the complete Local Dynamic Perception is detailed associating the meta-knowledge extracted from the semantic context to the evidential grid. Results demonstrated the viability of this new approach.
- In Chapter 5, it is discussed the navigation approach using virtual tentacles to determine the best trajectory as well as the detailed experimental platform. The approach considers a local navigation task, and gives a framework to bind the route planning in the context of trajectory planning. The experimental result was implemented using the PACPUS framework and tested with the ZOE platform in the test track of the Heudiasyc Laboratory.
- In Chapter 6, the conclusions and perspectives are presented.

CHAPTER 2

Related Works

We cannot solve our problems with the same thinking we used when we created them.

(Albert Einstein)

This work is more centralized at modeling the local environment at the time that the navigation is happening, using visual strategies. It necessarily falls in the problem of SLAM. Once the robot has handled properly these two tasks (localization and mapping), it can perform other tasks more or less complex depending on what it is designed (CHAPOULIE, 2012). So, this chapter presents the principal works in literature, taking into account the three fundamental tasks that the autonomous vehicles must do for driving, it means localization methods, perception tasks and trajectory planning, as described in section 1.4. It starts with an overview of the development of localization systems followed by environment perception which highlights the research in road detection, semantic image recognition and geometric representation of the environment. Finally, it is presented approaches developed to trajectory planning.

2.1 Localization

To locate a vehicle, the primordial system is the GPS. It provides a georeferenced position of the receiver by measuring the time of reception of signals transmitted by satellites, enabling its localization. Despite this system be employed in a extensive range of applications such as maritime, aerial and terrestrial transports, low-cost sensors commonly found in commerce have

a low precision in order of few tens of meters. This effect restricts its direct application in autonomous vehicle due to the requirement on accuracy of the trajectory planning in the order of centimeters, as example to distinguish the traffic lane from the sidewalk.

A feasible solution to estimate the 3D pose of the vehicle can be performed by cameras. This kind of exteroceptive sensor provides a rich information (geometric and photometric), is precise, low-cost sensor and easily embedded in vehicle. Visual localization can be estimated considering the camera displacement relative to a fixed referential. This referential might be, for example, the first image of a sequence or, then, a known object that is tracked along the image sequence, enabling the extraction of the relative position between the camera and the object. According to Meilland (2012), the strategies to estimate the camera movement between two sequence of images can be divided to two groups, feature-based methods and image-based methods.

In Feature-based methods, a sparse set of points is extracted per image frame and the points are matched temporally based on their feature descriptor. Frequently, the feature extraction is based on the points of Harris and Stephens (1988), SIFT (Scale Invariant Features Transform) proposed by Lowe (2004) or SURF (Speeded-Up Robust Features) demonstrated in Bay et al. (2008). The matching process is performed using SSD (Sum of Squared Differences) or SAD (Sum of absolute differences) and also employing some technique to improve the robustness, such as RANSAC (Random Sample Consensus) established in (FISCHLER; BOLLES, 1981). This strategy is the most commonly used for pose estimation between images, thanks to the reduction of feature points contained in the images (GEIGER et al., 2011; BADINO; KANADE, 2011; BADINO et al., 2013). Differently from the feature-based methods, Image-based methods are not used for the feature extraction. The displacement of the camera is obtained directly from the raw image data, minimizing the error between the two images by a process of image registration (TYKKALA; COMPORT, 2011). This kind of strategy is frequently applied for both planar surface and multi planar surfaces (SILVEIRA et al., 2008; DAME; MARCHAND, 2010).

To reach the problem of SLAM in the scope of localization, classically, the methods employ variations of these two strategies explained before, adding some estimator such as Kalman filter, Particle filter or then changing the configuration of the cameras to gather some benefit for example using mono, stereo or omnidirectional approach (MONTEMERLO et al., 2002;

DAVISON; MURRAY, 2002; THRUN, 2002; DURRANT-WHYTE; BAILEY, 2006; KLEIN; MURRAY, 2007; COMPORT et al., 2010; WU; QIN, 2011). However, these approaches present a problem of drift, generated by the integration error, for long distances. To avoid it, some approaches like Williams et al. (2009) and Chapoulie et al. (2011) use the loop-closure detection to identify places by where the vehicle has already passed. Other approaches divide the localization into two steps. Firstly they consider a model of map built off-line using a technique of learning such as image memory, where the map model is represented by a graph containing key-images extracted from a learning process before the navigation. Secondly, it is performed an localization ego-centered (relative localization), associating the image acquired on-line with the database modeled, enabling a global localization inside this learned model (MEILLAND et al., 2010; MEILLAND, 2012).

Based on these aforementioned approaches, this thesis uses a solution of localization ego-referenced to perform an accurate navigation, avoiding propagation errors related to reconstruction and geometric approximation. The global localization can be expressed by a common GPS, where the required precision to navigate is already solved by the ego-localization.

2.2 Environment Perception

One major challenge for intelligent autonomous driving systems is the requirement to perceive and interpret the environment (GEIGER, 2013). It is aimed on cheap and reliable visual perception and this section surveys the current state-of-the-art in this field. The spectrum of the referenced works ranges from very task-specific methods (e.g. road detection) to more general urban scene understanding approaches (e.g. semantic image recognition), positioning this work among the existent approaches concerned to geometric representation of the environment.

2.2.1 Road Detection

Research into Intelligent Transportation Systems (ITSs) has seen considerable development over the last few decades, particularly in fields of vehicle safety and autonomous driving. The latest driving-assistance systems include many vision-based applications such as lane detection (ASSIDIQ et al., 2008), road detection (KUEHNL et al., 2012) and pedestrian detection (BROGGI et al., 2009), providing drivers with useful information. Current vision-based solutions for Intelligent Vehicles are mostly focused on the detection of obstacles such as pedestrians, bicyclists and other cars (BEHLEY et al., 2013). However, research into Advanced Driving Assistance Systems (ADAS) is making steady progress toward the implementation of systems for urban road scene understanding and not merely obstacle detection. The system for urban road scene understanding can be done by archiving both object recognition and semantic segmentation in inner-city scenarios. A method to automatically classify each pixel and then extract the meaningful regions of a inner-city scene would be particularly helpful in ADAS and also in application of autonomous driving.

Several approaches specifically related to vision-based for road detection have been proposed, varying from mono to stereo or omnidirectional vision sources. These specific methods are useful in cases of less structured roads, for example when driving on dirty roads, as required during the navigation in inner-city scenarios. They aim at solving several challenging problems such as the continuous change of backgrounds in different environments (inner-city, highway, off-road), the different road types (shape and color), the presence of different objects (signs, vehicles, pedestrian) and also the different imaging conditions (variation of illumination and weather conditions).

Early approaches directly classify each pixel using the gray value structure tensor as feature (ZHANG; NAGEL, 1994). To increase robustness, approaches using Mono vision are based on color (ALVAREZ et al., 2009; ALVAREZ; LOPEZ, 2011; TAN et al., 2006), texture (RASMUSSEN, 2004; KONG et al., 2010) or the combination of both (STURGESS et al., 2009) (YUN et al., 2007). Despite the rich amount of information present in the color feature, which imposes less physical restrictions by performing the detection according to some similarity, it may fail due to high intra-class variability present in the dynamic nature of the scenes. The texture is scale-dependent and is affected by the strong perspective in road image. Techniques

such as V-disparity Map (LABAYRADE et al., 2002; BROGGI et al., 2005) are examples of approaches that use 3D information as a source. However, these techniques must deal with different noise sources, such as shadow, road texture, light variations, that make hard to create the disparity map. An alternative to overcome these constraints is to combine 2D and 3D data, enhancing the detection and reducing the drawback of each approach running separately. Some works that use this approach are shown in (SOQUET et al., 2007a; VITOR et al., 2013).

This work details a novel method aimed to inner-city scenarios, merging 2D and 3D information. In fact, it is presented three approaches that were developed during this PhD course that are a extension of the works (VITOR et al., 2014a) and (VITOR et al., 2014c).

2.2.2 Semantic Image Recognition

While the approaches described so far are largely rooted in the domain of robotics and intelligent vehicles, the perceptual side of scene understanding has received a lot of attention in the computer vision and machine learning communities as well.

Seeking to enlarge this road detection concept to scene understanding, the idea is to expand this pixel based classification to be used to identify multiple scene elements in the field of view, including the road surface. Therefore, it is introduced the most important developments in this field, where the principle of all works are to partition the input image into disjoint regions and assign an unique class label (e.g., car, build, vegetation, sidewalk, sky, etc.) to each of them (KUMAR; HEBERT, 2003a; KUMAR; HEBERT, 2003b; TU et al., 2005; BILESCHI, 2006; WOJEK; SCHIELE, 2008; BROSTOW et al., 2008; LADICKY et al., 2009; BACHMANN; LULCHEVA, 2009; STURGESS et al., 2009; DESAI et al., 2009; ESS et al., 2009; GOULD et al., 2009; ZHANG et al., 2010; LADICKÝ et al., 2010; FLOROS; LEIBE, 2012; GUO; HOIEM, 2012; ZHU et al., 2012). These approaches commonly employ offline training or learning using a variety of feature representations like appearance, color, shape or depth to model each class in the scene. These features are then combined to build a spatial smoothness prior for a Markov Random Field (MRF) or Conditional Random Field (CRF) optimization to infer the semantic information for each pixel or super-pixel (i.e., pixel sets with homogeneous

attributes).

While these models handle directly at the pixel and/or super-pixel-level, they provide useful cues which are exploited as features in the same proposed framework of road detection, developed in this thesis. It is aimed to infer the full layout of the scene into an occupancy grid representation including the accurate position and meta-knowledge of buildings, the street, the sidewalk and all vehicles, as reviewed in the next section.

2.2.3 Geometric representation model of the environment

All these road and obstacle detection methods presented, including localization methods, require a special modeling of the environment. It is denoted by mapping in the context of SLAM. Although most of these representations are metric, there are also approaches using topological representations in 2D space and also in 3D space that have demonstrated outstanding results recently, as can be seen in the work of Meilland et al. (2014). Here, it is presented the main modes of geometric representation and their use into existing perception systems, which are divided in Feature-based and grid-based approaches. It should be mentioned that the map-building is directly associated with vehicle location, as previously presented in section 2.1. Therefore, these approaches have the same meaning as before.

Feature-based approach: This method uses geometric features to represent the environment. The type of feature used is dependent on the target application, the environment considered, the accuracy required and the computational power. In many cases, the bird's eye view modeling is employed. As explained, this method depends on the feature extraction and matching to ensure the consistency of the mapping at each instant of time.

Many Feature-based SLAM systems, such as Montemerlo et al. (2002), use a representation of the environment based on natural features. In this kind of approach, the environment is represented by a state vector containing the coordinates of these landmarks. The state vector is filtered over time using Kalman filter (KALMAN, 1960) or particle filter (JULIER; UHLMANN, 2004). The upgrade process between detections is resumed to the problem already referred of data association, that is processed in different ways (SHALOM et al., 2000; HäH-

NEL et al., 2003). In others, this representation concerns the obstacle detection using 2D or 3D shape-models. It depends on the application. In Petrovskaya and Thrun (2009) and Fayad and Cherfaoui (2007), they are interested in vehicle detection using laser sensors, which objects are modeled as rectangular boxes and also performing the tracking over time. As observed, the problem of SLAM considers that the keypoints are fixed. To avoid the impropriety with the inclusion of moving entities, the SLAMMOT algorithm, as presented in (LIN; WANG, 2010), is proposed as a form to improve the mapping, avoiding the usage of these keypoints in the localization process. According to Moras (2013), some pros and cons can be highlighted for feature-based representation:

- Advantages:
    - Simple representation
    - Easy propagation over the time
    - Mobile objects are considered
    - Low memory consumption

- Disadvantages:
    - A non exhaustive representation, and therefore, inadequate to navigation
    - Very high sensitivity to the results of the matching process.

Grid-based approach: This approach works without an object-model: the environment is modeled as a grid of cells, where each one contains information whether the given associated portion of the environment is occupied or not. The occupancy state of each cell is independently evaluated. The update process takes into account all modeled cell of the grid. In general, the cells are square but some work such as Herrmann et al. (2010) considers different geometry to the grid.

Works on occupancy grids using 2D grid to build and update the map of the environment were early proposed by Elfes (ELFES, 1989a; ELFES, 1989b; ELFES, 1991). Initially restricted by the computational complexity, this approach has been widely used for navigation recently. In the works of Thrun et al. (1998), Bourgault et al. (2002), Steux and Hamzaoui (2010), Levinson and Thrun (2010), they have modeled a fixed grid that allows the correction of the position from each new measurement. In Coué et al. (2006), the mobile object tracking using a grid referenced in the vehicle is done. The update process is performed taking into account the occupation and

also the speed vector of each cell. Then, considering the occupation, speed vector and position of the cells, the object can be regrouped. In the work of Gate (2009) is used the grid to perform the SLAMMOT. Alternatively, there are some works that consider the 3D space in which the grid is represented as a cube (MIYASAKA et al., 2009), or then, defining the grid as being a QuadTree (XIE et al., 2010), with the objective of reducing the memory space and the calculus for homogeneous areas. Predominantly, the approaches for grid-based representation have the following aspects (MORAS, 2013):

- Advantages:
    - A comprehensive representation that allows autonomous driving
    - No assumption on the geometry of the elements in the environment

- Disadvantages:
    - It has a complex propagation over time
    - It is difficult to take into account moving objects
    - It has considerable computational cost and memory usage

Almost all previous works use the probabilistic model to represent the occupancy uncertainties in the grid. According to work of Moras (2013), how to represent these uncertainties has important implications on how to process the information contained in the grids. In this sense, in addiction to the probabilistic model, two more approaches are presented in literature such as accumulation methods and evidential methods. For completeness, a brief introduction is given considering the three methodologies.

The formalism of accumulation is quite simple, inspired on the principle of voting: more the cell is seen as occupied, more it is likely to be occupied. Although relatively little used, it is possible to find original contributions using this formalism. In Borenstein and Koren (1991), the accumulation grids are used for the navigation of an experimental indoor robot equipped with sonar. In Xie et al. (2010), the accumulation grids are used to perform the mapping of an external environment equipped with a scanning laser rangefinder. The localization is online guaranteed by grid matching using multiresolution grids (QuadTree).

The probabilistic approach is based on Bayes' theory (BAYES; PRICE, 1763) and is the most used in the field of robotics. It was the first formalism of uncertainty management used in occupancy grids. The first one that proposed this scheme was Elfes in his works (ELFES,

1989b; ELFES, 1989a). This kind of approach defines the state of a cell from two exclusive possibilities, occupied O or free F. Each cell of the occupancy grid contains a probability of occupancy $P(O)$ and/or non-occupancy $P(F)$, and it assumes that all cells are independent of each other. Different formulations exist, using either a direct sensor model or a inverse sensor model (THRUN et al., 2005), and also, considering static or dynamic environment as previously explained.

The third one, the evidential approach, is derived from the theory of Dempster-Shafer and Transferable Belief Model (TBM) (DEMPSTER, 1986; SHAFER, 1976; SMETS; KENNES, 1994), which is a generalization of probabilities. The underlying problem of all grid-based approaches is related to the conflicts generated by the sensitive presence of moving objects in the scene (MORAS et al., 2011a)(MORAS et al., 2011b). The approaches proposed by (MORAS et al., 2012)(KURDEJ et al., 2014) have presented satisfactory results using heuristics combining several sources of information. Nevertheless, these approaches consider the following hypothesis: (i) They are restricted to places where prior digital map information of the environment should be given. (ii) The precise pose estimation of the ego-car should also be supplied (using differential GPS) in order to combine and update the evidential grid. (iii) The perception system is not able to accurately distinguish the feasible navigable area in urban scenarios, it means the street area.

A reliable perception with the annotation of relevant objects could be used as a source to improve safety in urban scenarios. In general, the proposed vision-based approaches to the perception lacks the ability to annotate the environment with semantic information and keep a satisfactory level of precision. Based on preceding works and considering these assumptions that are required to perform the mapping, the proposed work uses the evidential method to deal with the uncertainties in a grid-based approach. This grid is built online, performing a local ego-centered mapping to avoid drift errors and providing the required accuracy to autonomous navigation. It is also introduced the Semantic Context associating automatic meta-knowledge on grid, enabling to manage uncertainties of different entities in a complex urban scenario using only a pair of stereo cameras.

2.3 Trajectory Planning

As introduced in chapter 1, several approaches use a conception of path planning divided to two stages, the route planning or global path planning and trajectory planning or local navigation. According to Victorino (2002), the task of navigation for intelligent systems must use both approaches, where these two processes are complementary. The route planning is concerned with long range planning and is a slow, deliberative process which finds the most efficient path to a long term goal (LUETTEL et al., 2012). The trajectory planning operates with the available information acquired from sensors, at real-time, leading into an effective and appropriate configuration that enables a reactive mechanism to take into account the environment uncertainty (GIESBRECHT, 2004). Some authors denote this conception as being a hybrid approach having two levels: reactive and deliberative (ARKIN, 1990; OREBäCK; CHRISTENSEN, 2003). Based on the principle depicted by Fulgenzi et al.(2008, 2009), the deliberative level has the task of determining and offering to the reactive level, those behavioural patterns that are required for the vehicle to achieve its objectives. The reactive level has to execute these behavioural patterns by ensuring safe and real-time constraints.

Interested on trajectory planning, many reactive approaches have been proposed to deal with the problem of autonomous vehicle navigation for urban scenario. As depicted in the work of (AYARI et al., 2012), among all methods proposed to trajectory planning, the most important methods are denoted by Potential Fields, Vector Field Diagram, Dynamic Window, Curvature Velocity, Nearness Diagram, Elastic Bands and the Velocity Obstacles method. Each one of them has a formalism that is briefly described in the following.

The Potential Field methods (PFM) has been suggested by Andrews and Hogan (1983) and Khatib (1985). The idea of these methods are imaginary forces acting on a robot, where obstacles exert repulsive forces onto the robot while the goal applies an attractive force to the robot. The sum of all forces, the resultant force, determines the subsequent direction and speed of its displacement.

The Vector Field Histogram (VFH) was firstly proposed by Borenstein and Koren (1991) and then enhanced by Ulrich and Borenstein (2000). It uses a statistical representation of the robot's environment through a histogram to represent the obstacle's density and therefore the ro-

bot can move in the direction where there are less obstacles, in order to minimise its interaction with them.

The Dynamic Window approach (DWA) and Curvature-Velocity method (CVM), developed respectively by Fox et al. (1997) and Simmons (1996), are based on the steer angle field approach (FEITEN et al., 1994). They assume that the robot moves in circular paths where a search in the space of translational and rotational velocities is employed to determine the motion commands. This search considers only motion commands which obstacles would not cause a collision within a certain time period, taking into account the kinematic and/or the dynamics of the robot. In accordance with these restrictions, the motion command is designated based on an objective function that maximizes both aspects, the robot safety and the goal.

The Nearness Diagram (ND) observed in the works of Minguez and Montano (2000), Minguez and Montano (2004), consists on an strategy called divide and conquer to simplify the navigation. In this sense, two polar diagrams are used. The first one extracts information of the environment and identifies the immediate goal valley, and the second one defines the safety level between the robot and the obstacles by classifying the closest one.

The Elastic Bands approach (EBA) was proposed by Quinlan and Khatib (1993) with the idea of closing the gap between global path planning and real-time robot control. This method provides a global trajectory that can be modified by using a deformable collision-free path considering artificial forces which depend on the layout of the obstacles in the path.

The Velocity Obstacle (VO) proposed by Fiorini and Shillert (1998), maps the dynamic environment into the robot velocity space. Using this velocity space, the distance difference between robot and obstacle is applied in a function that minimizes its conflict at real-time.

In addition, considerable works have been arisen to overcome some drawbacks observed in these previous methods such as infinite loop or local minimum, or then new conception about reactive approaches. Some of them include behavior-based Fuzzy Logic (ZHU; YANG, 2004; WANG; LIU, 2005; MOTLAGH et al., 2009), a improved ND navigation (MINGUEZ, 2005), obstacle-restriction method (MINGUEZ, 2005), the virtual wall (ORDONEZ et al., 2008), the Closest gap (MUJAHAD et al., 2010; MUJAHED et al., 2013) and a improved DWA (MAROTI et al., 2013).

Looking at the finalists of the DARPA Urban Challenge, an interesting technique that called attention for this field of research was the method named "tentacles", which had its implementation proposed by the team AnnieWAY (HUNDELSHAUSEN et al., 2009). This method is driven by its simplicity, although safe and attends with its objective. Its idea consists on a set of virtual antennae (arc of circle) associating the kinematic constraints of the vehicle projected in an occupancy grid. The best command is defined at real-time, minimizing a linear function composed by the called perceptual primitives, which are functions that perform a given task based on the occupancy grid, such as the use of a longitudinal histogram for classifying tentacles, determining the distance to the first obstacle along the tentacle, and a speed-dependent evaluation length (noted crash distance). Recent developments using this method can be observed in the works of Yu et al. (2012), Ke-ke et al. (2011), Cherubini et al. (2012). They differentiate their approach by changing or adding some perceptual primitive function that composes the linear system, consequently impacting in a different result and functionality.

The proposed method used in this thesis is mainly inspired from Hundelshausen et al. (2009) and Cherubini et al. (2012). In this sense, the tentacle generation follows the definitions explained in the work of Cherubini et al., but, instead of considering three areas of risk (collision, dangerous central and dangerous external areas), here, it is considered only two areas. Based on Hundelshausen et al., it is used the definitions of the crash distance as well the Clearance Value that composes the linear system. However, in this work two functions are implemented to deal with the specific constraints of the proposed system. The first function models a gaussian distribution based in the previous "best" tentacle to perform a smooth transition. The second function defines a "mutant" point that ponders the proposed method, recognizing the information delivered by the route planning, distinguishing from Hundelshausen et al. that uses the direct information of the GPS and also from Cherubini et al. that left the tentacle method purely reactive, since the visual task and the obstacle avoidance task are defined in different state spaces.

CHAPTER 3

Semantic Road Scene Understanding

As far as the laws of mathematics refer to reality, they are not certain; as far as they are certain, they do not refer to reality.

(Albert Einstein)

This chapter addresses the Urban Road Scene Understanding for autonomous driving. In this case, two phases are proposed to accomplishment of this task, the Road Detection and the Urban Scene Understanding. The first one is motivated by the autonomous navigation, i.e, the basic and principal capability of such system is to recognize precisely the road in front of the ego-car where any driving maneuver or vehicle control should be performed. After that, the scene understanding could be done to infer potential risks associated with obstacles, as previously mentioned.

Based on previous works, the proposed approach takes into account all these techniques to build the solution diagram presented in Figure 3.1. In this context, this chapter presents, in section 3.1, a detailed description of the Image Segmentation module used to obtain the super-pixel images. Furthermore, in-depth details the composed feature descriptor provided by Texton and Dispton maps, merging appearance, color, shape, context and depth information are explained in section 3.2 and section 3.3. Section 3.4 presents three different methods based on machine learning to perform the Road Detection and also the Urban Road Scene Understanding. Additionally, in section 4.5 is provided the quantitative and qualitative results about these proposed machine learning methods, followed by the conclusion presented in section 4.6.

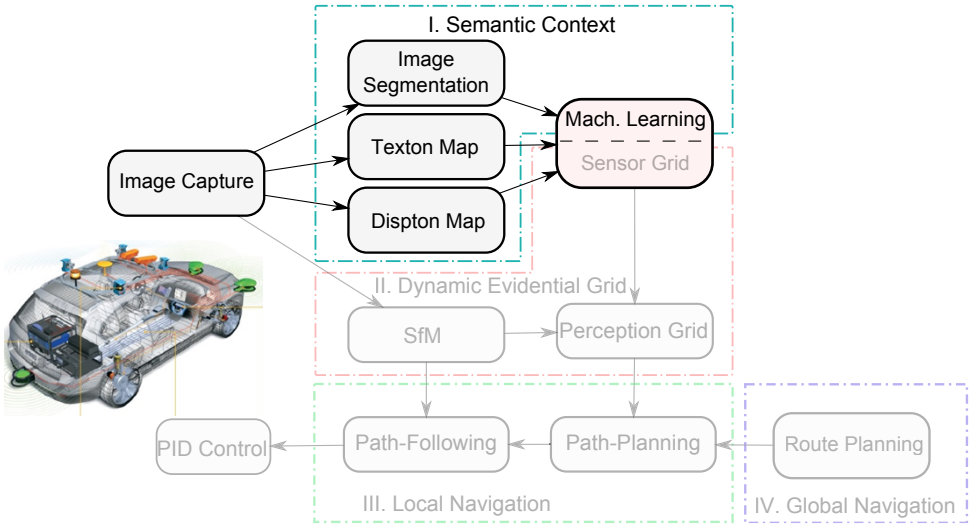


Figure 3.1 - The proposed solution to Urban Road Scene Understanding.

3.1 Super-Pixel Segmentation

Among some approaches used for generating super-pixel images, the most observed nowadays in literature is the mean-shift algorithm (COMANICIU; MEER, 2002; LADICKY et al., 2009) and SLIC approach (ACHANTA et al., 2012). In this thesis work however, it is explored another methodology based on Watershed Transform that since (BEUCHER; BILODEAU, 1994) has been applied to semantic context, specially for road detection. The combination of Watershed Transform with other filters has presented encouraging results in (VITOR et al., 2013). It is performed in a gradient image and its direct application produces a constraint of over-segmentation. In order to avoid this constraint and obtain a reasonable flexibility to determine the segmentation level, three pre-filters were added: the *Morphological Gradient Adjusted*, the *Morphological Reconstruction Area Closing* and the *Morphological Reconstruction Hmin*. The proposed image segmentation sequence can be seen in the diagram of Figure 3.2. These three filter algorithms take advantage in the domain of morphological image processing (SERRA, 1982), which has received considerable attention in the past few decades, especially after their theoretical foundations have been demonstrated. These theoretical foundations are known as connected operators (SERRA; SALEMBIER, 1993) and geodesic reconstruction process (VINCENT, 1993b). The interest above all is due to its fundamental property of simplifying an image without corrupting contour information (DARBON; AKGUL, 2005). Therefore, the next subsections explain these algorithms used to process the Super-Pixel Segmentation Module.

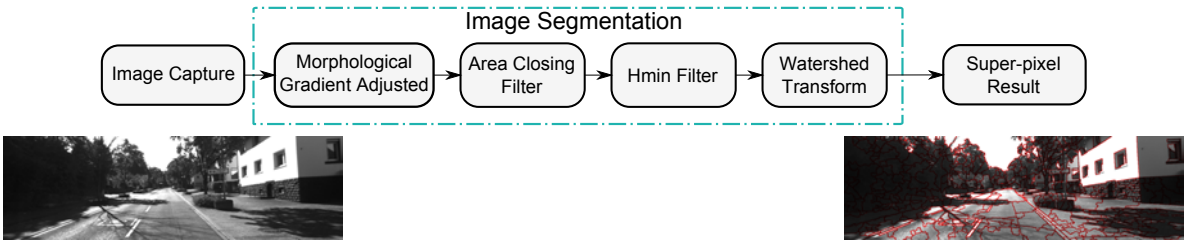


Figure 3.2 - The proposed solution to Super-pixel Segmentation module.

3.1.1 Morphological Gradient Adjusted

In general, gradient operators are used for image segmentation because they enhance intensity variations, also called as *edge detectors* (RIVEST et al., 1993). Between many gradient operators as Sobel, Prewitt and Roberts, the chosen one was the Morphological gradient due its good approximation of edges in urban environments, focus of this application.

This filter detects the intensity variations of pixel values in a given neighbourhood. It is obtained by the arithmetic difference between an extensive operator and an anti-extensive operator. In other words, it is given by the difference between dilation and erosion (DOUGHERTY; LOTUFO, 2003), where the classical formulation is defined as (3.1):

$$gradMorph(f) = (f \oplus g_e) - (f \ominus g_i), \tag{3.1}$$

where here f is the image function, g_e and g_i are structuring elements centered at the origin, and the operators $\oplus$ and $\ominus$ are respectively dilation and erosion.

In the work of (VITOR et al., 2013), was observed that the low-contrast of higher frequency in shadow areas of the image provides a not correct segmentation by merging different regions such as road, sidewalk and/or obstacles present in these areas. To avoid this drawback, it is performed an enhancement on the gradient where the shadow occurs, as illustrated in Figure 3.3, by applying a non-linear transformation defined in (3.2):

$$MG_{Adj} = \begin{cases} c\left[(f \oplus g_e) - (f \ominus g_i)\right]^{\gamma} & \text{, if } \{\forall x | f(x) < \rho\} \\ (f \oplus g_e) - (f \ominus g_i) & \text{, otherwise} \end{cases} \tag{3.2}$$

As can be seen in Equation (3.2), the non-linear transformation at low-contrast of high frequency is dependent on three parameters, the threshold ρ, the factor γ and the constant of normalization c. These parameters control the quality of the final enhancement gradient for the dark areas of the current image, at the same time keeping the maximum intensity variation of the image unchanged. The first parameter is the threshold intensity value ρ, that is responsible for creating a set of pixels with grayscale lower intensities than the chosen limit, which is correspondent to a dark area that may be withholding information. Thus, the estimation of those dark areas, the grayscale lower intensities can be obtained applying a threshold value heuristically fixed to $\rho = 50$, as can be seen in Figure 3.3(d). The second parameter is the factor γ, that is responsible for controling the intensity of the gradient on the dark areas. This enhancement is obtained by applying a function for all pixels belonging to the dark regions. The value of the factor γ can range from 1.0 to 0.01, where $\gamma = 1.0$ produces a linear transformation in the input data and values of γ tending to 0.01 will produce an abrupt exponential transformation in the gradient input value, like these ones depicted in Figure 3.3(e). Based on the output result observed, a satisfactory value was fixed to $\gamma = 0.45$ for all images. The third parameter is the constant of normalization c defined by the Equation (3.3) which is used to keep the same maximum intensity variations obtained from the gradient image calculated by (3.1) .

$$c = \frac{max(gradMorph(f))}{max(gradMorph(f))^{\gamma}} \tag{3.3}$$

The developed method can be summarized as follows. The grayscale current image, exemplified by Figure 3.3(a), is divided into two groups by a threshold value ρ separating the image in dark and light regions. After obtained the morphological gradient from the current image (Figure 3.3(b)), the pixels of the dark regions receive a non-linear transformation to highlight their values of high frequency. As a consequence of this process with the heuristic values set, it is possible to obtain the final enhancement gradient for the dark areas of the current image as shown in Figure 3.3(c).

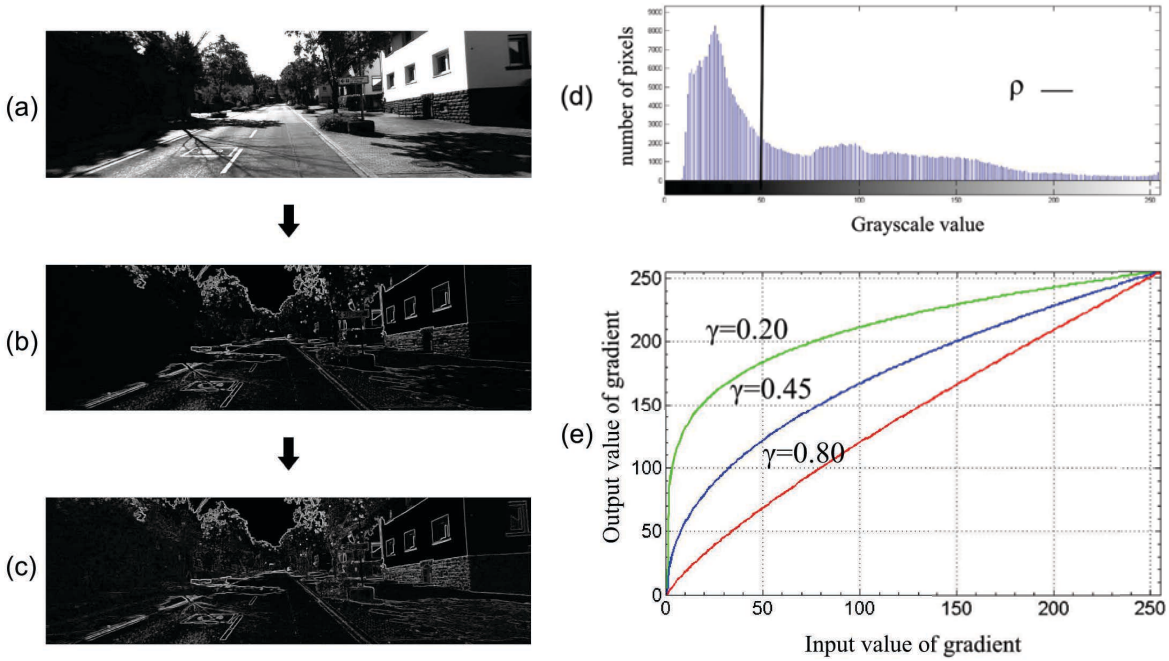


Figure 3.3 - Enhancing the contrast of higher frequency in shadow areas. (a) Original image, (b) gradient image, (c) the gradient image with shadow area enhanced, (d) the division of the two groups based on the image histogram and (e) presents different output results ranging the parameter γ.

3.1.2 Morphological Reconstruction Area Closing

Mathematical Morphology is a Set Theory applied to image processing. Its conception is linked to connexity and connected components. The connexity concept is defined by pixel adjacency and the connected component is defined by the union of these adjacent pixels (DOUGHERTY; LOTUFO, 2003). Since area closing is seen as the complement of area opening, by simplicity, here is presented the definitions for area opening (MEIJSTER; WILKINSON, 2002). The conception of this Area Opening filter is to remove from a binary image, the connected components of areas with number of pixels smaller than a parameter λ. Before demonstrating the grayscale area opening, some definitions should be done in the binary area opening. First, the connected opening $C_x(X)$ of a set $X \subseteq M$ at point $x \in X$ is the connected component of X containing x if $x \in X$ and 0 otherwise, and M being the binary image, $M \subset \mathbb{R}^2$. Thus, the binary area opening is defined over subsets of M (VINCENT, 1993a).

$$\xi^a_\lambda(X) = \{x \in X \mid Area(C_x(X)) \geq \lambda\} \tag{3.4}$$

The term $\xi^a_\lambda(X)$ denotes the morphological area opening with respect to the structure

element a and the parameter λ. The function $Area(.)$ calculates the number of elements in a connected component of $C_x(X)$. Its dual binary area closing is obtained as:

$$\phi^a_\lambda(X) = [\xi^a_\lambda(X^c)]^c \tag{3.5}$$

where X_c denotes the complement of X in M. Extending the filter for a mapping $f : M \to \overline{\mathbb{R}}$, in a grayscale image, the area opening $\xi^a_\lambda(f)$ is given by:

$$(\xi^a_\lambda(f))(x) = sup\{h \leq f(x) \mid x \in \xi^a_\lambda(T_h(f))\} \tag{3.6}$$

In Equation 3.6, $T_h(f)$ represents the threshold of f at value h:

$$T_h(f) = \{x \in M \mid f(x) \geq h\} \tag{3.7}$$

As mentioned before, the complement of the Equation 3.4 can be similarly extended to the conception of area closing to mappings from $M \to \overline{\mathbb{R}}$.

3.1.3 Morphological Reconstruction Hmin

Among several descriptions, the morphological grayscale reconstruction can be obtained by successive geodesic dilations. This principle employs two subsets of $\mathbb{R}^2$, called *mask image* and *marker image*. Both subsets must have the same size. Moreover, the *mask image* must have intensity values higher than or equal to those from *marker image* (VINCENT, 1993b).

In the geodesic dilation, the *marker image* is dilated following the dilation part demonstrated in Equation (3.1), where the resulting image is restricted to maintain under the *mask image*. Mathematically, defining *mask image* as I and *marker image* as M, observing that both have the same size and $M < I$, the geodesic dilation process considering only one amplitude, is defined as (3.8):

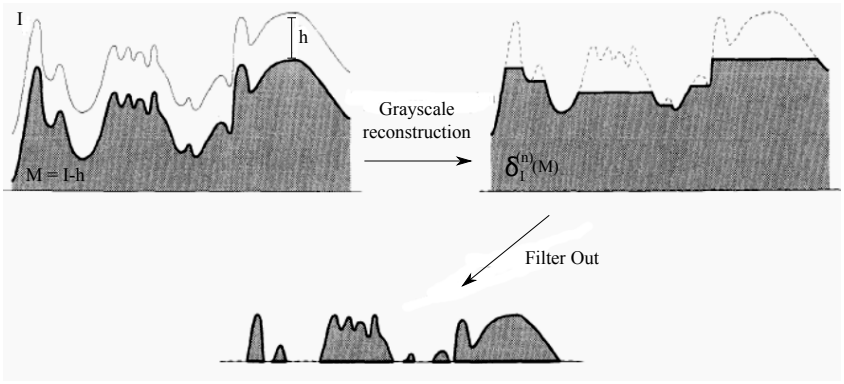


Figure 3.4 - Resulting morphological reconstruction based on height attribute. Adapted source: (VINCENT, 1993b)

$$\delta_I^{(1)}(M) = (M \oplus g_e) \wedge I \tag{3.8}$$

where $\wedge$ operator denotes the minimum of the intersection between the dilated M and I. Therefore, the geodesic dilation taking n amplitudes of M, considering the *mask image* I, is obtained by processing n successive geodesic dilations of amplitude 1 (3.9):

$$\delta_I^{(n)}(M) = \underbrace{\delta_I^{(1)}(M) \circ \delta_I^{(1)}(M) \circ \delta_I^{(1)}(M) \circ \ldots \circ \delta_I^{(1)}(M)}_{n \text{ times}} \tag{3.9}$$

The morphological reconstruction by geodesic dilation is defined in eq (3.9). According to (VINCENT, 1993b), the desired reconstruction of a image by its marker, is performed by n successive dilation until its stabilization. Figure 3.4 shows an example of the morphological reconstruction to 1D having a signal I and ist marker obtained by $M = I - h$.

In their book, (DOUGHERTY; LOTUFO, 2003) explain that the attribution of the *mask image* subtracted by a value h to *marker image* M, defines then the morphological reconstruction by a height attribute h, being called *H-maxima* or *Hmax* and defined by Equation (3.10).

$$Hmax_h^a(I) = I\Delta_a(I - h) \tag{3.10}$$

In this definition, Δ stands for morphological reconstruction with the structure element or connexity neighborhood a. As seen in Figure 3.4, this operator removes domes of an image based on a given attribute, in this case the height attribute. Figure 3.5 demonstrates the conception of possible filters using volume, area or height attribute. When the dome extraction with

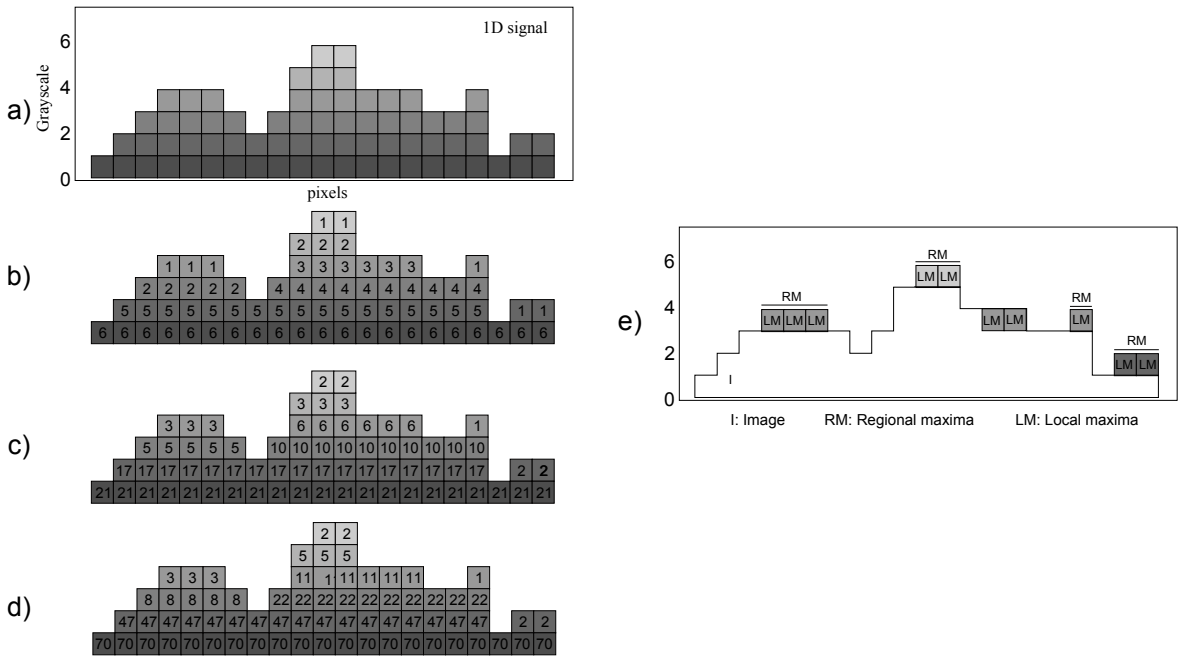


Figure 3.5 - Characteristics of connected component in a grayscale image: a) Input signal b) height attribute c) area attribute d) volume attribute e) regional maxima and local maxima of the grayscale image.

attribute h is set to 1, it is considered a regional maxima. A regional maxima RM of a grayscale image I is defined as a set of pixels with a given value which all pixels in its neighborhood must have a value strictly minor.

The regional maxima should not be mistaken with local maxima, as can be seen in Figure 3.5e. Every pixel belonging to a regional maxima necessarily is a local maxima, once there is no other neighbor pixel with higher value. The converse is not true, as can be seen in the same Figure. Such definitions are also valid to characterize regional minima.

Properly performing the reconstruction *H-maxima* or *Hmax*, it is possible to take the *H-minima* or *Hmin* from its complement. Then, by duality, the *Hmin* is defined by eq (3.11), where an example of this filter applied to 1D signal can be seen in Figure 3.6.

$$Hmin_h^a(I) = [I^c \Delta_a(I^c - h)]^c \tag{3.11}$$

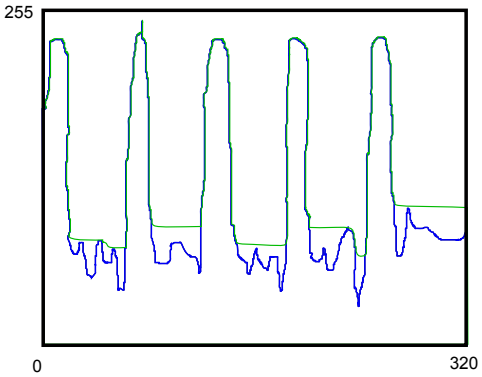


Figure 3.6 - Example of morphological reconstruction Hmin. The 1D input signal in blue and its result in green.

3.1.4 Super-Pixel by Watershed Transform

Studying different approaches, the Watershed Transform has different definitions at the literature, each one producing a different solution set, presented in (ROERDINK; MEIJSTER, 2001). The definitions are based on regional or global elements, such as influence zones and shortest-path forests with maximum or sum of weights of edges, or on local elements, such as the steepest descent paths. In this work, the local condition definition was used, called LC-WT (Local Condition Watershed Transform) (AUDIGIER; LOTUFO, 2010), with the purpose of simulating the behaviour of a drop of water on a surface. This definition seems to be the steepest descent paths, where the information from neighbours is used to create a path to the corresponding minimum, through an arrowing technique.

The arrowing is the algorithmic representation of the drop of water, where, for every pixel in the image (taking an image as a graph), an arrow is drawn from the current to the next one, which creates a path that ultimately leads to a regional minimum. The arrow points to the direction that a drop of water would flow, considering the image as a surface. For the LC-WT definition, every pixel that does not belong to a regional minima will have one and only one arrow. The union-find technique for Watershed transform is based on the algorithm for disjoint sets. Given that the regions of the output image form a partition and these are disjoint by definition, the union-find algorithm process paths to identify the roots - or representatives - for every pixel. At the end, all pixels that falls into the same minimum are labeled with its representative.

The goal of the filters presented in subsection 3.1.2 and subsection 3.1.3 is to control the

segmentation level of the Watershed Transform by acting on the regional minimum of GM_{Adj}. The procedure is similar to (VITOR et al., 2013), where the parameter λ of *AreaClose* determines the area of regional minimum to be cut out, and the parameter h of *Hmin* determines the height from regional minimum to be also cut out. The influence of these parameters exemplified by a synthetic image can be observed in Figure 3.7. Considering this image as a surface (Figure 3.7b), the Watershed Transform simulates the behavior of a drop of water slipping on it, which creates a path that lastly leads to a regional minimum. At the end, all path of pixels that fall into the same minimum are labeled with its representative (Figure 3.7c). When the regional minimum is removed, the drop of water's path is changed. The last two rows show the result of the synthetic image after applying the *Hmin* set to $h = 1$ and *AreaClose* set to $\lambda = 32$. A slight change on the signs could be observed when the regional minimum was removed, resulting in an variation in the image segmentation.

After demonstrating the conception and influence of these filters jointly with Watershed Transform, the final outcome applied in a sample image can be observed in Figure 3.8. Notice that the machine learning's procedure, to be described in subsection 3.4, is performed using each super-pixel of the resulting segmented image. Therefore, the analysis of these parameters is done to understand their sensibility in the final result of the segmentation.

3.2 Texton Maps

In the last decade, *Textons* have been proven effective for generic feature representation of object (LADICKY et al., 2009)(KRÄHENBÜHL; KOLTUN, 2012), where a class demands different appearances to have a compact representation maintaining their efficiency. Thereby, the methodology applied for this mapping module is to learn a dictionary of Textons using a textonization technique (SHOTTON et al., 2009), which allows to perform a dense-texture-based feature extraction for all pixels. The process of *textonization* generates the *Texton Map*, having the same size of the image. The Textons contained in the dictionary have their value associated with all pixels in this map. It can be seen as a pre-classification or a transformation from feature's space to the texton's space. Thus, this process is done by applying the K-Means algorithm on a given feature's space. Denoting a dictionary as D, each texton's element $x_j \in$

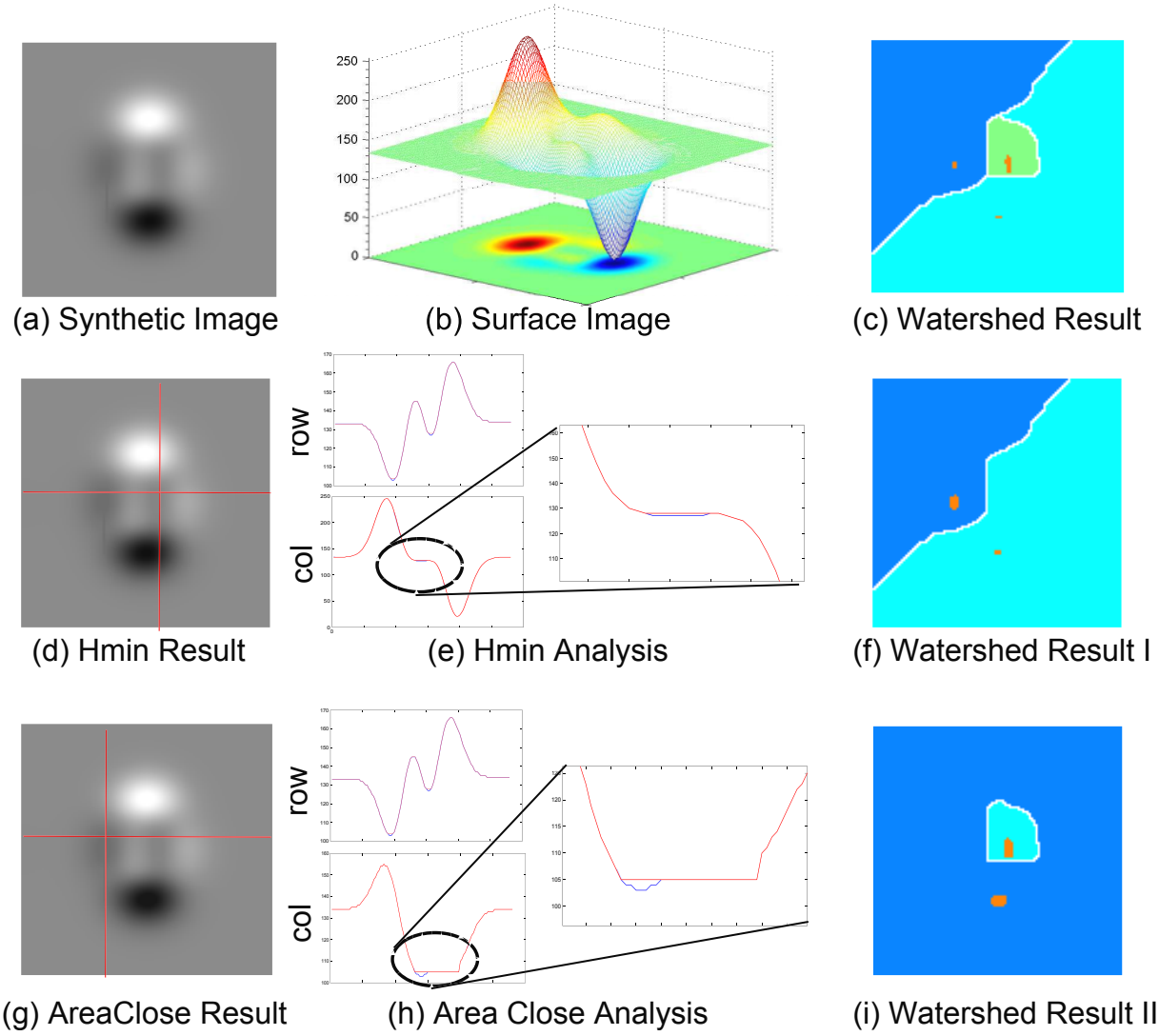


Figure 3.7 - Process of Image Segmentation. (a) Shows the synthetic image, (b) its surface perception and (c) the segmentation result. Figure (d and g) show the hmin result setted $h = 1$ and AreaClose setted $\lambda = 32$, (e and h) the signs alteration on regional minimum taking from a row and a col. Figure (f and i) show the final impact of these filters on segmentation, where the orange color represents the regional minimums

$D = \{x_1,x_2,...,x_K\}$ represents a cluster generated by the algorithm, employing the Euclidian-distance as a metric. Finally, it is obtained the *Texton Map* $T \in \mathbb{N}^2$ with the pixel i having value $x_j \in D$.

In this work, the *textonization* is executed in various feature's spaces. As it can be seen in (LADICKY et al., 2009), four methods were used to extract the feature descriptor: (i) a set of Filter Bank, (ii) the histograms of oriented gradient (HOG), (iii) the normalized pixel location (NpL) and (iv) CIELAB color.

Method (i) is defined by convolutions with gaussian kernel (G), first derivative of gaussian

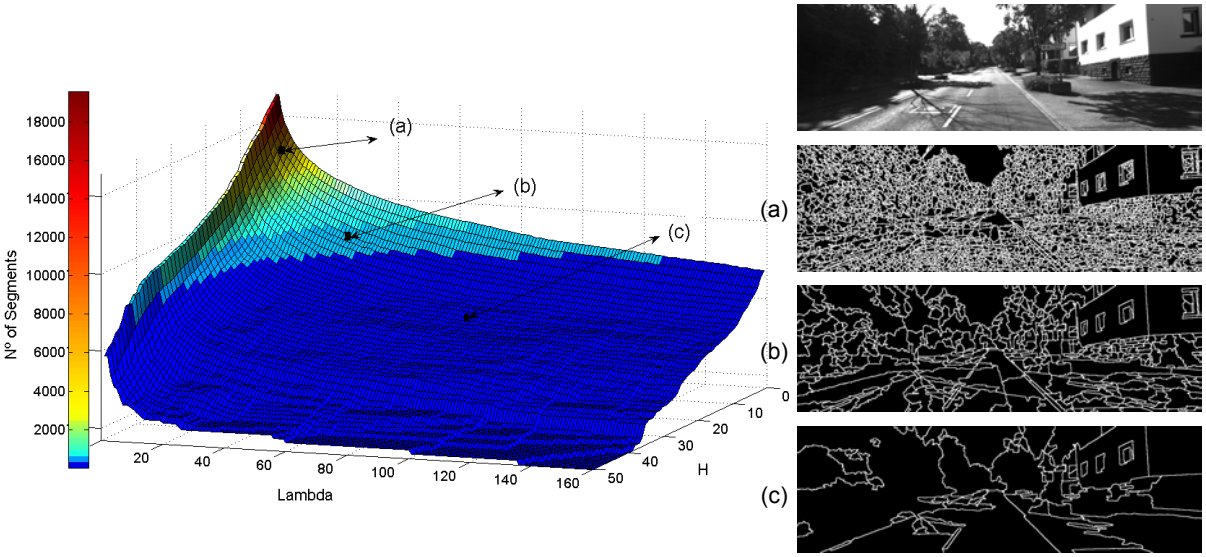


Figure 3.8 - Example of an influence surface for the parameters λ and h on the number of segments in an given image. In (a) $\lambda_1 = 5$, $h_1 = 2$ and 4090 segments; (b) $\lambda_2 = 30$, $h_2 = 5$ and 427 segments; (c) $\lambda_3 = 80$, $h_3 = 15$ and 73 segments.

kernel (DoG) and laplacian of gaussian kernel (LoG), totalizing a set of 17-dimensional features. The gaussian kernel is modeled according to Equation (3.12), having three different scales to standard deviations (σ). These different scales are convolving with each channel of the Lab Color space, extracting in this case, 9 features.

$$G(x,y;\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.12}$$

The second one, DoG kernel, is applied on lightness (L) channel of Lab Color space considering two different σ, to x-direction and also to y-direction. Its model is given by Equation (3.13).

$$DoG(x,y;\sigma) = \frac{xy}{\sigma^4}\left[\frac{1}{2\pi\sigma^2}\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.13}$$

To conclude the 17-dimensional feature extraction of filter bank, the LoG is modeled by Equation (3.14), following the same condition explained before, considering the (L) channel and applying four different σ to build this kernel. Globally, the composed filter banks are variations of these three kernels presented in Figure 3.9.

$$LoG(x,y;\sigma) = \frac{-1}{\pi\sigma^4}\left[-1\frac{x^2+y^2}{2\sigma^2}\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.14}$$

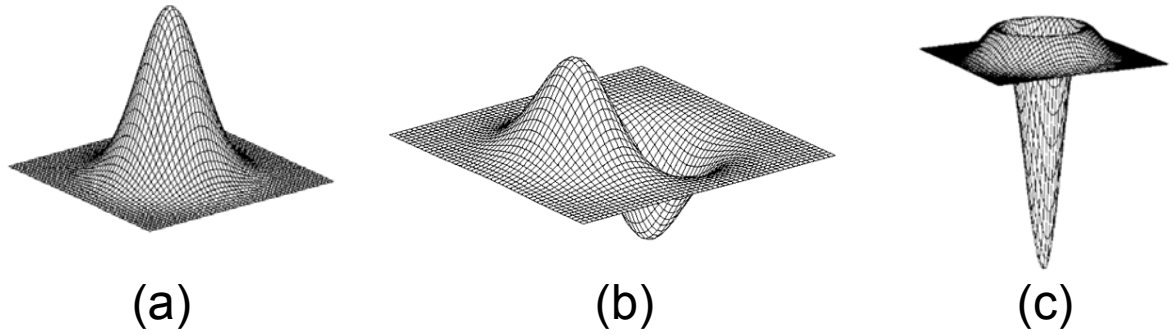


Figure 3.9 - Composing filter bank set that extracts 17-dimensional features of (a) Gaussian kernel (G), (b) Derivative of Gaussian kernel (DoG) and (c) Laplacian of Gaussian kernel (LoG).

In (ii), the method uses the algorithm proposed by (DALAL; TRIGGS, 2005). The algorithm extracts a feature vector containing 81 features for each pixel of whole image. Each element of this vector is a histogram bin of gradient orientations. The pseudo algorithm 1 summarizes the Histogram of Oriented Gradient feature descriptor process.

Algorithm 1 HOG algorithm:

1: Compute gradients for each pixel of an image.
2: Perform binning of gradients orientation (from 0 to 180 degrees, opposite directions count as the same).
3: Collect the histogram within a cell of pixels.
4: Weight the histogram by blocks and cells for local normalization of the contrasts.
5: Normalize the histogram by L2-norm or L1-norm.

In (iii), the feature descriptor is extracted taking the normalized location of pixels in a image. Considering an image resolution of (H) rows by (W) columns, the extracted feature is modeled dividing the x,y positions of the pixel by its resolution in each direction (eq (3.15)). In this case, the resulting feature descriptor defined by (NpL), has a vector containing 2 elements.

$$NpL(x,y;H,W) = \left[\frac{x}{W};\frac{y}{H}\right] \tag{3.15}$$

The last feature space mentioned (iv), takes direct access of the pixel values obtained by each channel of the CIELAB Color image.

After detailing the extraction of vector's features in their respective spaces, all feature's descriptor are whitened (to give zero mean and unit covariance). The textonization process is applied to learn the dictionaries of textons for each feature's descriptor. The learning procedure using K-means algorithm was configured assigning $D^b = 400$ clusters to Filter Bank, $D^g = 150$ clusters to HOG, $D^l = 144$ clusters to NpL and $D^c = 128$ clusters to CIELAb color. The output

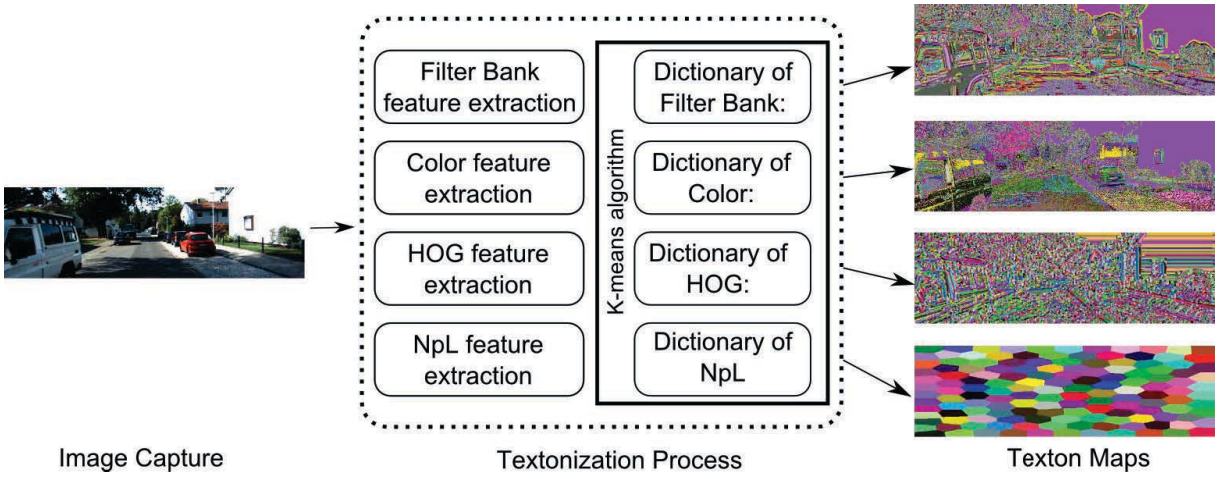


Figure 3.10 - The texton maps resulting from the textonization process using different features.

result for this module can be seen in Figure 3.10.

3.3 Dispton Maps

Based on the approach explained in section 3.2, *Texton Maps* are able to discriminate between class of similar textures. However this technique lacks spacial information. This section presents an approach to build two additional dictionaries over the 3D information from the Stereo Vision. This method is called *Dispton Map* and it aims at creating meaningful clusters based on the Disparity Map, denoted by I_Δ. Attempting to have the same functional advantage provided by the usage of U-Disparity and V-Disparity algorithm to filter and extract the navigable area and obstacles in literature (LABAYRADE et al., 2002)(SOQUET et al., 2007b), this work addresses another way to embed these information in a dictionary of Dispton, generating the *Dispton Maps* from I_Δ. This approach applies the Semi-Global Block Matching algorithm (SGBM) to compute the I_Δ, where its result can be seen in Figure 3.11.

Firstly, the technique consists in putting in evidence the peaks of the U-V Disparity maps, which concentrates the relevant information to start the process of *Disptonization*. Defining U-Disparity as $I_u\Delta$ and V-Disparity as $I_v\Delta$, they are obtained from a histogram for each column, $I_u\Delta = \{hist(I\Delta(:,u))|\forall u \in \{0..width-1\}\}$ and for each row, $I_v\Delta = \{hist(I\Delta(v,:))|\forall v \in \{0..height-1\}\}$. Like Watershed Transform, they can be seen as surfaces to apply the *Hmin* and filter the regional minimum (in this case considered as noise) of these surfaces. The result

Figure 3.11 - The disparity map result applying SGBM algorithm. (a) Left image (b) Right image (c) Disparity result applying SGBM algorithm.

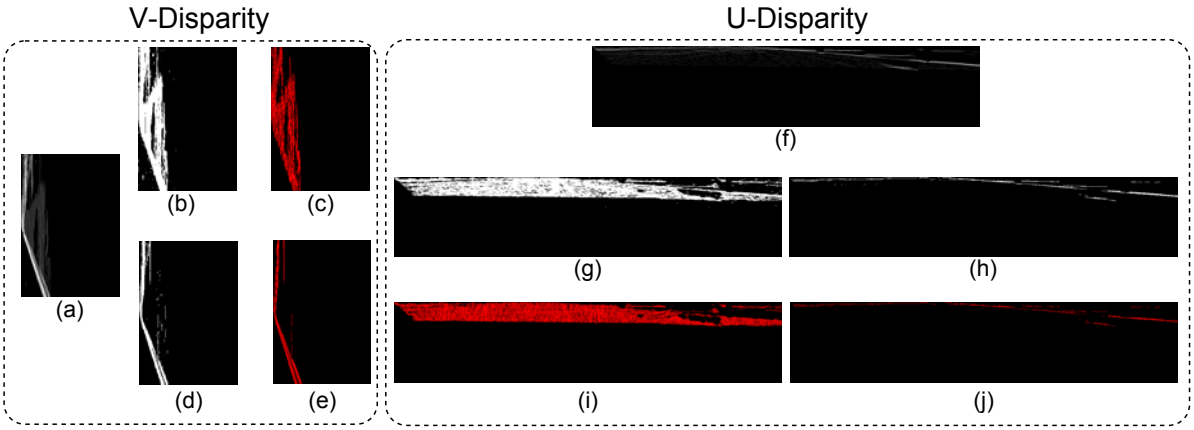


Figure 3.12 - The desired lines detection applying the Hmin filter. (a) and (f) show the V-Disparity and U-Disparity maps, (b) and (c) present the binarization and the detected line segments without applying the Hmin filter. (d) and (e) present the results using the Hmin filter. The same process was performed to U-Disparity map.

maps, denoted $I_u^h\Delta$ and $I_v^h\Delta$ are then binarized. Additionally, the Hough Transform is executed to detect line segments, characterized by l^u and l^v. The result of this procedure is depicted in Figure 3.12. The V-Disparity map and the U-Disparity map in (a) and (f) are obtained from the disparity map showed in Figure 3.12. Where (b) and (c) present the binarization result and the detected line segments without apply the *Hmin*. (d) and (e) present the binarization result and the detected line segments applying the *Hmin* filter. The same process can be seen at U-Disparity map, where (g) and (i) without applying the *Hmin* filter and, (h) and (j) with applying the *Hmin* filter. This technique notably improves the detection of the desired lines.

After that, to build the dictionary of *U-Dispton* (D^u) is applied the clusterization where the points from each line segment l_j^u , supplies seeds to perform the clusterization of dispton's element $j \in \{1..NumberOfLines\}$. In Equation (3.16), the clusterization process of a line segment l_j^u, denoted by $\Lambda(I_u^h\Delta)(l_j^u)$, is given by:

$$\Lambda(I_{\eta}^{h}\Delta)(l_{j}^{\eta}) = \begin{cases} C_{j}^{\eta} & \text{, if } \{x\} \in N_x(I_{\eta}^{h}\Delta) \neq 0 \\ 0 & \text{, otherwise} \end{cases} \tag{3.16}$$

where the cluster C_j^{η} is defined in Equation (3.17). Note that the variable (η) can be either u or v.

$$C_{j}^{\eta} = \{j | x_1 W + x_2 \in \{l_j^{\eta}\} \subset N_x(I_{\eta}^{h}\Delta)\} \tag{3.17}$$

Here, W is the number of image columns, and x_1 and x_2 are the row and column coordinates respectively, and the term $N_x(.)$ represents the neighbors of the x element. There are two additional clusters given by $C_1^u = \{I_u\Delta(v, :) | v \in \{0,1,..,\tau\}\}$ and $C_2^u = (I_u\Delta \cap C_1^u)'$, where ($'$) represents its complement. Therefore, the D^u is given by the union of all clusters (3.18). Each color of Figure 3.13a represents a defined cluster, which the dictionary was built.

$$D^u = \{ \bigcup_{\forall j \in l^u} \Lambda(I_u^h\Delta)(l_j^u) \cup C_1^u \cup C_2^u \} \tag{3.18}$$

Finally, with the *U-Dispton* dictionary, it is possible to obtain the *U-Dispton map* (Dm_u) by applying the following Equation (3.19), where this map is presented in Figure 3.13b.

$$Dm_{\eta} = \{D^{\eta}(I\Delta(x)) | \forall x; I\Delta(x) \neq 0\} \tag{3.19}$$

In order to build the *V-Dispton* dictionary (D^v), the clusters are created separately. Following Hu and Uchimura (2005), a road is modeled as a plane so that it can be represented by straight slope line segments in the V-disparity map. In this sense, the goal of the first cluster is to provide a curve fitting of these line segments to represent the surface of navigable area. From l^v set, a subset l^{vs} is obtained by filtering out the line segments with a given vertical orientation (3.20):

$$l^{vs} = \{l_i^v | \forall l_i^v; Ang(l_i^v) < 90° - \psi\} \tag{3.20}$$

The function $Ang(.)$ represents the angle of inclination with the reference defined on the

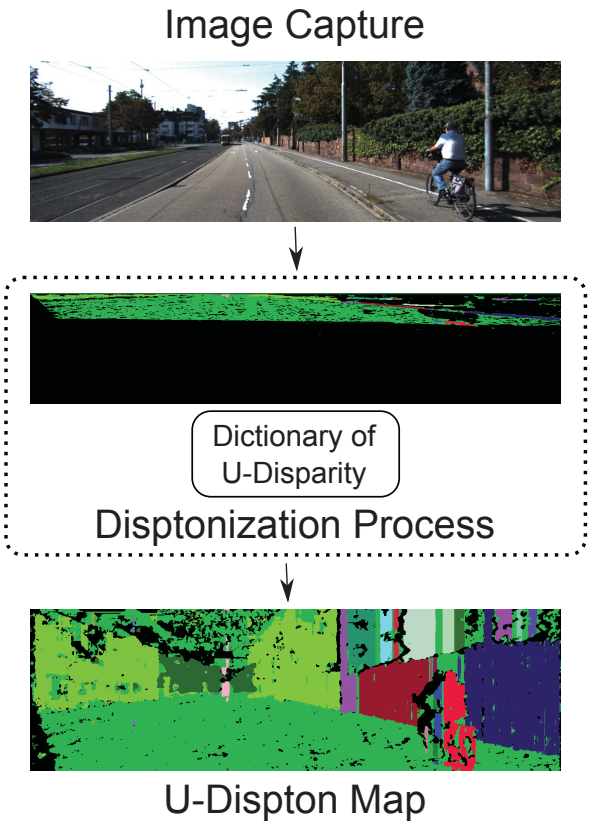


Figure 3.13 - Disptonization process to U-Dispton map. a) the U learnt dictionary and b) the U-Dispton map result.

bottom-left image and ψ an input parameter. Thus, the road surface can be formed by a succession of plane's parts, being projected as a piecewise linear curve (HAUTIERE et al., 2006). In order to connect the line segments that represent the surface, the algorithm sorts the l^{vs} set based on the distance from the line segment to the reference. Starting from the first line segment l_0^{vs} to the last one l_n^{vs}, the constraints that define whether two line segments can be connected, are given by Equation (3.21), where l^{vc} is the set of connected line segments,

$$l^{vc} = min(dist(l_i^{vs}, l_j^{vs})) \begin{cases} \forall l_j^{vs} \in \{l^{vs} > l_i^{vs}\} \text{ and} \\ \text{if } l_j^{vs} \subset AreaSupport(l_i^{s-}, l_i^{s+}) \end{cases} \tag{3.21}$$

and the function $dist(.)$ between two line segments is calculated considering the Euclidian distance from the nearest points of the current line segments, limited by a given maximum distance ε between them. The function $AreaSupport(.)$ delimits the search area by two line segments as given by Equation (3.22).

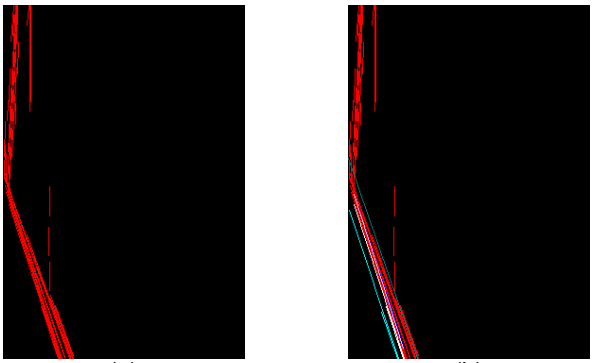


Figure 3.14 - Merging lines approach. (a) The detected line segments by Hough transform and (b) the linked lines in white and the support lines in blue

$$AreaSupport(l_{s1},l_{s2}) = \begin{cases} 1 & \text{, if } right(l,l_{s1}) \text{ and } left(l,l_{s2}) \\ 0 & \text{, otherwise} \end{cases} \tag{3.22}$$

This area is defined by a translation from l_i^{vs} given by the σ parameter, then $l_i^{s-} = l_i^{vs} - \sigma$ and $l_i^{s+} = l_i^{vs} + \sigma$. The other two functions in this Equation return true for the case when the line segment is on right and left of the reference lines. Figure 3.14 shows this process. As a result, the cluster C_1^v is obtained applying the Equation (3.16) on the l^{vc} set ($\Lambda(I_v^h\Delta)(l_j^{vc})$) with one more constraint, where all pixels cannot cross out the line (l_{lim1}) formed by the first and last points of l^{vc} (added a small shift constraint). In addiction, the second cluster is generated taking those pixels which cross out the first one and is restricted to another shifted line $l_{lim2} = l_{lim1} + \sigma_2$, resulting the Equation (3.23):

$$C_2^v = \{\Lambda(I_v^h\Delta)(l_j^{vc}) | \text{ if } AreaSupport(l_{lim1},l_{lim2})\} \tag{3.23}$$

To finish the *Disptonization*, the last two clusters are generated by $C_3^v = \{I_v\Delta(:,u) | \forall u \in \{1,..,\tau\}\}$, where τ defines the max disparity to be considered as background or infinity, and $C_4^v = \{(I_v\Delta \cap (C_1^v \cup C_2^v \cup C_3^v))'\}$. The *V-Dispton* dictionary (D^v) is defined as:

$$D^v = \{C_1^v \cup C_2^v \cup C_3^v \cup C_4^v\} \tag{3.24}$$

The generation of *V-Diston map* is obtained by the Equation (3.19) and its map is presented in Figure 3.15. Algorithm 2 summarizes the *Disptonization* process.

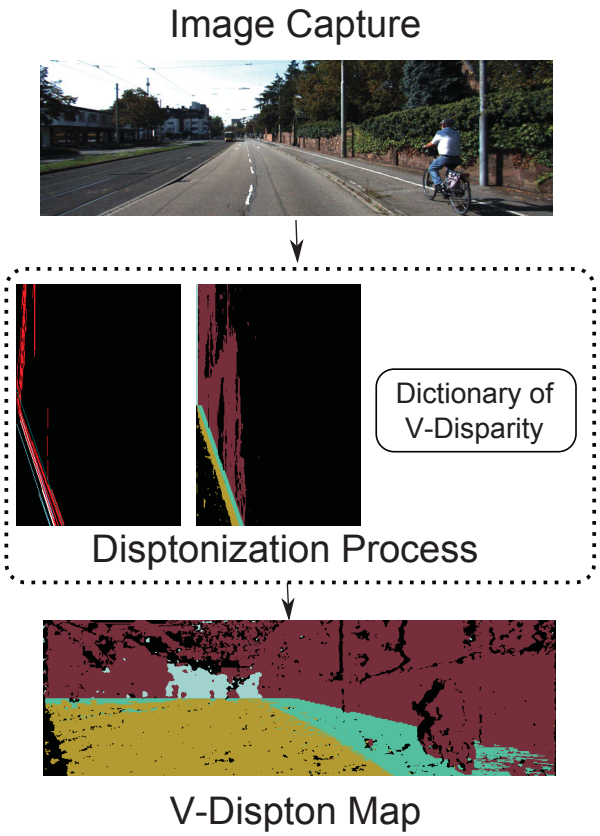


Figure 3.15 - Disptonization process to V-Dispton map. a) the V learnt dictionary and b) the V-Dispton map result.

Algorithm 2 Disptonization algorithm:

1: Process $I_u\Delta$ and $I_v\Delta$ from $I\Delta$;
2: Apply the *Hmin* filter on $I_u\Delta$ and $I_v\Delta$;
3: Binarize and obtain the line segments by *Hough Transf.* for l^u and l^v;
4: Determine the *U-Dispton* dictionary D^u by Equation (3.18):
5: - Apply the clusterization on l^u, Equation (3.16);
6: Determine the *V-Dispton* dictionary D^v by Equation (3.24):
7: - Filter out the vertical lines to take l^{vs}, Equation (3.20);
8: - Find out the connected lines l^{vc}, Equation (3.21);
9: - Define the clusterization to C_1^v,C_2^v,C_3^v,C_4^v;
10: Generate the *UV-Dispton map* by Equation (3.19)

Note that the *V-Dispton map* has 4-dimensional clusters and the *U-Dispton map* has N-dimensional clusters. Intuitively, they aim at storing important information such as navigable area, sidewalk, obstacles and background. The N-dimensional structure from *U-Dispton map* dynamically retrieve the representation of all possible different obstacles in the scene. With the *Texton maps* and *Dispton maps*, the next section explains how they are combined with the super-pixel image to perform the classification.

3.4 Machine Learning for Urban Road Scene Understanding

Based on super-pixel image, Texton maps and Dispton maps previously described in sections 3.1, 3.2 and 3.3, this section introduces three methods that were implemented in this PhD thesis. Basically, these algorithms can be seen as an evolution conditioned on experiences acquired with several research and learning. The first method implemented, the Artificial Neural Network, is presented in subsection 3.4.1, followed by the Histogram-Based Joint Boosting, that is presented in subsection 3.4.2. The last proposed algorithm, the Probabilistic Joint Boosting, is shown in subsection 3.4.3.

3.4.1 Artificial Neural Network

Artificial Neural Network (ANN) has been utilized in several applications as a good tool for data classification. Its wide utilization is based in three fundamentals properties: adaptability, ability to learn by examples and ability of generalization (SHINZATO; WOLF, 2011). This section describes the implementation of a Multilayer Perceptron (MLP) to perform a non-linear input-output mapping (HAYKIN, 1998) that models the road pattern. The algorithm executes the 2D image segmentation, presented in section 3.1, and a specific approach using 3D image processing to compose a feature's descriptor (different from those explained in sections 3.2 and 3.3). The specific approach will be briefly explained in this section and was not described before because it was used only for this method.

The 3D image processing algorithm, specific for the ANN implementation, uses the rectified pair of stereo images to allow an easy pre-classification of the image data into drivable and non-drivable areas. The approach takes advantage of the Epipolar geometry (FAUGERAS, 1993). First, the disparity map ($I_{\triangle}$) of the stereo pair is built using the Sum of Absolute Differences (SAD) correlation algorithm. As described in (LABAYRADE et al., 2002), the world environment can be approximated by horizontal and vertical planes related to the camera. Based

on this assumption, each disparity map ($I_{\triangle}$) is then projected in the v-disparity map ($I_{v\triangle}$) (SOQUET et al., 2007a), where each line on $I_{v\triangle}$ represent these planes. The next step is the execution of the Hough Transform with an auxiliary algorithm to detect the lines associated with the drivable area. In $I_{v\triangle}$, the continuous slope lines represent the drivable area, while the non-drivable areas are approximately represented by vertical lines. As a result, the lines from Hough transform are filtered out considering a given parameter of inclination performed by the auxiliary algorithm. Finally, the detected lines are reprojected into the image, as can be seen in Figure 3.16b, which generates an estimation of drivable and non-drivable area. Just for information, this technique was the basis for arising the development of the Dispton concept, inspired by the Texton concept. A detailed description of the 3D image processing can be found in (VITOR et al., 2013).

Over the results extracted from the 2D and 3D modules, other 2D and 3D features were extracted to build the final feature's descriptor of each super-pixel segment, which is the input data of the ANN. The first set of features consider the percentage of the drivable and non-drivable areas. It is based on the intersection of the resulting pre-classification of 3D image processing with the respective super-pixel. The second set of features is based on statistical measures, as proposed by (SHINZATO; WOLF, 2011), like mean, probability, entropy and variance. These values were calculated by the RGB[1] and HSV[2] color space values of each super-pixel segment.

The network training is based on the Backpropagation technique and its structure was projected with three layers, namely input, output and hidden layers. The size of the input layer corresponds to the number of features extracted which was defined in the experiments with six neurons in its final version. The hidden layer was defined with 30 neurons, where each neuron uses the sigmoid activation function. Finally, the output layer has two neurons to classify the feature's descriptor as road surface or non-road surface, which response vary from 0.0 to 1.0. The setup of the output layer containing two neurons is chosen to model the uncertainty for each class separately. The methodology for training, validation and testing for this ANN uses the cross validation method and its configuration is detailed in subsection 3.5.1. Figure 3.16 presents all processes performed by this first approach.

[1] Abbreviation for red, green and blue color space.
[2] Abbreviation for hue, saturation and value color space.

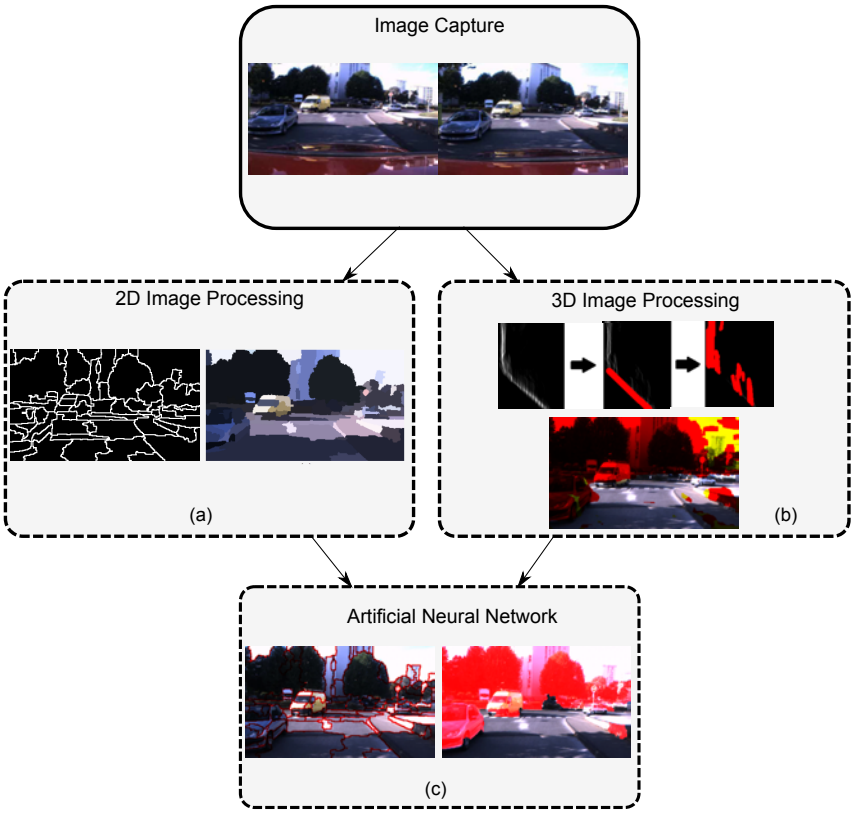


Figure 3.16 - The process overview for the ANN approach. (a) The 2D segmentation process to get the super-pixel. (b) The 3D process to obtain an estimation of the drivable and non-divable areas. (c) The final result applying the ANN, which the left image represents the lines of the super-pixel and the right image has the classification of non-road surface (red) and road surface (original color).

3.4.2 Histogram-Based Joint Boosting

This section presents the second method to represent and compute the classification of semantic context, including in this case not only the road class, but also more elements present in the scene such as vehicle, sidewalk, vegetation, building, etc. Therefore, here will be described the conception applied for a set of classes, where the road recognition can be thought as a set of classes containing only two classes, road and non-road.

The Histogram-based Joint Boosting, or simply HistonBoost, is executed using an adapted version of the Joint Boosting algorithm (TORRALBA et al., 2007). In fact, the algorithm is derived from *TextonBoost* (SHOTTON et al., 2007), which iteratively builds a *strong classifier* as a sum of *week classifiers*, simultaneously selecting discriminative features. The difference in this implementation is the shape filters utilized and consequently the representation of weak classifiers. Thus, the novelty proposed in this thesis is to build a multi normalized-histogram of the

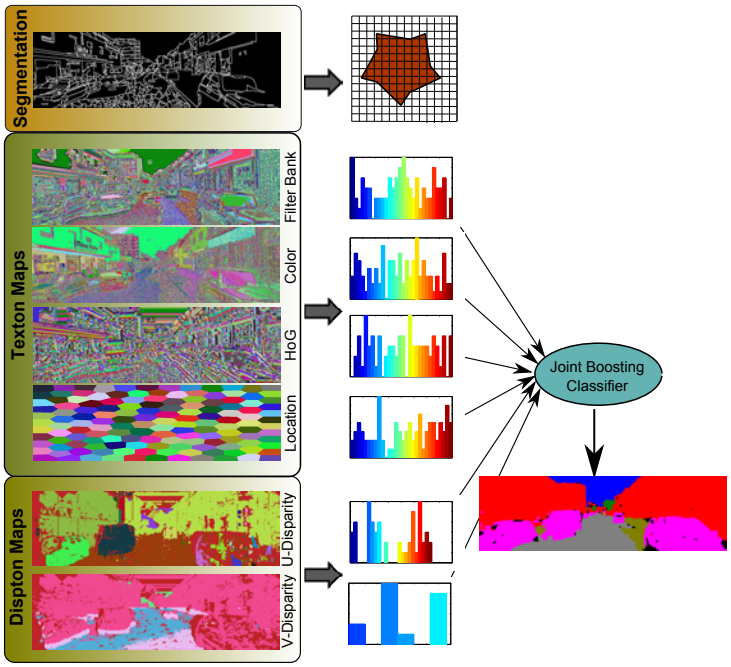


Figure 3.17 - The process of Road Recognition. For each sp_i is constructed the multi-normalized histogram from the Texton maps and the Dispton maps, which are classified with Joint Boosting algorithm.

clusters features from the specific super-pixel segment as shown in section 3.1. The representative *Textons* and *Disptons* histograms for all segments are generated in which each histogram bin represents the cumulative number of a given cluster on their respective regions, as can be seen in Figure 3.17. According to (LADICKY et al., 2009), the distributions of dense feature responses are more discriminative than any feature alone, representing many classification and recognition problems. Formally, the super-pixel segment $\{sp_i|i \in \{1,..,NumOfSegments\}\}$ has 6-tuples containing the histogram $\{g_f|f \in \{1,..,NumOfHistFeature\}\}$.

In this method, the *weak classifiers* are modeled as comparisons of this multi normalized-histogram to a decision stump based on a threshold, where each *weak classifier* is shared between a set of classes, allowing a single normalized-histogram response to help classify several classes at once. In case of only two classes are being considered, it is highlighted that this sharing does not have effect in the classification. Therefore, the weak classifiers are defined by wc containing 3-tuples [type feature f, random normalized histogram g, threshold θ], where the comparison response is given by the χ^2 metric (Equation (3.25)) to express how well the data feature and weak classifier match:

$$d(wc,sp_i) = \sum_I \frac{[wc.g(I) + sp.g_f(I)]^2}{wc.g(I) + sp.g_f(I)} \tag{3.25}$$

Thereby, the Joint Boosting algorithm is an additive model of the form $H(c_l) = \sum_{m=1}^{M} h_m(c_l)$, that sum the classification confidence of M joint weak classifiers. In this case, $H(c_l)$ represents the output value to the 'strong' learned classifier, given the class c_l, and each weak-learner is modeled as a decision stump of the form:

$$h(c_l) = \begin{cases} a\delta(d(wc,sp_i) > \theta) + b & \text{, if } \{c_l \in L\} \\ \kappa_{c_l} & \text{, otherwise} \end{cases} \tag{3.26}$$

Where $\delta(.)$ is a 0-1 indicator function. The share is given by those classes ($c_l \in L$), that the weak learner gives $h(c_l) \in \{a + b, b\}$ depending on the comparison of $d(wc,sp_i)$ to a threshold θ. The constant k_{c_l} ensures that asymmetrical sets of positive and negative training examples do not adversely affect the learning procedure, considering those classes that do not share the normalized-histogram feature ($c_l \notin L$). Thus, the resulting classification output is defined by the probability conversion given by (3.27):

$$P = \frac{1}{Z} \exp^{-H(c_l)} \tag{3.27}$$

In Equation 3.27, Z represents the normalization factor into the classes $c_l \in L$. Once explained this method, the learning procedure will be briefly described here, taking into account that an excellent detailed treatment of the learning algorithm is given in (TORRALBA et al., 2007) and has an optimized version with some improvements explained in (SHOTTON et al., 2007). Each training example sp_i (a super-pixel in a training image) is paired with a target value $z_i^{c_l} \in \{-1, +1\}$ (+1 if the majority of pixels belonging to sp_i vote to ground truth class c_l, -1 otherwise) and assigned a weight $w_i^{c_l}$ specifying its classification accuracy for class c_l after $m-1$ rounds of boosting. Round m chooses a new weak learner by minimizing an error function J_{wse} incorporating the weights eq (3.28):

$$J_{wse} = \sum_{l} \sum_{i} w_i^{c_l} (z_i^{c_l} - h_i^m(c_l)) \tag{3.28}$$

The training examples are then re-weighted (3.29)

$$w_i^{c_l} = w_i^{c_l} e^{-z_i^{c_l} h_i^m(c_l)} \tag{3.29}$$

to reflect the new classification accuracy and maintain the invariant $w_i^{c_l} = e^{-z_i^{c_l} H_i(c_l)}$. This procedure emphasizes poorly classified examples in subsequent rounds, and ensures that over many rounds, the classification for each training example approaches the target value.

Minimizing the error function J_{wse} unfortunately requires an expensive brute-force search over the possible weak learners h_i^m to find the optimal combination of the sharing set L, multi normalized-histogram features, and thresholds θ. However, given these parameters, a closed form solution does exist for a, b and $\{k_{c_l}\}_{c_l \notin L}$:

$$b = \frac{\sum_{c_l \in L} \sum_i w_i^{c_l} z_i^{c_l} [d(wc,sp_i) \leq \theta]}{\sum_{c_l \in L} \sum_i w_i^{c_l} [d(wc,sp_i) \leq \theta]} \tag{3.30}$$

$$a + b = \frac{\sum_{c_l \in L} \sum_i w_i^{c_l} z_i^{c_l} [d(wc,sp_i) > \theta]}{\sum_{c_l \in L} \sum_i w_i^{c_l} [d(wc,sp_i) > \theta]} \tag{3.31}$$

$$k_{c_l} = \frac{\sum_i w_i^{c_l} z_i^{c_l}}{\sum_i w_i^{c_l}} \tag{3.32}$$

3.4.3 Probabilistic Joint Boosting

This section presents the third approach to compute the classification of the semantic context. As shown in previous section 3.4.2, this third method is a variant of Histogram-based Joint Boosting, defined by Probabilistic Joint Boosting or its short name, ProbBoost. The representation of weak classifiers using another shape filter has been improved. Thus, the novelty in this case, is to build a set of probability distribution of the Texton and Dispton maps from the decomposition of the scene into a number of semantically consistent regions, supplied by the super-pixel segmentation result shown in section 3.1, to model the *weak classifier*.

The process could be formally explained taking into account the maps $\{M^f : f \in \{F\}\}$ where $F = \{b,c,g,l,v,u\}$ is the set of Textons and Disptons. Each element i in the map $M^f \in \mathbb{N}^2$ belongs to exactly one region, identified by its region-correspondence variable $S_r \in \{1,...,NumSegments\}$. The r-th region is then simply the set of elements i_r whose region-correspondence variable equals r, i.e., $i_r = \{i : M_i^f = r\}$. We use $X_i^f = \{X_1^f, X_2^f, ..., X_N^f\}$ to denote the set of random variables corresponding to the f-th value of i-th element into M^f. Any possible assignment to the random variables $X_i^f = x_j^f$ takes values from $j \in D^f$, which D^f is

defined by the constructed dictionary for each $f \in F$ generated in the sections 3.2 and 3.3.

The probability of the X_i^f if given by $P(X_i^f = x_j^f)$, and the associated set of probability distribution under the S_r is denoted by $P(X_r)$, as can be seen in the Equation (3.33):

$$P(X_r) = \{\bigcup_{f \in F} \left\{ \frac{1}{Z} \sum_{i_r} P(x_j^f) \right\} | \forall j \in D^f\} \tag{3.33}$$

In Equation (3.33), Z is a normalization factor for each probability distribution set. Using the probability representation of Textons and Disptons, the *weak classifiers* are modeled as comparisons of this probability distribution response to a decision stump based on a threshold, where each *weak classifier* is *shared* between a set of classes, allowing a single probability to help classify several classes at once. As mentioned before, in case of only two classes are being considered, it is highlighted that this sharing does not have effect in the classification. Accordingly, the weak classifiers are defined by wc containing 2-tuples $[x_{rand}, P(x_{random})]$, where the first component represents a random possible assignment $\{x_{rand} : x_j^f \in D^f\}$ and its value of probability randomly defined. To express how well the probability distribution of $P(X_r)$ at a given x_j^f matches the weak classifier, a comparison response is given by Equation (3.34):

$$d(wc,S_r) = 1 - \sqrt{[P(x_{rand}) - P(X_r = x_j^f)]^2} \tag{3.34}$$

As explained previously, the Joint Boosting algorithm is an additive model of the form $H(c_l) = \sum_{m=1}^{M} h_m(c_l)$, that sum the classification confidence of M joint weak classifiers. $H(c_l)$ represents the output value to the 'strong' learned classifier, given the class c_l, and each weak-learner is modeled as a decision stump of the form:

$$h(c_l) = \begin{cases} a\delta(d(wc,S_r) > \theta) + b & \text{, if } \{c_l \in L\} \\ \kappa_{c_l} & \text{, otherwise} \end{cases} \tag{3.35}$$

where $\delta(.)$ is a 0-1 indicator function. The share is given by those classes ($c_l \in L$), that the weak learner gives $h(c_l) \in \{a + b, b\}$ depending on the comparison of $d(wc,S_r)$ to a threshold θ. The constant k_{c_l} ensures that asymmetrical sets of positive and negative training examples do not adversely affect the learning procedure, considering those classes that do not share the

probability distribution feature ($c_l \notin L$). As before, the resulting classification output is defined by the probability conversion given by Equation (3.27). The learning procedure is identical as presented in the previous section (3.4.2), being in this case omitted here.

3.5 Experimental Results

Experiments were carried out in real conditions. A fair performance comparison among those proposed methods was done, using the common Kitti benchmark[3] (GEIGER et al., 2013). Aimed in applications to Advanced Driver Assistance Systems (ADAS) and any driving maneuver or vehicle control to autonomous navigation, the experiments were conducted in two phases. The first phase, shown in subsection 3.5.1, performs a meaningful evaluation centered at road recognition. The second one, presented in subsection 3.5.2, extends this pattern recognition to multi-class classification in urban scenarios.

The validation platform has been implemented in C++ and the experiments were executed in one equipment with an Intel Xeon E5-1650 processor with 3.20Ghz and with 16Gb DDR3, running the version 7 of the Windows OS.

In order to judge the quality of the proposed algorithms for use in automotive applications, all evaluations are based on the perspective space and also in the metric space, called Bird's Eye View (BEV). An exemplary detailed treatment of performance measure and evaluation is given in (FRITSCH et al., 2013), but here, it will be briefly described for completeness. The pixel-based analysis is evaluated employing the *F-measure* derived from the *Precision* and *Recall* values (Equation 3.36-3.37). It makes use of the harmonic mean (F1-measure, $\beta = 1$), while an unbalanced *F-measure* using a different weighting of *Precision* and *Recall* could also be applied. In addition, accuracy is also evaluated.

[3] <http://www.cvlibs.net/datasets/kitti/> accessed on: 05/07/2014

$$Precision = \frac{TP}{TP + FP} \tag{3.36}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.37}$$

$$F - measure = (1 + \beta^2)\frac{PrecisionRecall}{\beta^2 Precision + Recall} \tag{3.38}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.39}$$

For methods that provide outputs as confidence maps (in contrast to binary road classification), the classification threshold τ is chosen to maximize the F-measure, yielding F_{max}:

$$F_{max} = \underset{\tau}{\text{argmax}} \quad \text{F-measure} \tag{3.40}$$

According to (FRITSCH et al., 2013), in order to provide insights into the performance over the full recall range, the average precision (AP), based on the work of (EVERINGHAM et al., 2010), is computed for different recall values r:

$$AP = \frac{1}{11} \sum_{r \in 0, 0.1, \cdots, 1} \underset{\widetilde{r}:\widetilde{r}>r}{max} \quad \text{Precision}(\widetilde{r}) \tag{3.41}$$

Considering both measures it provides insights into an algorithm's optimal (F_{max}) and average (AP) performance.

3.5.1 Road Recognition

In this section the results are presented, using a specific dataset of the Kitti Benchmark, called Urban Kitti-Road.In brief, this dataset consists of $\simeq 600$ frames (375x1242 px) recorded in five different days and containing relatively low traffic density (FRITSCH et al., 2013), representing a typical road scene in inner-city. Data is categorized in three sets having each one a subset of training images and a subset of test images exclusively for evaluation performance. Table 3.1 shows the dataset statistics of Urban Kitti-Road benchmark showing the number of frames used in each case and Figure 3.18 illustrates some examples images for each category.

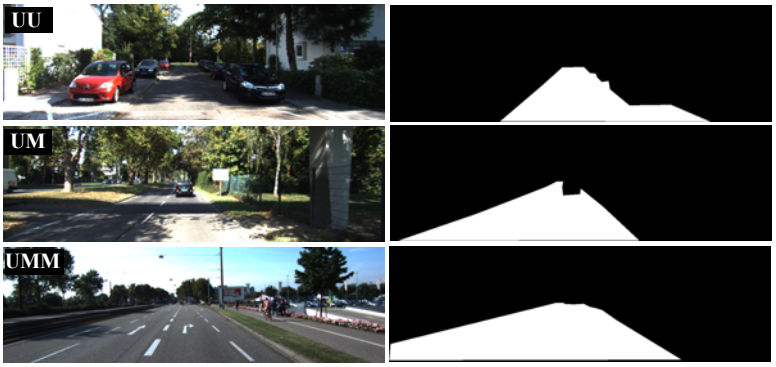


Figure 3.18 - Example test images from the different categories of the Urban Kitti-Road dataset. Note the high variability of the dataset. Source: (FRITSCH et al., 2013)

Emphasizing that the evaluation process is done on the metric space in order to capture the fact that vehicle control happens in the 2D environment. Further, the evaluation in perspective space is biased by the fact that the pixel's value in near range is more homogenous and covers a larger area of the evaluated perspective pixels.

Table 3.1 - Number of the frames of the KITTI-ROAD dataset. Source: (FRITSCH et al., 2013)

Abbreviation	train	test	description
UU	98	100	urban unmarked
UM	95	96	urban marked two-way road
UMM	96	94	urban marked multi-lane road
URBAN	289	290	all three urban subsets

Each proposed algorithm had its own learning process for each category. A sample set for each category was built, extracting $\simeq$ 128000 samples from the UU image training set, $\simeq$ 126000 samples from the UM image training set and $\simeq$ 117000 samples from the UMM image training set. These sets were used to perform the feature's descriptor using the ANN approach, the HistonBoost approach and the ProbBoost approach.

The methodology for training the ANN algorithm uses the cross validation method, splitting the training images in training, validation and test subsets, containing $\simeq 15\%$,$\simeq 7\%$ and $\simeq 78\%$ respectively. The time for training each category takes around 24 hours and the classification time for a single image takes around 3 seconds. It should be mentioned that some adaptations from the first version proposed in (VITOR et al., 2013) were required: (i) it does not use the moving average technique, because it is applicable only for image sequences, (ii) the learning process does not use the strategy of training with subclass (shadow area, normal area and land marks), where it would improve the final result of road detection. In the case of the learning process to HistonBoost, the training images were splitted into training subset containing $\simeq 40\%$, and the test subset with $\simeq 60\%$. The time for training each category of this approach

takes around 72 hours and the classification time for a single image takes around 2.5 minutes in a naive implementation. The learning procedure applied to ProbBoost considers the complete sample training set. To this approach, the time for training each category takes around 7 hours and the classification time for a single image takes the same time as HistonBoost, around 2.5 minutes.

Table 3.2 - Results (%) of pixel-based for URBAN UNMARKED road area evaluation.
Source: The KITTI Vision Benchmark Suite.<http://www.cvlibs.net/datasets/kitti/eval_road.php> accessed: 05/07/2014

Metric space - Testing set								
Rank	**Method**	**Setting**	**MaxF**	**AP**	**PRE**	**REC**	**FPR**	**FNR**
1	RES3D-Velo (SHINZATO et al., 2014)		83.78 %	73.29 %	78.63 %	89.65 %	8.11 %	10.35 %
2	SPRAY (KUEHNL et al., 2012)		82.63 %	87.30 %	82.32 %	82.94 %	5.93 %	17.06 %
3	ProbBoost (VITOR et al., 2014c)		80.29 %	69.05 %	85.58 %	75.61 %	4.24 %	24.39 %
4	MDE		79.34 %	80.04 %	82.25 %	76.63 %	5.50 %	23.37 %
5	BM (WANG et al., 2014)		78.15 %	62.68 %	71.06 %	86.82 %	11.76 %	13.18 %
6	RES3D-Stereo		78.15 %	73.55 %	77.89 %	78.42 %	7.41 %	21.58 %
7	SP + BL (EINECKE; EGGERT, 2014)		74.42 %	80.10 %	66.04 %	85.24 %	14.58 %	14.76 %
8	SP (EINECKE; EGGERT, 2014)		73.63 %	69.87 %	65.43 %	84.18 %	14.80 %	15.82 %
9	HistonBoost (VITOR et al., 2014a)		73.51 %	63.07 %	77.36 %	70.03 %	6.82 %	29.97 %
10	CN (ALVAREZ et al., 2012)		71.48 %	66.30 %	72.09 %	70.88 %	9.13 %	29.12 %
11	BL (FRITSCH et al., 2013)		69.49 %	73.84 %	65.73 %	73.70 %	12.78 %	26.30 %
12	ANN (VITOR et al., 2013)		54.17 %	36.86 %	39.50 %	86.19 %	43.92 %	13.81 %

For all quantitative evaluation, a distinguished study including a majority of proposed algorithm found in literature is demonstrated, comparing these developed algorithms with the state-of-art approach in this field. A remark is that the baseline (BL) to road area is provided as a lower bound (FRITSCH et al., 2013), by averaging all ground truth road maps from the present validation set. The Table 3.2 presents the quantitative evaluation for the UU category. As can be seen by the F_{max} measure, the ProbBoost reaches 80.29% for this category, showing an expressive evaluation compared with HistonBoost and ANN. It has a gain of 6.78% from HistonBoost and 26.12% from ANN. To this category, the RES3D-Velo (SHINZATO et al., 2014), which is the first in the rank at the moment (looking to the site [4]), shows an improvement of 3.49% with respect to ProbBoost. It is important to mention that this technique uses the Velodyne laser as sensor to detect the road, which is more robust to noise (this information is showed on the third columns of table called "Setting", where denotes Laser sensor, denotes stereo camera sensor and empty denotes mono-camera sensor). Considering algorithms that use camera as sensor, the SPRAY algorithm (KUEHNL et al., 2012) presents a gain of 2.34% from ProbBoost, being this proposed method comparable with the state-of-art approaches to this category.

[4] <http://www.cvlibs.net/datasets/kitti/eval_road.php> accessed on:10/09/2014

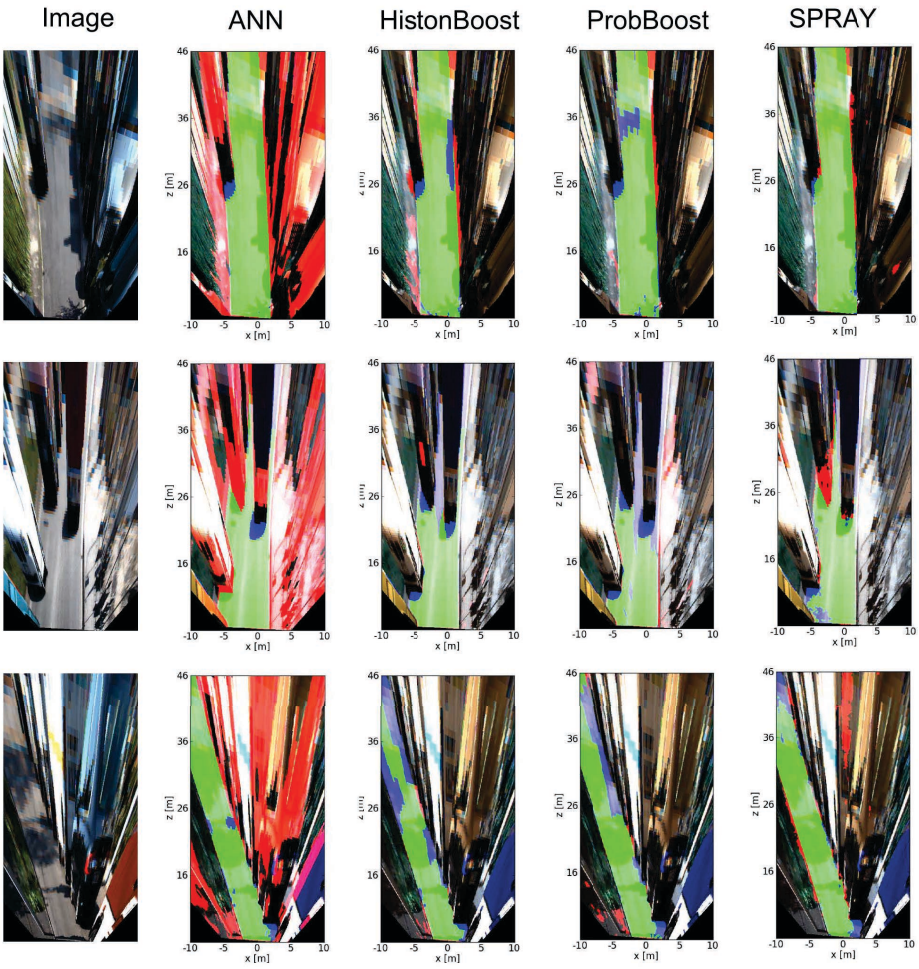


Figure 3.19 - The result process for Urban Unmarked (UU) in the BEV space. The columns present the original image, followed by ANN, HistonBoost, ProbBoost and SPRAY results. Green color represents True Positive, red one represents True Negative and blue defines False Negative.

Qualitative results to UU category can be seen in Figures 3.19 and 3.20. The Figure 3.19 presents the final road recognition in the metric space, where the evaluation process was done, comparing these three algorithms together with the SPRAY algorithm. Figure 3.20 is showing the same result in perspective space, where a slight difference between the SPRAY algorithm and the ProbBoost algorithm may be observed.

With respect to UM category, Table 3.3 shows that the evolution between these proposed algorithm remained the same, but with different gains. As can be seen by the F_{max} measure, ProbBoost had a improvement of 3.89% from HistonBoost and 24.96% from ANN. A curious phenomenon was that SPRAY and ProbBoost algorithms overcame the RES3D-Velo in 3.97% and 3.35% respectively. To this category, the difference between SPRAY and ProbBoost is only 0.62%. Example processing results for the different evaluated methods are depicted in Figure 3.21 to metric space and Figure 3.22 to perspective space.

Figure 3.20 - The result process for Urban Unmarked (UU) in the perspective space. The rows present the original image, followed by ANN, HistonBoost and ProbBoost results. Green color represents True Positive, red one represents False Negative and blue defines True Negative.

Table 3.3 - Results (%) of pixel-based for URBAN MARKED road area evaluation.
Source: The KITTI Vision Benchmark Suite.<http://www.cvlibs.net/datasets/kitti/eval_road.php> accessed: 05/07/2014

		Metric space - Testing set						
Rank	**Method**	**Setting**	**MaxF**	**AP**	**PRE**	**REC**	**FPR**	**FNR**
1	SPRAY (KUEHNL et al., 2012)		88.22 %	91.32 %	88.63 %	87.80 %	5.20 %	12.20 %
2	ProbBoost (VITOR et al., 2014c)		87.60 %	76.04 %	85.92 %	89.36 %	6.76 %	10.64 %
3	SP + BL (EINECKE; EGGERT, 2014)		85.66 %	88.98 %	84.11 %	87.28 %	7.62 %	12.72 %
4	RES3D-Velo (SHINZATO et al., 2014)		84.25 %	74.95 %	76.07 %	94.41 %	13.71 %	5.59 %
5	HistonBoost (VITOR et al., 2014a)		83.71 %	73.31 %	82.58 %	84.87 %	8.27 %	15.13 %
6	MDE		83.40 %	86.61 %	83.45 %	83.35 %	7.63 %	16.65 %
7	BL (FRITSCH et al., 2013)		82.53 %	85.59 %	79.24 %	86.11 %	10.41 %	13.89 %
8	BM (WANG et al., 2014)		79.19 %	66.78 %	70.29 %	90.66 %	17.69 %	9.34 %
9	RES3D-Stereo		79.01 %	80.21 %	76.64 %	81.54 %	11.47 %	18.46 %
10	SP (EINECKE; EGGERT, 2014)		78.49 %	76.85 %	72.77 %	85.20 %	14.72 %	14.80 %
11	ICF		74.14 %	58.41 %	64.37 %	87.40 %	22.34 %	12.60 %
12	CN (ALVAREZ et al., 2012)		73.97 %	73.64 %	69.93 %	78.51 %	15.59 %	21.49 %
13	ANN (VITOR et al., 2013)		62.64 %	46.80 %	50.18 %	83.34 %	38.21 %	16.66 %

Taking into account the UMM category, two attractive results are necessary to be emphasized, as depicted in Table 3.4 by the F_{max} measure. The first one, that should be highlighted, is that the results of ANN approach overcome the baseline(BL) approach in 4.92%. The second one is that the response of ProbBoost algorithm outperforms all other methods for this specific category. It obtained a gain of 1.7% from the SPRAY method. The result for this category can be seen in Figures 3.23 and 3.24.

To conclude the evaluation process, Table 3.5 presents the final results merging UU, UM and UMM categories. As can be seen by the F_{max} measure, considering all categories together,

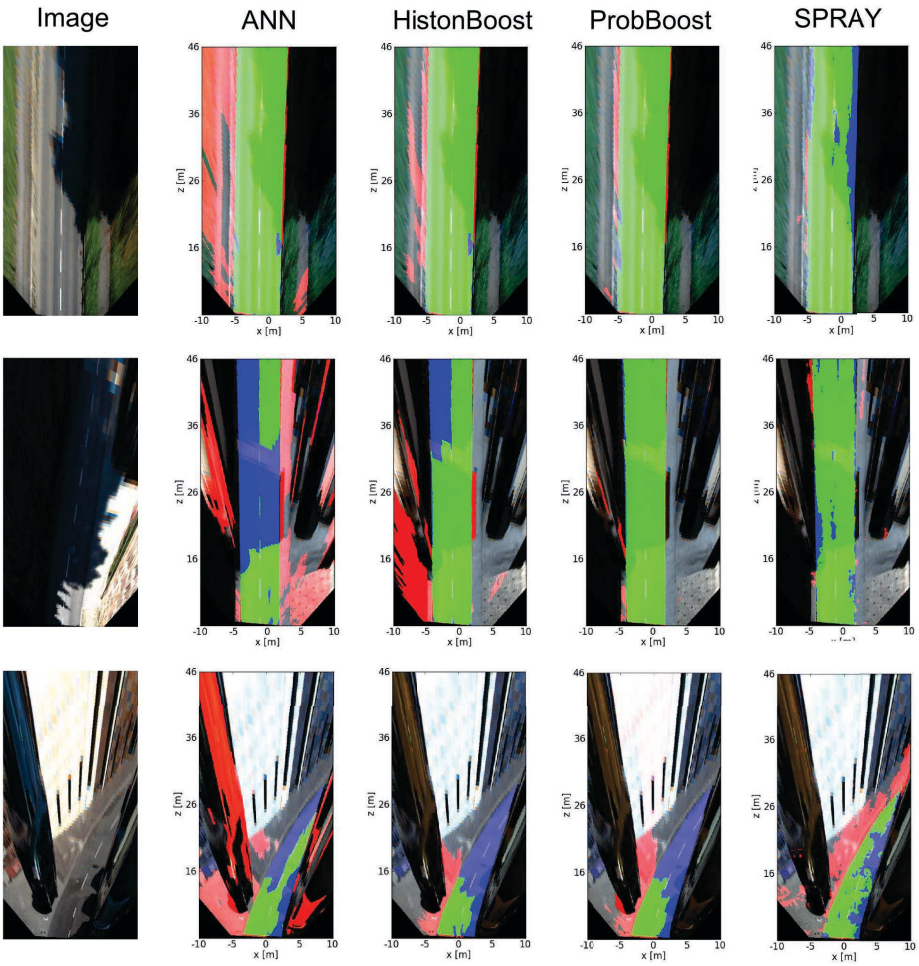


Figure 3.21 - The result process for Urban Marked (UM) in the BEV space. The columns present the original image, followed by ANN, HistonBoost, ProbBoost and SPRAY results. Green color represents True Positive, red one represents True Negative and blue defines False Negative.

the ProbBoost algorithm has presented expressive results, becoming the first one of the ranking at the present moment. It represents the cutting-edge approach to field of road detection considering this challenging urban Kitti-road benchmark, where can be considerable an impressive result given the complexity of the scenarios. However, in this current implementation, the processing time is still an issue if compared with other approaches and also if aimed to autonomous navigation which demands real-time processing. It should be mentioned that improvements on real-time processing can be reached using for example the benefits of GPU cards.

Compared to other approaches, ANN outcomes presents the worst results. It can be explained by the higher complexity of the Kitti-road dataset, or, the training process used for ANN was not adequate, having low expressiveness if observed the strategy of subclasses used in (VITOR et al., 2013).

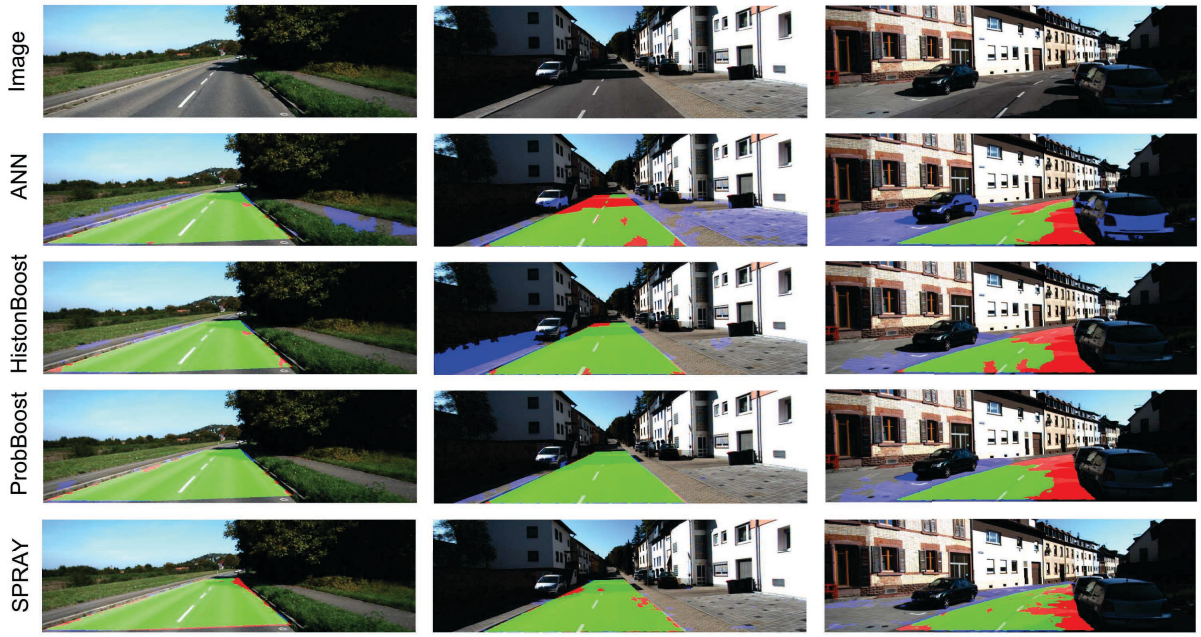


Figure 3.22 - The result process for Urban Marked (UM) in the perspective image. The rows present the original image, followed by ANN, HistonBoost, ProbBoost and SPRAY results. Green color represents True Positive, red one represents False Negative and blue defines True Negative.

Table 3.4 - Results (%) of pixel-based for URBAN MARKED MULTI-LANE road area evaluation. Source: The KITTI Vision Benchmark Suite.<http://www.cvlibs.net/datasets/kitti/eval_road.php> accessed: 05/07/2014

		Metric space - Testing set						
Rank	**Method**	**Setting**	**MaxF**	**AP**	**PRE**	**REC**	**FPR**	**FNR**
1	ProbBoost (VITOR et al., 2014c)		90.12 %	85.04 %	88.15 %	92.18 %	14.50 %	7.82 %
2	SPRAY (KUEHNL et al., 2012)		88.42 %	93.56 %	88.31 %	88.53 %	13.71 %	11.47 %
3	HistonBoost (VITOR et al., 2014a)		87.70 %	81.59 %	84.36 %	91.32 %	19.81 %	8.68 %
4	RES3D-Velo (SHINZATO et al., 2014)		87.64 %	85.81 %	86.70 %	88.60 %	15.91 %	11.40 %
5	BM (WANG et al., 2014)		86.56 %	80.49 %	83.15 %	90.26 %	21.40 %	9.74 %
6	CN (ALVAREZ et al., 2012)		85.77 %	84.91 %	83.37 %	88.30 %	20.61 %	11.70 %
7	MDE		84.49 %	89.57 %	88.24 %	81.04 %	12.63 %	18.96 %
8	SP (EINECKE; EGGERT, 2014)		81.95 %	83.09 %	76.77 %	87.88 %	31.11 %	12.12 %
9	SP + BL (EINECKE; EGGERT, 2014)		81.62 %	85.53 %	75.65 %	88.62 %	33.38 %	11.38 %
10	RES3D-Stereo		81.31 %	85.43 %	80.04 %	82.62 %	24.11 %	17.38 %
11	ANN (VITOR et al., 2013)		81.09 %	68.93 %	70.43 %	95.56 %	46.94 %	4.44 %
12	BL (FRITSCH et al., 2013)		76.17 %	78.42 %	65.02 %	91.95 %	57.89 %	8.05 %

3.5.2 Urban Street Scene Understanding

Giving continuity to Semantic Context recognition, this section explores the applicability of two algorithms, the HistonBoost and the ProbBoost, in the task of Urban Street scene understanding. To demonstrate the viability of extending the results to multi-class classification, this experiment uses another dataset publicly available at KITTI benchmark. This dataset is composed by 446 images acquired in a inner-city scenario, having a sequence image of 0:45'min.

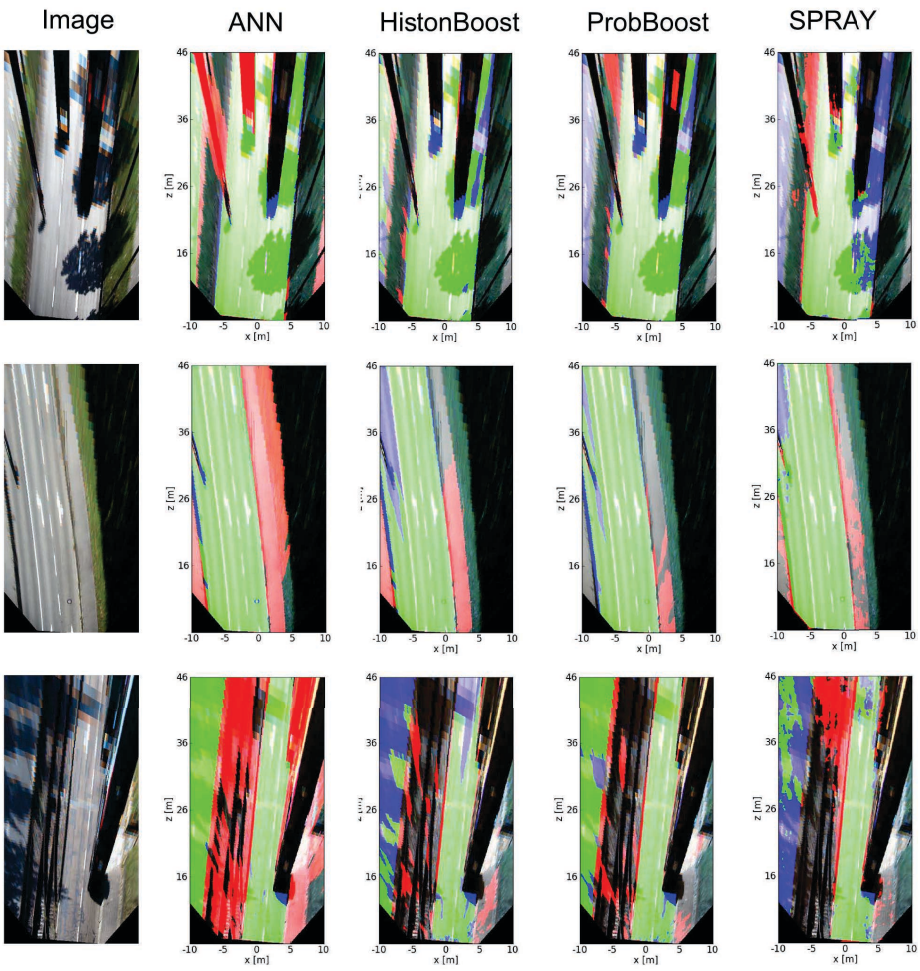


Figure 3.23 - The result process for Urban Marked Multi-lane (UMM) in the metric space. The columns present the original image, followed by ANN, HistonBoost, ProbBoost and SPRAY results. Green color represents True Positive, red one represents True Negative and blue defines False Negative.

The images include common objects such as cars, trees, building at a resolution of 1392x512 pixels. It was manually annotated 90 images, randomly selected, to supply the quantitative evaluation and also for training the final classifiers. The scene was labeled into seven semantic classes, i.e., Building, Vehicle, Sky, Sidewalk, Road, Vegetation and Void. This specific dataset is quite challenging and even objects of the same class in the scene have different appearance.

The quantitative evaluation was performed using a strategy defined by *One against All*. It means that each class is evaluated separately. Based on this approach, it was applied the same evaluation measures explained before, and the final evaluation was given by the average obtained from all classes measures output. Restrictively for evaluation analysis, the 90 images that were hand-labelled to provide the ground truth for training procedure, 67% of these images were taken for training the algorithms and 33% to comparative evaluation. In this case, the learning process to both algorithms use a sample set extracted from these training set, having

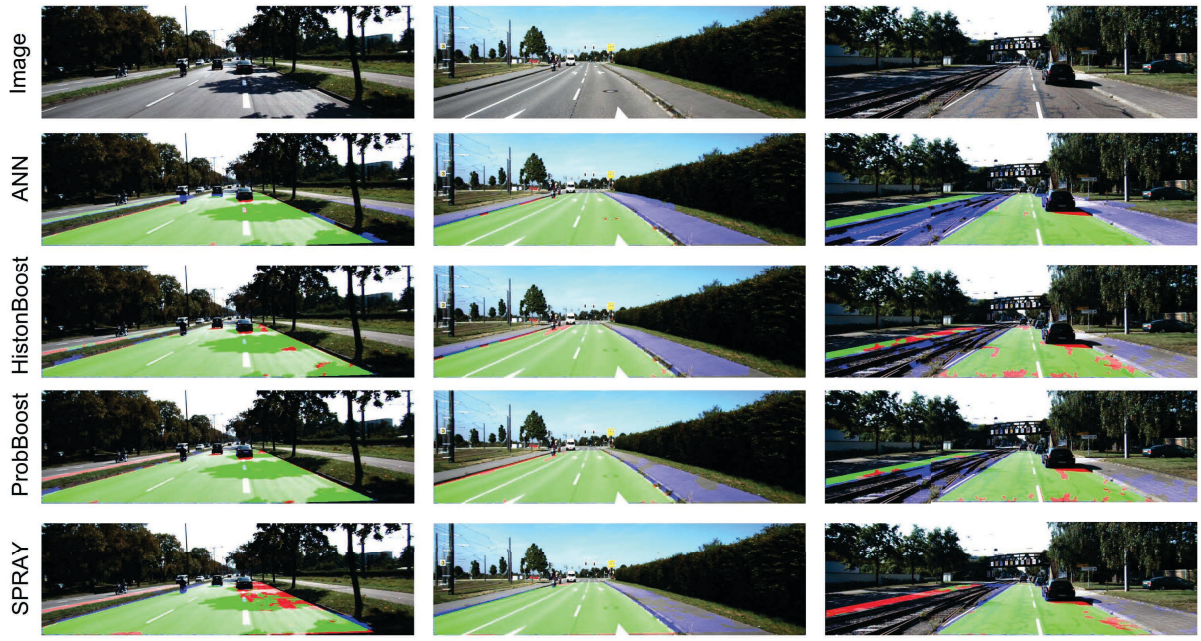


Figure 3.24 - The result process for Urban Marked Multi-lane (UMM) in the perspective image. The rows present the original image, followed by ANN, HistonBoost, ProbBoost and SPRAY results. Green color represents True Positive, red one represents False Negative and blue defines True Negative.

Table 3.5 - Results (%) of pixel-based for complete URBAN ROAD area evaluation, performed on the metric space (BEV).
Source: The KITTI Vision Benchmark Suite.<http://www.cvlibs.net/datasets/kitti/eval_road.php> accessed: 05/07/2014

Rank	Method	Setting	MaxF	AP	PRE	REC	FPR	FNR
1	ProbBoost (VITOR et al., 2014c)		87.21 %	77.79 %	86.96 %	87.47 %	7.55 %	12.53 %
2	SPRAY (KUEHNL et al., 2012)		86.33 %	90.91 %	86.78 %	85.89 %	7.53 %	14.11 %
3	RES3D-Velo (SHINZATO et al., 2014)		85.49 %	79.03 %	79.93 %	91.88 %	13.28 %	8.12 %
4	HistonBoost (VITOR et al., 2014a)		83.41 %	74.06 %	82.39 %	84.46 %	10.39 %	15.54 %
5	MDE		82.72 %	87.58 %	85.44 %	80.17 %	7.87 %	19.83 %
6	BM (WANG et al., 2014)		82.32 %	68.95 %	76.15 %	89.56 %	16.15 %	10.44 %
7	RES3D-Stereo		79.91 %	81.56 %	78.55 %	81.32 %	12.79 %	18.68 %
8	SP + BL (EINECKE; EGGERT, 2014)		79.48 %	83.93 %	73.59 %	86.40 %	17.85 %	13.60 %
9	CN (ALVAREZ et al., 2012)		78.92 %	79.14 %	76.25 %	81.79 %	14.67 %	18.21 %
10	SP (EINECKE; EGGERT, 2014)		78.75 %	77.66 %	72.41 %	86.30 %	18.93 %	13.70 %
11	BL (FRITSCH et al., 2013)		75.61 %	79.72 %	68.93 %	83.73 %	21.73 %	16.27 %
12	ANN (VITOR et al., 2013)		68.12 %	51.52 %	54.85 %	89.85 %	42.59 %	10.15 %

$\simeq 750000$ samples to perform the procedure. Take note that the other subset was used only for evaluation proceeding.

Table 3.6 and 3.7 depict the evaluation result to each class of both algorithms respectively. Based on the same principle presented before, it is proposed a baseline for each class to provide the lower bound comparison. It is obtained by averaging all ground truth road maps from the present validation set. and its results are depicted in Table 3.8. Considering these Tables, as can be seen by the F_{max} measure, the ProbBoost approach reached values more than 88% for all classes (except to sidewalk that receives 74.99%). This behavior is not the same to HistonBoost,

Table 3.6 - HistonBoost Results (%) of pixel-based for Urban Street evaluation, performed on the perspective space.

Class	MaxF	AP	PRE	REC	FPR	FNR
road	87.50%	75.58%	87.80%	87.20%	3.13%	12.80%
sidewalk	45.80%	33.21%	76.24%	32.73%	0.96%	67.27%
vehicle	69.30%	56.67%	81.94%	60.04%	1.88%	39.95%
building	76.30%	63.58%	79.21%	73.60%	5.41%	26.40%
sky	57.89%	49.74%	99.22%	40.87%	0.03%	59.13%
Vegetation	84.40%	75.63%	86.37%	82.52%	4.89%	17.48%

Table 3.7 - ProbBoost Results (%) of pixel-based for Urban Street evaluation, performed on the perspective space.

Class	MaxF	AP	PRE	REC	FPR	FNR
road	93.32%	86.64%	93.25%	93.40%	1.75%	6.60 %
sidewalk	74.99%	54.92%	81.37%	69.55%	1.50%	30.45%
vehicle	88.20%	73.98%	87.65%	88.75%	1.78%	11.25%
building	88.02%	76.82%	89.02%	87.05%	3.01%	12.95%
sky	95.33%	87.07%	94.93%	95.74%	0.48%	4.26%
Vegetation	90.73%	85.00%	90.77%	90.69%	3.46%	9.31%

Table 3.8 - Baseline Comparator (%) to pixel-based for Urban Street evaluation, performed on the perspective space.

Class	MaxF	AP	PRE	REC	FPR	FNR
road	81.52%	87.95%	82.42%	80.64%	4.45%	19.36%
sidewalk	45.65%	37.76%	33.81%	70.22%	12.98%	29.78%
vehicle	52.53%	47.90%	38.32%	83.51%	19.12%	16.49%
building	61.53%	60.57%	52.79%	73.76%	18.49%	26.24%
sky	52.48%	47.56%	40.12%	75.88%	10.52%	24.12%
Vegetation	56.64%	51.83%	43.74%	80.33%	38.77%	19.67%

that its threshold to all classes is 57.89% (except to sidewalk that receives in this case 45.80%). Observing the baseline, which restricts the lower bound for each class, the both algorithms performed better than the baseline taking into account all classes.

In general, Table 3.9 demonstrates the results for the three approaches extracted from the average of all classes. In this case, is possible to see the performance among them. Comparing the F_{max} measure, the ProbBoost algorithm reached an improvement of 18.23% from HistonBoost and from Baseline this percentage is approximately 30% in the perspective space. The results express a satisfactory performance to ProbBoost even considering the task of multi-class recognition.

Figure 3.25 presents the result process for two scenarios. The first row shows the original image followed by the ground truth and these approaches. As can be seen, these approaches were able to detect several obstacles such as vehicles, trees and sidewalks. Even with differ-

Table 3.9 - Comparison Results (%) of pixel-based for Urban Street evaluation, performed on the perspective space.

Method	MaxF	AP	PRE	REC	FPR	FNR
HistonBoost	70.20 %	59.07%	85.13 %	62.83%	2.72%	37.17 %
ProbBoost	88.43%	77.41%	89.50%	87.53%	2.00%	12.47%
BaseLine	58.39%	55.60%	48.53%	77.39%	17.39%	22.61%

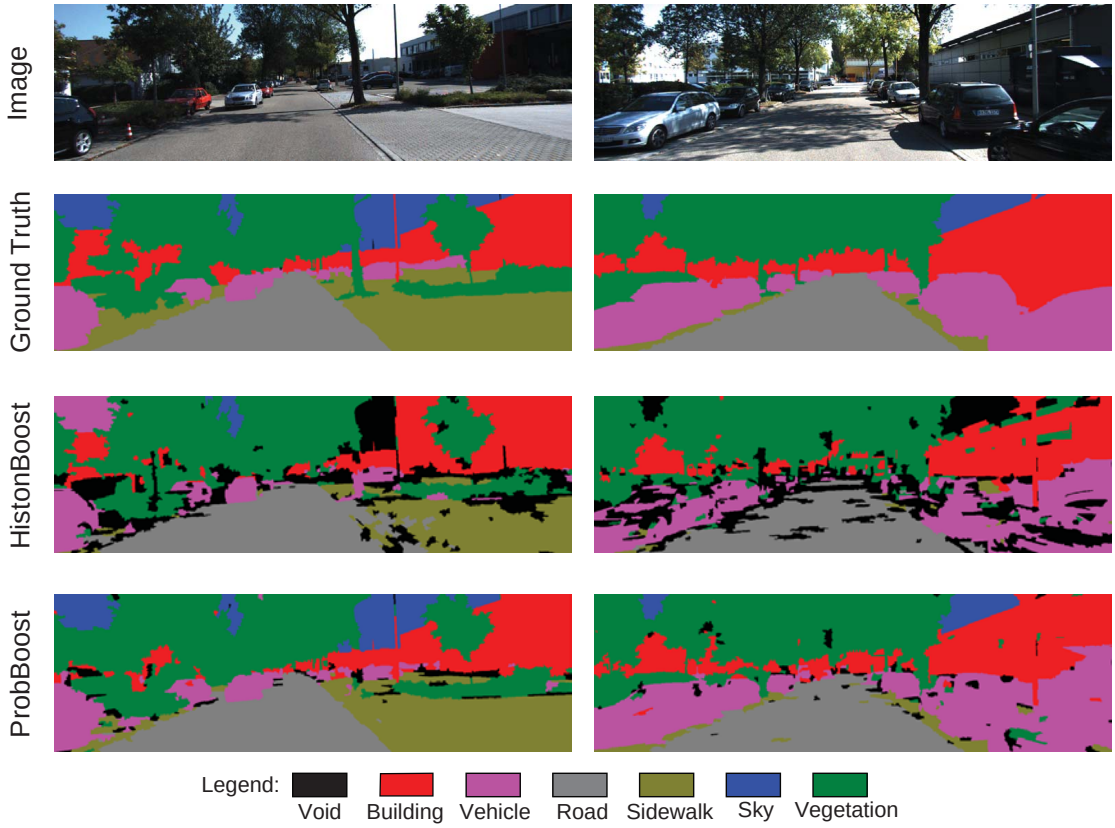


Figure 3.25 - The Urban Scene Understanding results. The first row shows the original image followed by the ground truth, HistonBoost and ProbBoost results.

ent conditions such as shadow areas and areas with high luminosity provided by the sun, the algorithms could generate outcomes with satisfactory approximations for the road area. The complete video results using the ProbBoost algorithm is public available in Vitor (2014f) [5] and the HistonBoost algorithm in Vitor (2014e) [6].

[5] <http://youtu.be/JxXIhpsC9Gk> accessed on: 05/08/2014
[6] <http://youtu.be/MsR3U5RUtxk> accessed on: 05/08/2014

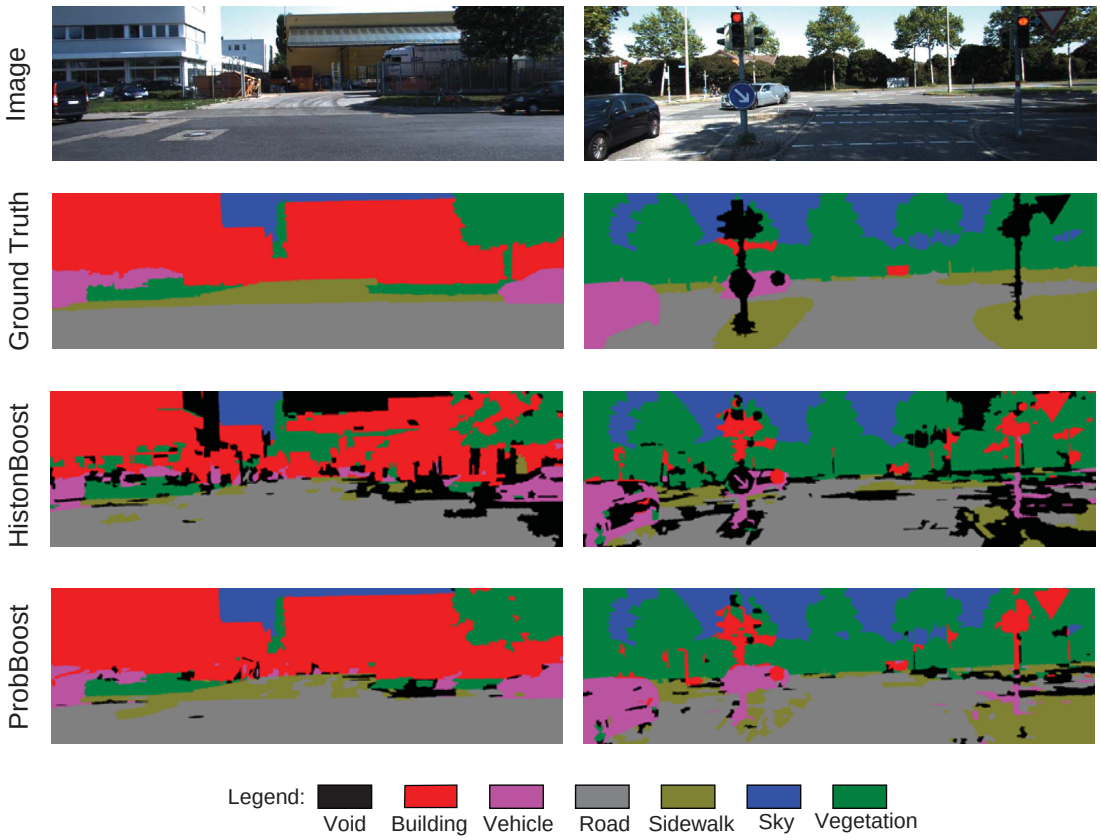


Figure 3.26 - Other Urban Scene Understanding results. The first row shows the original image followed by the ground truth, HistonBoost and ProbBoost results.

3.6 Conclusion

In this chapter was presented a new semantic context recognition method for urban traffic scenes. Experiments were divided into two phases. The first phase, focused at the task of road recognition for automotive application such as ADAS systems and autonomous navigation. In this phase, the meaningful performance analysis merging three distinct proposed algorithms with other excellent methods found in literature were validated. The first proposed algorithm, called ANN, constructs a feature's descriptor extracting statistical measures of 2D and 3D information to be classified using an Artificial Neural Network. The second algorithm, defined as HistonBoost, constructs a multi normalized-histogram feature's descriptor using by extracting information from Texton and Dispton maps, and then the output of this process supplies the multi normalized-histogram Joint Boosting classifier. The third proposed algorithm, called ProbBoost, was an improvement with respect of the second one, where this one replaced the multi normalized-histogram features by the creation of a probabilistic distribution based on Texton and Dispton maps to model the probabilistic weak classifiers used in the Joint Boosting

classifier.

The experiments conducted on real driving situations demonstrate the qualitative and quantitative evaluation of these algorithm to detect road despite the presence of shadows and other objects in the scene, inherent from the complexity of inner-city environments. The ProbBoost method copes better than HistonBoost and ANN methods for all dataset used. Further, considering all categories unified, the results also provide the benefits over existing methods, becoming the cutting-edge approach in urban road recognition found in literature until the present moment (10/2014). Details are available at Urban Kitti-Road dataset[7].

The second phase, focused at the task of multi-class recognition in urban scenario. Using the same principle of evaluation applied to road detection, in this phase was studied the meaningful performance analysis extending the HistonBoost and ProbBoost algorithms to act in this task of multi detection. Presented in the results, the ProbBoost performs better than HistonBoost even considering multiple classes. So, this chapter presented a solid algorithm that plays important role in the task of Urban Road Scene Understanding.

[7] <http://www.cvlibs.net/datasets/kitti/eval_road.php> Accessed on: 10/09/2014

CHAPTER 4

Dynamic Evidential Grid using Semantic Context

I believe in evidence. I believe in observation, measurement, and reasoning, confirmed by independent observers. I'll believe anything, no matter how wild and ridiculous, if there is evidence for it. The wilder and more ridiculous something is, however, the firmer and more solid the evidence will have to be.

(Isaac Asimov)

This chapter presents an approach dedicated to local perception mapping and characterization of static and moving obstacles using only a pair of stereo vision cameras to model the occupancy grid (evidential grid). This kind of sensor provides different measurement characteristics if compared to laser scanners. Moreover, cameras are already onboarded on standard vehicles differently from laser scans, in general, more expensive. Figure 4.1 shows a rude block diagram of the proposed system linked with the previous diagram described in chapter 3 which is responsible to provide the semantic context of the image. For the second part (II) in Figure 4.1, a Structure from Motion (SfM) is applied to get the rigid transformation between two successive images, the Sensor Grid models an inverse sensor grid and the Perception Grid performs the temporal fusion and mobile cells detection.

Differently from others works, it does not require some prior digital map information, neither a precise pose estimation is needed, nor the tracking of vehicles. The main contributions of this chapter are the proposed technique to build a new sensor model that provides reliable urban environment sensing despite the uncertainty in distance measurement associated with

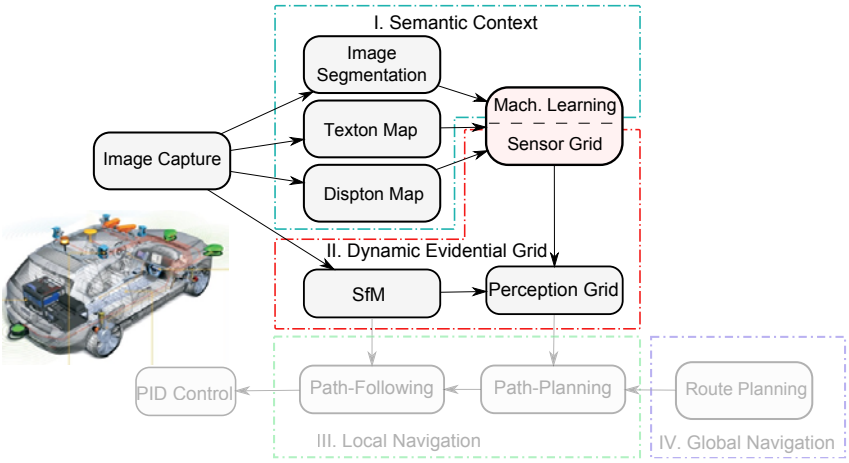


Figure 4.1 - The proposed solution to Evidential grid adding the semantic context.

the stereo geometry and the combination/update rules for meta-knowledge that incorporate the semantic context on evidential grids.

4.1 Occupancy Grid to robotic perception

An autonomous robot must be able to understand the transversality and occupation of space around itself. Among representation models of the environment, discrete approaches have frequently been used in the field of robotic, particularly through of the occupancy grid representation shown in Elfes (1989b). The main idea of this formalism is to model the environment without making any assumptions about the geometry of the elements present into the scene. Thus, this section starts by providing a detailed description about principle and definitions involving the representation of metric grids (section 4.1.1). Then, the conception of occupancy grid associated with this metric grids is described in section 4.1.2. Finally, in section 4.1.3 is demonstrated the system architecture based on egocentric referential approach that binds the occupancy grid with the like-car robot, providing intuition about the system evolution along the time.

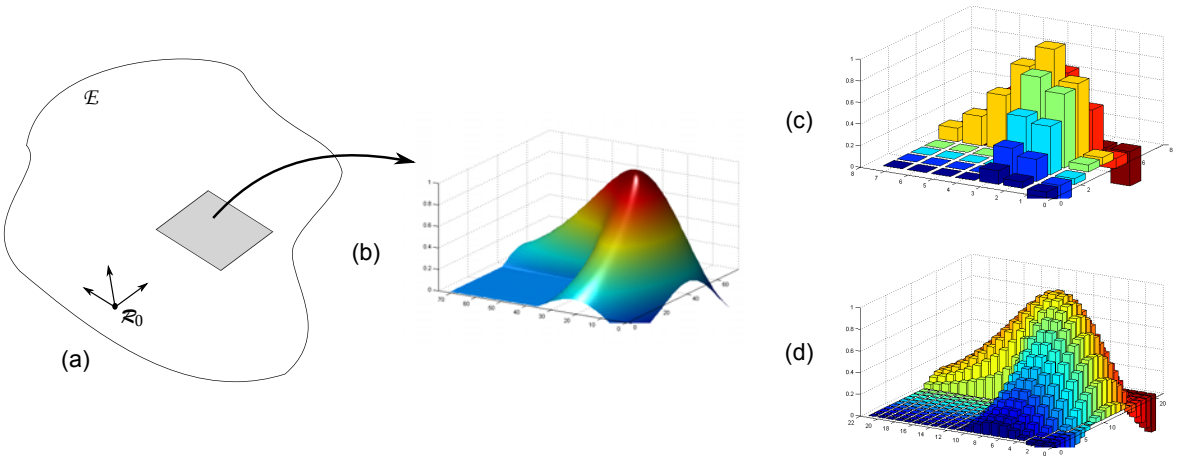


Figure 4.2 - The approximative representation of a 2D space. (a) The 2D space E associated to a reference R_0; (b) The continuous function $f(x,y) :\rightarrow [0,1]$ (c) and (d) represent different approximation in function of the resolution step r.

4.1.1 Principle and Definitions

As previously mentioned, in this work is used a representation of the environment based on a grid. This grid is modeled considering a 2D dimensional space E associated to a reference $R_0(0,\vec{x},\vec{y})$. Given a continuous function $f(x,y) : E \rightarrow [0,1]$ defined by every point $P(x_p,y_p)$ of E, a model can be defined to estimate this function f. A possible method consists of modeling a discrete approximation performing a sampling in E. This discretization permits the construction of a matrix representation denoted by G, which performs a regular sampling in f, having its own referential relative to the axes of R_0 in the space. Considering a regular sampling, the discretization step has a dimension of a cell, where its size expresses the quality of an approximation, i.e, how smaller the cell size is, better will be the approximation of a function. In this case is defined the grid resolution as being the number of cell by unit length. The resolution r is similar to a spatial frequency that can be expressed as a function of sampling step $r = \frac{1}{\Delta}$, where Δ is the size step. Figure 4.2 depicts a practical example considering different approximation to a function.

Taking into account that the function f is not constant along the time, it is made an approximation to each instant t of sampling (temporal). Then, it is considered that all cells are simultaneously sampled and independents in the time ($f(x,y,t)$). In the same way to spatial sampling, the temporal refresh rate is determinant with respect to the dynamics of the function.

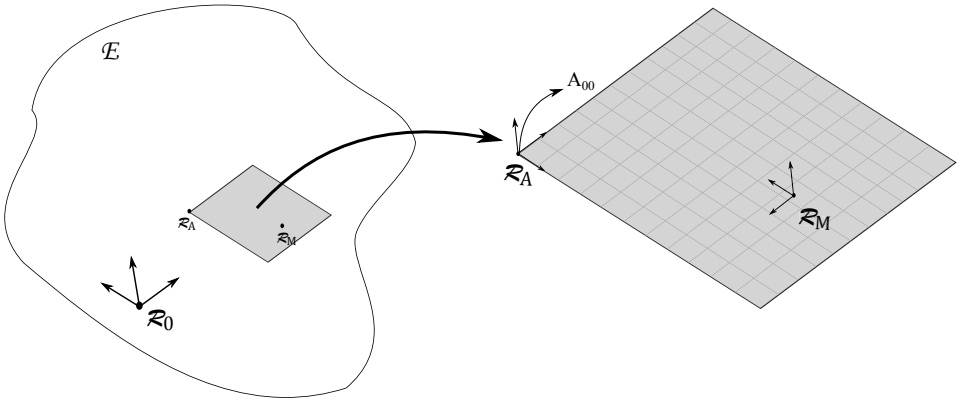


Figure 4.3 - The referential correspondence between 2D space and the defined grid. R_A represents the oriented reference of the grid and R_M represents the oriented reference of the 2D space.

The definition of the grid can be analogously represented like image processing, and its formalism explained here, was inspired and based on the work of Moras (2013). Considering a rectangular domain $D = [a,b] \times [c,d] \subset \mathbb{E}^2$ having a reference R_M to allow the localization of all cells in the space, and given the function $f(x,y) : D \rightarrow [0,1]$ to be approximated by a discrete method. The grid G^A, where A represents the reference, can be defined as a partitioning of D in an set of rectangular cells G^A_{ij} with dimension $\Delta_x \times \Delta_y$ arranged, by convention, in a matrix along the axes of reference R_A:

$$G^A = \left\{ G^A_{ij} | \forall (i,j) \in \mathbb{Z}, \begin{bmatrix} x_{ij} \\ y_{ij} \end{bmatrix} \in D \right\} \tag{4.1}$$

where each G^A_{ij} corresponds to the index cell (i,j). The grid is defined by its intrinsic characteristics:

- The reference R_A to localize the cells;
- The relative position of R_A from a point M of the spatial reference $R_M(M,\vec{x},\vec{y})$.
- A spatial domain D^A corresponding to rectangle of work according to the axes of the reference R_M.
- A resolution $\Delta^A(\Delta^A_x,\Delta^A_y)$ defined as being the cell size.

The referential R_M is by convention oriented as addressed in Figure 4.3, observed its relative position from R_A.

Given a cell G^A_{ij} of the grid A with index (i,j), is defined:

- C_{ij}^A as being its center
- S_{ij}^A as being its surface, $S_{ij}^A = [c_{ij}^A.x - \frac{\Delta_x^A}{2}, c_{ij}^A.x + \frac{\Delta_x^A}{2}] \times [c_{ij}^A.y - \frac{\Delta_y^A}{2}, c_{ij}^A.y + \frac{\Delta_y^A}{2}]$
- I_{ij}^A as being its value.

Similarly to the grid G^A, is defined:

- $C^A = \{C_{ij}^A\}$ as being the whole set of centers,
- $S^A = \{S_{ij}^A\}$ as being the whole set of Surfaces,
- $I^A = \{I_{ij}^A\}$ as being the whole set of values.

Finally, it is defined the belonging notion of a point P to a cell $P \in S_{ij}^A$, or to abuse of notation $P \in G_{ij}^A$ as follows:

$$\forall P = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{E}^2, P \in S_{ij}^A \iff \begin{array}{l} c_{ij}^A.x - \frac{\Delta_x^A}{2} < x \leq c_{ij}^A.x + \frac{\Delta_x^A}{2} \\ c_{ij}^A.y - \frac{\Delta_y^A}{2} < y \leq c_{ij}^A.y + \frac{\Delta_y^A}{2} \end{array} \tag{4.2}$$

which can be also expressed by:

$$\forall P = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{E}^2, P \in G_{ij}^A \iff \begin{array}{l} round({}^M x) =^0 x - i.\Delta_x \\ round({}^M y) =^0 y - j.\Delta_y \end{array} \tag{4.3}$$

In Equation (4.3), $round()$ is the function to round a number to the nearest integer. $({}^M x, {}^M y)$ are coordinates of the point in the referential R_M, and $({}^0 x, {}^0 y)$ are coordinates of point A_{00} in the referential R_M.

To extend the previous notation, it is defined to the point P with coordinates[1] $[x\ y\ 1]^T$:

- $C^A(x,y)$ as being the center of the cell containing P,
- $S^A(x,y)$ as being the surface of the cell containing P,
- $I^A(x,y)$ as being the value of the cell containing P.

[1] Homogeneous coordinates

4.1.2 The conception of Occupancy grid

The defined metric grid is the basis to represent the environment around the car, in the discrete form. Occupancy grid is a way to model the dynamic environment using the metric grid basis. In this first moment, the Occupancy grid is defined using a deterministic approach to formalize the conception about it, but, in section 4.2 is demonstrated the formalism that considers the uncertainties associated with the environment.

The function that represents a occupied space in the environment is characterized by $O(P,t)$ at time instant t. Taken into account the metric grid, the Occupancy grid G^O is defined considering the referential R_O of domain D and resolution Δ, where the estimated function is the occupied function $O(x,y,t)$. Each cell G^O_{ij} has a binary value meaning its occupation in the space, that is denoted by extension $O(G^O_{ij},t)$.

Each occupancy cell of the grid has a surface which is referenced in the physic world and it generalizes all information of the points therein contained. In this sense, a function F is defined to represent the occupation of that cell, taking all points belonging to the surface S^O_{ij} as depicted in Equation (4.4):

$$I^O_{ij} = O(G^O_{ij},t) = F(\{O(P,t)|P \in S^O_{ij}\}) \tag{4.4}$$

Considering the definition explained in the work of Moras (2013), each point belonging to the same cell has a different value of occupation, arising three possible cases:

- The cell is empty: All its points are free. $\forall P \in S^O_{ij}, O(P,t) = 0$. In this case $F = (\{0,\ldots,0\}) = 0$.

- The cell is occupied: All its points are occupied. $\forall P \in S^O_{ij}, O(P,t) = 1$. In this case $F = (\{1,\ldots,1\}) = 1$.

- The cell is partially occupied: $\exists P_1,P_2 \in S^O_{ij}, O(P_1,t) = 0 \& O(P_2,t) = 1$. In this case, an approximation should be done to describe the situation of the cell, carrying with it an error of approximation. Various strategies are considered as:

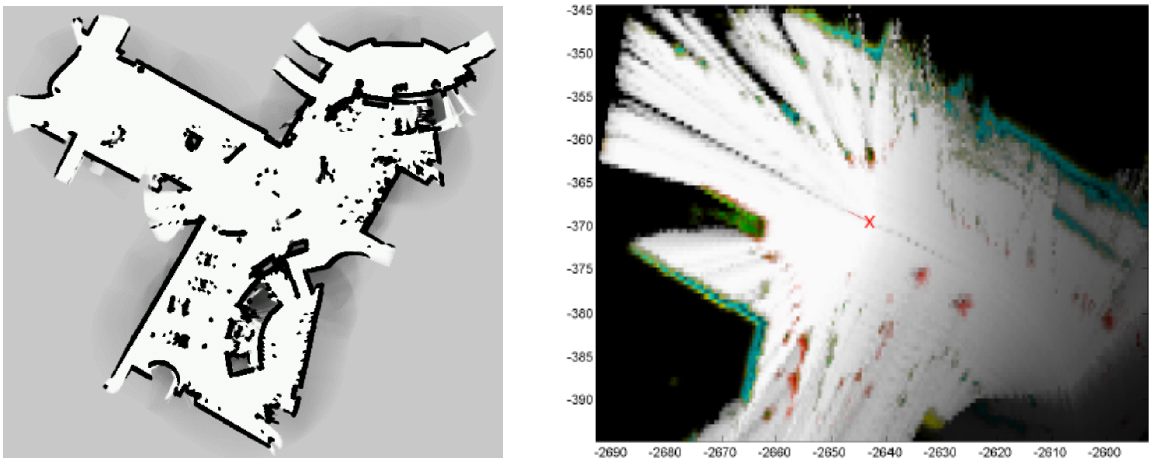


Figure 4.4 - Example of occupancy grids found in literature. (a) probabilistic Occupancy grid proposed by Thrun et al. (2005) and (b) Evidential Occupancy grid proposed by Moras (2013).

- The cell is free if it contains at least a free point, $O(G_{ij}^{O}) = 0 \Leftrightarrow \exists P \in S_{ij}^{O}, O(P) = 0$
- The cell is occupied if it contains at least a occupied point, $O(G_{ij}^{O}) = 1 \Leftrightarrow \exists P \in S_{ij}^{O}, O(P) = 1$
- The cell is occupied if it contains the majority of occupied points, $O(G_{ij}^{O}) = 1 \Leftrightarrow \begin{cases} 1 & \text{, if } \{\int_{S_{ij}^{O}} O(P)dP > \int_{S_{ij}^{F}} F(P)dP\} \\ 0 & \text{, otherwise} \end{cases}$

Generally, in applications to autonomous navigation where the safety must be considered, the second strategy is more appropriate and prudent for these applications. Following the words of Moras (2013), this kind of model is a powerful geometric modeling to represent the environment. since it makes no assumption about the shape of elements that compose the environment. Figure 4.4 presents some examples of the occupancy grid used in robotics.

Observed the errors of approximation and discretization, as well as the uncertainties of measures, in section 4.2 is presented a different formalism for improving the modeling and representation of these uncertainties, estimating a value of belief for every cells in the Occupancy grid.

4.1.3 The system architecture based on egocentric referential approach

In robotics, two strategies are usually used to define the spatial position of a robot and the elements that compose this spatial environment. These two strategies are denoted by Allocentric frame of reference and Egocentric frame of reference.

In allocentric frame of reference, all objects in the environment have a spatial position with reference to a fixed point, where this point of reference does not move along of the time. This kind of strategy is quite used in cartography or in approaches based on simultaneous localization and mapping (SLAM). In the Egocentric frame of reference, all objects in the environment have a spatial position with reference to a relative point that moves along of the time. In this work, the egocentric frame of reference is adopted due to the navigation procedure that considers the planning of navigation composed by two layers, the local navigation and global navigation, which is better detailed in Chapter 5.

Consider a car-like robot that operates within a finite domain D of a world plan. As previously explained, this domain is defined in an euclidian space having two dimensions $\mathbb{E}^2$, as can be seen in Figure 4.5. In this case, the Egocentric frame of reference is determined associating the reference R_M to the fixed point M defined as being the center of the car-like robot in the spatial environment. By simplicity, it will be called and referenced by ego-car. So, the Occupancy grid that is relative to the reference R_M, will move together with the ego-car. This approach has the advantage of always covering the same area around the ego-car and does not limit its field of evolution.

To make the perception more reliable while the ego-car is displacing in the environment, a temporal filter can be applied to take into account observed redundancies providing the Occupancy grid more robust to noises and with more complete information. This temporal filter uses the formalism of the predictor-corrector type, as showed in Figure 4.6.

Based on predictor-corrector approach, the principle of the system proposes the utilization of two distinct Occupancy grids denoted by Sensor Grid and Perception Grid to perform the sequential updating, as depicted in Figure 4.7. The Sensor Grid (SG) is built from the measured

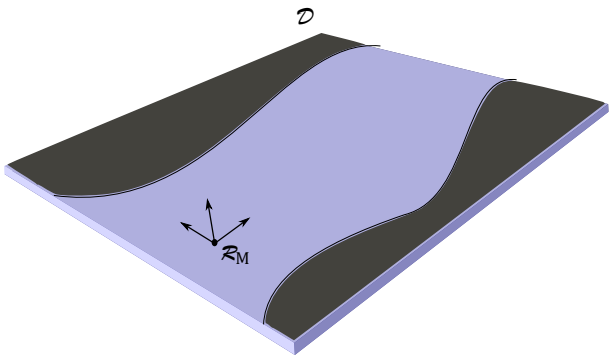


Figure 4.5 - Egocentric frame of reference for car-like robot. The irregular surface symbolizes the evolution field of the car-like robot embedded in the domain D, having a relative point of reference fixed at the center of the robot.

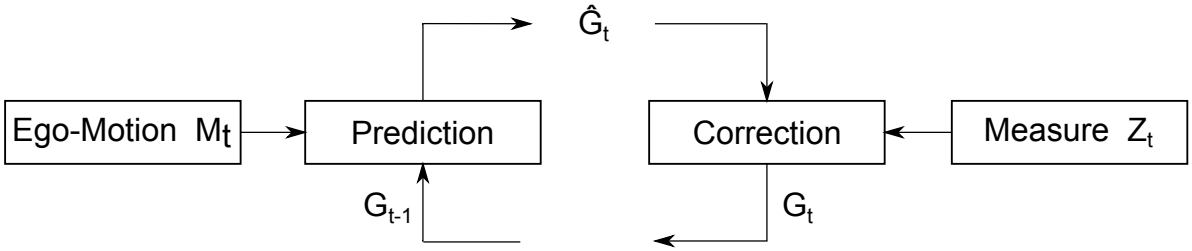


Figure 4.6 - Temporal filter based on predictor-corrector approach.

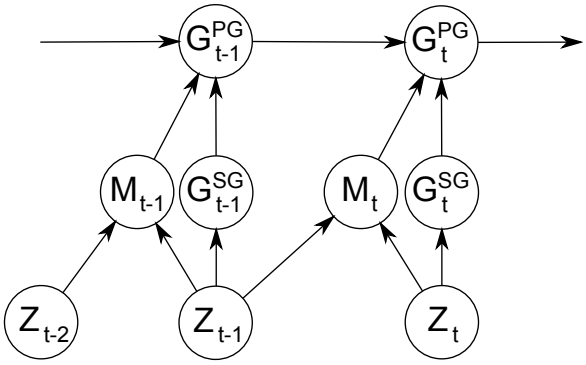


Figure 4.7 - The sequential update to Occupancy Grid.

Z_t and merged in the Perception Grid. The Perception Grid (PG) is a grid that maintains the cell state between the different instants of time, realizing in this way a Dynamic Local Perception. To perform the merging of the SG and the PG, it should be ensured that these two grids have a spatial and temporal coherence. In each time that a new SG is disposed, the current position is estimated and the prediction of PG must be done to the current time. The variable M_t represents the rigid transformation $M = [R|T]$ and the measure Z_t is obtained by a Stereo camera.

Prediction of the grid is realized when a new measure is available. It is required due to the displacement of the ego-car and also the elements of the scene that compose the dynamic environment. Improper prediction may cause inconsistency to PG. Therefore, the vehicle movement

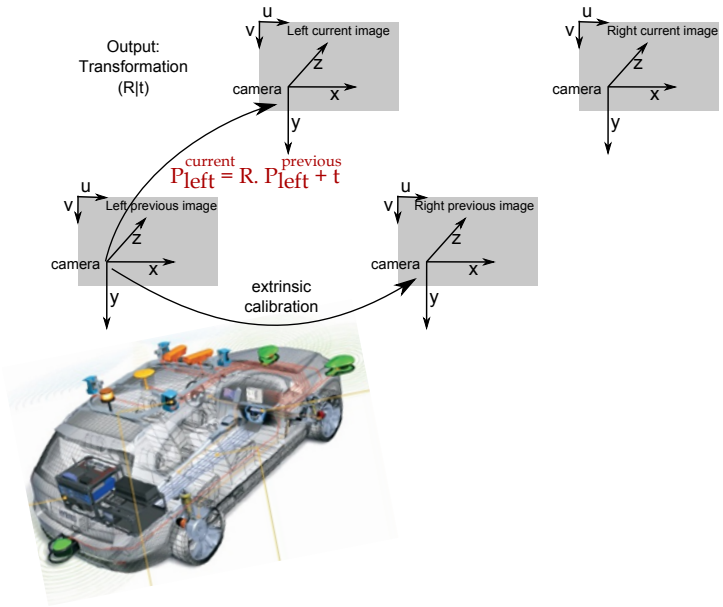


Figure 4.8 - Visual Odometry approach Geiger et al. (2011). This method uses the active feature match search strategy by leveraging the motion estimate from the last frame. This reduces the problem of ambiguities and improves performance especially in challenging scenarios.

(ego-motion) between two sequential measure sampling should be compensated. This is reached applying the rigid transformation that includes the car geometry and the disposition of camera sensors. The bilinear interpolation method is used to better fit the values among the cells of the grid. In fact, the matrix of the rotation and translation that composes the rigid transformation is obtained employing a technique of Visual Odometry based on Structure from Motion (SfM) proposed in the distinguished work of Geiger et al. (2011) and applied to this work. Figure 4.8 presents the principle to recover the rigid transformation between two pair of consecutive stereo images acquired from a calibrated stereo system.

Despite the ego-motion compensation has been done, the information contained in the previous grid is not more completely valid for the reason that the scene has changed from an instant to other. The dynamic of the scene between these two instant of time is kept fixe, increasing in this way the uncertain of the grid. An advantage of this action is that when the SG and PG are merged, a conflict between the values of a same cell will happen. These conflicting cells signalize a possible moving object in the scene. The mechanism that manages uncertainties in the form of evidences, is the same that provides this detection procedure.

A brief notion about the system architecture was established. Details about specific methods is given in the next sections. In section 4.2 is presented the formalism based on evidence theory to manage the uncertainties and also to detect the mobile cells. Section 4.3 presents the

inverse model sensor to model the SG, followed by section 4.4 that explains the complete Local Dynamic Perception proposed in this chapter.

4.2 Fundamentals of Evidential Grid

In the previous subsections are presented the principle of occupancy grids to navigation. These grids are used to estimate the occupation of space with uncertainties. How to represent these uncertainties have important implications on how to handle the information in the grid and then reasoning about the data. In this thesis, the tool to manage uncertainties associated to the occupancy grid reposes under a mathematical theory of evidence (SHAFER, 1976). Specifically, the occupancy grid uses the formalism of the Dempster-Shafer Theory (DST) to model the uncertainties, which is a generalization of the Bayesian Theory of subjective probability (DEMPSTER, 1968). The DST model associated in the Occupancy grid takes the name of Evidential Occupancy grid or then Evidential grid. As previously explained, there are some works which take advantage of evidential grids in the context of mobile perception (MORAS et al., 2011a; KURDEJ et al., 2014) and autonomous vehicles (PAGAC et al., 1998; YANG; AITKEN, 2006). The reasons for this choice are that the approach allows faster convergence (CANAS; MATELLáN, 2006), conflict detection, fusion of unreliable sources, etc (MORAS, 2013).

In the evidential grid, uncertainties are modeled as a belief function. The concerned proposition is defined by Free and Occupied, having a set composed by $\Omega = \{Occupied(O), Free(F)\}$. The frame of discernment (FOD) of Ω is the set of all possible subsets of Ω and denoted by $2^{\Omega} = \{Occupied(O), Free(F), unknown(\Omega), conflict(\emptyset)\}$ There are various forms to represent the belief function such as mass, belief, plausibility and communality, and all these representation are equivalents (MORAS, 2013). In this approach is used the mass function m^{Ω}, having the following property given by Equation (4.5):

$$\begin{aligned} m &: 2^{\Omega} \rightarrow [0...1] \\ \textstyle\sum_{A \in 2^{\Omega}} m(A) &= 1 \end{aligned} \tag{4.5}$$

The mass function of all cells of the evidential grid is a vector containing four masses that

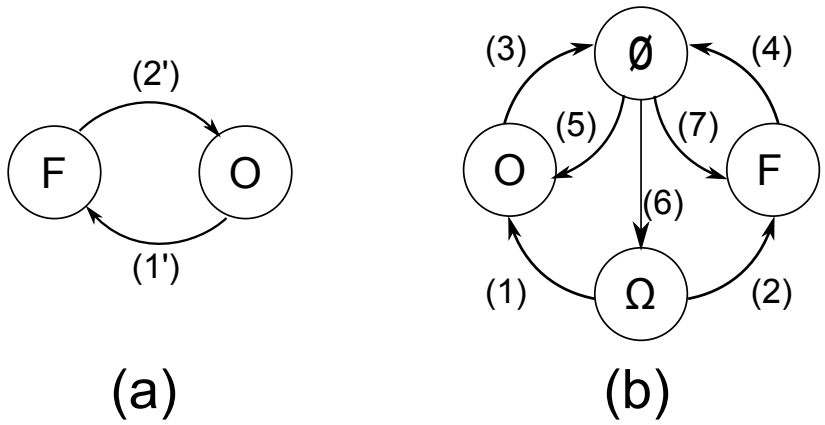


Figure 4.9 - Comparison of belief transition in Bayes's theorem and DST. (a) The belief transition in probabilistic approach. (b) In DST approach, the transitions are: fusion(1,2), conflict generation(3,4) and conflict normalization(5,6,7). Adapted source from Moras (2013)

represent the belief for each element in 2^{Ω}:

$$I_{ij}^{O} = \{\ m(F)\ m(O)\ m(\Omega)\ m(\emptyset)\ \} \tag{4.6}$$

In Equation (4.6), the mass of each element corresponds a belief level that the cell (G_{ij}) is in a given state. All cells of the grid is initialized with a mass function called Basic Belief Assignment (BBA).

Before explaining the mechanism of decision and fusion in evidential grid, an brief comparative between DST and Bayes' theorem should be mentioned. In Bayes' theorem, the probabilistic method describes the occupation of a cell using only one probability value by cell, being that the belief function requires the computation of three masses (the fourth mass is obtained by the condition of Equantion 4.5) (YAGER et al., 1994). In this case, the computation cost required by the evidential approach is higher, considering the memory usage and processing time. However, in probabilistic approach, the belief transition is possible only between the two states and symmetrically restricted. In evidential approach, the belief can be transferred among the four states, which each one of these transition has a different meaning, dynamic and importance. Figure 4.9 presents the comparison of belief transition between Bayes' theorem and DST.

As a simple example extracted of Nguyen et al. (2012), considering that one cell contains 3D points from obstacles, according to Bayes' theorem, $P(O)$ would be somewhere greater than 0.5. Let it assumes $P(O) = 0.6$. According to the DST, it has a belief mass $m(O) = 0.6$. The fewer 3D points from obstacles one cell has, the less certain it can be that the cell is occupied.

This uncertainty can be represented with $m(\Omega)$. Since there is no evidence detected that a cell is free, it has $m(\Omega) = 1.0 - m(O) = 0.4$. After Bayes' theorem, it has $P(F) = 1.0 - P(O) = 0.4$. This means that the uncertainty is automatically represented as free, which is not quite correct. Thus, these comparatives illustrate the applicability and interest of the DST approach.

In applications of autonomous vehicle, the navigation needs to take decisions based on a higher level of confidence provided by the perception module, as for example the navigable area where the vehicle could maneuver. However, evidential grid does not model directly the occupancy state of the cell, instead this estimation is given by a function that models its confidence. In this case, a decision function should be modeled to serve this purpose.

To that end, the formalism of belief function has several methods to supply this requirement, using the methodology denoted by Transferable Belief Model (TBM) described in the work of Yager et al. (1994) and Smets (1989). Accordingly with the TBM, a solution is to use the upper and lower bounds, also called belief $bel(A)$ and plausibility $pl(A)$. The belief $bel(A)$ for a set A is defined as the sum of all the masses of subsets of the set of interest (Equation 4.7), and this function states the certainty level to a given state.

$$bel(A) = \sum_{X|\emptyset \neq X \subseteq A} m(X) \tag{4.7}$$

The plausibility function $pl(A)$ states the plausibility level to a given state, and it is obtained as the sum of all the masses of the sets X that intersect the set of interest A, Equation (4.8):

$$pl(A) = \sum_{X|X \cap A \neq \emptyset} m(X) \tag{4.8}$$

These two functions represent the limits of the belief values to each state, verifying this property in Equation (4.9)

$$pl(A) = bel(\Omega) - bel(\overline{A}) = 1 - bel(\overline{A}) \tag{4.9}$$

Therefore, looking at the last example area, a decision maker can be given by Equation (4.10):

$$decision(F,th) = \begin{cases} 1 & \text{, if } pl(F) > th \\ 0 & \text{, otherwise} \end{cases} \tag{4.10}$$

In Equation (4.10), the function $pl(F)$ is given by $pl(F) = 1 - bel(\overline{F})$.

The updating procedure showed in Figure 4.14, is formalized using the Dempster's rule of combination. This fusion operator allows the merging of two independents mass function defined in the same FOD. Further, it assumes that all sources are reliable and its result leverages a more informative mass function than the two previous sources (DEMPSTER, 1968). In this sense, two reliable sources can be merged performing two steps: the conjunctive combination rule followed by the normalization of the conflicting mass function $m(\emptyset)$. In Equation 4.11, the result is denoted by $m_1 \otimes m_2(A)$, taking m_1 and m_2 as mass function of two reliable sources applying the conjunctive rule $\otimes$.

$$\begin{cases} m_{1,2}(\emptyset) & = 0 \\ m_{1,2}(A) & = (m_1 \otimes m_2)(A) = \frac{1}{1-(m_1 \otimes m_2)(\emptyset)} \sum_{B \cap C = A \neq \emptyset} m_1(B).m_2(C) \end{cases} \tag{4.11}$$

where

$$(m_1 \otimes m_2)(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B).m_2(C) \tag{4.12}$$

It is observed that the merging process using the Dempster's rule of combination works only with independent mass functions defined in the same FOD. To include the meta-knowledge extracted from Urban Road Scene Understanding, presented in Chapter 3, it is essential to extend the FOD to incorporate this information. A problem arises when the set $\Omega = \{F,O\}$ passes to represent $\Omega = \{F,O,...,C_n\}$, where in this case the FOD exponentially increase to 2^{Ω}, depending of the number of propositions C_n added in the set. This phenomenon impacts directly on the time to processing any calculus on the grid. A proposed alternative to avoid this problem, is to maintain the propositions of Free and Occupied, and then, generates a refinement in the Occupied proposition as follows.

The refinement consists in expanding a subset of propositions with respect to the Occupied

proposition. This subset is defined by $r(O) = \{V, B, T, S\}$ which denotes respectively (V) vehicle, (B) building, (T) vegetation and (S) sidewalk. It assumes that each proposition in the occupied refinement subset is represented by $\{or_i | \forall i \in r(O)\}$. It should be mentioned that the set $r(O)$ is not the same that is classified in the chapter 3, i.e, the set $r(O)$ considers only meta-knowledge defined as obstacle, avoiding in this case the classes sky, road and void. The combination rule to occupied refinement proposition, denoted by $Prop(O)$, is obtained by the ro argument of the highest mass function between the two evidences conditioned to the fact that the $argmax[m_1 \otimes m_2(A)] = O$. The fusion rule to proposition refinement is given by (4.13):

$$Prop_1 \otimes Prop_2(A) = \left\{ or \;\middle|\; \begin{array}{l} argmax[m_1(A), m_2(A)] \text{ and} \\ argmax[m_1 \otimes m_2(A)] = O \end{array} \right\} \tag{4.13}$$

After demonstrating the procedure to update two grids merging its evidences, at the moment is possible to introduce the conflict analysis used by Moras et al. (2011a) and Kurdej et al. (2014) to detect mobile cells. In fact, the conflict is determined when two information are merged. In DST, it is explicitly represented by $m(\emptyset)$. If the mass function resulting from the fusion of Dempster's rule (before the normalization) has $m(\emptyset) \neq 0$, it means that at least the merged information is partially in contradiction. According to Moras (2013), the conflict can have different causes: a difference of expert opinion or an improper system modeling. In the present case, the conflict source arises from two principal errors:

- The assumption that the grid is static, being that the observed scene contains dynamic elements.
- The geometric approximation due to the discretization in sensor model and during the grid propagation.

Considering these errors, Moras (2013) proposes in his work the decomposition of the term $(m_{t-1}^{PG} \otimes m_t^{SG})(\emptyset)$ in two others terms (Equation 4.14):

$$(m_{t-1}^{PG} \otimes m_t^{SG})(\emptyset) = \underbrace{m_{t-1}^{PG}(F).m_t^{SG}(O)}_{T_1} + \underbrace{m_{t-1}^{PG}(O).m_t^{SG}(F)}_{T_2} \tag{4.14}$$

The first term T_1 corresponds a cell that was previously free with a certain confidence level $m_{t-1}^{PG}(F)$ and, at current time t, it is observed as occupied with confidence level $m_t^{SG}(O)$.

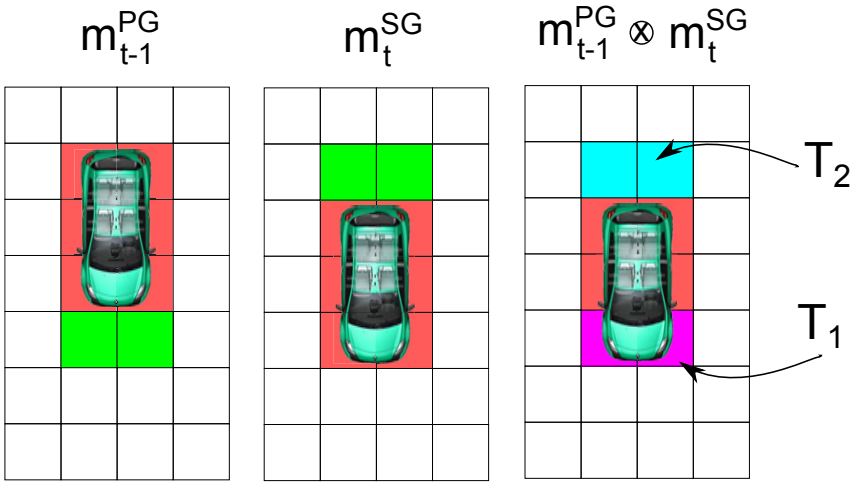


Figure 4.10 - The conflict analysis example generated by a mobile object. Red cells represent occupied area, green cells represent free area. Cyan cells represent the conflict depicted by T_2 and violet cells represent the conflict explained by T_1.

If it is considered that the conflict comes from a object in movement into the scene, the term T_1 means that a free cell becomes occupied and therefore, that an object is entering in that space represented by the cell. In the same way, the term T_2 means that an occupied cell becomes free and consequently, that an object is going out from that space represented by the cell.

Taking into account the generated conflict in the cells due to a moving object, it is possible to analyze the terms T_1 and T_2 to provide insights not only to conflict itself, but also to determine the direction and sense of an object in movement, as illustrated in Figure 4.10.

4.3 Sensor Grid Model

In this section the method to build the Sensor Grid (SG) is presented. The Sensor Grid is built at each instant that the sensor provides a new measurement. It transforms the acquired data information to its representation in the evidential grid. Therefore, it implements in some form, the sensor model used in the algorithm.

The general conception is depicted in diagram of Figure 4.11. Applying the SGBM algorithm in an stereo pair of rectified images, the disparity map I^{Δ} is obtained. Based on the Epipolar Geometry and camera calibration, 3D reconstruction is performed obtaining the 3D points with referential in the camera. After that, an affine transformation using homogeneous

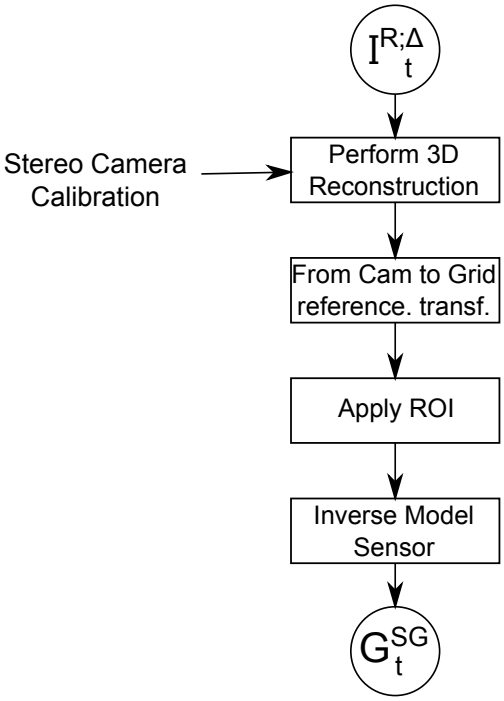


Figure 4.11 - The architecture of the Sensor Grid Model

coordinates is calculated to represent the points in the referential of the grid R_G. Due to restrictions on camera position and the grid dimensions, only 3D points that fall in this region of interest (ROI) are considered. An improvement is done in the ROI to consider also specific points that have a Semantic Context associated, observed in the Semantic Urban Road Scene Understanding and denoted by I^R. Finally, the Inverse Sensor Model using a gaussian to represent the uncertainties associated with the points is computed.

The principle of 3D reconstruction is to recover metric points from associated pixels of an rectified pair of stereo images, and also, incorporating the meta-knowledge linked with the Semantic Context. The 3D reconstruction applies the methodology explained in the book of Faugeras (1993), to get the points in the 3D cartesian space relative to the left camera, with homogeneous coordinates $[X^c, Y^c, Z^c, W]^T$. Incorporating the meta-knowledge associated with the semantic context, a 5-tuples denoted by P^c, which c represents the referential of the camera R_c, is defined containing the homogeneous 3D point and the information of the occupied refinement proposition denoted by ro. This transformation is shown in Equation (4.15):

$$P^c = \begin{bmatrix} X^c \\ Y^c \\ Z^c \\ W \\ or \end{bmatrix} = \begin{bmatrix} \frac{u.Z^c}{f_x} \\ \frac{v.Z^c}{f_y} \\ \frac{f.b}{d} \\ 1 \\ I_{ij}^R \end{bmatrix} \tag{4.15}$$

where

$$\begin{aligned} u &= i - c_x \\ v &= j - c_y \end{aligned} \tag{4.16}$$

In Equation 4.15, f,f_x,f_y represent the focal length in pixels and are obtained by the off-line calibration process. b represents the baseline of the stereo cameras (in meters). d represents the value of disparity obtained from I^{Δ}. I_{ij}^R represents the value of the semantic context at index position i,j. Finally, in Equation (4.16), c_x and c_y are the coordinates of the optical axis in the image plane.

The set points P^c should be expressed in coordinates of the grid to compute the following steps. To do this, two relations should be defined, from the referential of the camera to the referential of the vehicle and from the referential of the vehicle to the referential of the grid. The coordinate systems of these three referentials are defined as illustrated in Figure 4.12, i.e:

- **Camera**: x = right, y = down, z = forward
- **Vehicle**: x = forward, y = left, z = up
- **grid**: x = right, y = backward, z = up

The first transformation is obtained defining the position of the camera with respect to the center of the ego-car. Let that the left camera is fixed at point ${}^0P^m$ in the vehicle reference R_M. The affine transformation is built applying the translation of the point ${}^0P^m$ followed by two rotation, $\beta = -90$ degrees in the Y axis and $\alpha = 90$ degrees in the X axis. Therefore, the affine transformation from camera to vehicle is obtained by Equation (4.17):

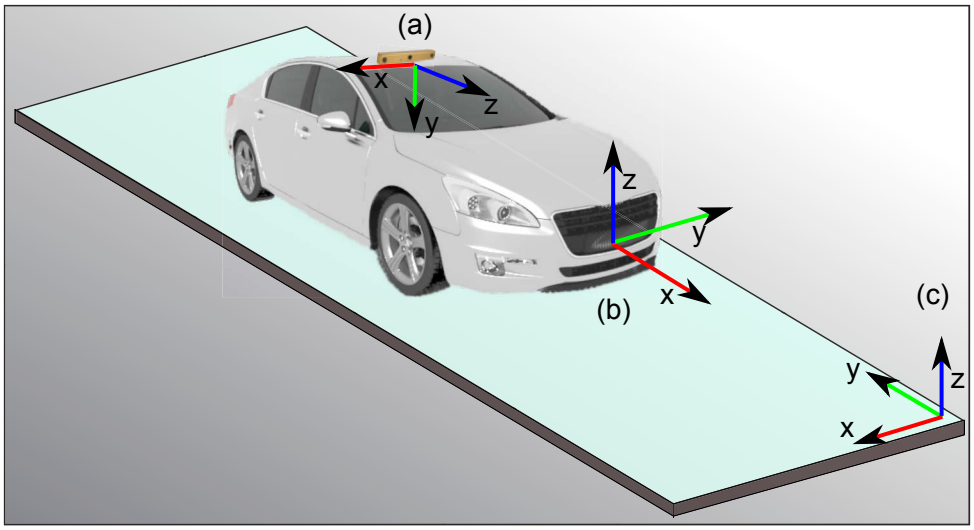


Figure 4.12 - The defined coordinate Systems. (a) Camera coordinate system, (b) vehicle coordinate system, (c) grid coordinate system.

$$M_{cam\_car} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(-\alpha) & -sin(-\alpha) & 0 \\ 0 & sin(-\alpha) & cos(-\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Rx}} * \underbrace{\begin{bmatrix} cos(-\beta) & 0 & sin(-\beta) & 0 \\ 0 & 1 & 0 & 0 \\ -sin(-\beta) & 0 & cos(-\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Ry}} * \underbrace{\begin{bmatrix} 1 & 0 & 0 & -{}^0P_x^m \\ 0 & 1 & 0 & -{}^0P_y^m \\ 0 & 0 & 1 & -{}^0P_z^m \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{T}} \tag{4.17}$$

The second transformation is obtained defining the position of the ego-car with respect to the origin of the grid. Let that the ego-car is fixed at point ${}^0P^g$ in the grid reference R_G. The affine transformation is built applying the translation of the point ${}^0P^g$ followed by a rotation of $\theta = -90$ degrees in the Z axis, then the Y axis should be inverted and finally, a scale factor considering the discretization (Δ_x,Δ_y) in the grid. Therefore, the affine transformation from vehicle to grid is obtained by Equation (4.18):

$$M_{car\_grid} = \underbrace{\begin{bmatrix} 1/\Delta_x & 0 & 0 & 0 \\ 0 & 1/\Delta_y & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{resolution factor}} * \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{invert Y axis}} * \underbrace{\begin{bmatrix} cos(-\theta) & -sin(-\theta) & 0 & 0 \\ sin(-\theta) & cos(-\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Rz}} * \underbrace{\begin{bmatrix} 1 & 0 & 0 & -{}^0P_x^g \\ 0 & 1 & 0 & -{}^0P_y^g \\ 0 & 0 & 1 & -{}^0P_z^g \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{T}} \tag{4.18}$$

To finish the affine transformation between the referential of camera and the referential of the grid, the final transformation is simply multiply these two previous matrices. So, the set points P^c is represented in the referential of the grid as being P^g and is obtained by Equation (4.19):

$$P^g = M_{car\_grid} * M_{cam\_car} * P^c \tag{4.19}$$

Due to the restrictions on camera position and the grid dimension, the set formed by all reconstructed points P_g suffer a filtering molded by a ROI. The ROI is defined considering the following restrictions:

- Assuming that the plane formed by the optical axis (Z axis) with the horizontal axis (X axis) is parallel to the road surface, a value of height from the road surface is defined where the points that transcend this threshold are not considered;
- Observing the grid dimensions, all points out of this condition: $0 \leq P^g \leq Grid_{size}$ are also discarded;
- For the remaining points inside the ROI, only those that have their semantic context associated with obstacles are considered.

So, taking into account the defined ROI, the selected set of points can be projected in the grid. Therefore, this selected set, denoted P^s, is defined by (4.20):

$$P^s = \{P^g \mid \forall P^g \subseteq ROI \text{ and } or \in r(O)\} \tag{4.20}$$

To perform the projection of P^s set in the grid, a inverse sensor model is described considering the noise in stereo measurements and also considering the uncertainty linked with stereo geometry reconstruction, where exponential error is observed when increasing the distance. This method approximate the uncertainties using a Gaussian distribution, as can be seen in Figure 4.13. The inverse sensor model defined by $\psi_O^{prob}(G^{SG}, P^s)$, has the $prob$ index representing the probability distribution, and O index representing the Occupied proposition. The function can be described by the following Equation (4.21):

$$\psi_O^{prob}(G^{SG}, P^s) = \left\{ min(\sum_{G_{ij}^{SG} \cap AG} k.exp^{(-\alpha.Dx^2 + 2\beta.Dx.Dy + \gamma.Dy^2)}, \vartheta_O) \,|\forall P_n^s \in \{P^s\} \right\} \tag{4.21}$$

where

$$\alpha = \frac{cos^2\theta}{2\sigma_x^2} + \frac{sin^2\theta}{2\sigma_y^2} \tag{4.22}$$

$$\beta = -\frac{sin2\theta}{4\sigma_x^2} + \frac{sin2\theta}{4\sigma_y^2} \tag{4.23}$$

$$\gamma = \frac{sin^2\theta}{2\sigma_x^2} + \frac{cos^2\theta}{2\sigma_y^2} \tag{4.24}$$

In Equation (4.21), κ is a constant representing the percentage that a single 3-D point could contribute to the occupancy level of a cell G_{ij}. ϑ_O is a parameter that belongs to $[0, 1]$ and reflects the confidence in the measurement (1 if confident). This confidence is linked to the principle of measurement (false alarm or miss detection). Dx and Dy are the difference between the coordinates of the P_n^s and C_{ij}, i.e, $Dx = C_{ij}.x - P_n^S.x$ and $Dy = C_{ij}.y - P_n^S.y$. The index $G_{ij}^{SG} \cap AG$ at the sum in Equation (4.21), represents the area of the gaussian AG that its distribution overlaps the cells of the grid G^{SG}. The parameters σ and θ model the dispersion of the distribution in function of the distance and its orientation with reference to ego-car. The dispersion of the gaussian considering the distance Z of the camera is modeled taking each value of disparity $\{d_i \in \{d\}\}$, as depicted in Equation (4.26):

$$\sigma_y = \left\{ \frac{\sqrt{\sum_{n \in N(d)} \left[\left(\frac{f.b}{d_i}\right) - \left(\frac{f.b}{d_n}\right) \right]^2}}{card(N(d)) - 1} | \forall d_i \in \{d\} \right\} \tag{4.25}$$

$$\sigma_x = \left\{ \frac{\sigma_y}{d_i} | \forall d_i \in \{d\} \right\} \tag{4.26}$$

In Equation (4.26), $N(d)$ represents the neighborhoods of the disparity and $card(N(d))$ represents the cardinality of the set $N(d)$. In the case of the parameter θ, it is modeled by Equation (4.27):

$$\theta = \left| arctan\left(\frac{{}^0P^g.x - P_x^s}{{}^0P^g.y - P_y^s} \right) \right| \tag{4.27}$$

The inverse sensor model proposed to project the meta-knowledge information in the grid is based on the principle of voting. This method assumes that the better semantic information which represents the cell is defined by the sum of votes that a given meta-knowledge received from those points belonging to the cell. Therefore, to model the occupied refinement subset with proposition $\{or \in r(O)\}$, in this case denoted as $\psi_O^{prop}(G^{SG}, P^s)$, is described by (4.28):

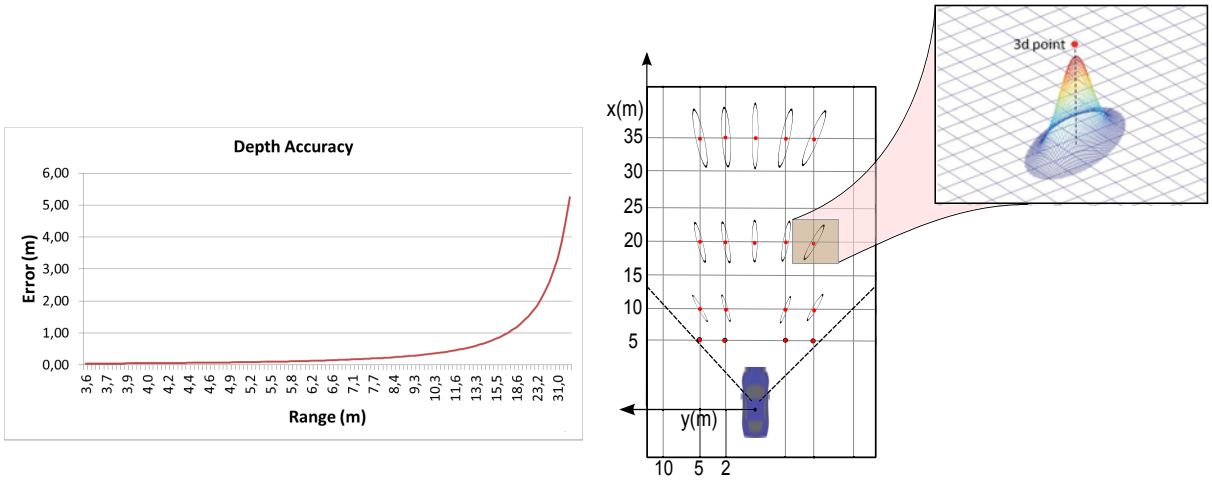


Figure 4.13 - The model of uncertainty associated with stereo geometry reconstruction and noise in stereo measurements.

$$\psi_O^{prop}(G^{SG}, P^s) = \left\{ \underset{ro}{\text{argmax}}(\omega(G_{ij}^{SG},P^s))|\forall G_{ij} \in G^{SG} \right\} \tag{4.28}$$

where

$$\omega(G_{ij}^{SG},P^s) = \left\{ \sum_{P_n^s \subseteq S_{ij}^{SG}} \delta(P_n^s.or,or_l)|\forall or_l \in r(O) \right\} \tag{4.29}$$

$$\delta(or,or_l) = \begin{cases} 1 & \text{, if } or = or_l \\ 0 & \text{, otherwise} \end{cases} \tag{4.30}$$

In Equation (4.29), the index $P_n^s \subseteq S_{ij}^{SG}$ represents all 3D points $P_n^s \in \{P^s\}$ that are contained in the surface of the cell S_{ij}^{SG}.

Until this point, the solution is able to manage the probability of occupied areas using 3D points that belong to obstacles. It is also able to perform the occupied refinement which determines the best proposition that represents those occupied areas. Based on the principle that the obstacles were already processed, a simple but effective method which model the free areas is applied. The objective of this method is that, if it exist a light ray from camera sensor to a detected obstacle point, it can be stated with a given probability that every cell that lies along this line is free. So, the solution for free areas, defined by $\psi_F^{prob}(G^{SG},FL)$, which $\{FL\}$ denotes the set of Free Lines, is modeled by a function that attributes the free probability for all cells which intercept the line generated from camera sensor position to all first obstacle detected. This technique is done by the Bresenham algorithm (BRESENHAM, 1965), and described by

Equation (4.31):

$$\psi_F^{prob}(G^{SG},FL) = \left\{ \max_{fl \cap G_{ij}^{SG}} (1,0 - \vartheta_F)) | \forall fl \in \{FL\} \right\} \tag{4.31}$$

In Equation (4.31), ϑ_F is a parameter that belongs to $[0,1]$ and reflects the confidence in the measurement of the Free area (0 if confident). As previously explained, this confidence is linked to the principle of measurement (false alarm or miss detection).

To conclude the inverse sensor model for all states, the function that models the unknown state (Ω) should respect the property presented in Equation (4.5), and is defined by the following Equation (4.32):

$$\psi_\Omega^{prob}(G^{SG}) = 1.0 - \psi_O^{prob}(G^{SG},P^s) - \psi_F^{prob}(G^{SG},FL) \tag{4.32}$$

Therefore,the resulting evidential grid modeled to represent the sensor grid at each instant of a measurement, has its BBA defined as Equation (4.33) :

$$\begin{aligned} m^{SG}(O) &= \psi_O^{prob}(G^{SG}, P^s) \\ m^{SG}(F) &= \psi_F^{prob}(G^{SG}, FL) \\ m^{SG}(\Omega) &= \psi_\Omega^{prob}(G^{SG}) \\ m^{SG}(\emptyset) &= 0 \\ Prop^{SG}(O) &= \psi_O^{prop}(G^{SG}, P^s) \end{aligned} \tag{4.33}$$

4.4 Dynamic Local Perception Grid

As introduced in previous sections, the *Perception grid* (PG) is responsible for building the final representation of the environment along the time. The complete Dynamic Local Perception (DLP) proposed in this work is a new conception involving the Semantic Urban Road Scene Understanding with Occupancy grids. Its architecture is detailed in Figure 4.14. At each instant t that a measure is acquired, the meta-knowledge I_t^R representing the Urban Road Scene Understanding is archived. By simplicity, this process is renamed to Semantic Context. The I_t^R

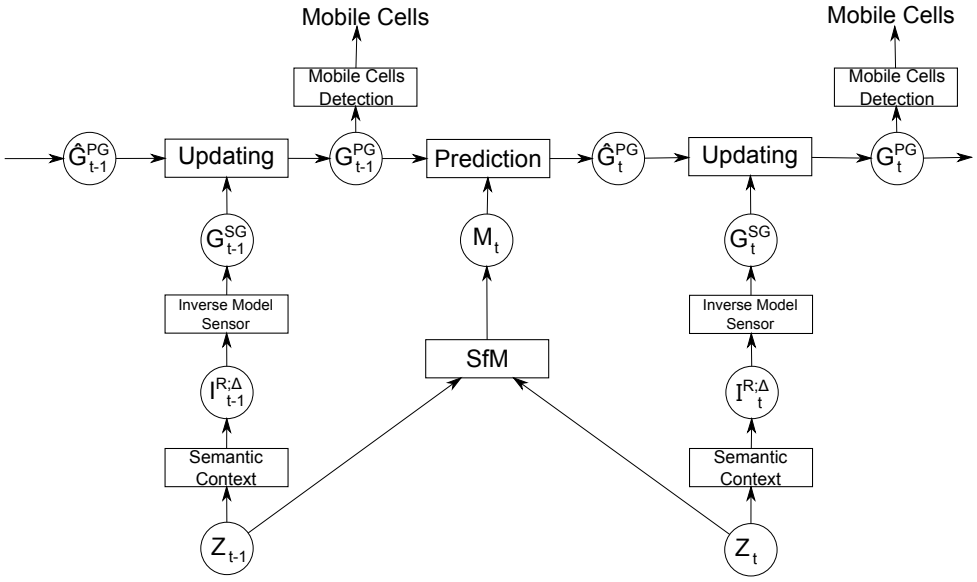


Figure 4.14 - Detailed system architecture proposed to Dynamic Local Perception.

jointly with the Disparity Image I_t^{Δ} supply the necessary information to construct the G_t^{SG}, performing the novel Inverse Sensor Model. Following, the prediction of the $\hat{G}_t^{PG}$ is estimated and then the G_t^{SG} is updated with this estimation using the methodology based on DST to manage the uncertainties of the system. Inside the updating procedure, it is possible to detect the mobile cells based on G_t^{PG}.

Launched the conception of the DLP, the details of the system architecture are presented. The Semantic Context is presented in Chapter 3. The Inverse Sensor Model is depicted in section 4.3.

The prediction process estimates the grid G^{PG} to $\hat{G}^{PG}$ in function of the displacement generated by the ego-car at instant $t-1$ to t. The rigid transformation ($M_t = [R|T]$) that represents this displacement is performed in two consecutive images, as explained in section 4.1.3. Thus, $G_{t-1}^{PG} \rightarrow G_t^{PG}$ uses the affine transformation function $f(G_{t-1}^{PG}, M_t)$ to evolute the cell information, given by Equation (4.34).

$$\begin{aligned} \hat{m}_t^{PG} &= B(f(m_{t-1}^{PG}, M_t)) \\ \hat{Prop}_t^{PG} &= \varpi(f(Prop_{t-1}^{PG}, M_t)) \end{aligned} \tag{4.34}$$

where

$$\varpi(f) = \{max(\sum_{x=N_{cell}} \delta(or_j, or_x)) \mid \forall j \in r(O)\} \tag{4.35}$$

In Equation (4.34), the function $B(.)$ applies the bilinear interpolation in the mass function, and the function $\varpi(.)$ realizes the same process as $B(.)$, but on occupied refinement proposition. Following Equation (4.35), the N_{cell} stands for all neighbors of the cell on grid, and $\delta(.)$ is defined in Equation (4.30). In this process, there are some cells that disappear and other cells that come within the scope of the new grid. These new cells are initialized with the unknown mass function ($\Omega = 1.0$).

When the $\hat{G}_t^{PG}$ is computed, the fusion process with G_t^{SG} can be done. Each cell refers to an occupancy mass function defined on 2^{Ω} plus the refinement $r(O)$, shown in Equations (4.11) and (4.13) of section 4.2. The mass function m_t^{PG} at time $t = 0$, has its values representing no prior information (4.36):

$$\begin{aligned} m_t^{PG}(O) &= 0.0 \\ m_t^{PG}(F) &= 0.0 \\ m_t^{PG}(\Omega) &= 1.0 \\ m_t^{PG}(\emptyset) &= 0.0 \\ Prop_t^{PG}(O) &= (\{\}) \end{aligned} \tag{4.36}$$

However, the updating mechanism is achieved in two steps in order to keep the conflicting information and also to combine the proposition refinement. The first step, the fusion process uses the Equation (4.11) without the normalization factor, to merge the mass function. The Equation (4.13) is applied to get de associated proposition, as demonstrated in Equation (4.37).

$$\begin{aligned} m'^{PG}_t &= \hat{m}_t^{PG} \otimes m_t^{SG} \\ Prop_t^{PG} &= \hat{Prop}_t^{PG} \otimes Prop_t^{SG} \end{aligned} \tag{4.37}$$

In Equation (4.37), m'^{PG}_t is the conjuctive fusion, i.e, the Dempster's rule without the normalization factor. The second step, the updating process is done normalizing the mass function by the conflict mass, as shown in Equation (4.38). It should be mentioned that the conflict mass $m'^{PG}_t(\emptyset)$ is stored for mobile detection analysis.

$$\begin{cases} m_t^{PG}(A) = \frac{{m'}_t^{PG}(A)}{1-{m'}_t^{PG}(\emptyset)} & A \neq \emptyset \\ m_t^{PG}(\emptyset) = 0 & A = \emptyset \end{cases} \tag{4.38}$$

The detection of a mobile object is reached by analyzing the conflict mass ${m'}_t^{PG}$. If G_{t-1}^{PG} and G_t^{SG} contradicts themselves, it means that a conflict occurred and it can be analyzed following the Equation (4.14). As previously explained, the first term T_1 detects the generated conflict when a moving object leaves the cell and the second term T_2 detects the generated conflict when a moving object appears in the cell. Due to noises and imprecise measurements arising from data acquisition, bad displacement estimation, etc, many false-positive detections may appear. In this case, using the meta-knowledge associated in the G_t^{PG}, the detection of mobile obstacles can be improved by restricting that only the $r(O) \supset V$ could generate this conflict.

$${m'}_t^{PGr}(\emptyset) = \{{m'}_{ij}^{PG} | Prop_t^{PG}(O) \subseteq \{V\}\} \tag{4.39}$$

4.5 Experimental Results

In this section are presented experiments using two dataset acquired in urban conditions. The first one is a continuation of the outcomes presented in the previous chapter 3, using the Kitti benchmark depicted in section 3.5.2. In this dataset, the used sensor camera is characterized by two PointGrey Flea2 color cameras model FL2-14S3C-C, with focal length of 4 mm and horizontal opening angle of ∼90 degrees. The baseline of the stereo camera rigs is approximately 54 cm. The left calibrated camera has the principal point in $c_x = 609.5593$ px. and $c_y = 172.8540$ px. as well as the focal length $f_x = f_y = 721.5377$ px. The mounting positions of the sensors with respect to the vehicle body is demonstrated in Figure 4.15(a), given in the work of Geiger et al. (2013).

The second dataset was acquired at the test track of the Heudiasyc Laboratory using a moving platform called ZOE. This dataset is composed of 473 sequential images with resolution of 640x480. The sensor camera used in this case is characterized by a PointGrey Bumblebee x3 color model BBX3-13S2C-60, with focal length of 6mm and horizontal opening angle of ∼43

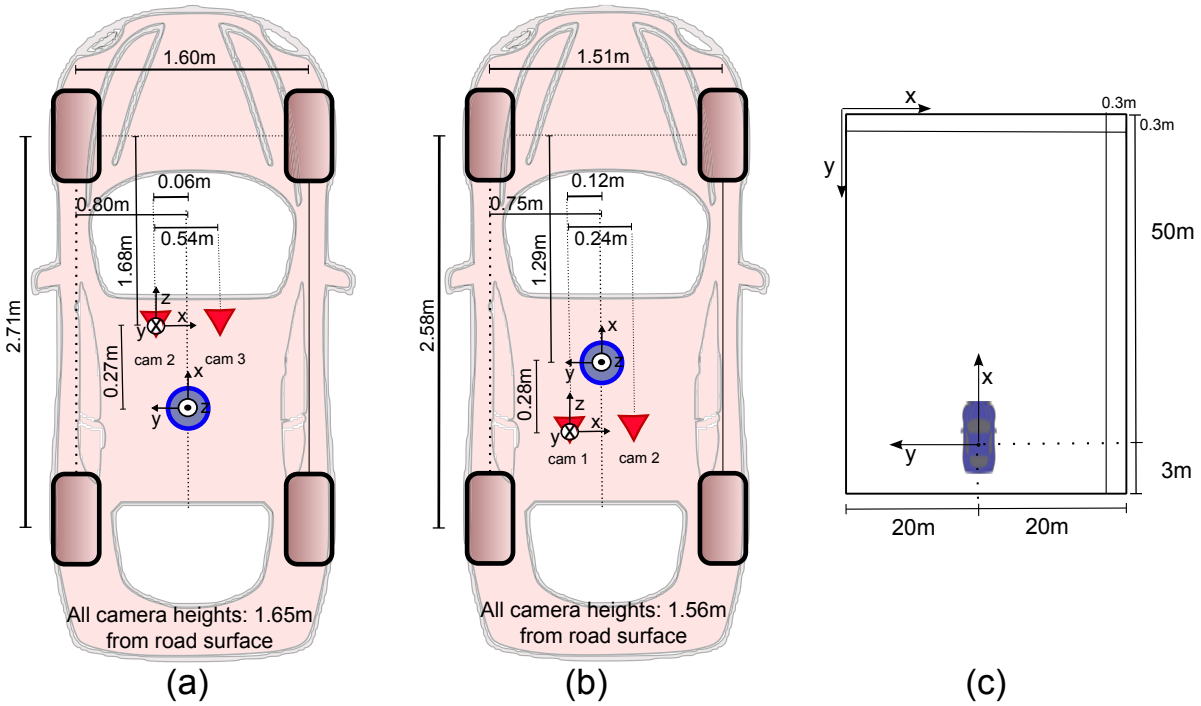


Figure 4.15 - Sensor Setup. This figure illustrates the dimensions and mounting positions of the sensors (red) with respect to the vehicle body. Heights above ground are measured with respect to the road surface. (a) Using the KITTI Benchmark (GEIGER et al., 2013) (b) The ZOE platform where the second dataset was acquired, at Heudiasyc Laboratory (c) Local Perception Grid setup.

degrees. The baseline of the stereo camera is approximately 24 cm. The calibrated left camera has the principal point in $c_x = 325.6$ px. and $c_y = 244.90464$ px. as well as the focal length $f_x = 807.7376$ px. and $f_y = 807.7392$ px. The mounting position of the sensor with respect to the vehicle body considers the same disposition as the previous one, but attempting to horizontal angle of vision that in this case is very restrict. It is demonstrated in Figure 4.15(b).

For both datasets, the grids are defined to cover an area of 39.9m x 53.1m with resolution of 0.3m x 0.3m. The center of the ego-car was positioned on the grids with coordinates (20m, 50m), remembering that the reference is fixed on left-up of the grid. The transformations between referential were modeled considering the car geometry and disposition of sensor cameras shown before.

Validation and performance analysis of this kind of perception systems is quite difficult to be archived. Moras (2013) proposed a scheme to perform the validation of results. As that work used a scanner laser rangefinder, its methodology allows a qualitative validation by projecting the estimated local grid as being a transparency in the image, at the same instant of time that the laser sensor acquired the measurement. So, the author argues that it enables a dense visual

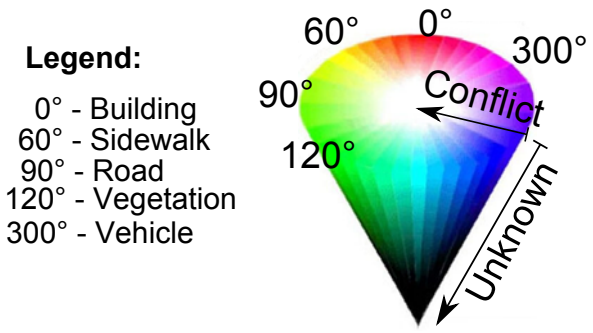


Figure 4.16 - The proposed visualization method using the HSV color space.

analysis. Therefore, the accuracy of the approach was analyzed by correspondence between the results obtained and the observed scene. It is noted that the hidden cells can not be evaluated in this way. Thus, the validation analysis performed in this work considers the same principle, although, instead of the grid to be built by the scanner laser rangefinder, it is built by stereo images.

A simple method is created in order to better visualize the various information contained in the DLP. This method proposes to give a different meaning for the HSV color space. The idea is to represent the four states of the evidential grid considering also the semantic context such as building, sidewalk, road, vegetation and vehicle. To do this, the colors in the Hue axis which ranges from 0 - 360 degrees is changed to represent the semantic context. Consequently, it is defined a fixed value of degree to each class. Therefore, 0°, 60°, 90°, 120° and 300° represent respectively building, sidewalk, road, vegetation and vehicle. The set of degrees {0°, 60°, 120°, 300°} belongs to occupied state (O) and the degree 90° belongs to free state (F). Following, the Saturation axis that ranges from 0 - 100% is changed to represent the conflict state ($\emptyset$), but inverting the values of the axis, i.e, the value of Saturation $S = 1 - m(\emptyset)$. It means that the higher is the conflict, the closest of the white color it will be. To finish the proposed conception of visualization, the unknown state (Ω) is presented by the Value axis, modeled in the same form of conflict state. It means that the complete ignorance is represented by the black color. In this sense, the variations of the colors are proportional to the value mass of these respective states. Figure 4.16 presents the HSV color space with its axis representing the arranged visualization.

To illustrate the results of the method, the following figures show the original left image, the semantic context result where the meta-knowledge is based and the final outcome of the DLP. Supplementary, the disparity map is added to contextualize the metric information in the grid. It should be mentioned that all semantic context results is generated by the ProbBoost

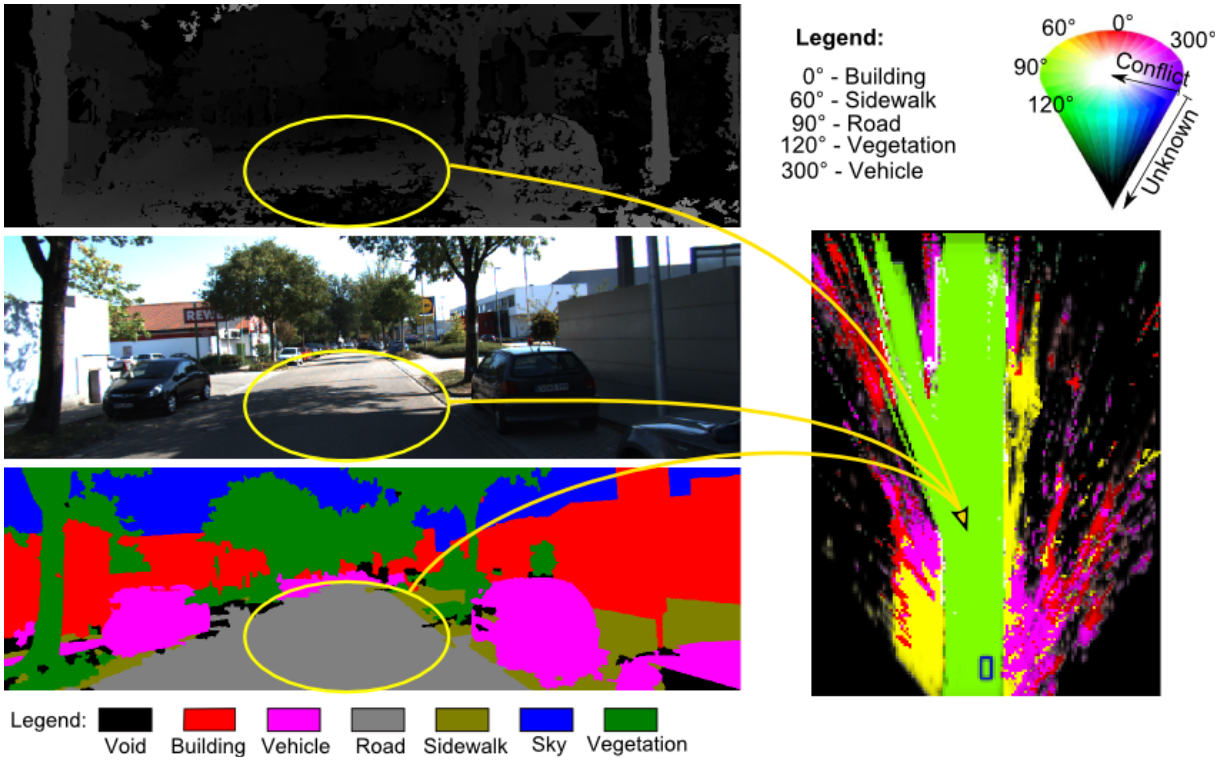


Figure 4.17 - DLP result highlighting the road detection. The yellow circle highlights the presence of shadow area jointly with higher influence of the sun, and more, the sidewalk that is quite similar to the road.

algorithm, due to the solid performance presented in the previous chapter.

In the first example of Figure 4.17 is presented the robustness to the detection of road, the principal element to perform the autonomous navigation. As can be seen, this figure shows a typical scenario frequently found in urban area. The yellow circle highlights a problem that a perception system using camera sensor should manage. The presence of shadow area jointly with higher influence of the sun provide a challenging task to be treated, considering also the sidewalk that is quite similar to the road. Taking into account these mentioned factors, the DLP system is able to maintain a high confidence about the free space without using any other sensor or prior digital map to build this perception.

In the example of Figure 4.18 is demonstrated the conflicting cells that are able to detect a mobile object. In this first case represented by number 1, a moving car is passing by the ego-car, also in movement. At this instant, the cells in white represent clearly a object in movement. Although, due to inaccuracies in the estimation of the ego-motion and errors produced by the phenomenon of discretization and transformation between grids, some wrong conflicting cells arise, as emphasized in the second case. In this case, using the semantic context to improve the detection (as presented by Equation 4.39), it is possible to distinguish these two kinds of

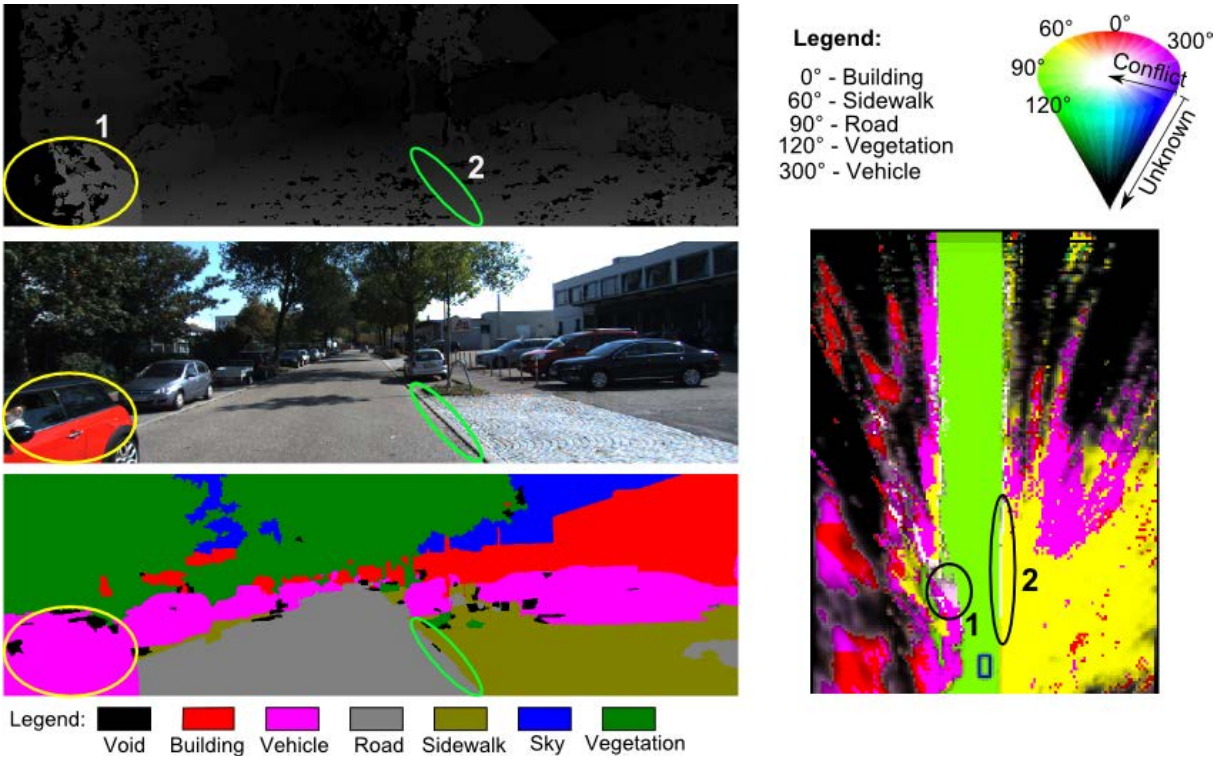


Figure 4.18 - DLP result highlighting the distinction between mobile and static cells. The yellow ellipse highlights the presence of a mobile object, and the green ellipse show the conflicting cells that can be filter out considering the meta-knowledge of the scene.

conflict considering that only cells recognized as vehicle could be in movement. The resulting output of the DLP presents the new conception to improve substantially the representation and understanding of dynamic urban environments.

The example illustrated in Figure 4.19 demonstrates the multi-detection of mobile objects in a challenging complex scenario. In this environment, the DLP system presents an outstanding approach, detecting all vehicles in the scene. As can be seen in the cases 1, 2 and 4, they are correctly detected as mobile vehicles. However two wrong cases should be mentioned. Analyzing the third case, the semantic context result correctly detected the occluded vehicle but it was not able to detect the vertical signs before the car. At the same time, looking the disparity map result, it is possible to see that the distance relation from the vertical signs to stereo camera was performed correctly. Consequently, the projection into the Perception Grid correctly maps the position of object but associating a wrong meta-knowledge, and therefore, impacting in the process of filter out the conflicting noise cells. To finish this example, the case represented by number 5 demonstrates an important factor of risk that was not recognized. At the moment, the proposed system is not able to deal with pedestrian recognition. In this case, same that it was detected as being an obstacle badly classified, its occupied area in the grid is not adequate to represent this kind of obstacle (as discretized), due to each cell to represent a space of 0.3x0.3m,

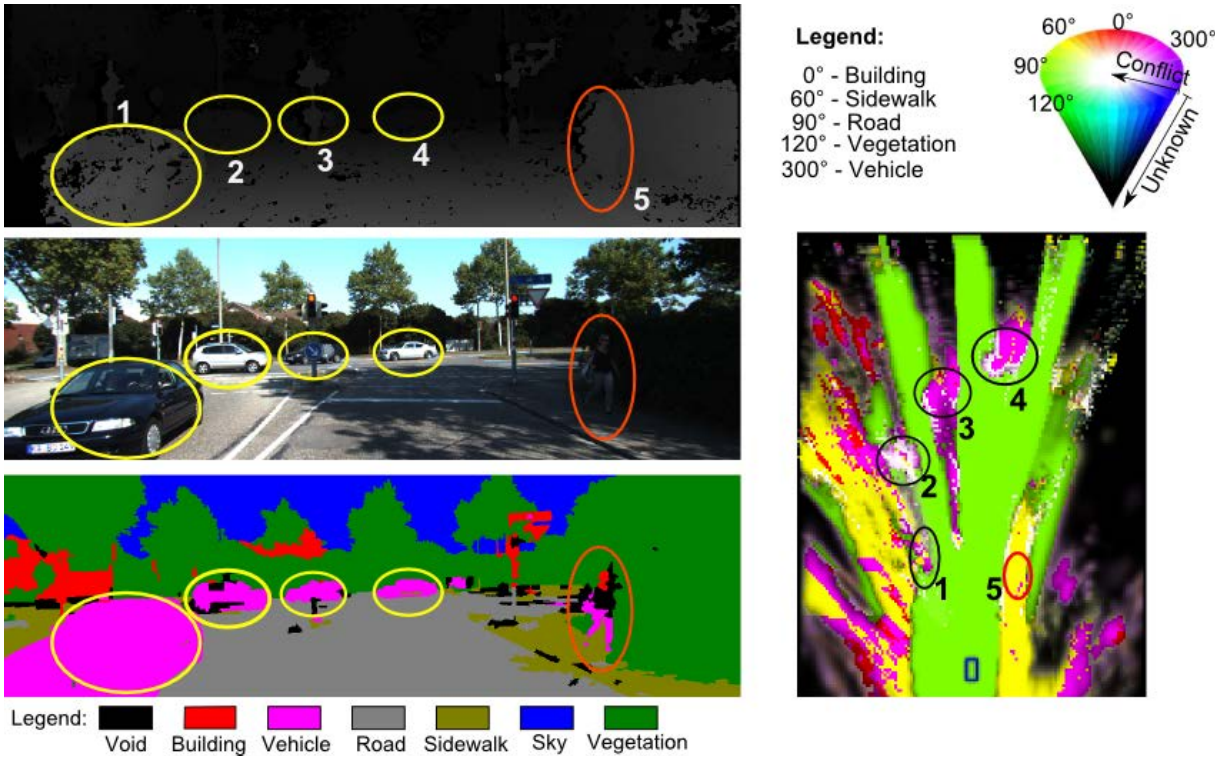


Figure 4.19 - DLP result highlighting the multi-detection of mobile objects. The yellow ellipse highlights the presence of mobile vehicles, and the red ellipse show the wrong case to pedestrian recognition.

and its projection is represented by only 2 or 3 cells, becoming hard its discrimination considering noises.

The last example is illustrated by Figure 4.20 using the dataset of the Heudiasyc Laboratory in the test track. The idea here is to show some restrictions when changing completely the experimental platform. As can be seen, the DLP is quite similar as obtained with the previous dataset, but, in this platform the sensor camera has a limitation with respect to the Horizontal Field of View (HFOV). This factor strongly affects the perception around the ego-car. A consequence of this restriction is that the road is not observed in sharp curves, becoming a hard task to the navigable process to generate a reliable trajectory.

The videos containing the complete result of the DLP for both experiments are available. The first one using the sequence of images from Kitti can be found in Vitor (2014d) [2], and the second one using the sequence of images from Heudiasyc can be seen in Vitor (2014c) [3].

[2] <http://youtu.be/H_zJjX8uMtI> Accessed: 20/07/2014
[3] <http://youtu.be/chLqC1r974k> Accessed: 04/08/2014

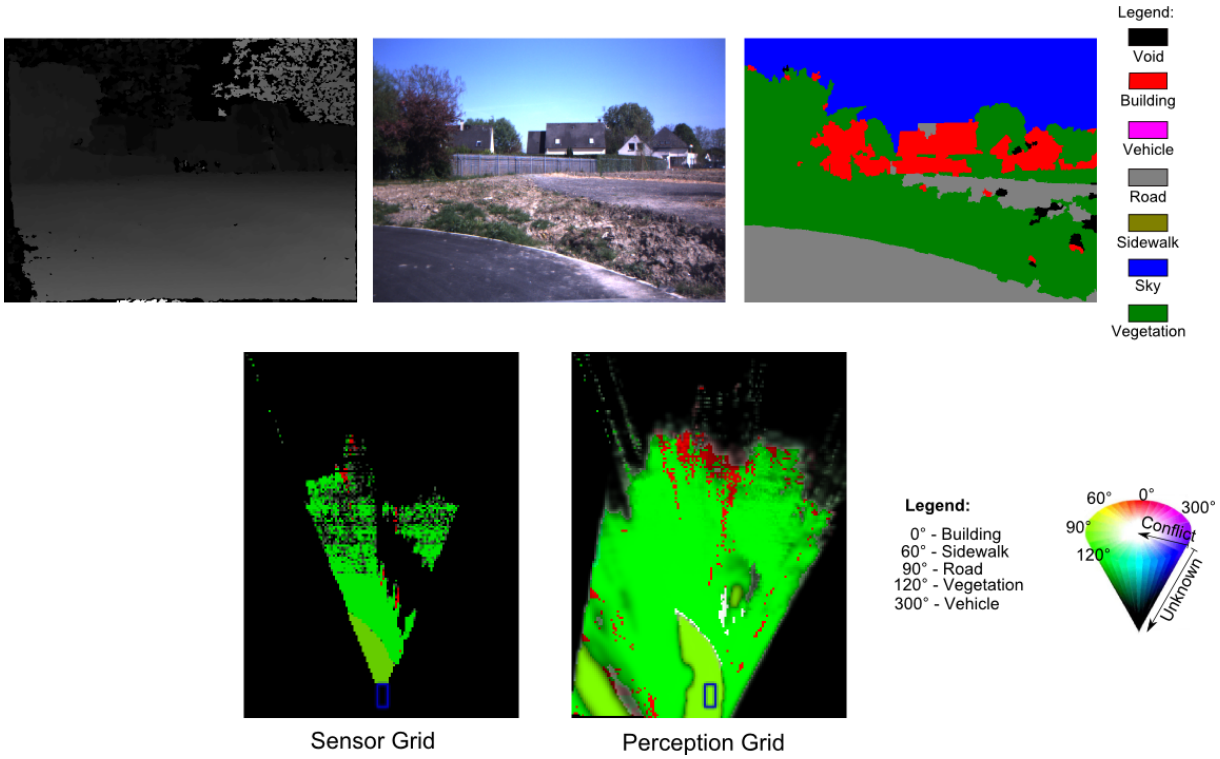


Figure 4.20 - DLP result highlighting the HFOV of the Sensor camera from ZOE platform at Heudiasic laboratory. The restricted HFOV of the camera limits the perception system to detect the road with sharp curves,becoming a hard task to the navigable process to generate a reliable trajectory.

4.6 Conclusion

In this chapter, a new perception scheme based on dynamic mapping and relative localization using only a pair of stereo cameras has been introduced. The advantages of using a stereo camera is, in addition to the possibility to measure distances, the availability of image information. Therefore, the proposed approach, denominated by Dynamic Local Perception, combines the evidential occupancy grid with the meta-knowledge molded by a machine learning to characterize the uncertainties of the occupied areas and incorporating also the semantic context, associated with these areas to improve the representation of dynamic urban environment along the time.

In synthesis, this chapter contributes to this line of research by offering a novel technique that does not require inertial sensors, laser sensors and also a prior digital mapping to leverage a robust system of perception. Further, a new inverse sensor model that considers uncertainties in distance measurements and improves the occupancy grid with associated meta-knowledge. At the end, working with the DST, the prediction and updating processes are modeled to combine

semantic context to discriminate static and mobile objects in the scene, making this solution a promising approach towards urban scene understanding.

CHAPTER 5

Autonomous Navigation

I am enough of an artist to draw freely upon my imagination. Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world

(Albert Einstein)

In this chapter the DLP is used to perform the autonomous navigation in urban environment. The proposed navigation adopts an hybrid approach. The route planning uses a topological representation to generate the complete path planning, and the path-planning uses the local metric representation to perform the precise navigation. Whilst the actual approach considers a local navigation task, it gives a framework to bind the route planning in the context of path-planning, performing in this way a deliberative-reactive navigation. Further, it is detailed the experimental platform as well as the complete automotive system integrating the proposed perception, planning and control modules, using the PACPUS framework embedded in the ZOE platform of the Heudiasyc Laboratory. Therefore, it starts by discussing the conception of the hybrid navigation approach in section 5.1. The kinematic model of the vehicle is then presented in section 5.2, followed by the path-planning using virtual tentacles (section 5.3) and its basic control (section 5.4). After that, the experimental platform is introduced in section 5.5, embedding the proposed system detailed in section 5.6. Finally, it is demonstrated the results for the path-planning as well as the autonomous navigation experiment carried out in the test track of the laboratory. A brief overview of the system is illustrated in Figure 5.1.

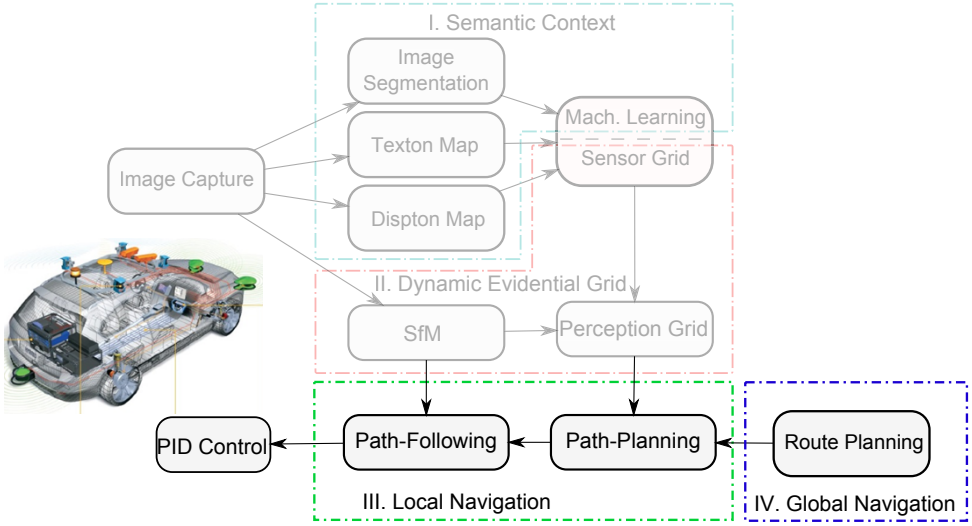


Figure 5.1 - The proposed solution to path-planning used in the autonomous navigation.

5.1 Conception of the proposed hybrid navigation

The proposed navigation has an hybrid approach. This hybrid approach mimics the conception like humans may drive in urban environment. A topological representation is used to perform the route planning, which is usually run as a planning phase before the robot-like car begins its journey. A metric representation is employed to local navigation, which is more appropriated to reactive issues and is more robust to localization errors.

The topological principle operates on landmarks and identifiable locations such as intersections. Its directions are modeled by some terms such as “go ahead”, “Take the 1st left onto Solferino avenue”, etc. Usually, its representation is based on graph theory, where nodes of this graph could be associated to landmarks and connected edges between these nodes representing the relative path to reach the final objective. In this sense, landmarks may be represented by waypoints of a global positioning by satellites or then by a principle of image memory. In both cases, the requirement to accomplish a navigation considering the present scope, is to model landmarks as being intersections in the environment. For example using image memory, one or a set of images may be used to identify if the vehicle is close to a intersection. Similarly using waypoints of a common GPS for example, the idea is to check if the local navigation is approaching an intersection and then to inform the direction from that current situation which the vehicle must follow.

An example regarding this concept is illustrated in Figure 5.2. As can be seen, the route

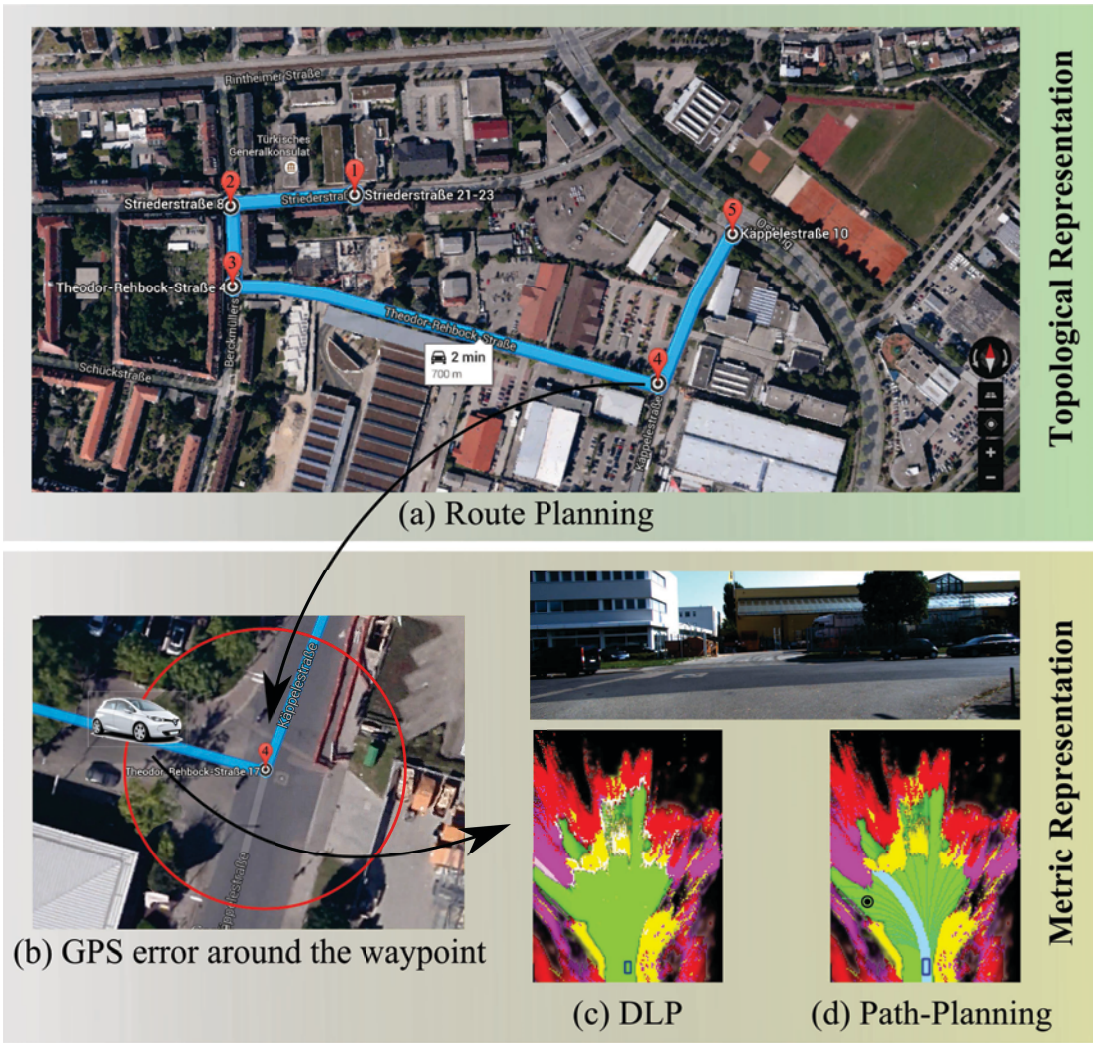


Figure 5.2 - Example of an hybrid navigation conception.

planning in Figure 5.2(a) uses the topological representation to model a graph containing five landmarks. In this specific example, these landmarks are associated with waypoints of the GPS fixed at the intersections. Considering that the vehicle's position is oriented in direction of the path and near from waypoint 1. The route planning generates the list of waypoints representing the graph, which each node has the attributes of latitude, longitude and direction, as depicted in Table 5.1. From this point, the local navigation is actioned. Assume in a given instant that the vehicle is navigating in the path between the nodes 3 to 4. At the moment that the system detects its approximation from the waypoint 4, independent of the position's error leveraged by this kind of sensor (as illustrated in Figure 5.2(b)), the local navigation knows that it should turn left. Analysing the local scenario of this intersection, and observed by the point of view of the vehicle, the proposed DLP represents the metric space of the environment around the vehicle exactly at this intersection, as demonstrated in Figure 5.2(c). It provides an interesting way to model the path-planning taking into account the information produced by the route planning

(i.e. Turn left).

Table 5.1 - Example of waypoints from topological route planning shown in Figure 5.2

Waypoint	Latitude	Longitude	Direction
1	49.010411	8.436024599999996	go ahead
2	49.0103407	8.434537999999975	turn left
3	49.0096828	8.434577600000011	turn left
4	49.0088916	8.439693799999986	turn left
5	49.0101118	8.440546100000006	stop

In this case, as the information has a semantic meaning, this work introduces a simple but effective form to convert this qualitative data into a quantitative measuring without losing the representativeness. Since all urban constraints are related in the local scenario, a point is created to represent the semantic meaning and attempting those urban constraints such as forcing the path-planning to generate a path that maintains the vehicle at the right side of the street for example, or then driving the path during a turn. This point is called *mutant point* because it changes its representativeness according to a given criteria. Figure 5.2(c) shown the result of this approach. Using the virtual tentacle is possible to apply the mutant point to ponder the elaboration of the path tending to turn left as can be seen in the example.

Although the procedure of route planning still was not implemented in this work, its principle was explained as support to contextualize the local navigation and their interaction. Introduced this conception of hybrid navigation, the next sections concern is how the path-planning models and selects the appropriated command and path to follow, considering geometric restrictions of a nonholonomic robot-like car, avoiding obstacles and contemplating the target task.

5.2 Kinematic Model of the vehicle

The vehicle used in the experiments is modeled using a front-wheel-steering that takes into account the Ackerman condition, as illustrated in Figure 5.3. The kinematic model is defined by Equation (5.1):

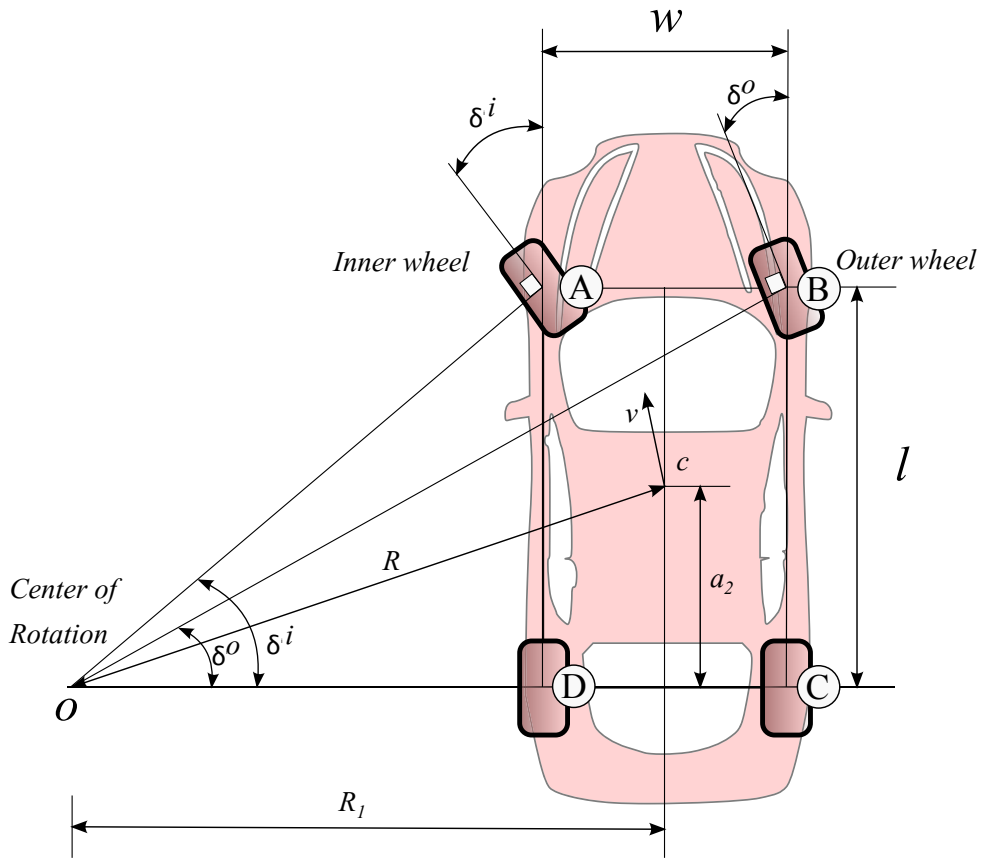


Figure 5.3 - A front-wheel-steering kinematic vehicle and steer angles of the inner and outer wheels.

$$\begin{cases} \dot{x} = V\cos(\psi) \\ \dot{y} = V\sin(\psi) \\ \dot{\psi} = \frac{V}{l}\tan(\delta) \end{cases} \tag{5.1}$$

where (x,y) are the cartesian coordinates, V corresponds to the longitudinal velocity of the vehicle, $\dot{\psi}$ is the angular velocity, ψ the orientation, δ is the steering angle of the wheels and l is the length between the axis of the wheels called of the *wheelbase* (JAZAR, 2008).

The method based on virtual tentacles projects arcs of circle considering the vehicle's geometry. Analysing Figure 5.3, the required parameters of the vehicle's geometry can be defined to give the basis of the tentacles generation in subsection 5.3.1. From the triangles ΔOAD and ΔOBC is possible to calculate the inner and outer steer angles δ^i and δ^o, as given by Equation (5.2):

$$\begin{aligned} \tan(\delta^i) &= \frac{l}{R_1 - \frac{w}{2}} \\ \tan(\delta^o) &= \frac{l}{R_1 + \frac{w}{2}} \end{aligned} \tag{5.2}$$

where

$$\begin{aligned} R_1 &= \frac{1}{2}w + \frac{l}{tan(\delta^i)} \\ &= -\frac{1}{2}w + \frac{l}{tan(\delta^o)} \end{aligned} \tag{5.3}$$

To find the vehicle's turning radius R, the bicycle model is considered. The radius of rotation R is perpendicular to the vehicle's velocity vector v at the mass center C, being then described by Equations (5.5):

$$\begin{aligned} R^2 &= a_2^2 + R_1^2 \\ cot(\delta) &= \frac{R_1}{l} \\ &= \frac{1}{2}(\cot(\delta^i) + \cot(\delta^o)) \end{aligned} \tag{5.4}$$

Therefore:

$$R = \sqrt{a_2^2 + l^2 cot^2(\delta)} \tag{5.5}$$

5.3 Path-planning approach

This section presents the detailed strategy for the path-planning. The proposed architecture is composed by three principal blocks, as can be seen in Figure 5.4. The path generation block is responsible to perform the requirement space of the vehicle during a turn, providing the representative set of cells in the grid (denoted by A), related to these areas. As it uses a ego-referenced grid, this process is calculated only at the initialization and considers a discretized set of steering angles δ that is represented by those virtual tentacles. During the navigation, the set of cells representing the intersection of obstacle with the area of the vehicle's path is obtained by the obstacle detection block, using the current DLP (G_t^{PG}) and defined by $OP_t = \{A \cap O | \forall O \in r(O)\}$. After that, in the tentacle selection block, a linear system is applied to extract the best tentacle that represents the longitudinal velocity and the steering angle of the wheels, denoted by $T_t^b = \{V_t, \delta_t\}$, where the complete set of tentacles is defined by $\{T\}$.

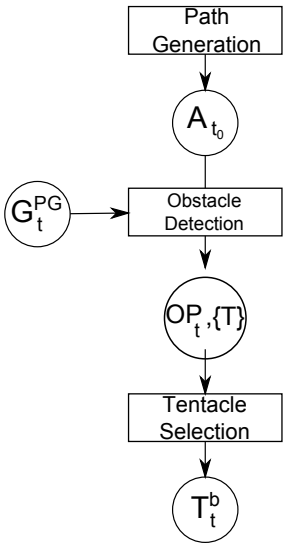


Figure 5.4 - The proposed architecture for path-planning.

5.3.1 Tentacle structure and generation

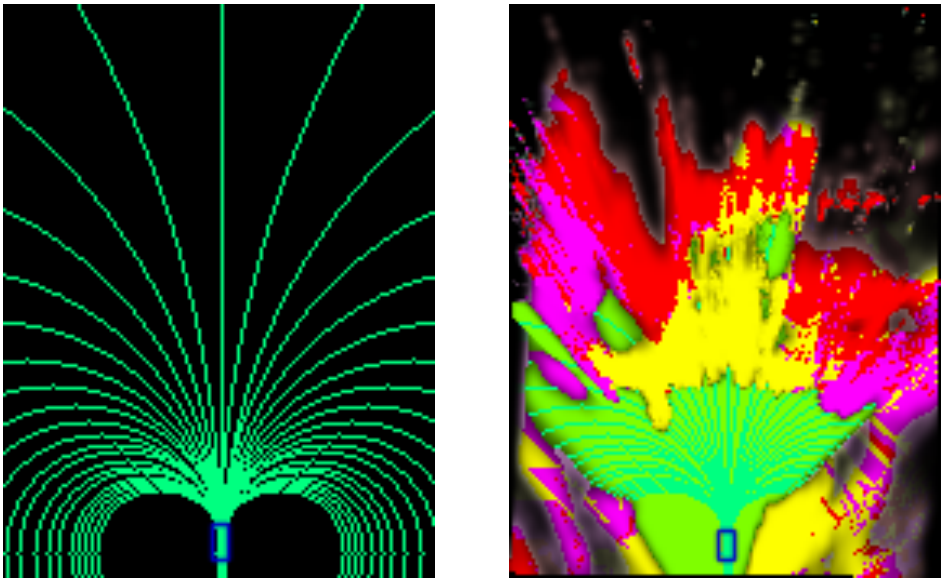
In this system, a set of tentacles containing 41 elements is modeled. As mentioned, all tentacles are represented in the local coordinate system of the vehicle. They start at the vehicle's center of gravity and take the shape of circular-arc areas. Each circular-arc segment represents the path corresponding to a specific steering angle. Thus, the motion path of the vehicle can be approximately disassembled to many circular-arc segments with different curvature.

The detailed geometry of the used tentacles is as follow. Based on Equation (5.5), the set of radius $\{R\}$, taking each steering angle δ_k of the kth tentacle ranging from 0 to number of tentacles n is given by Equation (5.6):

$$\{R\} = \begin{cases} \sqrt{a_2^2 + l^2 cot^2(\delta_k)} & |\forall k < \frac{n-1}{2} \in n \\ \infty & |k = \frac{n-1}{2} \in n \\ \sqrt{a_2^2 + l^2 cot^2(\delta_k)} & |\forall k > \frac{n-1}{2} \in n \end{cases} \tag{5.6}$$

After the initialization of n which must be a odd number, and $\delta \in [\delta_{-max},\delta_{max}]$ with δ_{max} being the maximal front wheel angle of the vehicle, the projected shapes of virtual tentacles on the grid are shown in Figure 5.5.

In order to ensure the safety and feasibility of the path, each tentacle k is characterized



(a) (b)

Figure 5.5 - The projected shapes of virtual tentacles on the grid. (a) the shapes of virtual tentacles and (b) its projection on the DLP.

by two classification areas, similar from the work of Cherubini et al. (2012) denoted by dangerous central area and dangerous external area. These areas represent an overestimated space required to the vehicle during its turn. In this way, the space requirement is modeled defining the minimum radius R_{min} and the maximum radius R_{max}. The R_{min} is estimated by the distance between the origin point and a point in the inner side of the vehicle at the location of the rear axle, whereas the R_{max} is estimated considering the outer point of the front of the vehicle. Therefore, the required space for turning is a ring with a width ΔR, which is a function of the vehicle's geometry illustrated in Figure 5.6. It is represented by Equation (5.7).

$$\begin{aligned} \Delta R &= R_{max} - R_{min} \\ &= \sqrt{(R_{min} + w)^2 + (l + g)^2} - R_{min} \end{aligned} \tag{5.7}$$

The required space ΔR can be calculated based on the steer angle by substituting R_{min}

$$\begin{aligned} R_{min} &= R_1 - \tfrac{1}{2}w \\ &= \frac{l}{\tan \delta^i} \\ &= \frac{l}{\tan \delta^o} - w \end{aligned} \tag{5.8}$$

and getting the Equation (5.9)(JAZAR, 2008).

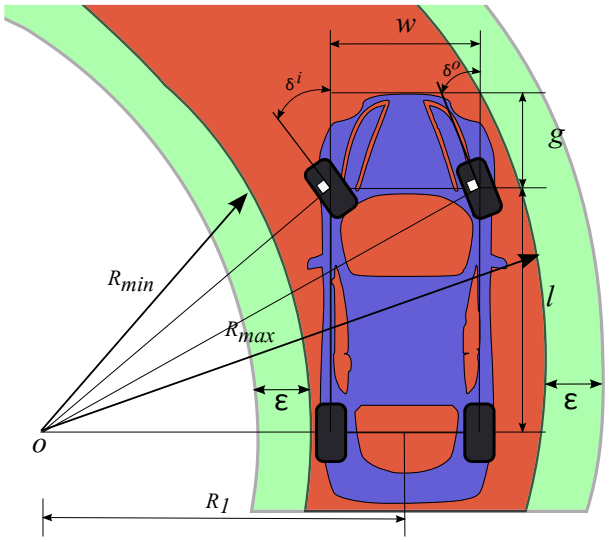


Figure 5.6 - The required space for a turning two-axle vehicle. In red the central dangerous area and green the external dangerous area.

$$\begin{aligned} \Delta R &= \sqrt{\left(\frac{l}{\tan(\delta^i)} + 2w\right)^2 + (l+g)^2} - \frac{l}{\tan(\delta^i)} \\ &= \sqrt{\left(\frac{l}{\tan(\delta^o)} + w\right)^2 + (l+g)^2} - \frac{l}{\tan(\delta^o)} + w \end{aligned} \tag{5.9}$$

To finish, the central dangerous area for each tentacle k is defined as being $\{CDA_k = \Delta R_k | \forall k \in \{T\}\}$. The external dangerous area EDA is calculated maintaining the height of the vehicle and increasing its width. In this case, it is estimated using a constant ε to ensure that, in the presence of disturbances, the current overestimated path is included in the classification area. The EDA is given by Equation (5.10) and the conception of these two areas are represented in the Figure 5.6.

$$\begin{aligned} EDA &= \left(\sqrt{\left(\frac{l}{\tan(\delta^i)} + 2w\right)^2 + (l+g)^2} + \varepsilon\right) - \left(\frac{l}{\tan(\delta^i)} - \varepsilon\right) \\ &= \left(\sqrt{\left(\frac{l}{\tan(\delta^o)} + w\right)^2 + (l+g)^2} + \varepsilon\right) - \left(\frac{l}{\tan(\delta^o)} + w - \varepsilon\right) \end{aligned} \tag{5.10}$$

5.3.2 The best tentacle classification

With the description of the tentacle-related data structures and their computation being completed, in each time cycle of the navigation, a set of obstacles that lies in the tentacle's area

is extracted. It means, among all obstacles present in the G^{PG}, only the ones which are contained in the central dangerous and external dangerous area are considered for further processing. This set is then obtained by (5.11). The set for each tentacle is characterized by OP_k^t and all sets are defined by OP_t.

$$OP_t = \{(CDA_k \cup EDA_k) \cap O | \forall O \in r(O) \text{ and } \forall k \in \{T\}\} \tag{5.11}$$

In this case, all tentacles (T) jointly with OP_t are used to select the best tentacle which derives the final path to drive at instant t. To decide on the best tentacle, three functions are linearly combined to derive a single decision value, which is maximized considering the safety and goal task provided by the route planning. So, this linear system composed by the Clearance function $F_{clearance}(T,OP_t)$, the Smoothness function $F_{smoothness}(T,T_{t-1}^b,\sigma)$ and the Target function $F_{target}(T,p_m)$, is modeled by Equation (5.12).

$$T_t^b = \underset{T_k \in T}{\operatorname{argmax}}\ \alpha F_{clearance}(T,OP_t) + \beta F_{smoothness}(T,T_{t-1}^b,\sigma) + \gamma F_{target}(T,p_m) \tag{5.12}$$

Where T_t^b is the best tentacle at current time, T_{t-1}^b is the best tentacle at instant $t-1$ and p_m is the mutant point explained in section 5.1. The α, β and γ are parameters that can be used to change the behavior of the selection mechanism. An important aspect after the best tentacle T_t^b to be selected, is to verify if its traversability is higher than a distance crash l_c required to stop the vehicle without hits the nearest obstacle lying in that path. Similar to work of Hundelshausen et al. (2009), the crash distance is the distance the vehicle needs to stop using a constant convenient deceleration a plus a security distance l_s. It depends on the longitudinal velocity V of the vehicle and is calculated by Equation (5.13).

$$l_c = l_s + \frac{v^2}{2a} \tag{5.13}$$

Experiments performed in this work, considered $l_s = 1.5$ m and $a = 1.5$ m/s^2, due the low velocity of the vehicle. Therefore, if the distance of the nearest obstacle is less than crash distance, the velocity parameter of the tentacle is setted to $T_t^b.V = 0.0$ m/s. Following, it is detailed the three functions used in Equation (5.12).

Clearance Function

This first function is responsible to express how far the vehicle could drive along a tentacle before hitting an obstacle, in others words, the risk of collision. The output value of this function is a normalized value ranging from 0 to 1, where a value of 1 designates a preference for such a tentacle. The first step to calculate the clearance value is to extract the minimum curvature length of tentacle's arc between the intersection of this arc with any limit border of the vehicle and the set of obstacles. This is reached by Equation (5.14):

$$H_k = \{min[D(OP_k^t,Pv)]|\forall k \in \{T\}\} \tag{5.14}$$

Where H_k is the resulting distance of the nearest obstacle lying in the kth tentacle, Pv is the intersection point of the arc with the limit border of the vehicle. $D(.)$ function is the length arc given by $D = 2\pi r(\frac{\theta}{360})$, having radius r and the central angle of the arc in degrees θ. After this step, the clearance value can be directly calculated using a sigmoid-like function defined in Equation (5.15).

$$F_{clearance}(T,OP_t) = \begin{cases} 1 & \text{, Free Tentacle} \\ 3 - \frac{2}{1+e^{c_{clearance} \cdot H_k}} & \text{, otherwise} \end{cases} \tag{5.15}$$

where the constant $c_{clearance}$ is given by Equation (5.16).

$$c_{clearance} = \frac{ln(\frac{1}{3})}{-\mu} \tag{5.16}$$

In Equation (5.16), $\mu = 15$m. This constant gives the sense of what is the middle distance in the path from a obstacle. Figure 5.7 illustrates the output value of clearance function ranging the distance H_k of the nearest obstacle lying in the kth tentacle.

Smoothness Function:

The second function is responsible by pondering the choice of the current tentacle based on the previous one, preferring smooth paths. This is reached modelling a normal distribution, taking into account the former selected tentacle T_{t-1}^b as being the mean of this distribution and

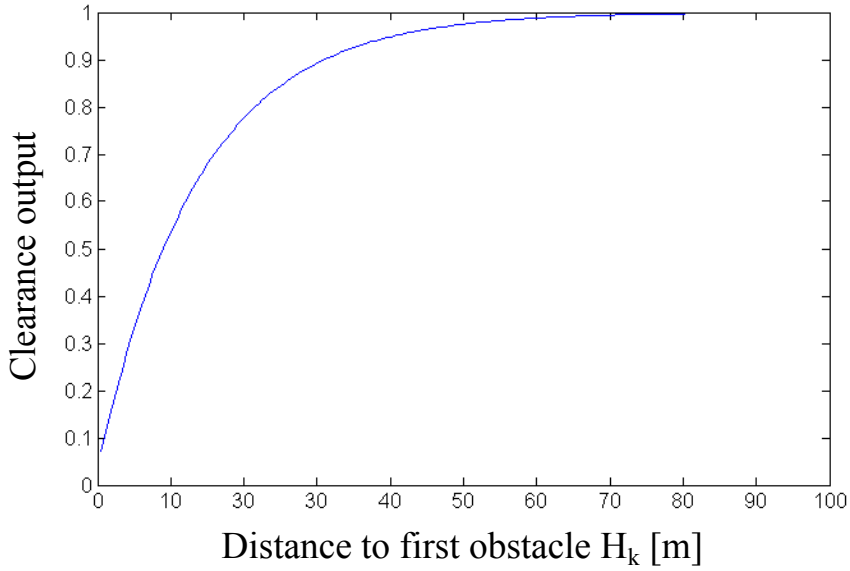


Figure 5.7 - Plot of Clearance function output shown in Equation (5.15). It shows that the greater the distance to the first obstacle, the higher is the change in the output result and the change of impact in the overall tentacle selection process.

defining a parameter σ to represent the standard deviation from T^b_{t-1}. Therefore, the smoothness function is obtained by Equation (5.17):

$$F_{smoothness}(T,T^b_{t-1},\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(T_k - T^b_{t-1})^2}{2\sigma^2}} \tag{5.17}$$

Target Function:

This last function has the objective of representing the semantic information leveraged by the route planning to accomplishment of the task. This is done positioning the mutant point in any place in the navigable area where it may represent the semantic information. As example, if the route planning is telling “go ahead”, the mutant point could be positioned further away from the navigable area. Another case, if it is detected that the vehicle reached an intersection area, and the route planning is telling to “turn right”, so the mutant point could be placed at the further right side of the navigable area. The idea behind it is that after this point be placed (different from waypoint of the GPS proposed in literature), a measure can be performed combining this information together with the previous functions to compose the linear system that results in an optimal solution, given the appropriate configuration of the weights.

In this sense, the measurement is modeled considering the euclidian distance from the mutant point to the tentacle’s closest point. As tentacles are represented by circular-arc seg-

ments, the Equation is given by the distance from the point to the circle, as depicted in Equation (5.18). In case where the tentacle represents the steering angle equal to zero, this function uses the distance to the straight line.

$$F_{target}(T,p_m) = \begin{cases} \frac{(m*p_m.x)-(p_m.y+b)}{\sqrt{1-m^2}} & \text{, if } \delta_i = 0 \\ \left| \sqrt{\sum (p_m - p_o)^2 - r} \right| & \text{, otherwise} \end{cases} \tag{5.18}$$

In Equation (5.18), the (x,y) are the coordinates of the mutant point p_m. m and b are the constants parameters of the line. p_o is the tentacle's origin point and r its radius.

5.4 Path-following Control

The path-following control for this system is implemented to ensure that the vehicle is following the designed path along its navigation. In this case, as the projected path is always referenced at the current pose of the vehicle inside the grid, the control is derived from the relative vehicle's orientation variation ψ.

It can be implemented based on the relative displacement of the vehicle between two consecutive time interval, assuming that the optical axis of the camera is parallel to the longitudinal axis of the vehicle. Given the measurement of the 3D rigid transformation obtained by the visual odometry, it is reduced to 2 dimensions and represented in the grid's coordinate. From rotation matrix R and the translation vector $\vec{t}$, it is possible to associate the path command generated by the virtual tentacle with this information, according to the block diagram model shown in Figure 5.8.

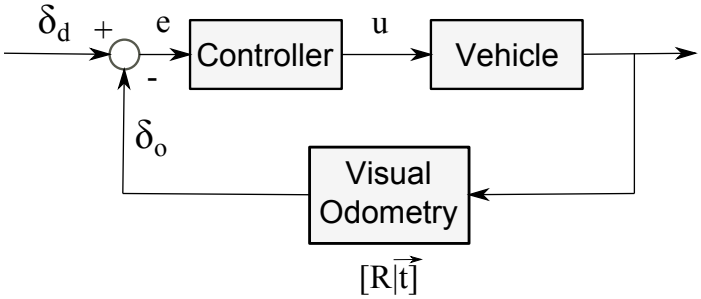


Figure 5.8 - Simplified block diagram of the closed-loop control developed to path-following.

As seen in Figure 5.8, the reference value, defined as desired reference of the wheels orientation δ_d, is set using the wheel orientation δ of the best tentacle T_t^b at instant t. Also, to apply the control it is necessary to obtain the observed wheels orientation δ_o that will be demonstrated here by a sequence of operations.

First it is possible to obtain the observed vehicle's orientation variation ψ_o directly measured from the rotation matrix R obtained from the extracted information $[R|\vec{t}]$ of the visual odometry, shown in Equation (5.19).

$$R_{\psi_o} = \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix} \tag{5.19}$$

Then it is necessary to obtain a relation between the vehicle's orientation variation ψ_o in function of the parameters of the tentacle to obtain the observed turning radius R_o, as shown in Figure 5.9.

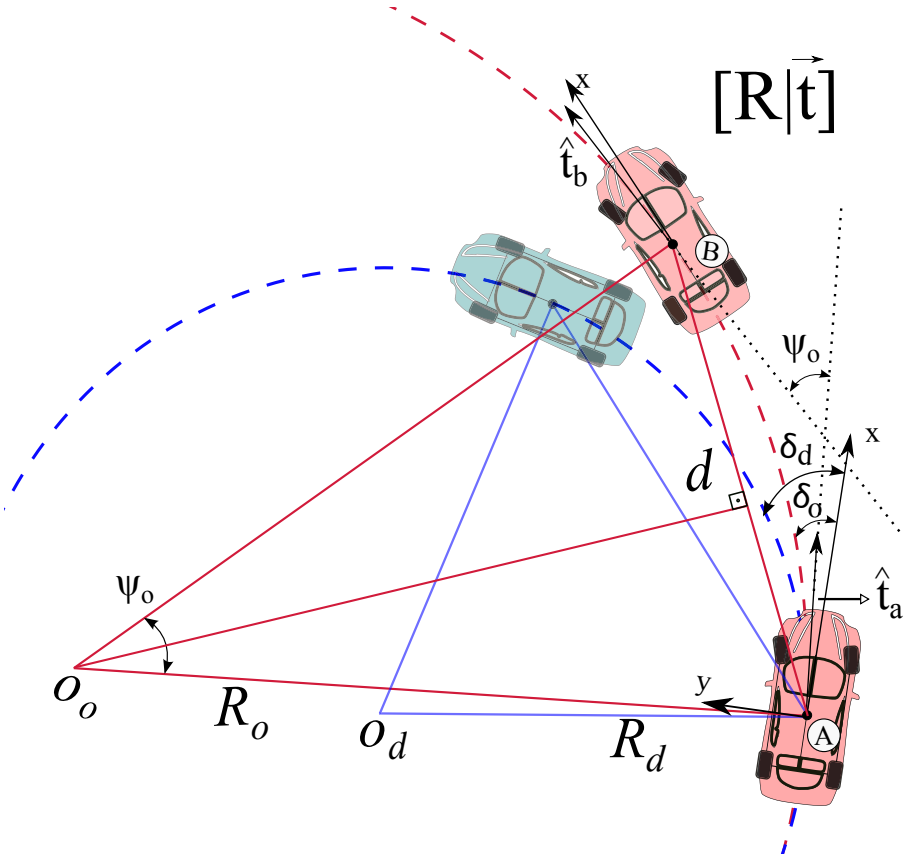


Figure 5.9 - Controlling of the path.

The observed vehicle's orientation variation ψ_o is formed assuming the distance between

two points expressed by $\vec{t}$ ($d = \sqrt{\vec{t_x^2} + \vec{t_y^2}}$), jointly with the observed turning radius R_o of the observed tentacle, and depicted in Equation (5.19) :

$$\frac{\psi_o}{2} = \sin^{-1}\left(\frac{d/2}{R_o}\right) \tag{5.20}$$

Finally, with the observed turning radius R_o and the Equation (5.5) it is possible to obtain the observed wheel orientation δ_o.

The operation procedure in the time domain can be describe as follows. If the vehicle starts from point A at a desired steering angle $\delta_d(t-1)$ of the wheels at a time instant $(t-1)$, it will follow the observed tentacle resulting to a point B at a time instant t. Therefore, if you manage to observe the vehicle's orientation variation $\psi_o(t)$ at a time instant (t) with Equation (5.19), it is possible to obtain by Equation (5.20) the observed turning radius $R_o(t)$. With the observed turning radius $R_o(t)$ it is possible to obtain by Equation (5.5) the observed steering angle $\delta_o(t)$ shown on Figure 5.10. So, with the observed steering angle $\delta_o(t)$ and the next desired steering angle $\delta_d(t)$ it is possible to apply the path-following control according to Figure 5.9 and calculate the error $e(t)$. Then using a proportional control with gain kp, it is possible to apply the steering output $u(t)$ calculated by Equation (5.21), which is proportional to the error $e(t)$ calculated by the difference between the desired reference $\delta_d(t)$ and the observed $\delta_o(t)$ orientations, to achieve point C.

$$u(t) = kp\ e(t) = kp(\delta_d(t) - \delta_o(t)) \tag{5.21}$$

5.5 Vehicle Platform

The experimental vehicle belonging to the Heudiasyc Laboratory is used for the implementation and the validation of the global system in real-life conditions. The experimental platform consists of an electric vehicle called ZOE and for the experiments, it was equipped with a stereo camera sensor, two embedded computers and two CAN-bus [1]. As previously presented in

[1] Controller area Network

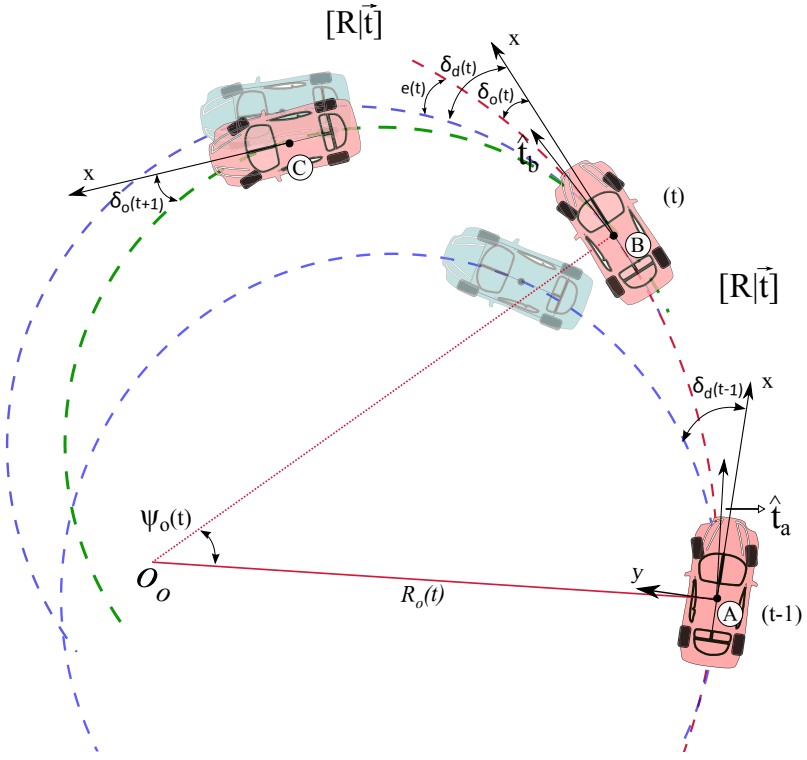


Figure 5.10 - Controlling of the path in the time domain.

Chapter 4, the used sensor camera is characterized by a PointGrey Bumblebee x3 color model BBX3-13S2C-60, with focal length of 6mm and horizontal opening angle of ∼43 degrees. The baseline of the stereo camera is approximately 24 cm. The first computer is responsible to process the complete system (except the low PID control) and has an Intel Xeon E5-1650 processor with 3.20Ghz and with 16Gb DDR3, running the version 7 of the Windows OS. The second computer is an EffiBox Intel Core i5-3610ME processor with 2.70Ghz and with 4Gb, also running the version 7 of the Windows OS. It is responsible to run the low PID control and to communicate via CAN-Bus with the embedded sensors and actuators used to drive the vehicle. A CAN-bus is employed only for listening the data information delivered from the car, and the other one is employed to send the steering angle, braking and acceleration data signal.

Figure 5.11 presents the experimental platform as well as the designed configuration for the hardware interaction. The stereo camera is connected via firewire IEEE-1394b with the first computer. Then, an User Datagram Protocol (UDP) communication was created to exchange information between these two computers. Lastly, the second computer sends the data signal by CAN-Bus to the car.

(a) (b)

Figure 5.11 - (a) The experimental platform Zoe. (b) The designed configuration for the hardware interaction.

5.6 Proposed System Architecture

The Heudiasyc laboratory supports a software framework called *Perception et Assistance pour une Conduite Plus sûre* (PACPUS), which provides standard operating system services such as hardware abstraction, low-level device control, implementation of commonly used functionality, message-passing between processes and package management.

The system architecture designed on the PACPUS is based on components. Components are process that run in parallel under the framework. As can be seen in Figure 5.12, it was developed 13 components that perform each one, one or two specific tasks of the proposed approach. The components starting from Image Caption until Path-Following Control are placed on the Xeon computer, including the component manager. Only the PID control component runs in the second dedicated computer. It was required due to restrictions on the communication with the CAN-Bus, which must maintain an exclusive process to exchange information on the port to preserve an open connection with the car. As these components run in parallel, the synchronization among them should be done to protect the temporal consistency of data exchange. The function of the component manager is exactly to ensure this requirement. In this way, a multi-processed pipeline is implemented to minimize the time cost in sequential data stream, it means, multiple independent tasks are overlapped in the time execution.

The synchronization was divided in four stages. The first stage is triggered by the Component Manager after it receives a response signal from the Image Caption component telling

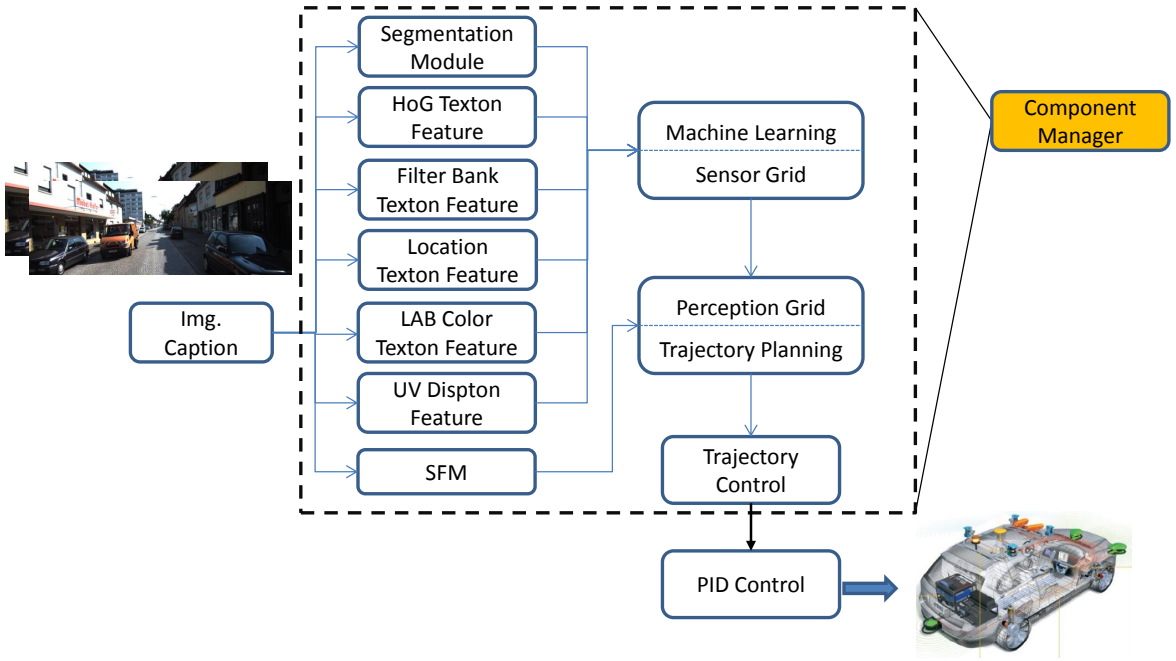


Figure 5.12 - The designed software architecture based on PACPUS framework.

that the current image t is available. It should be mentioned that the Image Caption component runs freely, without any synchronism. The first stage contemplates the components of the segmentation, texton, dispton and SFM processes. The second stage is triggered after almost all components of the first stage conclude their tasks, except the SFM process. This stage considers the component of the Machine Learning and Sensor Grid Tasks. Following, the third stage is triggered by the Component Manager that received the concluded signals from the SFM component and the second stage. Finally, the fourth stage is dependent only of the third stage. An execution example of this designed multiprocessed pipeline is illustrated in Figure 5.13.

5.7 Experimental Results

This section presents the results obtained with the proposed method detailed until here. The results are organized in two experiments. The first one, described in subsection 5.7.1, is

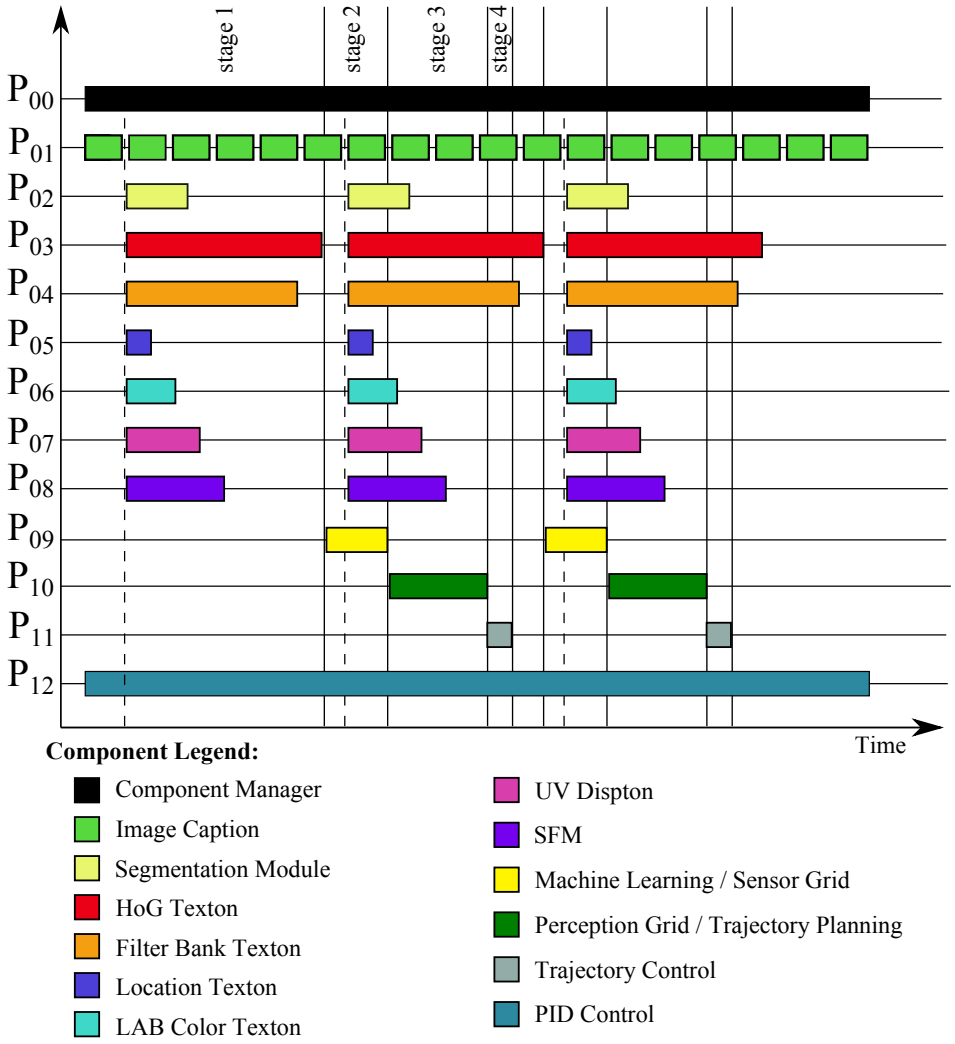


Figure 5.13 - Execution example of the designed multiprocessed pipeline.

aimed at the presentation of the path-planning using the proposed virtual tentacles, showing its response for two distinct scenarios as well as with different hardware. The second one, reported in subsection 5.7.2, shows the complete local navigation system obtained at the test track of the Heudiasyc Laboratory. These experiments highlight the capabilities of this kind of hybrid navigation.

These two experiments use the mutant point placed further away of the navigable area. As the route planning was not implemented yet, positioning the mutant point in this way, it means that the route planning is telling to always “go ahead” to path-planning, although, an analysis was done to get a feeling of this concept.

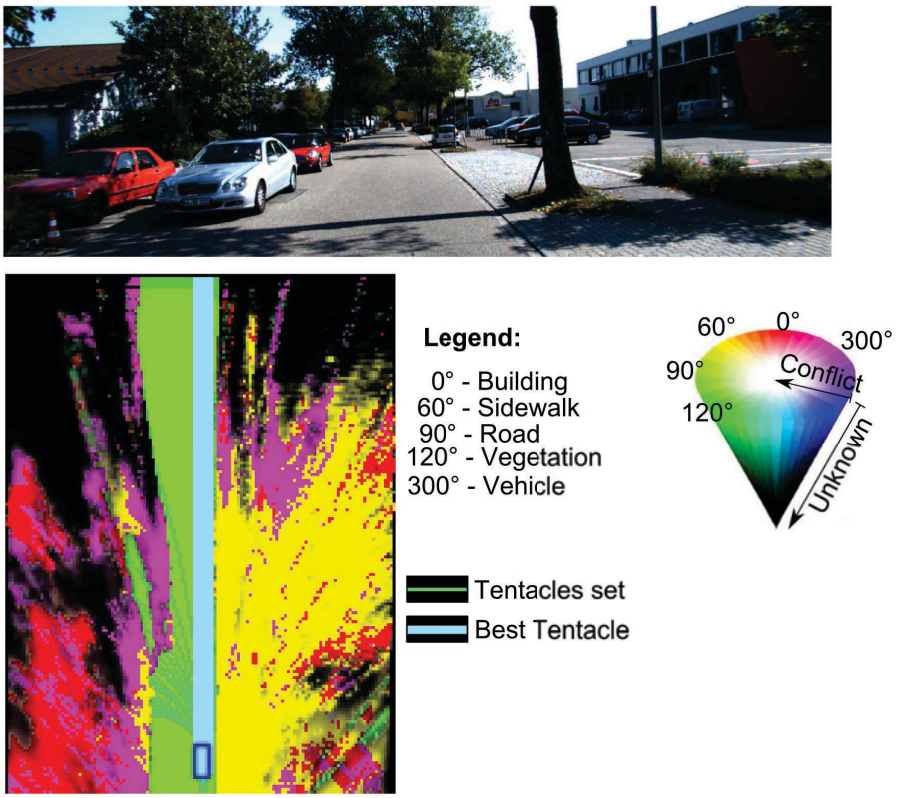


Figure 5.14 - Path-planning in a straight street.

5.7.1 Path-planning results

This section presents two essays. The first essay was carried out using the Kitti platform. As the path-planning does not require the effective execution of the command, the fundamental step is to analyse the coherent response that this module is delivering at real urban scenario. Therefore, a qualitative study is employed to validate the method.

The video sequence using the Kitti platform was acquired in German and represents the local scenario of the mentioned example given in section 5.1. In this case, this video sequence corresponds to the route planning indicated by the waypoints number 3, 4 and 5 in Figure 5.2.

For all the results presented below, figures contain the original image acquired at the same instant that the local navigation grid is displayed. In this grid, in which colors are represented as explained in chapter 4, the projected tentacles are signaled by circular-arcs into the navigable area defined by the SpringGreen color. It should be mentioned that each tentacle has its associated central dangerous and external dangerous areas which are not being displayed. Lastly, the best tentacle is displayed with silver color including the CDA.

Figure 5.14 shows a example of local navigation between the waypoints 3 and 4. This specific scenario is frequently observed in inner-city. The configuration of the parameters α, β and γ are setted respectively to 1.0, 0.5 and 0.0. With this configuration, the path-planning becomes a reactive system. As can be observed, the output response is appropriated for this situation, where the selected tentacle does not go in direction of collision with the other car, nor planning a path that goes up on sidewalk. The complete result for this sequence is public available in Vitor (2014h) [2].

In order to analyse the impact of the mutant point in the proposed mechanism of selection, the configuration presented in the previous example was changed to $\alpha = 1.0$, $\beta = 0.5$ and $\gamma = 0.5$. This current configuration is defined as being 1 and the previous as being 2. The idea is to verify if the mutant point exercises the influence under the selection of the best tentacle as a form to guide the path in function of the information delivered by the route planning. Figure 5.15 illustrates a sequence of images in a crucial area of intersection, confronting these two configurations side to side. As explained before, the mutant point is representing the information of "go ahead" placed further away of the navigable area. As can be seen in this example, in the way as the vehicle runs in the direction to the intersection, the configuration 2 results in the optimal solution without taking into account the target function. In this case the better output indicates that the vehicle should turn left to avoid possible collision in a precipitated time. The interesting fact is that the configuration 1 maintains the best tentacle appointing to "go ahead" until it is no more possible to do it. At the end, as this point is placed further way from the navigable area, in case where it does not have option, the solution is "go ahead" where the navigable area exists. It has the same human feeling when telling "go ahead" in a curve street. The semantic information of "go ahead" is not always to follow a straight line. The video result of this example can be found in Vitor (2014b) [3].

The second essay was carried out with a database acquired at the test track (Hediasyc lab.), as mentioned before. The setup car was defined as shown in chapter 4. The configuration of the tentacles' selection mechanism is the same as previously explained, in configuration 1. Figure 5.16 shows images in different parts of the test track. As can be seen, this is less challenging due to the observed scenario. However, an interesting factor is that this test track has two sharp curves and the HFOV of the used camera sensor is not able to get sufficient

[2] <http://youtu.be/dBXHgGdbk80> accessed on: 04/08/2014
[3] <http://youtu.be/5PGnfB3djzQ> accessed on: 01/10/2014

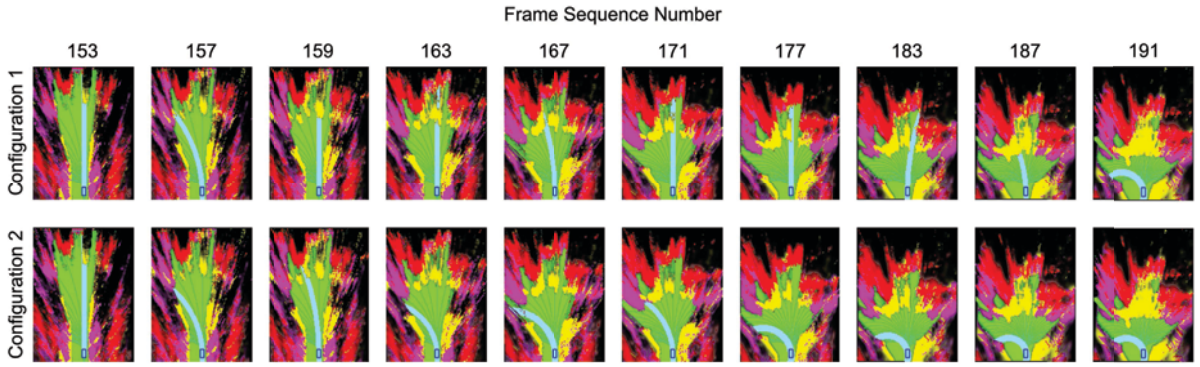


Figure 5.15 - The comparison of distinct configuration to Path-planning. The numbers indicate the frame sequence in video.

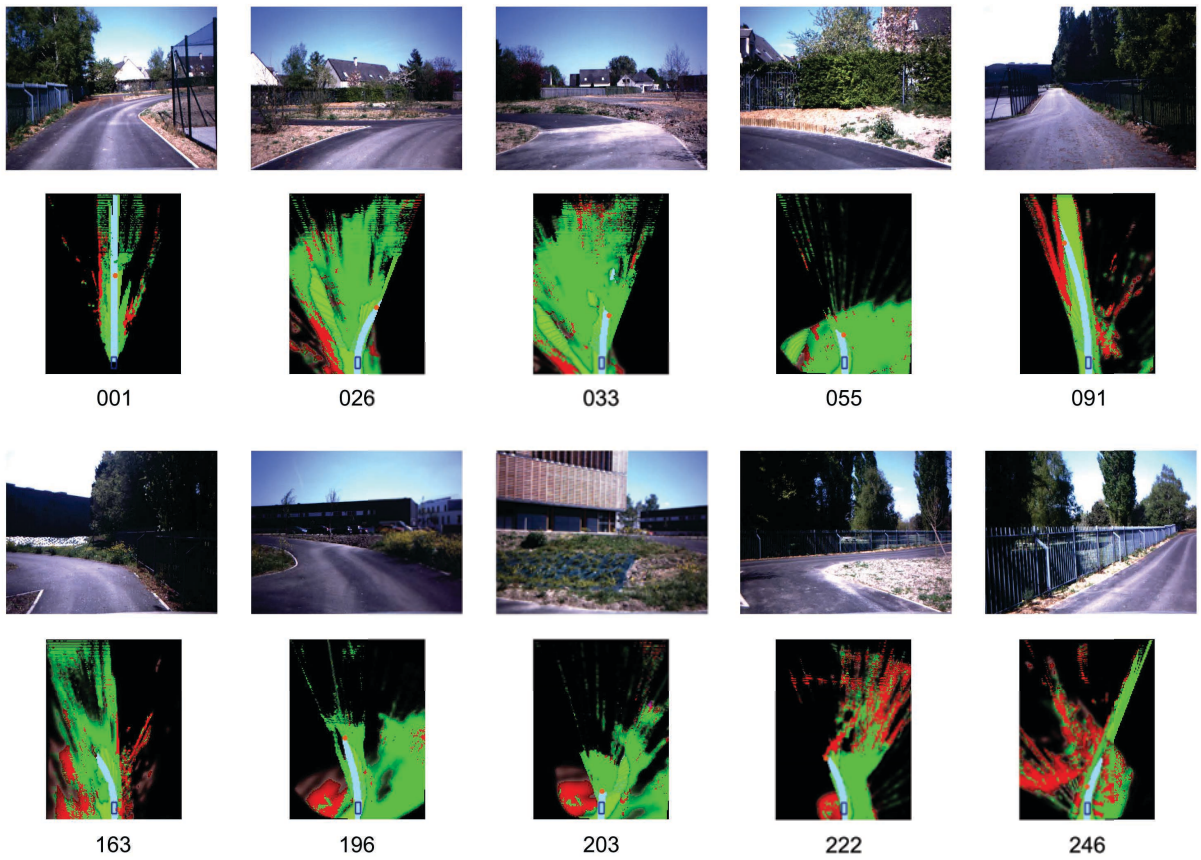


Figure 5.16 - Path-planning carried out in the test track of Heudiasyc Laboratory. The numbers indicate the frame sequence in video.

information in these areas, becoming a difficult task to path-planning. Despite this restriction, the output result is coherent with the expected response, given this partial occlusion in sharp curves. The video result of this experiment can be found in Vitor (2014g) [4].

[4] <http://youtu.be/YKYCwdUEaVs> accessed on: 04/08/2014

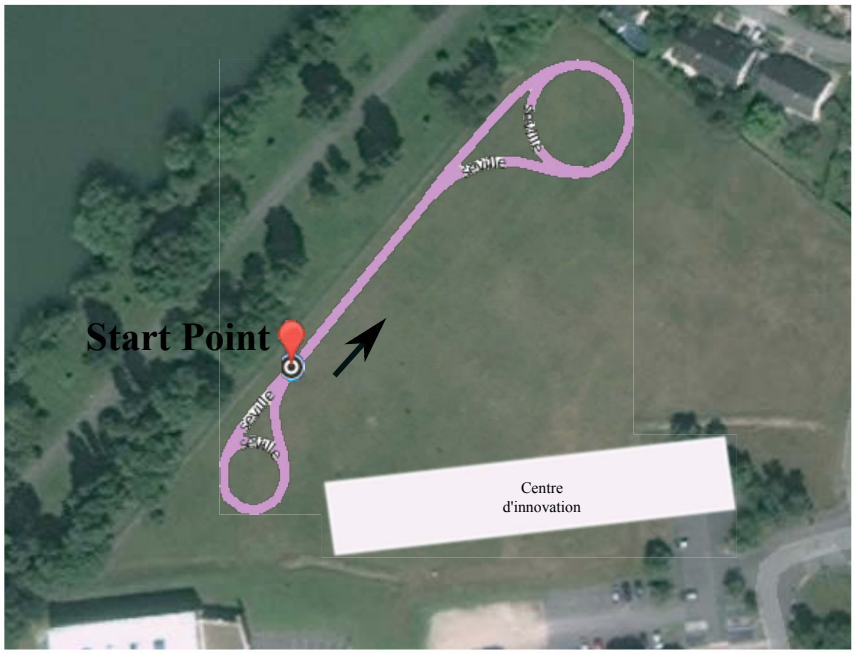


Figure 5.17 - Image Satellite of the test track.

5.7.2 Local autonomous navigation

After the path-planning was observed, this section presents the complete system using the PACPUS framework embedded in the ZOE platform. The autonomous navigation was performed in the test track which consists of a closed circuit where 1 loop corresponds to more than 250m, containing two turns of 360° degrees, with radius of approximately 4 and 7 meters, as can be seen in Figure 5.17.

Firstly, the performance analysis in function of the time consumption was validated. Figure 5.18 presents the time in seconds for each component that run in the main computer. These measurements were performed considering the mean time of a sequence of images with resolution 320x240px. Using the pipeline approach detailed in section 5.6, the overall time consumption was reduced more than 50%, leveraging a total time of approximately 2.76s for each image, considering the complete approach.

After the time of the system was validated, it enables the autonomous navigation system to run between 1 and 3km/h. The experiments presented here consider common and necessary tasks required to perform a satisfactory local navigation. The first example presents the autonomous vehicle changing its path in function of an obstruction in the path. As can be seen in Figure 5.19, the first time the vehicle follows the straight path observed in the environment. In

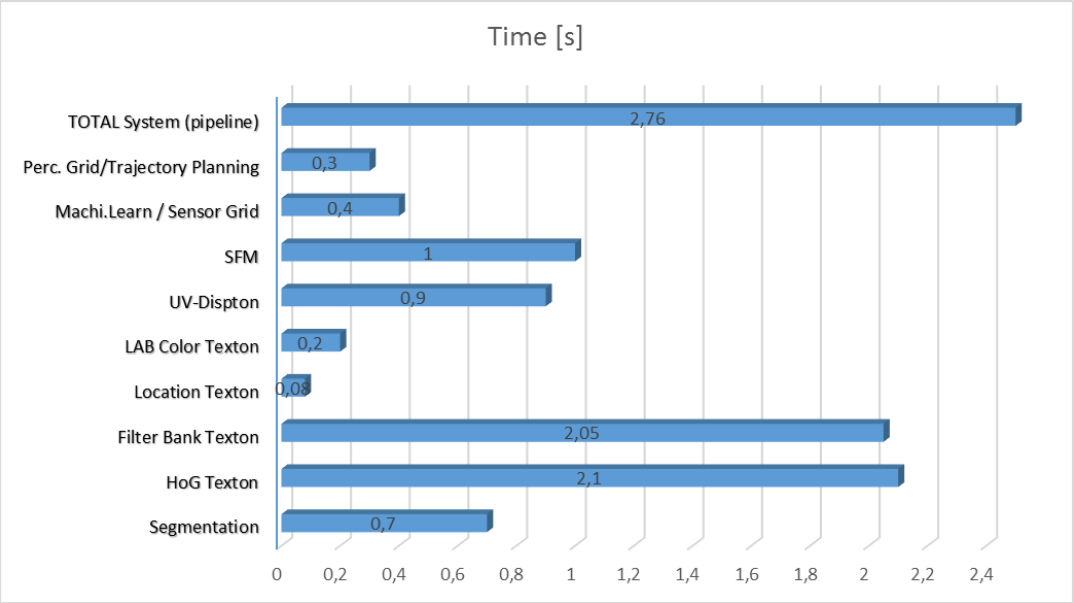


Figure 5.18 - Time consumption for each system component.

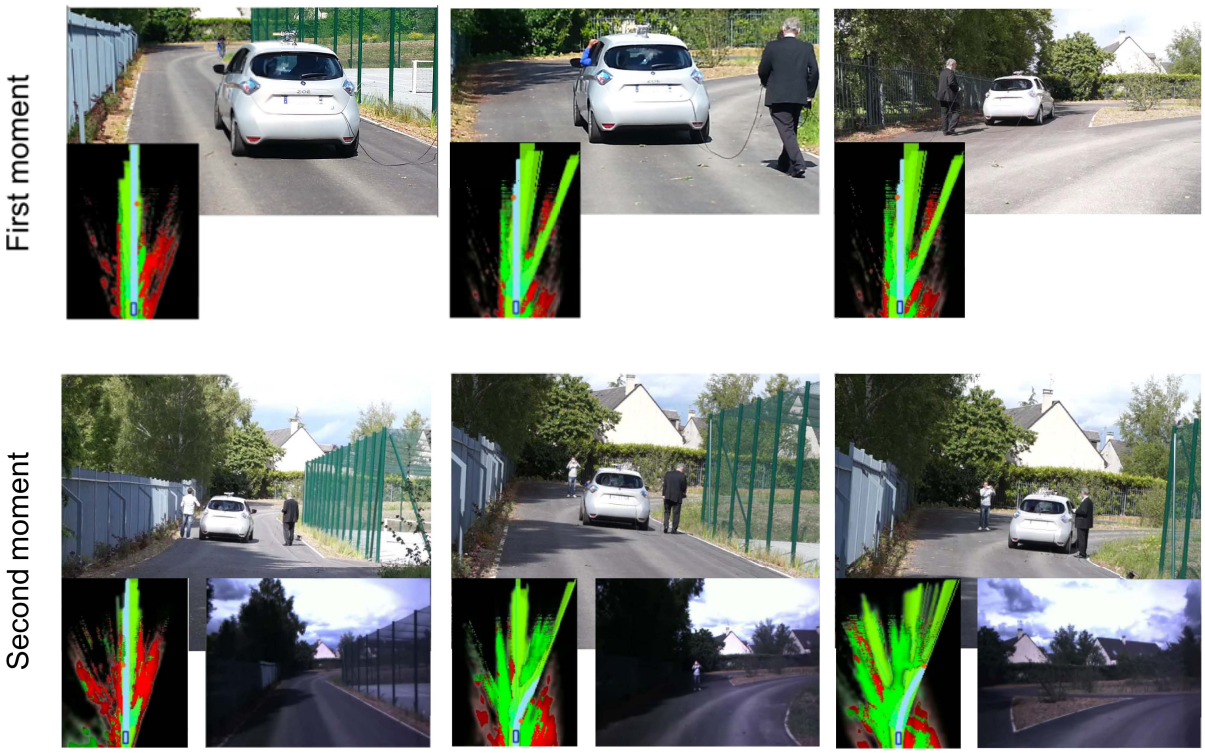


Figure 5.19 - The autonomous vehicle finding an alternative path.

a second time, the same path was blocked to avoid its passage. As a consequence, the autonomous vehicle found an alternative path avoiding that obstructed one.

The example illustrated in Figure 5.20 presents the autonomous system during right and left turns. As previously mentioned, even with the restrictions of the camera sensors, this present result was the first one at the Heudiasyc Laboratory to conclude the task using this kind of

Figure 5.20 - The autonomous vehicle enabled to perform sharp curves in presence of partial occlusions.

Figure 5.21 - The autonomous vehicle enabled to perform obstacle avoidance.

sensor. Even in sharp curves where occlusion frequently occur as it is observed in these experiments.

To finish, Figure 5.21 depicts the obstacle avoidance. During the autonomous navigation, a person stopped in front of the vehicle. In this case, independently if the system's classification is not trained to understand what is a pedestrian, it is capable to distinguish between obstacle and navigable area. It is the fundamental aspect required in an autonomous system, it means, safety. The complete video containing these results can be seen in Vitor (2014a) [5].

5.8 Conclusion

In this chapter, a hybrid navigation approach concerning global and local navigation is discussed. A new form to associate the route planning in the framework of the trajectory planning is proposed, using the conceit of mutant point to influence the selection mechanism of the

[5] <http://youtu.be/shw1IRYd2T4> accessed on: 04/08/2014

virtual tentacles, arising in this way a interesting deliberative-reactive system acting in the local navigation.

A new path-following control based on virtual tentacles using ego-centered conception is developed, where, jointly with the DLP and trajectory planning provide a robust system to automatically control a vehicle in real environment. Although the autonomous navigation has been tested in real conditions of a closed test track, since the trajectory planning and trajectory control are modeled directly using the proposed DLP, these obtained outcomes realized in the urban scenario using the database of the KITTI benchmark justify the application for inner-city scenes in urban environment. It means that the same autonomous vehicle using the proposed system will ensure its applicability in working out of the test track.

It is important to notice that the system is not totally dependent of a global and precise GPS, and also that it could use some image memory approach which indicates the direction in an intersection. It does not require any previous map to perform a safe navigation, although it could be interesting to improve robustness specially at intersections. It uses only a pair of stereo cameras to project the path avoiding obstacles. And lastly, this conception is closer from the human discernments. However, several improvements must be done to ensure the same robustness in real-time, since the complete system is running at 2.76s in the current implementation. A careful performance analysis should be done to verify the real advantages and drawbacks of the trajectory control, and finally, it should be necessary to include more sensors to improve errors given by the stereo geometry as well as in environments where cameras are not appropriated.

CHAPTER 6

General Conclusion and Perspective

If we knew what it was we were doing, it would not be called research, would it?

(Albert Einstein)

6.1 Conclusions

The main contributions of this thesis are described in the following.

Urban road scene understanding

The requirement to perform safe autonomous navigation in urban environment is directly related with its capability to recognize precisely the road in front of the car, where any driving maneuver or vehicle control should be done. And also, it needs to infer potential risks associated with obstacles present in the scene. These aspects promoted a strong study concerning the road recognition in inner-city scenario as well as the elements that compose this one. To be able to deal with this challenge, three algorithms were proposed, denominated ANN, HistonBoost and ProbBoost. These algorithms can be seen as an evolution of the understanding about the problem, becoming the ProbBoost a cutting-edge approach in this field. In this algorithm, a combination of 2D/3D information provided by the Texton maps and the new one Dispton maps, together with the segmentation procedure using the Watershed Transform showed to be an interesting feature descriptor technique for this kind of problem. Further, this ProbBoost

algorithm was extrapolated to work with not only road recognition, but also with a set of elements contained in the urban road environment. Based on a standard performance analysis, the method has proved to deal also with this challenge of urban road scene understanding, which has reached an improvement of 18.23% from HistonBoost and 30% from the Baseline.

Dynamic evidential grid using semantic context

Following, a dynamic local perception system is developed to build the representative model of the environment around the car. The metric representation based on occupancy grid is implemented. It uses the evidential formalism leveraged by the Dempster-Shafer Theory, which has been receiving considerable attention recently. This formalism permits to manage uncertainties associated with the grid discretization, partial observation of the environment and also dynamic elements present in the scene. In this context, it is created a new inverse model sensor using only a pair of stereo cameras, where meta-knowledge of the scene associated with depth information are taken into account to provide a better and reliable representativeness of navigable, infrastructure and obstacle areas. To the best of our knowledge, this approach is the first one to associate the semantic context extracted from machine learning procedure to build an inverse model sensor that uses the meta-knowledge to influence the elaboration of the belief mass into the evidential grid.

Hybrid navigation binding route planning with trajectory planning

An hybrid navigation approach is considered in this thesis. It is composed by two levels, the route planning and the trajectory planning. As discussed, the route planning uses a principle of topological representation and it is responsible to generate the complete path using graph theory. The trajectory planning uses the dynamic local perception grid to define the trajectory path by where the car should go, considering the fundamental requisites such as safety and objectivity. Therefore, in this work is implemented the virtual tentacle approach, creating a new selection mechanism to deal with reactive issues required in the local environment and including a measurement based on the defined mutant point, which is responsible to represent the semantic information provided by the route planning. Consequently, it introduces a novel

method to manage and bind the route planning information into the framework of trajectory planning.

Path-Planing Control

Boosted by the fact that all system has been done into a local environment and with the relative reference centered on the car, how to model a controller that ensures the execution of the path with respect to the best tentacle generated by the trajectory planning is the question. With this specific challenge in hands, a new method that controls the path execution is proposed. It considers the output result of the tentacle's selection mechanism, the visual odometry and the kinematic model of the car, always recognized in the local scope. A brief description is as follow. The tentacle's selection mechanism provides a desired steering angle of the wheels. After the displacement of the vehicle, the observed distance travelled and the observed vehicle's orientation are calculated by the visual odometry. With these two information, the real executed curvature of the vehicle is derived and the observed steering angle of the wheels is extracted using the radius of this curvature in the kinematic Equations. In this way, this observed feedback signal can be employed to close the loop of the controller.

Autonomous Navigation

The aforementioned approaches constitute a set of tasks desired for accomplishment of an autonomous system. Considering the overall system, the approach is done using only a pair of stereo images. It does not require any prior digital map to perform the navigation. It is not totally dependent on global and precise positioning system by satellites. The Module independency of this approach leverages the usage of parallelization and improvement on resources usage. And lastly, it has an interesting aspect to follow the human sense.

6.2 Perspectives

These proposed methods have been validated on a sufficient numbers of experiments which have allowed the identification of several points that should be improved. Among them, the principal ones are highlighted.

Concerning the road detection, the high level of complexity to build and test a recognition model is due to the different characteristics present in inner-city environments. We believe that an alternative to model a complex road pattern should take into account the learning of intra-classes pattern derived from the main class such as normal area, shadow area, horizontal land mark if exist, area with influence of sun, etc. Thereby, an strategy using any kind of unsupervised classifier to leave arise clusters of these intra-classes automatically, and after that, merging this result process in a supervised method could be interesting to improve the result of road recognition. In the case of multi-class recognition, an attractive method that has been demonstrating excellent results is to model a probabilistic approach based on CRF to better formalize the relationship among neighbours pixels, filtering out inconsistencies intra-classes.

Regarding the dynamic local perception system, two issues were observed. The first one is related to the formalism used to manage the meta-knowledge associated with the belief masses. Actually, the proposed method uses a principle based on voting, which is not quite appropriated for this end. An improvement might be to use a probabilistic formalism to upgrade and merge these information. The second one is related to the temporal information propagated into the grid. The mechanism of contextual discounting may be used in order to represent the variation in information lifetime of objects present in the environment.

To finish, the complete system should be optimized, for example, implementing this proposed solution in GPU architecture, leveraging the real-time required to perform an autonomous navigation at higher speeds, as well as integrating the route planning to validate the proposed hybrid conception.

References

ACHANTA, R. et al. Slic superpixels compared to state-of-the-art superpixel methods. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 34, n. 11, p. 2274–2282, Nov 2012. ISSN 0162-8828.

ALVAREZ, J.; GEVERS, T.; LOPEZ, A. Learning photometric invariance from diversified color model ensembles. In: **Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on**. [S.l.: s.n.], 2009. p. 565–572. ISSN 1063-6919.

ALVAREZ, J.; LOPEZ, A. Road detection based on illuminant invariance. **Intelligent Transportation Systems, IEEE Transactions on**, v. 12, n. 1, p. 184–193, 2011. ISSN 1524-9050.

ALVAREZ, J. M.; GEVERS, T.; LECUN, Y.; LOPEZ, A. M. Road scene segmentation from a single image. In: **ECCV 2012**. [S.l.]: Springer Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, v. 7578). p. 376–389.

ANDREWS, J. R.; HOGAN, N. Impedance control as a framework for implementing obstacle avoidance in a manipulator. In: **Control of Manufacturing Processes and Robotic Systems**. [S.l.: s.n.], 1983. p. 243–251.

ARKIN, R. C. Integrating behavioral, perceptual, and world knowledge in reactive navigation. **Robot. Auton. Syst.**, North-Holland Publishing Co., Amsterdam, The Netherlands, The Netherlands, v. 6, n. 1-2, p. 105–122, jun. 1990. ISSN 0921-8890. Available from: <<http://dx.doi.org/10.1016/S0921-8890(05)80031-4>>.

ASSIDIQ, A.; KHALIFA, O.; ISLAM, R.; KHAN, S. Real time lane detection for autonomous vehicles. In: **Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on**. [S.l.: s.n.], 2008. p. 82 –88.

AUDIGIER, R.; LOTUFO, R. de A. Relationships between some watershed definitions and their tie-zone transforms. **Image Vision Comput.**, Butterworth-Heinemann, Newton, MA, USA, v. 28, n. 10, p. 1472–1482, oct. 2010. ISSN 0262-8856. Available from: <<http://dx.doi.org/10.1016/j.imavis.2009.11.002>>.

AWASTHI, A. **Développement d'un système de routage hiérarchique pour les réseaux urbains**. PhD Thesis (PhD) — Université de Metz, 2004. Available from: <<http://tel.archives-ouvertes.fr/documents/archives0/00/00/77/51/index_fr.html>>.

AYARI, E.; HADOUAJ, S. E.; GHEDIRA, K. A reactive anticipation for autonomous robot navigation. **Serial and Parallel Robot Manipulators - Kinematics, Dynamics, Control and Optimization**, InTech Publishers, v. 1, n. 1, 2012. ISSN 978-953-51-0437-7. Available from: <<http://www.intechopen.com/books/serial-and-parallel-robot-manipulators-kinematics-dynamics-control-and-optimization/a-reactive-anticipation-for-autonomous-robot-navigation>>.

AZI, N.; GENDREAU, M.; POTVIN, J.-Y. An exact algorithm for a single-vehicle routing problem with time windows and multiple routes. **European Journal of Operational Research**, v. 178, n. 3, p. 755 – 766, 2007. ISSN 0377-2217. Available from: <<http://www.sciencedirect.com/science/article/pii/S0377221706001299>>.

BACHMANN, A.; LULCHEVA, I. Bayesian scene segmentation incorporating motion constraints and category-specific information. In: RANCHORDAS, A.; ARAúJO, H. (Ed.). **VISAPP (1)**. INSTICC Press, 2009. p. 291–298. ISBN 978-989-8111-69-2. Available from: <<http://dblp.uni-trier.de/db/conf/visapp/visapp2009-1.html#BachmannL09>>.

BADINO, H.; KANADE, T. A head-wearable short-baseline stereo system for the simultaneous estimation of structure and motion. In: **IAPR Conference on Machine Vision Application**. [S.l.: s.n.], 2011. p. 185–189.

BADINO, H.; YAMAMOTO, A.; KANADE, T. Visual odometry by multi-frame feature integration. In: **First International Workshop on Computer Vision for Autonomous Driving at ICCV**. [S.l.: s.n.], 2013.

BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Speeded-up robust features (surf). **Comput. Vis. Image Underst.**, Elsevier Science Inc., New York, NY, USA, v. 110, n. 3, p. 346–359, jun. 2008. ISSN 1077-3142. Available from: <<http://dx.doi.org/10.1016/j.cviu.2007.09.014>>.

BAYES, M.; PRICE, M. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s. **Philosophical Transactions (1683-1775)**, The Royal Society, v. 53, p. pp. 370–418, 1763. ISSN 02607085. Available from: <<http://www.jstor.org/stable/105741>>.

BEHLEY, J.; STEINHAGE, V.; CREMERS, A. Laser-based segment classification using a mixture of bag-of-words. In: **Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on**. [S.l.: s.n.], 2013. p. 4195–4200. ISSN 2153-0858.

BELENO, R.; VITOR, G. B.; FERREIRA, J. V.; MEIRELLES, P. Proposta de uma plataforma

de testes para o desenvolvimento de veículos autônomos. In: **XIX Congresso Brasileiro de Automática (CBA)**. [S.l.: s.n.], 2012. v. 1, p. 2735–2742. ISBN 978-85-8001-069-5.

BELENO, R. D. H.; VITOR, G. B.; FERREIRA, J. V.; MEIRELLES, P. S. Planeacion y seguimiento de trayectorias de un vehiculo terrestre com base en el control de dirección en un ambiente real. In: **Primer Congreso Internacional sobre Tecnologias Avanzadas de Mecatronica, Diseño y Manofactura (AMDM)**. [S.l.: s.n.], 2013.

BENENSON, R. **Perception for driverless vehicles in urban environment**. PhD Thesis (PhD) — Ecole des Mines de Paris, 2008. Available from: <<http://pastel.paristech.org/5327>>.

BEUCHER, S.; BILODEAU, M. Road segmentation and obstacle detection by a fast watershed transformation. In: **Intelligent Vehicles '94 Symposium, Proceedings of the**. [S.l.: s.n.], 1994. p. 296–301.

BHALLA, K. et al. **Transport for Health: The Global burden of disease from motorized road transport**. [S.l.], 12 2014.

BILESCHI, S. M. **StreetScenes : towards scene understanding in still images**. PhD Thesis (PhD) — Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2006. Available from: <<http://hdl.handle.net/1721.1/37896>>.

BORENSTEIN, J.; KOREN, Y. Histogramic In-Motion Mapping for Mobile Robot Obstacle Avoidance. **IEEE Transactions on Robotics and Automation**, v. 7, n. 4, p. 535–539, 1991. Available from: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.8.1317>>.

BOURGAULT, F.; MAKARENKO, A.; WILLIAMS, S.; GROCHOLSKY, B.; DURRANT-WHYTE, H. Information based adaptive robotic exploration. In: **Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on**. [S.l.: s.n.], 2002. v. 1, p. 540–545 vol.1.

BRESENHAM, J. E. Algorithm for computer control of a digital plotter. **IBM Systems Journal**, v. 4, n. 1, p. 25–30, 1965. ISSN 0018-8670.

BROGGI, A.; CARAFFI, C.; FEDRIGA, R. I.; GRISLERI, P. Obstacle detection with stereo vision for off-road vehicle navigation. In: **Proceedings of the International IEEE Workshop on Machine Vision for Intelligent Vehicles**. [S.l.: s.n.], 2005. p. 1–8.

BROGGI, A.; CERRI, P.; GHIDONI, S.; GRISLERI, P.; JUNG, H. G. A new approach to urban pedestrian detection for automatic braking. **IEEE Transactions on Intelligent Transportation Systems**, v. 10, n. 4, p. 594–605, 2009.

BROGGI, A.; MEDICI, P.; ZANI, P.; COATI, A.; PANCIROLI, M. Autonomous vehicles control in the vislab intercontinental autonomous challenge. **Annual Reviews in Control**, v. 36, n. 1, p. 161 – 171, 2012. ISSN 1367-5788. Available from: <<http://www.sciencedirect.com/science/article/pii/S1367578812000132>>.

BROSTOW, G. J.; SHOTTON, J.; FAUQUEUR, J.; CIPOLLA, R. Segmentation and recognition using structure from motion point clouds. In: FORSYTH, D.; TORR, P.; ZISSERMAN, A. (Ed.). **Computer Vision - ECCV 2008**. [S.l.]: Springer Berlin Heidelberg, 2008, (Lecture Notes in Computer Science, v. 5302). p. 44–57. ISBN 978-3-540-88681-5.

BUEHLER, M.; IAGNEMMA, K.; SINGH, S. **The 2005 DARPA Grand Challenge: The Great Robot Race**. 1rd. ed. [S.l.]: Springer Tracts in Advanced Robotics, 2007.

BUEHLER, M.; IAGNEMMA, K.; SINGH, S. **The DARPA Urban Challenge: Autonomous Vehicles in City Traffic**. [S.l.]: Springer Tracts in Advanced Robotics, 2010.

CANAS, J.; MATELLáN, V. Dynamic gridmaps: comparing building techniques. **Mathware and Soft Computing**, XIII, n. 1, p. 5–22, 2006. ISSN 1134-5632. Available from: <<http://gsyc.es/jmplaza/papers/gridmaps2006.pdf>>.

CHAPOULIE, A. **Contributions aux méthodes de détection visuelle de fermeture de boucle et de segmentation topologique de l'environnement.** Date de rédaction: 2012. PhD Thesis (THESE) — Université Nice Sophia Antipolis, dec. 2012. Available from: <<http://tel.archives-ouvertes.fr/tel-00764868>>.

CHAPOULIE, A.; RIVES, P.; FILLIAT, D. A spherical representation for efficient visual loop closing. In: **Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on**. [S.l.: s.n.], 2011. p. 335–342.

CHERUBINI, A.; CHAUMETTE, F. Visual navigation with obstacle avoidance. In: **Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on**. [S.l.: s.n.], 2011. p. 1593–1598. ISSN 2153-0858.

CHERUBINI, A.; SPINDLER, F.; CHAUMETTE, F. A new tentacles-based technique for avoiding obstacles during visual navigation. In: **Robotics and Automation (ICRA), 2012 IEEE International Conference on**. [S.l.: s.n.], 2012. p. 4850–4855. ISSN 1050-4729.

CHOSET, H. et al. **Principles of Robot Motion: Theory, Algorithms, and Implementations**. Cambridge, MA: MIT Press, 2005.

COMANICIU, D.; MEER, P. Mean shift: a robust approach toward feature space analysis. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 24, n. 5, p. 603–619, 2002. ISSN 0162-8828.

COMPORT, A.; MALIS, E.; RIVES, P. Real-time quadrifocal visual odometry. **The International Journal of Robotics Research**, v. 29, n. 2-3, p. 245–266, 2010. Available from: <<http://ijr.sagepub.com/content/29/2-3/245.abstract>>.

COUÉ, C.; PRADALIER, C.; LAUGIER, C.; FRAICHARD, T.; BESSIERE, P. Bayesian Occupancy Filtering for Multitarget Tracking: an Automotive Application. **International Journal of Robotics Research**, ?, v. 25, n. 1, p. 19–30, jan. 2006. Voir basilic : http://emotion.inrialpes.fr/bibemotion/2006/CPLFB06/. Available from: <<http://hal.inria.fr/inria-00182004>>.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: **In CVPR**. [S.l.: s.n.], 2005. p. 886–893.

DAME, A.; MARCHAND, r. Accurate real-time tracking using mutual information. In: **ISMAR**. IEEE, 2010. p. 47–56. Available from: <<http://dblp.uni-trier.de/db/conf/ismar/ismar2010.html#DameM10>>.

DARBON, J.; AKGUL, C. B. An efficient algorithm for attribute openings and closings. In: **In Proceedings of the 13th European Signal Processing Conference (EUSIPCO), Electronic proceedings**. [S.l.: s.n.], 2005.

DAVISON, A.; MURRAY, D. Simultaneous localization and map-building using active vision. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 24, n. 7, p. 865–880, Jul 2002. ISSN 0162-8828.

DEMPSTER, A. P. A generalization of bayesian inference. **Journal of the Royal Statistical Society**, v. 30, n. B, p. 205–247, 1968.

DEMPSTER, A. P. A generalization of bayesian inference. **Journal of the Royal Statistical Society**, v. 30, n. 1, p. 205–247, 1986.

DESAI, C.; RAMANAN, D.; FOWLKES, C.; IRVINE, U. C. Discriminative models for multi-class object layout. In: **In Proc. IEEE International Conf. on Computer Vision (ICCV)**. [S.l.: s.n.], 2009.

DICKMANNS, E.; MYSLIWETZ, B. Recursive 3-d road and relative ego-state recognition. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 14, n. 2, p. 199–213, Feb 1992. ISSN 0162-8828.

DOUGHERTY, E. R.; LOTUFO, R. A. **Hands-on Morphological Image Processing (SPIE Tutorial Texts in Optical Engineering Vol. TT59)**. [S.l.]: SPIE Publications, 2003. Paperback. ISBN 081944720X.

DURRANT-WHYTE, H.; BAILEY, T. Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms. **Robotics & Automation Magazine, IEEE**, IEEE, v. 13, n. 2, p. 99–110, jun. 2006. ISSN 1070-9932. Available from: <<http://dx.doi.org/10.1109/mra.2006.1638022>>.

EINECKE, N.; EGGERT, J. Block-matching stereo with relaxed fronto-parallel assumption. In: **IV**. [S.l.: s.n.], 2014. p. 700–705.

ELFES, A. A tesselated probabilistic representation for spatial robot perception and navigation. In: **In Proceedings of the NASA Conference on Space Telerobotics**. JPL, California Inst. of Tech: [s.n.], 1989. v. 2, p. 341–350.

ELFES, A. Using occupancy grids for mobile robot perception and navigation. **Computer**, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 22, n. 6, p. 46–57, 1989. ISSN 0018-9162.

ELFES, A. Occupancy grids: A stochastic spatial representation for active robot perception. In: IYENGAR, S. S.; ELFES, A. (Ed.). **Autonomous Mobile Robots: Perception, Mapping, and Navigation (Vol. 1)**. Los Alamitos, CA: IEEE Computer Society Press, 1991. p. 60–70.

ESS, A.; MUELLER, T.; GRABNER, H.; GOOL, L. J. V. Segmentation-based urban traffic scene understanding. In: **BMVC**. [S.l.: s.n.], 2009. p. 1–11.

EVERINGHAM, M.; GOOL, L. V.; WILLIAMS, C. K. I.; WINN, J.; ZISSERMAN, A. The pascal visual object classes (voc) challenge. **International Journal of Computer Vision**, v. 88, n. 2, p. 303–338, jun. 2010.

FAGNANT, D. J.; KOCKELMAN, K. M. **Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations**. [S.l.], 2013. Available from: <<https://www.enotrans.org/wp-content/uploads/wpsc/downloadables/AV-paper.pdf>>.

FAUGERAS, O. **Three-dimensional computer vision: A geometric view point**. Cambridge: MIT Press, 1993.

FAYAD, F.; CHERFAOUI, V. Tracking objects using a laser scanner in driving situation based on modeling target shape. In: **Intelligent Vehicles Symposium, 2007 IEEE**. [S.l.: s.n.], 2007. p. 44–49. ISSN 1931-0587.

FEITEN, W.; BAUER, R.; LAWITZKY, G. Robust obstacle avoidance in unknown and cramped environments. In: **Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on**. [S.l.: s.n.], 1994. p. 2412–2417 vol.3.

FIORINI, P.; SHILLERT, Z. Motion planning in dynamic environments using velocity

obstacles. **International Journal of Robotics Research**, v. 17, p. 760–772, 1998.

FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. **Commun. ACM**, ACM, New York, NY, USA, v. 24, n. 6, p. 381–395, jun. 1981. ISSN 0001-0782. Available from: <<http://doi.acm.org/10.1145/358669.358692>>.

FLOROS, G.; LEIBE, B. Joint 2d-3d temporally consistent semantic segmentation of street scenes. In: **Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on**. [S.l.: s.n.], 2012. p. 2823–2830. ISSN 1063-6919.

FOX, D.; BURGARD, W.; THRUN, S. The dynamic window approach to collision avoidance. **IEEE Robotics and Automation Magazine**, v. 4, p. 23–33, March 1997.

FRITSCH, J.; KUEHNL, T.; GEIGER, A. A new performance measure and evaluation benchmark for road detection algorithms. In: **International Conference on Intelligent Transportation Systems (ITSC)**. [S.l.: s.n.], 2013.

FULGENZI, C. **Navigation autonome en environnement dynamique utilisant des modèles probabilistes de perception et de prédiction du risque de collision.** PhD Thesis (THESE) — Institut National Polytechnique de Grenoble, jun. 2009. Available from: <<http://tel.archives-ouvertes.fr/index.php?halsid=0s6i4gljuuuql8u3l5adoceh02&view_this_doc=tel-00398055&version=1>>.

FULGENZI, C.; TAY, C.; SPALANZANI, A.; LAUGIER, C. Probabilistic navigation in dynamic environment using rapidly-exploring random trees and gaussian processes. In: **Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on**. [S.l.: s.n.], 2008. p. 1056–1062.

GATE, G. **Reliable Perception of Highly Changing Environments: Implementations for Car-to-Pedestrian Collision Avoidance Systems**. PhD Thesis (THESE) — Ecole Nationale Supérieure des Mines de Paris, dec. 2009. Available from: <<http://tel.archives-ouvertes.fr/docs/00/50/14/59/PDF/These_Gate.pdf>>.

GEIGER, A. **Probabilistic Models for 3D Urban Scene Understanding from Movable Platforms**. PhD Thesis (PhD) — Karlsruher Institut für Technologie (KIT), 2013. Available from: <<http://www.cvlibs.net/publications.php>>.

GEIGER, A. et al. Team annieway's entry to the 2011 grand cooperative driving challenge. **Intelligent Transportation Systems, IEEE Transactions on**, v. 13, n. 3, p. 1008–1017, Sept 2012. ISSN 1524-9050.

GEIGER, A.; LENZ, P.; STILLER, C.; URTASUN, R. Vision meets robotics: The kitti dataset.

International Journal of Robotics Research (IJRR), 2013.

GEIGER, A.; ZIEGLER, J.; STILLER, C. Stereoscan: Dense 3d reconstruction in real-time. In: **IEEE Intelligent Vehicles Symposium**. Baden-Baden, Germany: [s.n.], 2011.

GELAS, J.; HILUY, J.; MOTA, J. **Anales du III Forum BRAFITEC: CAPES/CDEFI**. CAPES, 2007. Available from: <<http://books.google.fr/books?id=HS9tMwEACAAJ>>.

GIESBRECHT, J. **Global Path Planning for Unmanned Ground Vehicles**. PO Box 4000, Station Main,Medicine Hat, AB,CA,T1A 8K6, December 2004.

GOULD, S.; FULTON, R.; KOLLER, D. Decomposing a scene into geometric and semantically consistent regions. In: **Computer Vision, 2009 IEEE 12th International Conference on**. [S.l.: s.n.], 2009. p. 1–8. ISSN 1550-5499.

GUO, R.; HOIEM, D. Beyond the line of sight: Labeling the underlying surfaces. In: **ECCV (5)**. [S.l.: s.n.], 2012. p. 761–774.

HäHNEL, D.; BURGARD, W.; WEGBREIT, B.; THRUN, S. Towards lazy data association in SLAM. In: **Proceedings of the 11th International Symposium of Robotics Research (ISRR'03)**. Sienna, Italy: Springer, 2003.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: **In Proc. of Fourth Alvey Vision Conference**. [S.l.: s.n.], 1988. p. 147–151.

HAUTIERE, N.; LABAYRADE, R.; PERROLLAZ, M.; AUBERT, D. Road scene analysis by stereovision: a robust and quasi-dense approach. In: **Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on**. [S.l.: s.n.], 2006. p. 1–6.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation (2nd Edition)**. 2. ed. Prentice Hall, 1998. Hardcover. ISBN 0132733501. Available from: <<http://www.worldcat.org/isbn/0132733501>>.

HERRMANN, D.; KAMPHANS, T.; LANGETEPE, E. Exploring simple triangular and hexagonal grid polygons online. **CoRR**, abs/1012.5253, 2010.

HU, Z.; UCHIMURA, K. U-v-disparity: an efficient algorithm for stereo vision based scene analysis. In: **Intelligent Vehicles Symposium (IV),IEEE**. [S.l.: s.n.], 2005. p. 48–54.

HUNDELSHAUSEN, F.; HIMMELSBACH, M.; HECKER, F.; MUELLER, A.; WUENSCHE, H.-J. Driving with tentacles - integral structures for sensing and motion. In: BUEHLER, M.; IAGNEMMA, K.; SINGH, S. (Ed.). **The DARPA Urban Challenge**.

Springer Berlin Heidelberg, 2009, (Springer Tracts in Advanced Robotics, v. 56). p. 393–440. ISBN 978-3-642-03990-4. Available from: <<http://dx.doi.org/10.1007/978-3-642-03991-1_10>>.

INTERMEDIATE Report on the Security of the Connected Car. [S.l.], 09 2012. Available from: <<http://www.syssec-project.eu/m/page-media/3/syssec-d6.2-SecurityOfTheConnectedCar.pdf>>.

JAZAR, R. Steering dynamics. In: **Vehicle Dynamics: Theory and Application**. Springer US, 2008. p. 379–454. ISBN 978-0-387-74243-4. Available from: <<http://dx.doi.org/10.1007/978-0-387-74244-1_7>>.

JULIER, S.; UHLMANN, J. Unscented filtering and nonlinear estimation. **Proceedings of the IEEE**, v. 92, n. 3, p. 401–422, Mar 2004. ISSN 0018-9219.

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Transactions of the ASME, Journal of Basic Engineering**, v. 82, p. 35–45, 1960.

KE-KE, W.; HAN-QING, Z.; QIANG, L.; WEI, Z. A motion planning algorithm for autonomous land vehicle based on virtual tentacles. In: **Control and Decision Conference (CCDC), 2011 Chinese**. [S.l.: s.n.], 2011. p. 2431–2435.

KHATIB, O. Real-time obstacle avoidance for manipulators and mobile robots. In: **Robotics and Automation. Proceedings. 1985 IEEE International Conference on**. [S.l.: s.n.], 1985. v. 2, p. 500–505.

KLEIN, G.; MURRAY, D. Parallel tracking and mapping for small AR workspaces. In: **Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07)**. Nara, Japan: [s.n.], 2007.

KOJIMA, K.; RYAN, L. **TRANSPORT ENERGY EFFICIENCY**. [S.l.], 09 2010.

KONG, H.; AUDIBERT, J.-Y.; PONCE, J. General road detection from a single image. **Image Processing, IEEE Transactions on**, v. 19, n. 8, p. 2211–2220, 2010. ISSN 1057-7149.

KORBES, A.; VITOR, G.; LOTUFO, R. de A.; FERREIRA, J. Advances on watershed processing on gpu architecture. In: SOILLE, P.; PESARESI, M.; OUZOUNIS, G. (Ed.). **Mathematical Morphology and Its Applications to Image and Signal Processing**. [S.l.]: Springer Berlin / Heidelberg, 2011, (Lecture Notes in Computer Science, v. 6671). p. 260–271. ISBN 978-3-642-21568-1.

KRÄHENBÜHL, P.; KOLTUN, V. Efficient inference in fully connected crfs with gaussian edge potentials. **CoRR**, abs/1210.5644, 2012.

KUEHNL, T.; KUMMERT, F.; FRITSCH, J. Spatial ray features for real-time ego-lane extraction. In: **Proc. IEEE Intelligent Transportation Systems**. [S.l.: s.n.], 2012.

KUMAR, S.; HEBERT, M. Discriminative random fields: a discriminative framework for contextual interaction in classification. In: **Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on**. [S.l.: s.n.], 2003. p. 1150–1157 vol.2.

KUMAR, S.; HEBERT, M. Man-made structure detection in natural images using a causal multiscale random field. In: **Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on**. [S.l.: s.n.], 2003. v. 1, p. I–119–I–126 vol.1. ISSN 1063-6919.

KURDEJ, M.; MORAS, J.; CHERFAOUI, V.; BONNIFAIT, P. Controlling Remanence in Evidential Grids Using Geodata for Dynamic Scene Perception. **International Journal of Approximate Reasoning**, v. 55, n. 1, p. 355–375, jan. 2014.

LABAYRADE, R.; AUBERT, D.; TAREL, J. P. Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation. In: **Proceedings of the IEEE Symposium on Intelligent Vehicles**. [S.l.: s.n.], 2002. v. 2, p. 646–651.

LADICKY, L.; RUSSELL, C.; KOHLI, P.; TORR, P. H. S. Associative hierarchical crfs for object class image segmentation. In: **Computer Vision, 2009 IEEE 12th International Conference on**. [S.l.: s.n.], 2009. p. 739–746. ISSN 1550-5499.

LADICKÝ, L'. et al. Joint optimisation for object class segmentation and dense stereo reconstruction. In: **Proceedings of the British Machine Vision Conference**. [S.l.]: BMVA Press, 2010. p. 104.1–104.11. ISBN 1-901725-40-5. Doi:10.5244/C.24.104.

LATOMBE, J.-C. **Robot Motion Planning**. Norwell, MA, USA: Kluwer Academic Publishers, 1991. ISBN 079239206X.

LAUMOND, J.-P. **Robot Motion Planning and Control**. Berlin: Springer-Verlag, 1998. Available online. Available from: <<http://www.laas.fr/$\sim$jpl/book.html>>.

LEVINSON, J.; THRUN, S. Robust vehicle localization in urban environments using probabilistic maps. In: **Robotics and Automation (ICRA), 2010 IEEE International Conference on**. [S.l.: s.n.], 2010. p. 4372–4378. ISSN 1050-4729.

LEVITT, T. S.; LAWTON, D. T. Qualitative navigation for mobile robots. **Artif. Intell.**, Elsevier Science Publishers Ltd., Essex, UK, v. 44, n. 3, p. 305–360, jul. 1990. ISSN 0004-3702. Available from: <<http://dx.doi.org/10.1016/0004-3702(90)90027-W>>.

LIMA, D. A.; VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. A disparity map

refinement to enhance weakly-textured urban environment data. In: **International Conference on Advanced Robotics (ICAR), 2013 IEEE**. [S.l.: s.n.], 2013.

LIN, K.-H.; WANG, C.-C. Stereo-based simultaneous localization, mapping and moving object tracking. In: **Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on**. [S.l.: s.n.], 2010. p. 3975–3980. ISSN 2153-0858.

LITMAN, T. **Autonomous Vehicle Implementation Predictions: Implications for Transport Planning**. [S.l.], 01 2014. 36-42 p. Available from: <<http://www.vtpi.org/avip.pdf>>.

LOWE, D. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, Kluwer Academic Publishers, v. 60, n. 2, p. 91–110, 2004. ISSN 0920-5691. Available from: <<http://dx.doi.org/10.1023/B%3AVISI.0000029664.99615.94>>.

LUETTEL, T.; HIMMELSBACH, M.; WUENSCHE, H. J. Autonomous ground vehicles - concepts and a path to the future. **Proceedings of the IEEE**, v. 100, n. Special Centennial Issue, p. 1831–1839, May 2012. ISSN 0018-9219.

MAROTI, A.; SZALOKI, D.; KISS, D.; TEVESZ, G. Investigation of dynamic window based navigation algorithms on a real robot. In: **Applied Machine Intelligence and Informatics (SAMI), 2013 IEEE 11th International Symposium on**. [S.l.: s.n.], 2013. p. 95–100.

MEIJSTER, A.; WILKINSON, M. H. F. A comparison of algorithms for connected set openings and closings. **IEEE TRANS. PATT. ANAL. MACH. INTELL**, v. 24, n. 4, p. 484–494, 2002.

MEILLAND, M. **Cartographie RGB-D dense pour la localisation visuelle temps-réel et la navigation autonome**. PhD Thesis (THESE) — Ecole Nationale Supérieure des Mines de Paris, mar. 2012. Available from: <<http://tel.archives-ouvertes.fr/tel-00686803>>.

MEILLAND, M.; COMPORT, A. I.; RIVES, P. A spherical robot-centered representation for urban navigation. In: **IROS'10**. [S.l.: s.n.], 2010. p. 5196–5201.

MEILLAND, M.; COMPORT ANDREW, I.; RIVES, P. Dense omnidirectional RGB-D mapping of large scale outdoor environments for real-time localisation and autonomous navigation. **Journal of Field Robotics**, Wiley, 2014. Available from: <<http://hal.inria.fr/hal-01010429>>.

MINGUEZ, J. The obstacle-restriction method for robot obstacle avoidance in difficult environments. In: **Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on**. [S.l.: s.n.], 2005. p. 2284–2290.

MINGUEZ, J.; MONTANO, L. Nearness diagram navigation (nd): A new real time collision avoidance approach. In: **In Proc. of the ieee/rsj international conference on intelligent robots and systems (iros'00**. [S.l.: s.n.], 2000. p. 2094–2100.

MINGUEZ, J.; MONTANO, L. Nearness diagram (nd) navigation: collision avoidance in troublesome scenarios. **Robotics and Automation, IEEE Transactions on**, v. 20, n. 1, p. 45–59, Feb 2004. ISSN 1042-296X.

MIYASAKA, T.; OHAMA, Y.; NINOMIYA, Y. Ego-motion estimation and moving object tracking using multi-layer lidar. In: **Intelligent Vehicles Symposium, 2009 IEEE**. [S.l.: s.n.], 2009. p. 151–156. ISSN 1931-0587.

MONTEMERLO, M. et al. Junior: The stanford entry in the urban challenge. **Journal of Field Robotics**, John Wiley and Sons Ltd., Chichester, UK, v. 25, n. 9, p. 569–597, sep. 2008. ISSN 1556-4959. Available from: <<http://dx.doi.org/10.1002/rob.v25:9>>.

MONTEMERLO, M.; THRUN, S.; KOLLER, D.; WEGBREIT, B. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In: **Proceedings of the AAAI National Conference on Artificial Intelligence**. Edmonton, Canada: AAAI, 2002.

MORAS, J. **Grilles de perception évidentielles pour la navigation robotique en milieu urbain**. PhD Thesis (THESE) — Université de Technologie de Compiègne, Jan 2013. Available from: <<http://tel.archives-ouvertes.fr/tel-00866300>>.

MORAS, J.; CHERFAOUI, V.; BONNIFAIT, P. Credibilist occupancy grids for vehicle perception in dynamic environments. In: **Robotics and Automation (ICRA), 2011 IEEE International Conference on**. [S.l.: s.n.], 2011. p. 84–89. ISSN 1050-4729.

MORAS, J.; CHERFAOUI, V.; BONNIFAIT, P. Moving objects detection by conflict analysis in evidential grids. In: **Intelligent Vehicles Symposium (IV), 2011 IEEE**. [S.l.: s.n.], 2011. p. 1122–1127. ISSN 1931-0587.

MORAS, J. et al. Drivable space characterization using automotive lidar and georeferenced map information. In: **Intelligent Vehicles Symposium (IV), 2012 IEEE**. [S.l.: s.n.], 2012. p. 778 –783. ISSN 1931-0587.

MOTLAGH, O. R. E.; HONG, T. S.; ISMAIL, N. Development of a new minimum avoidance system for a behavior-based mobile robot. **Fuzzy Sets and Systems**, v. 160, n. 13, p. 1929 – 1946, 2009. ISSN 0165-0114. Theme: Information Processing and Applications. Available from: <<http://www.sciencedirect.com/science/article/pii/S0165011408004417>>.

MUJAHAD, M.; FISCHER, D.; MERTSCHING, B.; JADDU, H. Closest gap based (cg) reactive obstacle avoidance navigation for highly cluttered environments. In: **Intelligent**

Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. [S.l.: s.n.], 2010. p. 1805–1812. ISSN 2153-0858.

MUJAHED, M.; JADDU, H.; FISCHER, D.; MERTSCHING, B. Tangential closest gap based (tcg) reactive obstacle avoidance navigation for cluttered environments. In: **Safety, Security, and Rescue Robotics (SSRR), 2013 IEEE International Symposium on**. [S.l.: s.n.], 2013. p. 1–6.

NGUYEN, T.-N.; MICHAELIS, B.; AL-HAMADI, A.; TORNOW, M.; MEINECKE, M. Stereo-camera-based urban environment perception using occupancy grid and object tracking. **Intelligent Transportation Systems, IEEE Transactions on**, v. 13, n. 1, p. 154–165, March 2012. ISSN 1524-9050.

NILSSON, N. J. **Shakey The Robot**. 333 Ravenswood Ave., Menlo Park, CA 94025, Apr 1984.

ORDONEZ, C.; JR., E. G. C.; SELEKWA, M. F.; DUNLAP, D. D. The virtual wall approach to limit cycle avoidance for unmanned ground vehicles. **Robotics and Autonomous Systems**, v. 56, n. 8, p. 645 – 657, 2008. ISSN 0921-8890. Available from: <<http://www.sciencedirect.com/science/article/pii/S0921889007001741>>.

OREBäCK, A.; CHRISTENSEN, H. Evaluation of architectures for mobile robotics. **Autonomous Robots**, Kluwer Academic Publishers, v. 14, n. 1, p. 33–49, 2003. ISSN 0929-5593. Available from: <<http://dx.doi.org/10.1023/A%3A1020975419546>>.

PAGAC, D.; NEBOT, E.; DURRANT-WHYTE, H. An evidential approach to map-building for autonomous vehicles. **Robotics and Automation, IEEE Transactions on**, v. 14, n. 4, p. 623–629, 1998. ISSN 1042-296X.

PETROVSKAYA, A.; THRUN, S. Model based vehicle detection and tracking for autonomous urban driving. **Autonomous Robots**, Springer US, v. 26, n. 2-3, p. 123–139, 2009. ISSN 0929-5593. Available from: <<http://dx.doi.org/10.1007/s10514-009-9115-1>>.

POLIMENI, A.; VITETTA, A. A procedure for an integrated network and vehicle routing optimisation problem. **Procedia - Social and Behavioral Sciences**, v. 54, n. 0, p. 65 – 74, 2012. ISSN 1877-0428. Proceedings of {EWGT2012} - 15th Meeting of the {EURO} Working Group on Transportation, September 2012, Paris. Available from: <<http://www.sciencedirect.com/science/article/pii/S1877042812041882>>.

POMERLEAU, D. Ralph: Rapidly adapting lateral position handler. In: **IEEE Symposium on Intelligent Vehicles**. [S.l.: s.n.], 1995. p. 506 – 511.

QUINLAN, S.; KHATIB, O. Elastic bands: connecting path planning and control. In:

Robotics and Automation, 1993. Proceedings., 1993 IEEE International Conference on. [S.l.: s.n.], 1993. p. 802–807 vol.2.

RASMUSSEN, C. Grouping dominant orientations for ill-structured road following. In: **Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on**. [S.l.: s.n.], 2004. v. 1, p. I–470–I–477 Vol.1. ISSN 1063-6919.

RIVEST, J.-F.; SOILLE, P.; BEUCHER, S. Morphological gradients. **Journal of Electronic Imaging**, v. 2, p. 326–336, 1993.

ROERDINK, J. B. T. M.; MEIJSTER, A. **The Watershed Transform: Definitions, Algorithms and Parallelization Strategies**. 2001.

SANTA, J. et al. Comprehensive vehicular networking platform for v2i and v2v communications within the walkie-talkie project. **International Journal of Distributed Sensor Networks**, v. 2013, 2013. Available from: <<http://www.hindawi.com/journals/ijdsn/2013/676850/cta/>>.

SERRA, J. **Image Analysis and Mathematical Morphology**. 1rd. ed. [S.l.]: Academic Press, 1982.

SERRA, J. C.; SALEMBIER, P. Connected operators and pyramids. In: **Image Algebra and Morphological Image Processing IV**. [S.l.: s.n.], 1993. p. 65–76.

SHAFER, G. **A Mathematical Theory of Evidence**. [S.l.]: Princeton University Press, 1976. (Limited paperback editions).

SHALOM, Y.; BLAIR, W.; CALIFORNIA, L. A. U. E. University of. **Multitarget/Multisensor Tracking: Applications and Advances**. Artech House, Incorporated, 2000. (Multitarget-multisensor Tracking). ISBN 9781580530910. Available from: <<http://books.google.fr/books?id=-QB0RQAACAAJ>>.

SHINZATO, P. Y.; WOLF, D. F. A road following approach using artificial neural networks combinations. **J. Intell. Robotics Syst.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 62, n. 3-4, p. 527–546, jun. 2011. ISSN 0921-0296. Available from: <<http://dx.doi.org/10.1007/s10846-010-9463-2>>.

SHINZATO, P. Y.; WOLF, D. F.; STILLER, C. Road terrain detection: Avoiding common obstacle detection assumptions using sensor fusion. In: **Intelligent Vehicles Symposium (IV)**. [S.l.: s.n.], 2014.

SHOTTON, J.; WINN, J.; ROTHER, C.; CRIMINISI, A. **TextonBoost for Image**

Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context. 2007.

SHOTTON, J.; WINN, J. M.; ROTHER, C.; CRIMINISI, A. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. **International Journal of Computer Vision**, v. 81, n. 1, p. 2–23, 2009.

SILVEIRA, G.; MALIS, E.; RIVES, P. An efficient direct approach to visual slam. **Robotics, IEEE Transactions on**, v. 24, n. 5, p. 969–979, Oct 2008. ISSN 1552-3098.

SIMMONS, R. The curvature-velocity method for local obstacle avoidance. In: **Proceedings of the IEEE International Conference on Robotics and Automation**. [S.l.: s.n.], 1996. p. 3375–3382.

SMETS, P. Constructing the pignistic probability function in a context of uncertainty. In: **Uncertainty in Artificial Intelligence 5 Annual Conference on Uncertainty in Artificial Intelligence (UAI-89)**. Amsterdam, NL: Elsevier Science, 1989. p. 29–39.

SMETS, P.; KENNES, R. The transferable belief model. **Artif. Intell.**, v. 66, n. 2, p. 191–234, 1994. Available from: <<http://dblp.uni-trier.de/db/journals/ai/ai66.html#SmetsK94>>.

SMITH, R. C.; CHEESEMAN, P. On the representation and estimation of spatial uncertainty. **The International Journal of Robotics Research**, v. 5, n. 4, p. 56–68, 1986. Available from: <<http://ijr.sagepub.com/content/5/4/56.abstract>>.

SOQUET, N.; AUBERT, D.; HAUTIERE, N. Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation. In: **Proceedings of the IEEE Symposium on Intelligent Vehicles**. [S.l.: s.n.], 2007. p. 160–165. ISSN 1931-0587.

SOQUET, N.; PERROLLAZ, M.; LABAYRADE, R.; AUBERT, D. Free Space Estimation for Autonomous Navigation. In: **5th International Conference on Computer Vision Systems**. Bielefeld, Allemagne: [s.n.], 2007. Available from: <<http://hal.inria.fr/hal-00780658>>.

SPIES, R.; ABLAßMEIER, M.; BUBB, H.; HAMBERGER, W. Augmented interaction and visualization in the automotive domain. In: JACKO, J. (Ed.). **Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction**. Springer Berlin Heidelberg, 2009, (Lecture Notes in Computer Science, v. 5612). p. 211–220. ISBN 978-3-642-02579-2. Available from: <<http://dx.doi.org/10.1007/978-3-642-02580-8_23>>.

STEUX, B.; HAMZAOUI, O. E. tinyslam: A slam algorithm in less than 200 lines c-language program. In: **Control Automation Robotics Vision (ICARCV), 2010 11th International Conference on**. [S.l.: s.n.], 2010. p. 1975–1979.

STURGESS, P.; ALAHARI, K.; LADICKY, L.; TORR, P. H. S. Combining appearance and structure from motion features for road scene understanding. In: **BMVC**. [S.l.]: British Machine Vision Association, 2009.

TAN, C.; HONG, T.; CHANG, T.; SHNEIER, M. Color model-based real-time learning for road following. In: **Intelligent Transportation Systems Conference,ITSC**. [S.l.: s.n.], 2006. p. 939–944.

THRUN, S. Probabilistic robotics. **Commun. ACM**, ACM, New York, NY, USA, v. 45, n. 3, p. 52–57, mar. 2002. ISSN 0001-0782. Available from: <<http://doi.acm.org/10.1145/504729.504754>>.

THRUN, S.; BURGARD, W.; FOX, D. A probabilistic approach to concurrent mapping and localization for mobile robots. **Autonomous Robots**, Kluwer Academic Publishers, v. 5, n. 3-4, p. 253–271, 1998. ISSN 0929-5593. Available from: <<http://dx.doi.org/10.1023/A%3A1008806205438>>.

THRUN, S.; BURGARD, W.; FOX, D. **Probabilistic Robotics**. [S.l.]: The MIT Press, 2005.

TORRALBA, A.; MURPHY, K.; FREEMAN, W. Sharing visual features for multiclass and multiview object detection. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 29, n. 5, p. 854–869, May 2007. ISSN 0162-8828.

TU, Z.; CHEN, X.; YUILLE, A.; ZHU, S.-C. Image parsing: Unifying segmentation, detection, and recognition. **International Journal of Computer Vision**, Kluwer Academic Publishers, v. 63, n. 2, p. 113–140, 2005. ISSN 0920-5691. Available from: <<http://dx.doi.org/10.1007/s11263-005-6642-x>>.

TYKKALA, T.; COMPORT, A. A Dense Structure Model for Image Based Stereo SLAM. In: **IEEE International Conference on Robotics and Automation, ICRA'11**. Shanghai, China: [s.n.], 2011. Available from: <<http://www.i3s.unice.fr/~comport/publications/2011_ICRA_Tykkala.pdf>>.

ULRICH, I.; BORENSTEIN, J. Vfh*: local obstacle avoidance with look-ahead verification. In: **Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on**. [S.l.: s.n.], 2000. v. 3, p. 2505–2511 vol.3. ISSN 1050-4729.

URMSON, C. et al. **Tartan Racing: A Multi-Modal Approach to the DARPA Urban Challenge**. Pittsburgh, PA, April 2007.

VICTORINO, A. C. **La commande référencée capteur: une approche robuste au problème de navigation, localisation et cartographie simultanées pour un robot d'intérieur**. PhD Thesis (THESE) — L'Université de Nice-ophia Antipolis, sep. 2002.

Available from: <<http://opac.inria.fr/record=b1068912>>.

VINCENT, L. Morphological area openings and closings for grey-scale images. In: **Proc. of the Workshop on Mathematical Morphology and its Applications to Signal Processing**. Barcelona, Spain: [s.n.], 1993. p. 22–27.

VINCENT, L. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. **Image Processing, IEEE Transactions on**, v. 2, n. 2, p. 176–201, Apr 1993. ISSN 1057-7149.

VITOR, G. B. video containing the result of the thesis, **Autonomous Navigation using only a pair of stereo cameras - Heudiasyc laboratory**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 04/08/2014. Available from: <<http://youtu.be/shw1IRYd2T4>>.

VITOR, G. B. video containing the result of the thesis, **Comparison of the selection mechanism using virtual tentacle - KITTI database**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 01/10/2014. Available from: <<http://youtu.be/5PGnfB3djzQ>>.

VITOR, G. B. video containing the result of the thesis, **Dynamic Evidential Grid using Semantic Context - Heudiasyc database**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 04/08/2014. Available from: <<http://youtu.be/chLqC1r974k>>.

VITOR, G. B. video containing the result of the thesis, **Dynamic Evidential Grid using Semantic Context - KITTI database**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 20/07/2014. Available from: <<http://youtu.be/H_zJjX8uMtI>>.

VITOR, G. B. video containing the result of the thesis, **HistonBoost algorithm to urban scene understanding - KITTI database**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 05/08/2014. Available from: <<http://youtu.be/MsR3U5RUtxk>>.

VITOR, G. B. video containing the result of the thesis, **ProbBoost algorithm to urban scene understanding - KITTI database**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 05/08/2014. Available from: <<http://youtu.be/JxXIhpsC9Gk>>.

VITOR, G. B. video containing the result of the thesis, **Trajectory planning using virtual tentacle - Heudiasyc database**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 04/08/2014. Available from: <<http://youtu.be/YKYCwdUEaVs>>.

VITOR, G. B. video containing the result of the thesis, **Trajectory planning using virtual tentacle - KITTI database**. 2014. (Urban Environment Perception and Navigation using Robotic Vision). Accessed on: 04/08/2014. Available from: <<http://youtu.be/dBXHgGdbk80>>.

VITOR, G. B.; BELENO, R. D. H.; FERREIRA, J. V. Fuzzy application for mobile robot navigation using computer vision in real-time. In: **In proceedings on Congreso Internacional de Innovación y Tecnología: Sistemas Mecatrónicos.** [S.l.: s.n.], 2011.

VITOR, G. B.; KöRBES, A.; LOTUFO, R. D.; FERREIRA, J. V. Analysis of a step-based watershed algorithm using cuda. **International Journal of Natural Computing Research (IJNCR)**, IGI Global Disseminator of Knowledge, v. 1, p. 16–28, 2010. Available from: <<http://www.igi-global.com/article/analysis-step-based-watershed-algorithm/52612>>.

VITOR, G. B.; LIMA, D. A.; VICTORINO, A. C.; FERREIRA, J. V. A 2d/3d vision based approach applied to road detection in urban environments. In: **Intelligent Vehicles Symposium (IV), 2013 IEEE**. [S.l.: s.n.], 2013. p. 952–957.

VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. Comprehensive performance analysis of road detection algorithms using the common urban kitti-road benchmark. In: **Workshop on Benchmarking Road Terrain and Lane Detection Algorithms for In-Vehicle Application on IEEE Intelligent Vehicles Symposium (IV)**. [S.l.: s.n.], 2014.

VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. A histogram-based joint boosting classification for determining urban road (extended abstract). In: **International IEEE Conference on Intelligent Transportation Systems (ITSC)**. [S.l.: s.n.], 2014.

VITOR, G. B.; VICTORINO, A. C.; FERREIRA, J. V. A probabilistic distribution approach for the classification of urban roads in complex environments. In: **Workshop on Modelling, Estimation, Perception and Control of All Terrain Mobile Robots on IEEE International Conference on Robotics and Automation (ICRA)**. [S.l.: s.n.], 2014.

WANG, B.; FREMONT, V.; FLOREZ, S. A. R. Color-based road detection and its evaluation on the kitti road benchmark. In: **Workshop on Benchmarking Road Terrain and Lane Detection Algorithms for In-Vehicle Application, IEEE Intelligent Vehicles Symposium**. [S.l.: s.n.], 2014. p. 31–36.

WANG, M.; LIU, J.-K. Fuzzy logic based robot path planning in unknown environment. In: **Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on**. [S.l.: s.n.], 2005. v. 2, p. 813–818 Vol. 2.

WEIß, C. V2x communication in europe - from research projects towards standardization and field testing of vehicle communication technology. **Computer Networks**, v. 55, n. 14, p. 3103 – 3119, 2011. ISSN 1389-1286. Deploying vehicle-2-x communication. Available from:

<<http://www.sciencedirect.com/science/article/pii/S1389128611001198>>.

WHO. **Global status report on road safety: supporting a decade of action.** [S.l.], 2013. Available from:
<<http://www.who.int/iris/bitstream/10665/78256/1/9789241564564_eng.pdf?ua=1>>.

WIKIPEDIA. **Google driverless car - Wikipedia, the free encyclopedia**. aug. 2012. Accessed: 2014-07-25. Available from:
<<http://en.wikipedia.org/wiki/Google_driverless_car>>.

WILLIAMS, B. et al. A comparison of loop closing techniques in monocular {SLAM}. **Robotics and Autonomous Systems**, v. 57, n. 12, p. 1188 – 1197, 2009. ISSN 0921-8890. Inside Data Association. Available from:
<<http://www.sciencedirect.com/science/article/pii/S0921889009000876>>.

WOJEK, C.; SCHIELE, B. A dynamic conditional random field model for joint labeling of object and scene classes. In: FORSYTH, D.; TORR, P.; ZISSERMAN, A. (Ed.). **Computer Vision - ECCV 2008**. Springer Berlin Heidelberg, 2008, (Lecture Notes in Computer Science, v. 5305). p. 733–747. ISBN 978-3-540-88692-1. Available from:
<<http://dx.doi.org/10.1007/978-3-540-88693-8_54>>.

WU, H.; QIN, S.-Y. An approach to robot slam based on incremental appearance learning with omnidirectional vision. **International Journal of Systems Science**, v. 42, n. 3, p. 407–427, 2011. Available from: <<http://dx.doi.org/10.1080/00207720903572422>>.

XIE, J.; NASHASHIBI, F.; PARENT, M.; FAVROT, O. A real-time robust global localization for autonomous mobile robots in large environments. In: **Control Automation Robotics Vision (ICARCV), 2010 11th International Conference on**. [S.l.: s.n.], 2010. p. 1397–1402.

YAGER, R. R.; KACPRZYK, J.; FEDRIZZI, M. (Ed.). **Advances in the Dempster-Shafer Theory of Evidence**. [S.l.]: John Wiley & Sons, Inc., 1994. 5–34 p. ISBN 0-471-55248-8.

YANG, T.; AITKEN, V. Evidential mapping for mobile robots with range sensors. **Instrumentation and Measurement, IEEE Transactions on**, v. 55, n. 4, p. 1422–1429, Aug 2006. ISSN 0018-9456.

YU, H.; GONG, J.; IAGNEMMA, K.; JIANG, Y.; DUAN, J. Robotic wheeled vehicle ripple tentacles motion planning method. In: **Intelligent Vehicles Symposium (IV), 2012 IEEE**. [S.l.: s.n.], 2012. p. 1156–1161. ISSN 1931-0587.

YUN, S.; GUO-YING, Z.; YONG, Y. A road detection algorithm by boosting using feature combination. In: **Intelligent Vehicles Symposium, 2007 IEEE**. [S.l.: s.n.], 2007. p. 364–368. ISSN 1931-0587.

ZHANG, C.; WANG, L.; YANG, R. Semantic segmentation of urban scenes using dense depth maps. In: DANIILIDIS, K.; MARAGOS, P.; PARAGIOS, N. (Ed.). **Computer Vision - ECCV 2010**. Springer Berlin Heidelberg, 2010, (Lecture Notes in Computer Science, v. 6314). p. 708–721. ISBN 978-3-642-15560-4. Available from: <<http://dx.doi.org/10.1007/978-3-642-15561-1_51>>.

ZHANG, J.; NAGEL, H.-H. Texture-based segmentation of road images. In: **Intelligent Vehicles '94 Symposium, Proceedings of the**. [S.l.: s.n.], 1994. p. 260–265.

ZHU, A.; YANG, S. X. A fuzzy logic approach to reactive navigation of behavior-based mobile robots. In: **Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on**. [S.l.: s.n.], 2004. v. 5, p. 5045–5050 Vol.5. ISSN 1050-4729.

ZHU, L.; CHEN, Y.; LIN, Y.; LIN, C.; YUILLE, A. Recursive segmentation and recognition templates for image parsing. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE Computer Society, Los Alamitos, CA, USA, v. 34, n. 2, p. 359–371, 2012. ISSN 0162-8828.