# Wide-baseline Omnidirectional Stereovision for Intersection Monitoring

Sokèmi René Emmanuel Datondji

Normandie Université

# THÈSE

**Pour obtenir le diplôme de doctorat**

**Discipline :** *Informatique* **Spécialité**: *Vision par Ordinateur*

**Préparée au sein de l'Université de Rouen**

# Stéréovision Omnidirectionnelle Large Entraxe pour la Supervision d'Intersections Routières

Présentée et soutenue par

Sokèmi René Emmanuel DATONDJI

| **Thèse soutenue publiquement le 03/10/2017 devant le jury composé de** | | |
|---|---|---|
| Roland CHAPUIS | Professeur, Institut Pascal, Clermont Ferrand | Rapporteur |
| El Mustapha MOUADDIB | Professeur, laboratoire MIS, Université de Picardie Jules Verne | Rapporteur |
| Peter STURM | Directeur de Recherche, INRIA Grenoble, équipe STEEP | Examinateur |
| Sylvie TREUILLET | MCF-HdR, PRISME, Polytech Orléans | Examinatrice |
| Jean-Philippe TAREL | Chargé de Recherche, Lepsis, IFSTTAR | Examinateur |
| Peggy SUBIRATS | Responsable du groupe ESM, Cerema | Encadrante |
| Yohan DUPUIS | Chef de Projets, Responsable de l'Unité MTT, Cerema | Encadrant |
| Pascal VASSEUR | Professeur, LITIS, équipe STI, Université de Rouen | Directeur de thèse |

**Thèse dirigée par Pascal VASSEUR**

Cerema — Université de Rouen Normandie — LITIS — MIIS — Région Normandie

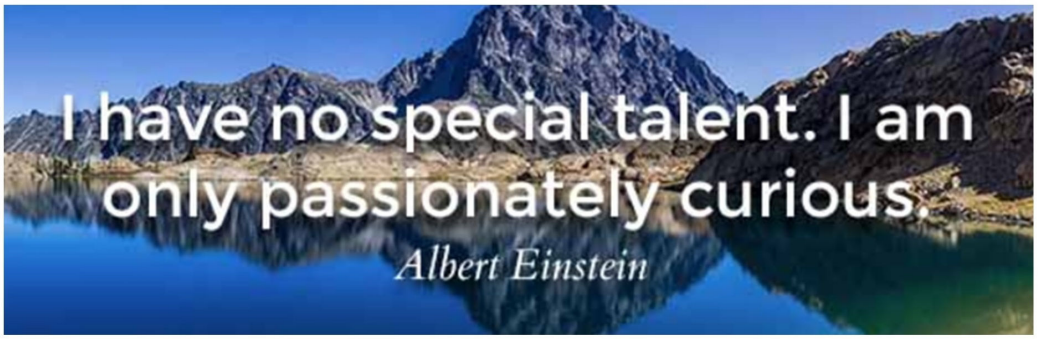
# Wide-Baseline Omnidirectional Stereovision for Intersection Monitoring

---

by

S. René E. DATONDJI

# Dedication




To my Family, with Love,

to my little angel,

and to all victims of terror and war

# Acknowledgement

After several years of intensive work, driven by a strong passion for innovation, I have finally reached a major step in my life and career. Last three years were thrilling, with succession of extreme deadlines, over-pressured sleepless nights, dramatic hours of scientific analysis, followed sometimes by incredible relief or euphoria at each new problem solved. Indeed, the road toward the Ph.D. has been a hard one. I did learn about science, computer vision, intelligent transportation systems, but even more about myself while often dealing with mixed feeling, doubts and fears. At each step, I have always found a positive way-out to overcome many challenges with optimism. Despite all the difficulties encountered, this experience was definitely fun, with plenty moments of happiness and enthusiasm. My dream would not have been possible without so many amazing people, who share the credit for the success of this project.

I would like to express my gratitude to my scientific director, Pascal VASSEUR, for his advice and expertise. None of this would have been possible without his genuine wisdom and exigency, his incredible moral support and encouragements. Whenever, I was struggling in my work, our discussions always led to bright ideas. Thank you Pascal, for your confidence and your availability. Thank you also for your precious proof-reading of our conference and journal publications. Finally, my special gratitude to you, for pushing me to my limits and even further.

I would like to thank my thesis advisors from the expertise center Cerema, Peggy SUBIRATS and Yohan DUPUIS for their continuous guidance. Particularly to you Yohan, I would like to express my gratitude for our excellent and inspiring talks, and for your active mentoring. I thank you both, dear advisors, for your kindness and your help on this adventure.

I would also like to express my gratitude to all the colleagues from the Department of Metrology of Traffic and Trajectories at Cerema, for their kind words and encouragements all over these years. My gratitude goes specially to all the super-colleagues, Alexandre, Azzedine, David, Olivier, Cyrille, who helped me for data acquisition, and with whom I've had constructive analysis.

I address my special thanks to my bright fellow doctoral colleagues and friends for their general advice and kindness: Rawia, Fabien, Kamaldeep, Danut, Imen, Marc-Aurele, Aymeric, Ahmad. Together we have shared our smart or crazy ideas, our anxieties, but also several joyful moments during pleasant afterwork evenings. I am also grateful to the members of the Intelligent Transportation Systems team at the Litis Lab, with whom I have had great discussions.

A special mention to Roland CHAPUIS, El Mustapha MOUADDIB, Peter STURM and Sylvie TREUILLET, Jean-Philippe TAREL, who accepted to review my dissertation and examine my research work. I am deeply honored to have you as jury members for my Ph.D defense.

I am also grateful to my previous advisors and colleagues at the research institute IRSEEM where, during my internships, my young passion for engineering, research and innovation, has been strengthened. Your advice have been of a great help at the early stage of my application process to pursue a Ph.D. degree, and even afterward. A very special thanks to Jean-Jacques DELARUE, for his confidence and for allowing me to gain a considerable teaching experience at ESIGELEC during my doctoral thesis.

I address my deep gratitude to my parents, Innocent and Joséphine, the best ever. Dear dad and mum, thank you for always pushing me forward and accepting nothing but excellence from me. Thank you again for instilling in me great set of moral, ethical and family values. You are both my greatest source of inspiration, personally and professionally. You always supported me to follow my dreams, throughout this thesis and my life in general. I owe this achievement all to you. I would also like to thank my brother and sisters, Chantal, Gaspard, Doris, Corine, for their support, encouragements, and for having my back. Despite the distance, you were all, always there for me, and I will always be there for you. Thank you infinitely.

To you my love, Esther, who had the patience to listen to me complain during these years, who stayed by my side as I was sometimes sick and exhausted by the work, who always helped me keep the balance in my life: thank you very much. Now I have to tell you, I will hopefully have more time to spend on weekends, so you'd better get ready to bear with me, more than usually. We will have fun!

Finally, I would like to thank all the people who, closely or remotely, have helped me throughout this amazing journey.

# Thank you all!

# Abstract

Visual surveillance of dynamic objects at road intersections has been an active research topic in the computer vision and intelligent transportations systems communities, over the past decades. Several projects have been carried out in order to enhance the safety of drivers in the special context of intersections. Our extensive review of related studies revealed that most roadside systems are based on monocular vision and provide output results generally in the image domain. In this thesis, we introduce a non-intrusive, wide-baseline stereoscopic system composed of fisheye cameras, perfectly suitable for rural or unsignalized intersections. Our main goal is to achieve vehicle localization and metric trajectory estimation in the world frame. For this, accurate extrinsic calibration is required to compute metric information. But the task is quite challenging in this configuration, because of the wide-baseline, the strong view difference between the cameras, and the important vegetation. Also, pattern-based methods are hardly feasible without disrupting the traffic. Therefore, we propose a points-correspondence-free solution. Our method is fully-automatic and based on a joint analysis of vehicles motion and appearance, which are considered as dynamic calibration objects. We present a Structure-from-Motion approach decoupled into the estimation of the extrinsic rotation from vanishing points, followed by the extrinsic translation at scale from a virtual-plane matching strategy. For generalization purposes we adopt the spherical camera model under the assumption of planar motion. Extensive experiments both in the lab and at rural intersections in Normandy allow to validate our work, leading to accurate vehicle motion analysis for risk assessment and safety diagnosis at rural intersections.

# Résumé

La surveillance visuelle des objets dynamiques dans les carrefours routiers a été un sujet de recherche majeur au sein des communautés de vision par ordinateur et de transports intelligents, ces dernières années. De nombreux projets ont été menés afin d'améliorer la sécurité dans le contexte très particulier des carrefours. Notre analyse approfondie de l'état de l'art révèle que la majorité des systèmes en bord de voie, utilisent la vision monoculaire. Dans cette thèse, nous présentons un système non-intrusif, de stéréovision-fisheye à large entraxe. Le dispositif proposé est particulièrement adapté aux carrefours ruraux ou sans signalisation. Notre objectif principal est la localisation des véhicules afin de reconstruire leurs trajectoires. Pour ce faire, l'estimation de la calibration extrinsèque entre les caméras est nécessaire afin d'effectuer des analyses à l'échelle métrique. Cette tâche s'avère très complexe dans notre configuration de déploiement. En effet la grande distance entre les caméras, la différence de vue et la forte présence de végétation, rendent inapplicables les méthodes de calibration qui requièrent la mise en correspondance d'images de mires. Il est donc nécessaire d'avoir une solution indépendante de la géométrie de la scène. Ainsi, nous proposons une méthode automatique reposant sur l'idée que les véhicules mobiles peuvent être utilisés comme objets dynamiques de calibration. Il s'agit d'une approche de type Structure à partir du Mouvement, découplée en l'estimation de la rotation extrinsèque à partir de points de fuite, suivie du calcul de la translation extrinsèque à l'échelle absolue par mise en correspondance de plans virtuels. Afin de généraliser notre méthode, nous adoptons le modèle de caméra sphérique sous l'hypothèse d'un mouvement plan. Des expérimentations conduites en laboratoire, puis dans des carrefours en Normandie, permettent de valider notre approche. Les paramètres extrinsèques sont alors directement exploités pour la trajectographie métrique des véhicules, en vue d'évaluer le risque et procéder à un diagnostic des intersections rurales.

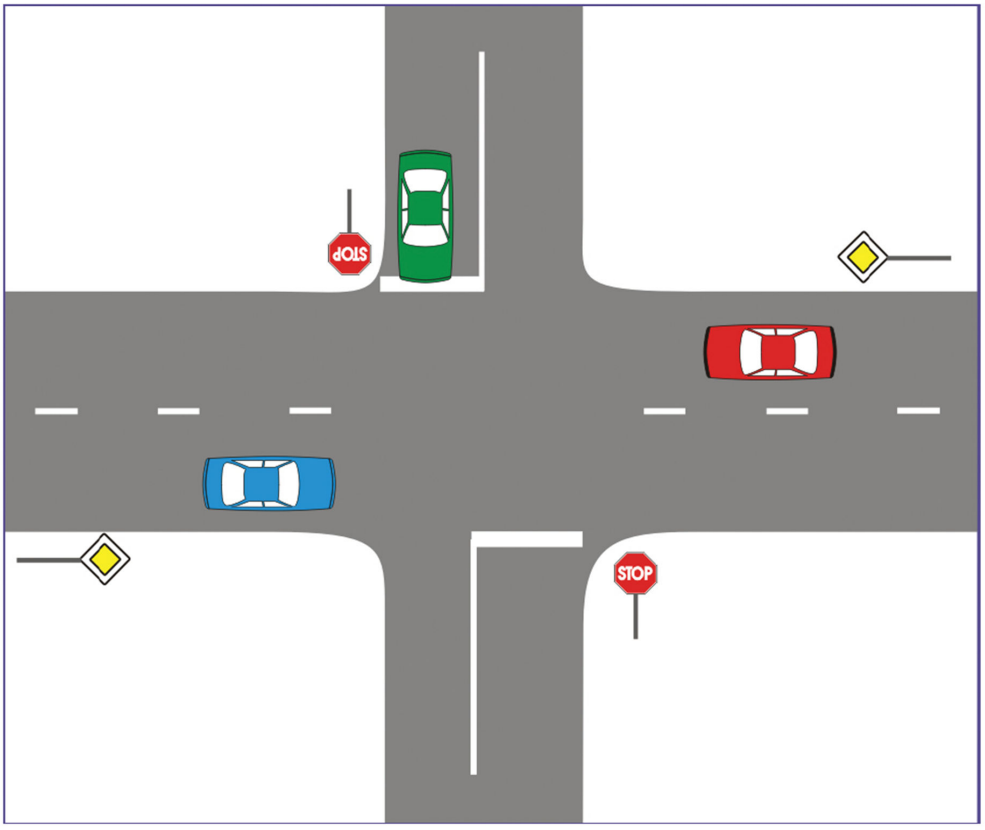
# List of Figures

# List of Tables

# Contents

# Part I

## Motivation and Background



# Introduction 1

> *If I have seen further it is by standing on the shoulders of giants*
>
> **— Isaac Newton**

Visual surveillance is widely used in many areas often for safety reasons. Several projects have led to important advances for vision-based traffic monitoring applications. In 1986, the European Research Program PROMETHEUS [201] was launched by the European automotive industry. It involved more than thirteen vehicle manufacturers and several research institutes from nineteen countries. The objectives of this pioneer project were to reduce road fatalities and improve traffic efficiency [190]. Later, the project VSAM was launched by the Defense Advanced Research Projects Agency, with the objective to develop an automated video understanding technology for use in future urban and battlefield surveillance applications [35]. Within this framework, it was reported an end-to-end testbed system demonstrating a wide range of advanced surveillance techniques such as real-time moving object detection and tracking from stationary and moving camera platforms or active camera control and multi-camera cooperative tracking. About two decades after these pioneer projects, the cooperative effort remained active with new European frameworks involving visual monitoring systems for intelligent transportations and road safety. As an example, the project ADVISOR [133], standing for Annotated Digital Video for Intelligent Surveillance and Optimized Retrieval, was successfully carried out in the early 2000s with the goal to develop a monitoring system for public transportation, in order to detect abnormal behaviors of users [126] [127]. However, despite the remarkable progress and efforts achieved by researchers, enhancing the safety of drivers is still a challenging issue, especially at road intersections.

Intersection safety is a critical worldwide issue. In fact, accidents at intersections represent an important cause of road fatalities [106]. Intersections are particularly dangerous compared to highways because of their architectures which introduce several conflicts nodes [164, 161] (Figure 1.1). Statistics from the U.S. Department of Transportation reveal that between 1998 and 2007, the number of fatalities at intersections exceeded 90,000 [186]. According to the European Road Safety Observatory, more than 62,000 people were killed in traffic accidents at intersections between 1997 and 2006 [76] [203]. According to the same statistics, cars and two-wheelers are more exposed to accidents at intersections than other vehicles. Moreover, the proportion of fatalities in intersection accidents in EU throughout the decade 2000-2010, remained slightly equal to 20% of

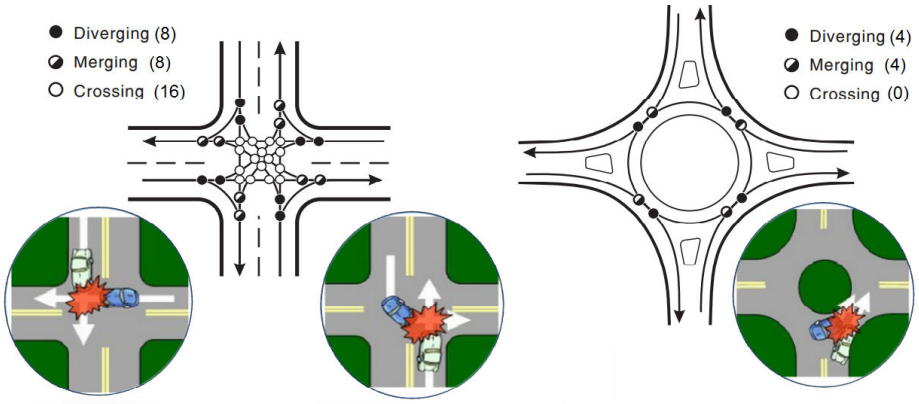



**Fig. 1.1:** Simplified illustrations of potential vehicle conflict points at a basic roundabout (8) vs. a 4-leg intersection (32) [144]

all cases [37]. More recently in EU, in 2013, more than 5.000 people were killed in road traffic accidents at intersections [38]. Cars, two-wheelers and pedestrians are particularly exposed to accidents at intersections [38]. Weather conditions are not a major cause of fatalities at intersections; because they affect accidents which occur away from intersections in a similar way [76], counting for less than 15% of cases. However, lighting conditions represent a major factor as almost a quarter of fatalities at intersections happens during night time [76]. In general, regardless of the geometry of intersections or the meteorologic conditions, human decisions remain one of the most critical factors. In fact, more than 80% of accidents at intersections are caused by driver errors [106] [174]. In order to reduce by half the number of road deaths by 2020 [36], it is therefore necessary to develop innovative vehicle monitoring systems especially at intersections.

Vehicle monitoring consists in two levels of interpretation (Figure 1.2). The first level consists in the actual scene modeling, vehicle detection and tracking. The output of this level provides data such as positions, speeds or classes of vehicles. In the literature, several papers have reported good performance for such tasks [80] [95]. The second level of interpretation consists in analyzing the interactions between vehicles to evaluate the risk-level [71] [91] [125]. This level allows to perform tasks such as predictions of specific behaviors, using the outputs of the previous level. Behavior interpretation has been actively investigated, not only to detect and analyze abnormal maneuvers, but also to prevent dangerous situations and anticipate conflicts [63] [209].

Vehicle monitoring at intersections is a special case which brings up several challenges. Vehicles at intersections can have variable and abrupt motions from different entry points. They can also occlude one another or be occluded by the infrastructure. Two categories of vehicle sensing

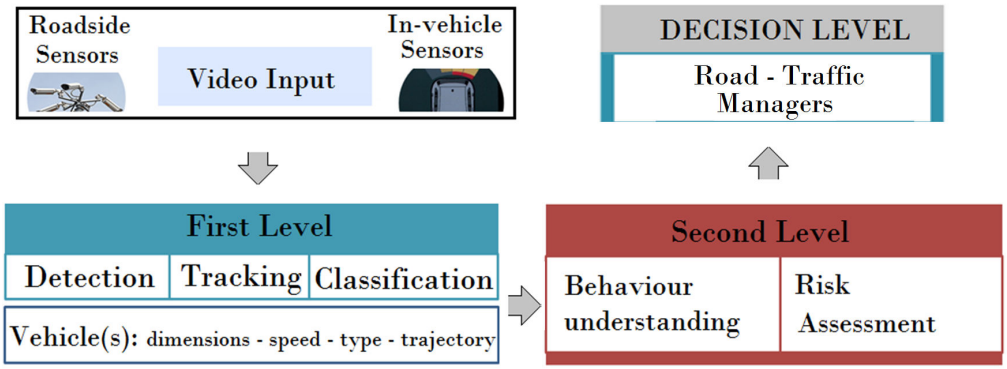



**Fig. 1.2:** Classic steps in video monitoring. First, vehicles are detected, tracked and sometimes classified. Second, outputs of the previous step are used to understand vehicle behavior and evaluate the risk-level

technologies are generally used for monitoring vehicles at intersections: active sensors (RADAR, LIDAR) and passive sensors (Cameras - monocular vision, stereo vision, wide angle or omnidirectional vision). In the recent literature, most intersection-related systems were developed for addressing collision avoidance and behavior analysis [42]. We can classify them into two categories: in-vehicle and roadside systems.

Roadside systems for intersections monitoring are stationary platforms generally consisting in pole-mounted cameras or cameras placed on elevated buildings and connected to a central processing unit [150] [121]. Today, cameras are cheaper, smaller and smarter [143]. In addition, the rising power of processors, as well as the emergence of new generation of embedded architectures allowing real-time implementations, have spawned a great interest for camera-based systems [157]. At road intersections, most of these systems require one or multiple cameras to be mounted at highly elevated positions, which can be a major drawback for the deployment. Single-camera based systems are generally preferred to monitor intersections. Most of the works on multiple cameras-based systems even treat the information of each camera independently, and then perform a high-level fusion [83] [181]. In addition, in most cases important preprocessing steps, such as intrinsic and extrinsic camera calibration, are necessary before further traffic analysis. Besides, despite the increasing use of omnidirectional cameras in general ITS systems [187], they have been less exploited for roadside intersection monitoring systems. Only few representative studies have used wide field of view cameras (Table 2.2) such as catadioptric [65] and dioptric [104] [88] vision sensors. However, these recent studies have shown the worthiness of omnidirectional cameras at intersections and have introduced several challenges. Despite the fact that they can enable to monitor an entire intersection, omnidirectional

images involve the necessity to analyze important amount of visual data at various scales as the vehicles moves through the scene.

In-vehicle systems for intersections monitoring are also generally based on cameras, often stereo setups with short baselines, installed or embedded in mobile platforms [61] [3]. Today, new generations of in-vehicle driving support systems such as cooperative advanced safety systems [30], based on sensor fusion, are being actively studied for applications at intersections (Figure 1.3) [212]. In the European project INTERSAFE [8], a mobile platform for accident detection at road intersections was developed combining a wide range of active and passive sensors. Recently, within the framework of the project Ko-PER [68] [173], a system combining laser scanners with low and high resolution cameras has been used to gather classification information about vulnerable road users. Thus, recent research works about intersections safety systems aims not only to reduce the risks of accidents, but also to provide solutions for reliable data collection, as well as innovative technologies for drivers behavior analysis.

To date, the latest research works on roadside intersection monitoring [104] [88] have led to the development of some commercial applications. For instance, the company GRIDSMART claim to built intersections monitoring systems, upon the foundations of simplicity, flexibility and transparency, with the goal to empower traffic managers. They pointed out the necessity for traffic professionals to get started easily with new technologies and use their own computers to design, organize, configure and manage intersections [70]. This is a perfect illustration of the growing challenges related to non-intrusive detection and tracking technologies for intersections monitoring with stationary omni-vision systems. Thus, the problem of monitoring traffic participants at intersections has moved toward another level. Nevertheless, the research is still opened, because to the best of our knowledge none of the existing systems are fully automated, can provide either complete reliability or robustness.

This thesis involves the development of non-intrusive roadside monitoring system, with the goal to observe traffic at intersections, analyze vehicle trajectories in order to identify potential safety issues. In the following sections, we present the origins of this doctoral project, then we outline the main scientific contributions of this thesis, and finally we introduce the thesis overview.

## 1.1 Thesis Project: roots and motivation

The severity of accidents in rural areas is 5.3 times higher than in urban areas. In Europe, above 80% of all fatal collisions occur on rural

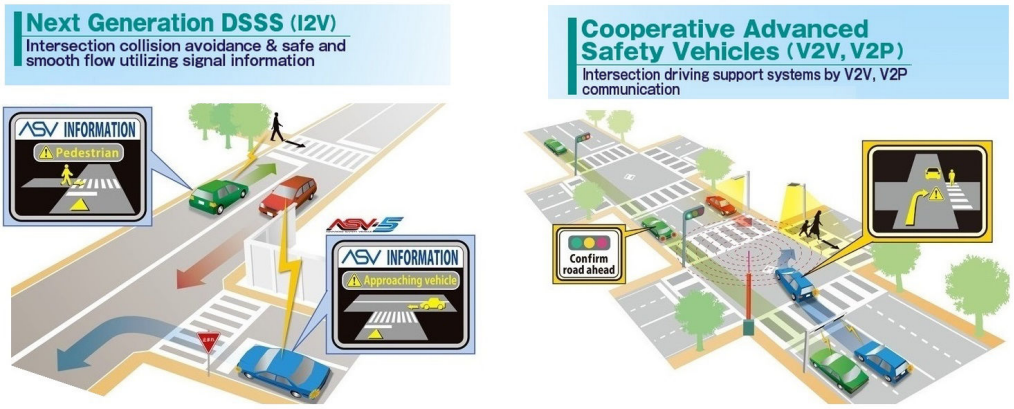



**Fig. 1.3:** Intersection safety systems and trends. New generations of driving support systems for infrastructure to vehicle communications, as well as cooperative advanced safety systems are being studied. [86]

roads. In France, intersections represent less than 1% of the distance traveled by traffic users, but over 10% of accidents. The risk of an accident is increased tenfold at a rural intersection. This alarming level of danger at rural intersections brings up a strong social pressure. Therefore road managers need to identify and understand the problems likely to entail further critical accidents. However the lack of complete police reports and the small number of injury accidents collected, make it difficult to analyze accidents processes and understand the causes. Besides, the audition of the drivers involved in the accidents is not possible or erroneous, because of their state of chock and the gravity of the accident. As a result, road managers have a lack of data to take appropriate decisions. Thus, there is a need for systems capable of performing fast preventive safety diagnosis at intersections, before the occurrence of dramatic accidents. Road managers also want tools that allow before-after accidents countermeasures efficiency evaluation.

The origins of these research works started about ten years ago with the project "CARREFOUR". In early 2007, experts-researchers from the Infrastructure and Multimodal Transportations Systems at Cerema[1], developed a tool to evaluate the safety at rural crossroads [122]. This precursor project involved the development of a non-intrusive near-miss traffic conflicts registration system (Figure 1.4). It uses traffic sensors, speed radars on the main road and pneumatic tubes on the minor axes, to detect vehicles. Data coming from both types of sensors, the radars and the road tubes, are transmitted through local wireless networks to a central processing unit. Then, an algorithm calculates the time to collision in order to identify near miss accidents. Finally, a video recorder framework is activated few seconds before and after the detected incident. The video

[1]Cerema, `http://www.cerema.fr/`

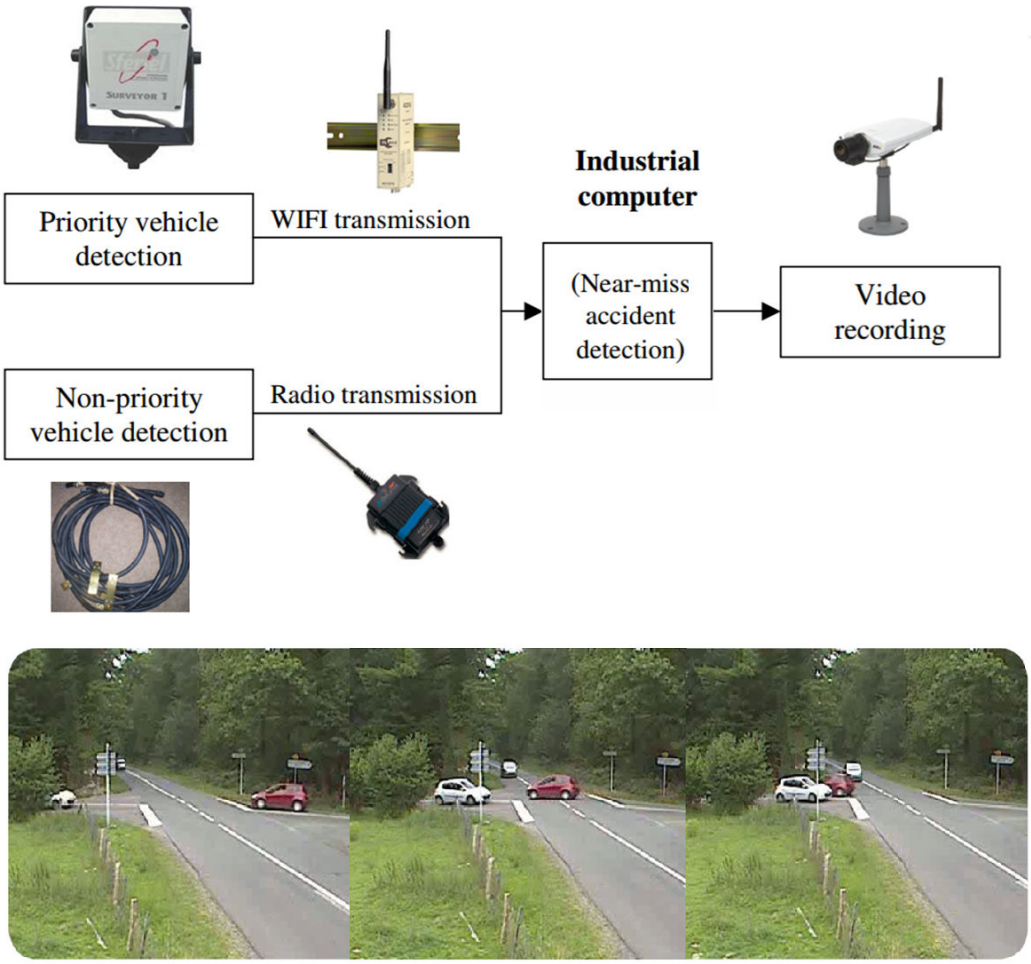



**Fig. 1.4:** Near-miss accident detection system diagram and example [122]

recorded allows a qualitative analysis to determine whether the near-miss accidents result from a driver error or a road design issue. Furthermore a risk indicator was proposed, mainly based on the number of conflicts in a given period and the speed of vehicles.

The system developed within the project "CARREFOUR", as described above, requires important sets of sensors. However it was prone to important false-positive recordings, often caused by the pneumatic tubes wrong detections or by abrupt vehicle motion. Therefore, a considerable amount of time was required by operators to verify the near-miss accidents and conflicts candidates recorded by the system. In addition, the results are also affected by parameters set at initialization, such as the collision confidence interval. Moreover, the video-recording framework only comes at the very end, as an output to the complete system.

This doctoral thesis has been co-funded by the French region Normandy and Cerema (projects RoadTrack and DAISI). This thesis is the follow-up - but with a novel philosophy - of the precursor project "CAR-

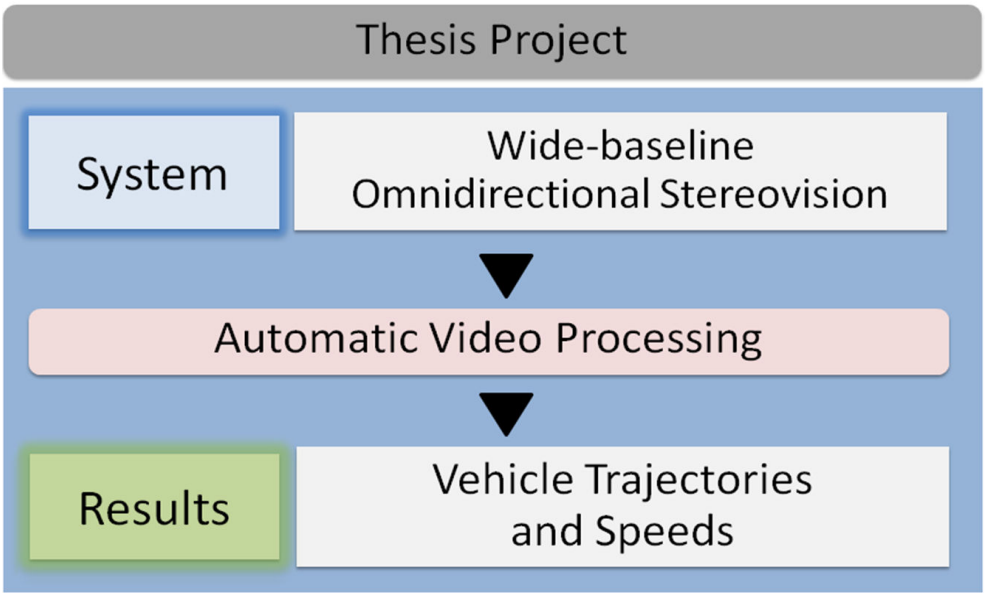



**Fig. 1.5:** System description and main objectives of this thesis

REFOUR" described above, for safety diagnosis at rural intersections. The first requirement in this thesis, is the exclusive use of omnidirectional cameras - in the place of speed radars and pneumatic tubes - for the complete pipeline: acquisition, monitoring, automatic analysis, toward risk-assessment. The automation of the proposed solution was also necessary to avoid time consuming manual setup or verifications by an operator, in order to ensure accurate processing. In this thesis, we have imagined and developed a wide-baseline stereoscopic system using fisheye cameras. The main outputs required are vehicle trajectories and speeds at the intersection. Our goal is to provide robust and accurate vehicle motion data, which can later be used by road managers to perform safety diagnosis at rural intersections.

## 1.2 Major Contributions

The ideas presented in this document can be interesting to the intelligent transportations systems and computer vision communities. Several topics are discussed ranging from pixel-level to object-level analysis (Figure 1.6). This thesis work presented herein, introduces several key contributions to the literature regarding traffic monitoring with wide-angle cameras in general, and specially in the context of rural intersections.

- Interesting reviews have been proposed about traffic monitoring and its applications in general [95] [176] [21] [167]. But to the best of our knowledge, none of these surveys paid a particular attention to intersection-like scenarios at a large scale. Our first contribution

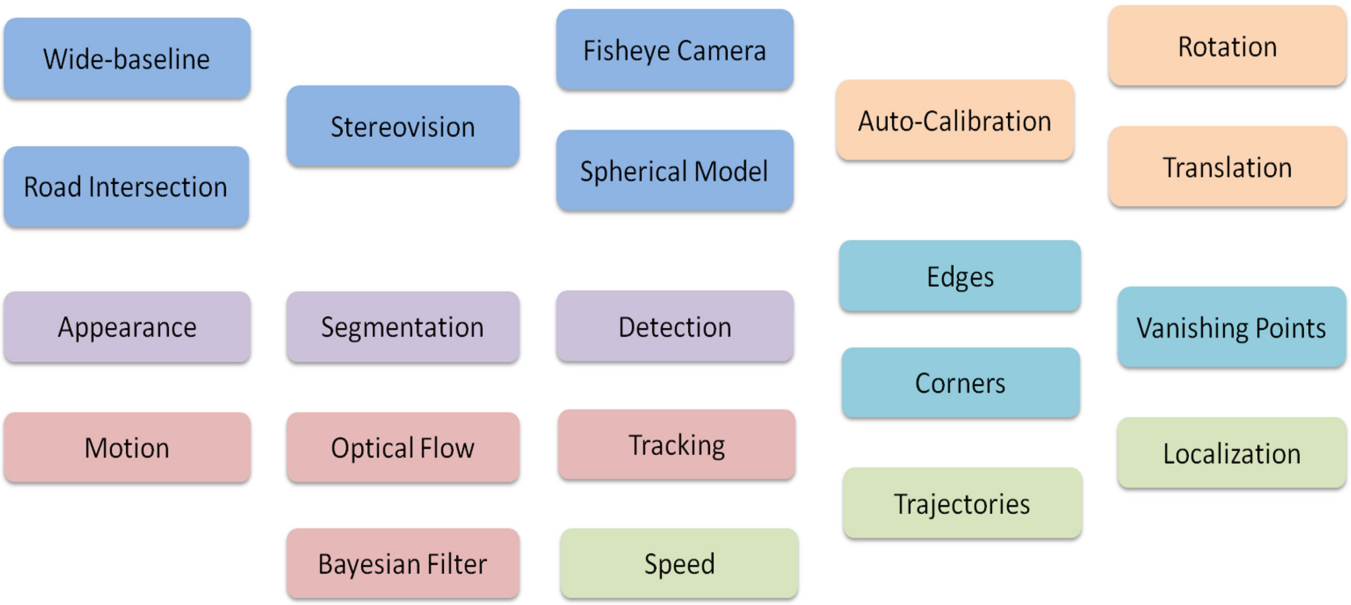



**Fig. 1.6:** Keywords map of different topic covered in this thesis

to the state-of-the art is a deep analysis of the specific challenges of vision-based traffic monitoring of intersections, with a special focus on roadside omnidirectional camera-based systems

- In order to compute 3D trajectory and motion information accurately, at least 2 cameras are needed. In this thesis, we introduce a non-intrusive and flexible Wide-Baseline Fisheye-Stereo monitoring system. The cameras are installed in the corners of the intersection directly on the available infrastructure. This flexible configuration allows a compromise between easiness of setup and monitoring quality and offers a multi-view perception which strengthens vehicles evidence.
- The large distance between the cameras in our system brings up important additional challenges, related to calibration, vehicle detection and tracking, and for which we have proposed innovative algorithms and solutions. In particular, the problem of the extrinsic calibration, meaning the estimation of the relative pose between the cameras with 6 degree-of-freedom, arises as a major aspect of this thesis. In fact, the extrinsic calibration is compulsory in order to achieve metric localization of vehicles, in the objective of trajectory and speed estimation at absolute scale. Therefore, a big part of our work involved the proposal and the development of a practical self-calibration approach based solely on traffic flow analysis. This

part of the work is based upon the key idea that vehicles assumed undergoing straight-planar motion, can be used as dynamic calibration objects. Thus, we have introduce an approach to recover automatically the rotation and the translation between the cameras once installed at the intersection. Our approach applies the spherical camera model and requires no prior knowledge of the scene geometry.

  - First of all, we propose a method to compute the rotation between the cameras by matching a trihedon of vanishing points estimated in each camera frame. For this, we also introduce two novel methods to compute these key features in our work: the Spherical RANSAC-based Vanishing Point algorithm (S-VP-RANSAC), and the scale-invariant pixel-wise Vanishing Point algorithm (SIP-VP).
  - Then, provided the rotation between the cameras, we introduce a vehicle-related virtual plane-like matching strategy to recover the translation at scale between the cameras.

- Extensive evaluations concerning the calibration results show the advantages of the proposed method. A first evaluation in the lab environment is achieved with comparison to a ground-plane induced homography method using a large-scale drawn pattern (only possible in the lab). A second calibration evaluation of the calibration was done based on vehicle localization error and distance measurement on the road surface in real traffic conditions. After quantitative and qualitative verifications of the extrinsic calibration, the last part of this thesis is dedicated to the development of a vehicle trajectory reconstruction module at intersections. Provided the extrinsic calibration between the cameras, vehicles can be localized at metric scale in the world reference. Consequently we have proposed a Bayesian framework that uses motion and appearance cues to estimate vehicle trajectories and thus compute their speeds. Vehicle trajectory and speeds are important data that can later be used for safety diagnosis at intersections.

## 1.3 Thesis Overview

This work describes a complete wide-baseline fisheye-stereo monitoring system, intended to observe vehicles, compute their speeds and trajectories as output in rural intersections. The rest of the thesis is organized as follows:

- Chapter 2, describes a deep literature review of vision-based traffic monitoring in general, with a focus on roadside intersections

monitoring setups. Challenges related to camera calibration, scene perception, vehicle detection and tracking, as well as behavior analysis are discussed. This analysis lead to the main contributions presented in this thesis.

- Chapter 3, is dedicated to our monitoring system (setup, modeling). We formulate the problem regarding the extrinsic calibration of the proposed wide-baseline fisheye-stereo. We present our algorithms that allow the estimation of vanishing points, which are key features in our solution. Then we describe points-correspondence-free approach to solve the problem, decoupled into the estimation of a pure rotation and a translation at scale.
- Chapter 4, describes different evaluation experiments carried out to validate the proposed extrinsic calibration method of the wide-baseline fisheye stereo. Afterward experiments regarding trajectory reconstruction and speed estimation, toward safety diagnosis are presented. The complete system is evaluated with different challenging datasets both in lab and in real traffic conditions. Results are discussed and analyzed.
- Chapter 5, is the summary of the work presented in this thesis. Limitations of system and possible future applications are also discussed.

# Vision-based monitoring of intersections: a review

# 2

> *To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science*
>
> **— Albert Einstein**

This chapter presents extensive literature review about several aspects of roadside traffic monitoring in general, especially at intersections. Previous reviews have been proposed regarding traffic monitoring [167] [21], but none of them gave a special attention to intersection monitoring issues. First of all we provide an overview of vehicle perception systems at road intersections, as well as representative related datasets. The reader is then given an introductory overview of general vision-based vehicle monitoring approaches. Then, we present a review of studies related to vehicle detection and tracking in intersection-like scenarios. Regarding intersection monitoring, we distinguish and compare roadside (pole-mounted, stationary) and in-vehicle (mobile platforms) systems. Then we focus on camera-based roadside monitoring systems, with a special attention to omnidirectional setups. Finally we present possible research directions likely to improve the performance of vehicle detection and tracking at intersections.

## 2.1 Observing vehicles at intersections

In this section, we present the different sensing technologies used for monitoring vehicles at road intersections (Figure 2.1). We can distinguish mainly between active and passive sensors. Active sensors such as RADAR and LIDAR, consist of a source that emits waves onto the target. The waves are reflected back by the object towards a receptor. Unlike active sensors, passive sensors only work as receptors that measure information either emitted or reflected by the object. In other words, sensors such as cameras, that use external energy sources to observe an object are passive. After a discussion of these sensing solutions, we will present several datasets that can be used for intersection monitoring studies.

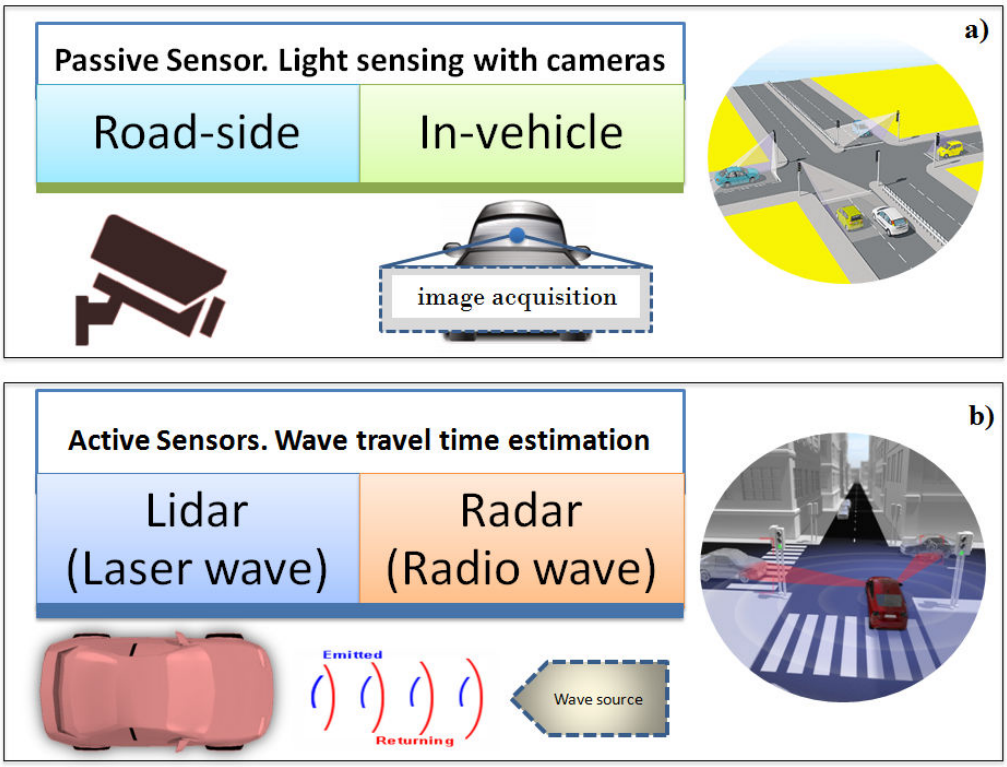



**Fig. 2.1:** Vehicle sensing modalities. a) Cameras are passive sensors that measure the reflected light by vehicles. b) Radar-Lidar are active sensor that measure a travel time of an emitted wave to the vehicle

## 2.1.1 Sensing technologies for intersection monitoring

### 2.1.1.1 Monitoring with RADAR

**RADAR** stands for RAdio Detection And Ranging, and has been used to detect objects in intersections-like scenarios [42]. In [8], four short range RADARs (SRR) placed at each corner of the vehicle and a long range RADAR (LRR) in front of the vehicle were used (Figure 2.21). RADAR can help determine the range, the altitude, direction, or speed of an object. Motion and dynamics parameters are the fundamental cues for RADAR-based vehicle detection and tracking systems. However it might cause false detections between different types of objects having the same motion model. As a consequence, RADAR data can be very noisy and require extensive post-refinements [8]. Though RADAR designed for traffic monitoring have a very narrow field of view, they can be robust to difficult weather and illumination[2].

### 2.1.1.2 Monitoring with LIDAR

**LIDAR** stands for Light Detection and Ranging. The use of this sensing technology for traffic applications and vehicle monitoring has

increased thanks to the reducing cost of the technology. It is an optical remote sensing technology which measures properties of scattered laser-wave to find range and additional parameters of a distant target. The measurement is based on the estimation of time-of-flight of a laser wave. Thus, it can be used to detect stationary as well as moving objects. In [8], laser beams are classified in a local grid map as static or dynamic in order to segment moving objects from static ones. In [156], single-row laser scanners data were used to perform trajectory and behavior analysis of vehicles passing at an intersection. The system classifies large amounts of trajectories based on a group of route models built from trajectory clustering. In [213], 3D clouds acquired from a dense beam scanning LIDAR mounted on the roof of an autonomous vehicle are used to detect when the vehicle reach an intersection. LIDAR has proved its efficiency, providing cleaner data than RADAR, but is sensitive to the environment and the weather.

#### 2.1.1.3 Monitoring with Cameras

**Cameras** are widely used for traffic monitoring at intersections and provide rich visual information. Objects are visible with the camera because of the light reflexion from their surface onto the vision-sensor. A point in 3D world reference is mapped onto the image plane reference, into a 2D point via a projection matrix. For daylight operations, visible light cameras can be used. However by night or during difficult weather, a visible light camera is unlikely to meet good performance needs over extended periods. The use of infrared cameras can be a good solution for night-time vision, as for instance in the night the temperatures of the wheels of the vehicle are higher than the isotherm temperature [117].

Camera networks can offer several advantages such as: acquiring richer data, solving occlusion issues, enabling redundancy. However, deploying a network requires to take into account critical points such as: mobility, power consumption, and compulsory spatial and temporal calibration [142]. In the context of intersections, camera networks are hardly used.

Omnidirectional cameras have been widely used in many areas as they possess wider field of view (FOV) than conventional cameras. However, they have been hardly used for intersection monitoring, as they introduce several challenges in this context such as: the real-time analysis of large image provided in a single shot, with highly distorted objects at variables scales [88]. Moreover, existing roadside systems with omnidirectional cameras cannot be easily adapted to different types of intersections [65] without important civil engineering.

### 2.1.2 Datasets for traffic analysis at intersections

The interest for traffic monitoring at intersections has kept increasing over the last decade, with an emphasis on environment modeling and vehicular behavior analysis. The cooperative effort contributes to the exchange of data and the advances in research. This has led to the emergence of challenging datasets for evaluation and benchmarking. We present under this subsection representative datasets that can be used for vehicle monitoring studies and applications at intersections.

**Fig. 2.2:** MIT dataset [198] [78]

**Fig. 2.3:** NGSIM dataset [134]

**Fig. 2.4:** CBSR dataset [81]

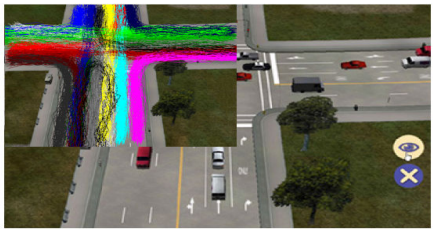


**Fig. 2.5:** CVRR dataset [125]

**Fig. 2.6:** QMUL dataset [78]

**Fig. 2.7:** Ko-PER dataset [173]

**Fig. 2.8:** KIT dataset [200]

**Fig. 2.9:** Urban tracker [89]

—**MIT** (Figure 2.2): the traffic dataset from the Massachusetts Institute of Technology, is for research on activity analysis of crowded scenes [160], and has become widely used. It holds nineteen minutes of raw video

recorded by a stationary camera, divided into 20 clips, including vehicle traffic at a four-way signalized intersections [198].

—**NGSIM** (Figure 2.3): the next-generation simulation program was initiated to develop a core of open behavioral models in support of microscopic modeling and traffic simulation [134] [183]. It includes validation data that have been used more recently for learning traffic behaviors and patterns [179] [73] [11] [1]. The overhead intersection cameras provide multi-view raw video data of intersections, with detailed vehicle trajectories, as well as supporting behavioral data.

—**CBSR** (Figure 2.4): the traffic dataset from the Center for Biometrics and Security Research, provides a single view of a complex intersection, with several possible entry and exit combinations. It has been essentially used for motion patterns analysis to predict abnormal behavior [81].

—**CVRR** (Figure 2.5): the traffic dataset from the Computer Vision and Robotics Research Lab, provides data for benchmarking unsupervised trajectory-based activity analysis algorithms: clustering, classification, prediction, and abnormality detection. This dataset consists of a simulated intersection, a real highway, as well as two indoor omnidirectional camera videos. However, the provided trajectories contain only spatial information. Besides, true activity labels are provided as well as full tracks based abnormality label and frame-by-frame unusual event information [125].

—**QMUL** (Figure 2.6): the dataset from Queen Mary University of London is specifically intended for activity analysis and behavior understanding [112] [78]. It contains one hour of traffic video collected at a busy intersection with ground truth.

—**Ko-PER** (Figure 2.7): the Ko-PER project datasets are meant to generate a comprehensive dynamic model of the ongoing traffic. The datasets consist of data from the laser-scanners network and cameras installed at a public intersection, as well as reference data and object labels. The datasets are shared for further research in the field of multi-object detection and tracking [173].

—**KIT** (Figure 2.8): the intersection traffic datasets from Karlsruhe Institute of Technology [200] provide eight videos recorded by a stationary calibrated camera at different location and in various conditions (rain, snow, fog). The dataset has been mainly used to study vehicle detection and tracking at intersections [138].

—**Urban Tracker** (Figure 2.9): the Urban Tracker dataset, shared by researchers from the École Polytechnique de Montréal, focuses on traffic research applications. It provides recordings of the traffic scenes, meta-

data, camera calibration, ground truth, protocols for comparing algorithms, software tools and libraries for reading the data [89]. There are four annotated videos sequences of intersections, recorded in different weather and lighting conditions (Sherbrooke, Rouen, St-Marc, René-Lévesque) from stationary cameras at different heights. This dataset is suitable for vehicle and pedestrian tracking at intersections, with several computer vision challenges.

## 2.2 General overview of vision-based vehicle monitoring

Under this section, we review vision-based monitoring in the general sense. In the first subsection, we present the major challenges in this field. In the second and third subsections, we give an introductory review of vision-based vehicle detection and tracking, respectively.

### 2.2.1 Challenges of vision-based vehicle monitoring

There are several challenges regarding vision-based vehicle monitoring. First of all, the quality of the sensor itself (noise, vibration, image format and resolution, lighting and optics, color representation, computation power available [130] [9]) might affect the monitoring process. Besides, there are important initialization and steps (camera calibration, image region of interest definition, initial maps) which are context-based. After a discussion about general initialization and preprocessing issues, we present challenges intrinsically related to vision-based vehicle detection and tracking.

#### 2.2.1.1 Initialization and preprocessing challenges

In many camera-based traffic monitoring systems, there are often compulsory and context-based initialization tasks, such as image formatting and rectification or initial map generation [93] [60]. For instance, in the particular case of omnidirectional cameras, image distortions and resolution are important issues for real-time systems. In fact, it is necessary to count a large number of pixels, to extract useful regions in the image, and monitor targets spanning nearly four order of magnitude [65]. Thus, in order to design a robust vision-based monitoring system for traffic applications, it is required to take into account several parameters and external factors.

Traffic monitoring systems require calibrated cameras to perform image to world mapping. Accurate calibration is necessary to estimate vehicle dimensions and motion parameters. It includes the estimation of

camera intrinsic parameters on the one hand, and the extrinsic parameters on the other hand. While intrinsic calibration can be determined beforehand, extrinsic calibration need to be assessed once the cameras are installed [85] [114].

The calibration of roadside cameras mostly relies on geometric constraints and prior knowledge of the scene. This task is mainly based on the estimation of vanishing points (VPs), calculated with parallel lines extracted from specific landmarks such as lane marking. The number of vanishing points required for calibration depends on the measurement available on the environment [85] [121] [192]. As a consequence, their accuracy highly depends on the user inputs and assumptions. A detailed taxonomy of roadside camera calibration methods is proposed in [94]. These methods cannot be efficiently applied when well-visible patterns with parallel lines on the road are hardly available. As an alternative, motion-based VPs estimation has been proposed as a good solution, provided the assumption of straight and planar motion [53].

#### 2.2.1.2 Vehicle Detection and tracking challenges

Vehicles have different characteristics such as size, color or shape. A fundamental question at initialization is the choice of an adequate and robust vehicle feature representation with respect to the application (blobs [194], set of points [197], or geometric models[22]). In general the vehicle representation defines the tracking strategy [204]. The most important goal is the ability to keep vehicle tracks active as long as possible in the scene. Reliable vehicle tracking is crucial, as it will allow posterior trajectory analysis and behavior recognition [126].

Through the literature, it appears that vehicle detection and tracking are somewhat related tasks. Some methods require vehicle to be detected first and then tracked, while other strategies use vehicle tracks as cues for detection. There are several issues that make vehicle detection and tracking challenging. The major challenges reported are: vehicle occlusion (vehicle-vehicle, infrastructure-vehicle), changes of vehicle perception (appearance, camera placement, distortion, size, scale [60]), sudden vehicle motion [149] or sudden change in the environment (illumination, weather, night time vision) [21], shadow detection and removal [13] [168] [147].

### 2.2.2 General review of vision-based vehicle detection

Under this section, we present the general approaches for vehicle detection. Though many papers in the recent literature have reported good performance of vehicle detection, it remains a major issue due to the

several challenges presented above. In general, vehicle detection is fully performed in two steps: finding foreground entities considered as vehicle hypothesis, and then verifying these candidates.

#### 2.2.2.1 Vehicle candidate localization

Vehicle candidate localization can require one or several frames. The different methods can be classified into four categories: background subtraction, model-based segmentation, feature-based segmentation, motion-based segmentation. The methods also vary with the system setups. In monocular vision, appearance-based cues are mostly used to obtain the vehicle hypothesis frame by frame. Adaptive motion models have also been often applied to differentiate vehicles from the background [23] [214]. Whereas in stereo-vision, motion-based methods are the current trend. In this case, multi-view geometry enables direct measurements of 3D information. Stereo-vision has been efficiently used to separate obstacles from the free space, thanks to the disparity maps [102], dynamic occupancy grids [44], or inverse perspective mapping [108].

##### Background subtraction

It consists in extracting foreground objects from a single image by removing a reference background model [196, 40]. The major difference between background subtraction approaches in the literature, relies in the way to obtain the background. It can be either static, dynamic, statistical or adaptive background estimation. There are several problems related to this task, because of the difficulties to define boundaries between background and foreground [168]. The most popular background modeling method is the Gaussian Mixture Model (GMM) [171]. It consists in modeling pixel values over time by a weighted mixtures of Gaussian [20]. Later, Barnich Olivier et al. introduced a novel approach called: the universal video background subtraction (ViBe). The model consists in a set of observed pixel values [13]. The pixels are classified as background by thresholding the distance from a given pixel and all samples. ViBe algorithm incorporates a memoryless update policy and is resilient to noisy data.

##### Model-based segmentation

It consists in identifying possible foreground vehicles in an image by fitting vehicle 2D-predefined or 3D-projected shapes to image regions, without any knowledge of a background model [121] [72]. However, these direct matching approaches are unrealistic because it is impossible to have a model for all possible vehicles that may be present in the scene. The use of probabilistic frameworks [82] [111] [121] or motion information

[138] [90] [22] along with vehicle models has been often used to reduce the matching search space.

#### Feature-based segmentation

Searching by analyzing the geometry or appearance features is a common method for vehicle candidate localization in the foreground. Researchers have used texture [74] [208], color [189], shadow [41] [136] and geometric elements such as corners [87], and symmetry analysis [77] [18] or fusion of several cues [28]. A good vehicle feature should be able to capture the distinctive characteristics and be robust enough to small variations over different background conditions [6]. This is necessary in order not only to reduce the dimensionality of the data to be processed, but also the computation time. More recently, new trends of descriptive features have been used because they enable a more direct vehicle hypothesis and localization: SIFT [210], ASIFT [79], SURF [175], Histogram of Oriented Gradient (HOG) [145] [33], Gabor features [177] [119], Haar-like Wavelets [110].

#### Motion-based segmentation

It consists in identifying moving items in the foreground by searching image regions having significant changes between successive frames [80]. It uses the pixel-level difference and the temporal information to extract moving regions. Motion-based hypothesis generation makes use of temporal information to detect vehicles, and obstacles by matching and grouping image pixels having the same motion characteristics over consecutive frames [87]. It can rapidly adapt to dynamic environments when temporal changes are important, but it may fail to extract all the representative pixels in complex scenes (appearance or scale change, shape change, variable motions, scale change, occlusion or clutter) [205]. Moreover, generating a displacement vector for each pixel is time consuming and computationally expensive. The discretization of the image gives better results and enables real time processing [175].

The optical flow is one of the most popular approach freely available [19]. It can be seen as the projection of a three-dimensional motion field onto the image plane. Dense stereo algorithms provide significantly more information on the 3D environment compared to sparse stereo methods. The gain in information and precision allows for improved scene reconstruction and object modeling. However, in order to reduce computation time, the sparse optical flow is preferred [99], as it considers only a sets of relevant pixels. The latter generally gives enough information to formulate the hypothesis about vehicles while being less sensitive to noise [26] [25] [148]. There have been several adaptations of the optical flow based on compact shapes such as stixels [10] [55] [141] in order

to obtain a more comprehensive representation, but also to speed up the processing. Besides, optical flow can also be used for understanding complex road scenarios. Geiger et al. [62] introduced the object flow descriptor which works as an urban traffic classifier, in particular to detect when vehicles reach an intersection.

#### 2.2.2.2 Vehicle candidate verification

Regardless of the approach selected for identifying potential vehicle candidates in the foreground, vehicle detection is completely achieved only after a verification procedure. The latter needs to be performed to discard false alarms, and ensure that the candidates are actually vehicles. We discuss vehicle verification techniques in two categories which are similarity estimation and discriminative classification. The former calculates correlation score between a given vehicle candidate and a given template, while the latter verifies the candidate after a learning process.

##### Similarity estimation

It consists in using predefined templates, and estimating their correlation between a vehicle candidate region. A similarity is expressed as the result of vehicle verification [199]. Once a vehicle hypothesis has been verified, it can be used dynamically as a new template if its correlation score or reliability exceeds a certain threshold [109]. Three-dimensional templates can be projected into 2D-templates and matched with images regions [72], along with probabilistic frameworks [111]. In [121], the convex hull for 3D vehicle models were generated in the image. The ratio between convex hull overlap of model and image normalized by the union of both areas generates a similarity score. Furthermore, 3D-models have been used as well to verify motion-based vehicle hypothesis [138] [90]. In [22], N. Buch presented a vehicle detection and classification system for urban traffic scenes in which: vehicles were detected in each frame using 3D models; then motion silhouettes were extracted and compared to a projected model silhouette, in order to identify the ground plane position and classes of vehicles. They tested the system successfully in three weather conditions and the full system including detection and classification for all data in various weather achieves a recall of 90.4% outperforming similar systems in the literature.

##### Discriminative classification

It consists in learning a decision boundary between two classes of features. The goal is to distinguish between non vehicles and vehicles objects. The classification is performed by first learning the appearance of a vehicle from a training dataset. The training is generally based on a supervised approach where a large set of labeled positive (vehicle)

and negative (non-vehicle) images are used. Common techniques used in recent literature for vehicle verification are: Support Vector Machines (SVM) [191], Artificial Neural Networks (ANN) [98] and AdaBoost [58].

SVM (Support Vector Machines) classifiers are developed based on the statistical learning theory described in [191]. The idea is to map the training data of two object classes from the input space into a higher dimensional feature space. Then an optimal separating hyperplane with maximum margin is constructed in the feature space to separate the two classes [119]. Based on a set of trained orientated HOG features, [145] successfully classified vehicle by orientation, with 88% accuracy. In [178], Gabor features, which capture the local lines and edges information at different orientations and scales, were trained on the SVM classifier for vehicle detection. The evaluation results show that the classifier can achieve 94.5% detection rate. In [31], the combination of boosted Gabor features enabled to reach 96% detection rate. By combining Gabor and Legendre moment features for training an SVM classifier, [211] reported a better performance of 99% for vehicle detection. In order to reduce the dimensionality of the data, [188] applied Principal Components Analysis (PCA). The authors build into a new sub-space a specific vehicle feature vector. Doing so, they obtain an excellent performance on the SVM training with 95% accuracy.

ANN (Artificial Neural Network) classifier are less used for vehicle detection in the late literature. In 2000, [155] proposed a vehicle classification system which calculates texture features using the co-occurrence matrix, within a classification scheme based on multilayer perceptron. More recently ANN were used as well for infrastructure detection [158] or for vehicle license plate recognition based on sliding concentric windows [50]. In [29], a probabilistic neural network framework has been proposed. The authors reported a maximum performance of 69.38% of successful automatic detection. They also claimed that the proposed approach has a substantially higher degree of performance, both qualitative and quantitative, than other state-of-the-art methods. However, ANN-based vehicle classification are less efficient than SVM-based methods.

Adaboost [57] (Adaptive Boosting) classification was first used to perform face detection, with a set of Haar wavelet features; in a constructed cascade of increasingly more complex classifiers [195]. In [172], a boosted cascade of weak classifiers is used to analyze the redness of tail lights by night, in order to detect vehicles on the road. Using a richer set of Haar features, [202] were able to detect cars and buses with 71% detection rate. Adaboost-based classification schemes have been less studies in the recent litterature for vehicle monitoring [166].

## 2.2.3 General review of vision-based vehicle tracking

### 2.2.3.1 Vehicle representation and tracking approaches

There are several approaches for vehicle tracking that have been discussed in the literature. Depending on the vehicle representation, which ranges from pixel level to object level, four general approaches can be distinguished: region-based tracking, contour-based tracking, feature-based tracking, model-based tracking.

#### Region-based tracking

Regions or blobs can be defined as connected image parts with distinguishing common properties, such as intensity, color or texture statistics. Region or blob based tracking aims at tracking vehicles according to variations of the image regions [204]. An entire region associated to a given vehicle is tracked over time using appearance, geometrical properties as well as motion cues. Region-based tracking is computationally efficient and works well in free-flowing traffic. However, under congested traffic conditions, complex deformation or a cluttered background, vehicles can partially occlude one another instead of being spatially isolated [215].

#### Contour-based tracking

Contour-based tracking algorithms represent objects by their contours, which are simply their boundary, and update these contours dynamically at every time increment. In fact, strong changes in image intensity generally occur at contours, which make it suitable for tracking purposes. Contour-based tracking algorithms provide more efficient description of vehicles than region-based algorithms: by reducing the computational time and complexity. However contour-based tracking does not perfectly solve occlusions. Even though vehicle contours can be extracted separately, a difficult task is to isolate vehicles being tracked if the contours of different vehicles are merged at some point. It requires an active track grouping and update policy [149].

#### Feature-based tracking

Feature-based tracking uses the principle that vehicles can be represented by a set of features, instead of an entire object. This refers to the group of methods that perform tracking by first extracting features in independent images and then matching the features over the frames. Features can be selected as representative parts of the vehicle, such as corners, lines, typical shapes. This technique is effective as long as the selected features are robust enough and can be distinguishable even if the vehicle is partially occluded at some point in the sequence. As suggested by [149], the feature-based tracking is suitable for daylight, twilight or

night-time conditions, as well as different traffic conditions. But since a vehicle can have multiple features, the major problem resides in defining conditions that allow proper grouping [60] or clustering of those features from a given vehicle. The main cues for grouping are spatial proximity and common motion. Feature-based algorithms can adapt successfully and rapidly, allowing real-time processing and tracking of multiple vehicles in dense traffic. Moreover, for a feature-based tracking approach to be reliable, it must minimize the risk of mismatches through robust estimation algorithms.

#### Model-based tracking

This approach consists in matching a projected model onto the image, as the vehicle moves frame by frame. This allows to recover trajectories and models, as well as the pose of the vehicle with higher accuracy [43] [15]. In [43], a 3D cuboid is used as the vehicle model. The matching between the measurement and the model is performed by an intersection of rectangles in the bird's-eye view, and a corner by corner matching in the image space. In [15], vehicles are modeled by a cloud of 3D points under the rigid body assumption. The vehicle model is updated by fusing stereo vision and tracking of image features. The major weakness is the need for accurate geometric object models. It is however unrealistic to have detailed models for all vehicles in the traffic, therefore additional cues are generally added upon initialization [138].

### 2.2.3.2 Popular algorithms for vehicle tracking

There are powerful mathematical tools for vehicle tracking, that can be either iterative or non-iterative. Iterative techniques can solve the correspondence problem between detected vehicles over frames; however without additional mechanisms they may not fulfill real-time purposes. Non-iterative approaches are used to overcome such limitations. Moreover, tracking can be applied based on single, or multiple hypothesis which are likely to improve the accuracy of the tracker at the cost of computational power [118]. Most of the trackers use the following steps: initialization and prediction; observation and data association; track update. We can distinguish between Matching and Bayesian tracking algorithms.

#### Matching algorithms

These algorithms use image features to steer a tracking hypothesis iteratively and therefore tend to refine the state estimation until convergence. In general the goal is to align a given template within the image frames in order to calculate the displacement iteratively. The popular Kanade–Lucas–Tomasi tracker uses an appearance-based model on a template to track the object. The tracker is based on the early work of Lucas

and Kanade [113] and was fully developed by Tomasi and Kanade [184]. KLT is a standard for vehicle tracking [149], and still often used in the recent literature [53] [60] [197] for complex road scenes.

#### Bayesian tracking algorithms

These algorithms model a state and the related observations as two stochastic processes. The goal is to estimate the probability of the next state, given all previous measurements, based on a conditional probability density function. The Kalman filter, also known as linear quadratic estimator [92] is one of the most popular methods [121] [170] [124] [107]. Because real-life dynamic problems are often non-linear, there have been several variations of the traditional Kalman filter [69]: the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF) aim to address the issue of (highly) non-linear dynamic systems [24]. Another Bayesian framework is the particle filter introduced in [84]. It aims to generalize the problem of the Kalman filter to non-linear systems [52] and overcomes the constraint of a single Gaussian distribution of Kalman filters. This allows to model more complex distributions as well as nonlinear transformations of random variables. The filter can enable multiple hypotheses propagation between frames by modeling arbitrary probability density functions based on particle sampling. Several applications of the particle filter for vehicle tracking can be found in the recent litterature [120] [27] [135] [12]. Bayesian frameworks for tracking remain however the trend over the recent literature.

So far, we have presented vehicle detection and tracking in general traffic monitoring applications. Earlier, we have also introduced vehicle sensing technologies as well as representative datasets for vehicle monitoring studies at road intersections. In the next part, we will focus on the special case of intersections. We will present challenges brought up by intersection monitoring systems, then we will review in-vehicle and roadside vehicle detection and tracking by monocular, stereo, and omni-directional vision.

## 2.3 Vision-based vehicle monitoring **at road intersections**

### 2.3.1 Challenges arising in the context of intersections

Meanwhile, vehicle monitoring for ITS has been an active research area for decades and achieved promising results, only a few studies have been attempted so far for intersections. Vehicle monitoring at intersections,

either with roadside or in-vehicle systems, is particularly challenging for many reasons.

#### 2.3.1.1 System setup challenges

These are essentially addressed to roadside systems, for which camera installation at the intersection needs to be flexible. Existing systems use generally a single camera looking at an intersection from variable positions (Figures 2.10, 2.11). Besides, vehicles need to be detected and tracked from many entry points [121] [104]. Moreover, it is also necessary that vehicles are visible from a relatively long distance, not just when they arrive exactly at the intersection; which is not always possible with traditional video-surveillance cameras. Thus, in order to observe the entire intersection in a single image, and monitor vehicles approaching on a long distance (Figures 2.12, 2.13), omnidirectional cameras have been recently used in this field [66] [88] [197]. However in all these works, cameras need to be pole-mounted at elevated positions and this task can be sometimes bulky [66]. Roadside camera systems should instead be designed to adapt to the infrastructure and must be easily installable. Thus, design challenges and constraints for camera-based roadside intersection monitoring are fundamentally: on the one hand, the easiness of deployment with flexibility to several types of intersections; on the other hand, the ability to monitor targets approaching from far distance, through the entire intersection.

#### 2.3.1.2 Computer vision challenges

As many general vehicle monitoring applications, almost all the existing systems for intersections monitoring require user input at initialization. Besides, the extrinsic calibration of roadside cameras is still a major challenge [51]. For a single camera, it refers to recovering the orientation of the sensor with respect to the road or traffic stream. Whereas in case of multi-camera systems, extrinsic calibration refers instead to the process of recovering the transformations that relate cameras to one another in the network. As defined, the extrinsic calibration is not straightforward. The main reason for that is the large baseline, which does not allow the use of traditional calibration grids or patterns on the ground. To the best of our knowledge multiple-camera networks are hardly used for roadside intersection monitoring, though they could offer several advantages in order to accurately track vehicles which can be easily occluded or have sudden motions [182].

The complexity of vehicle monitoring at intersections is also amplified due to vehicle motions. In fact, as opposite to highways, vehicles at

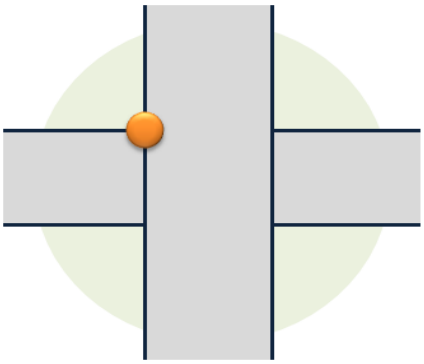


Fig. 2.10: Camera in a corner of the intersection

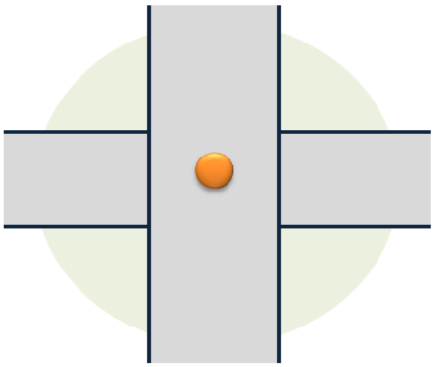


Fig. 2.11: Camera in the center of the intersection

Fig. 2.12: Example view from a fish-eye camera in a corner of an intersection [104]

Fig. 2.13: Example view from a fish-eye camera in the center of an intersection [88]

intersections can have variable, abrupt motions and can easily be occluded [193] [149]. Consequently vehicle tracking, trajectory prediction, as well as learning motion patterns in this context remain challenging and opened issues [81] [182] [63] [209].

## 2.3.2 Vision-based vehicle **detection at intersections**

### 2.3.2.1 Roadside intersection monitoring systems

#### Monocular vision

Early studies suggested that automatic monitoring algorithms at intersections should be based on local analysis of individual vehicle behavior [93]. In this context, adaptive background subtraction methods have been often used to detect vehicles [194] [193] [123] (Figure 2.15). This method has proven to be robust enough in case of illumination changes, ca-

mera noise, moving tree leaves, slow moving objects, but the performance is generally affected by shadows.

Messelodi et al. developed the system SCOCA which is a flexible real-time vision system for automating the detection of potentially dangerous situations at intersections [121]. In this work, vehicles are detected using background subtraction by binarizing the gradient of the moving map, with 90% of successful detection in good conditions. This strategy makes the process more robust to minor shadows or noise, and does not require the choice of an adaptive threshold. However, the system cannot work at night, because lights on the road surface have a major negative impact on the algorithm. To overcome these issues and develop a system that may work over long periods even by night, Arvind [117] proposed the use of infrared cameras (Figure 2.14), alongside two corners and in the center of the intersection.

In order to improve the performance of the monitoring at intersections, a multi-camera system has been developed in [181] (Figure 2.17). Thanks to the overlapping views, the intrinsic and extrinsic camera calibration are done by matching vanishing points and lines using visible pattern on the road. Vehicle are segmented based on the Mixture-of-Gaussian (MoG) algorithm for each camera. Then the segmented blobs inputs from each camera are matched, in order to give additional evidence of the vehicle from different angles.

#### Omndirectional vision

It has been recently introduced for intersection monitoring. A single omnidirectional camera can replace several traditional video-surveillance cameras normally required for monitoring large intersections. However for intersections scene monitoring, a large quantity of pixels is useless and does not cover the active areas of the scene. Therefore, Ghorayeb et al. [65] designed an optimized catadioptric vision sensor that increases the useful surface coverage in the image (Figure 2.20). The mirror-camera needs to be installed nine meters above the center of the crossroad. The performance of the system was demonstrated as they obtained 90% of useful image area, and proceeded to vehicle detection from virtual and real outdoor data [65]. However, despite this performance, the overall system deployment remains bulky, with the camera installed at an important height, and it might not be flexible for different intersections.

Fisheye optics have been more often used for intersection monitoring. In [104], the vehicle detection method consists in a background subtraction and a pixel displacement analysis. Jeffery et al. [88] proposed a fisheye vision system for data collection purposes about vehicles, motorcycles, bicycles, as well as pedestrians. The vision sensor is required to be

pole-mounted at an elevated position in a corner and directed downward the road surface. The system enables to view in a single distorted image all the roads of an intersection. The authors applied background subtraction and model-based verification. Only objects that move along consistent trajectories in appropriate directions are kept active and monitored [88] [197] (Figure 2.19). Lately, the research has evolved and led to the development of commercial applications for intersection safety, based on [88] [197], in order to ameliorate the user experience of traffic controllers [70].

#### 2.3.2.2 In-vehicle intersection monitoring systems

In-vehicle systems are mainly based on **stereo-vision** setups installed in front of the vehicle. These platforms with in-vehicle cameras have been the focus of research in recent years for Advanced Driver Assistance Systems and traffic modeling. Barth et al. [15] developed a stereo-vision in-vehicle system that calculates the optical flow to estimate trajectories of vehicles represented as rigid 3D points clouds. Aycard et al., [8] developed a stereo vision-based demonstrator, for conflict detection at intersections. In this work, cameras are placed in front looking forward in a region up to 35m in depth, with 70° horizontal field of view (Figure 2.21). A two-level architecture stereo sensor provides 6D-point information (3D position and 3D motion) using sparse optical flow of corners. Paromtchik et al. [139] worked on multimodal vehicle detection, based on data fusion from telemetric sensors and stereo-vision by means of the Bayesian Occupancy Filter (BOF). The authors showed that stereo-vision can enable vehicle detection at up to 10 m and discussed the advantage of sensor fusion. Muffert et al. [131] proposed to detect vehicles, represented by dynamic stixels, at a roundabout and warn the driver after time-to-contact computation (Figure 2.22). Though several demonstrators have been developed for in-vehicles systems, the technology itself is yet to be transfered and generalized for public use.

### 2.3.3 Vision-based vehicle **tracking at intersections**

#### 2.3.3.1 Roadside intersection monitoring systems

Many tracking approaches for roadside intersection monitoring are region-based and assume predictable vehicle motions [121]. Feature-based approaches can be used to handle partial occlusions in intersections [93]. However, the difficulties of grouping features (over-grouping), especially in far distance from the camera, can introduce important errors during vehicle tracking at intersections [149] (Figure 2.16).

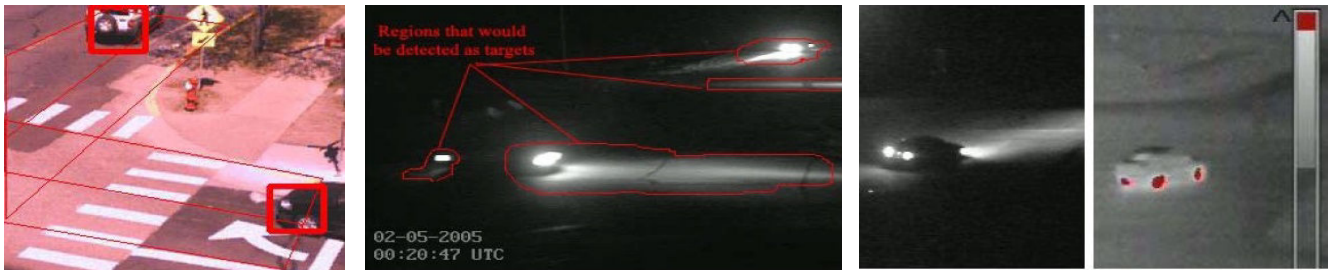
Tracking problems in intersections have been often addressed by the mean of Bayesian frameworks. The Kalman filter was used in [32] for a multiple-target tracking system at crossroad traffic. The proposed mechanism constructs candidates measurements list by first matching the sizes of the measurements and the targets. It is intended for tracking occluded vehicles without important computational complexity. For each object in the tracking list that has a vehicle ID, Kalman filtering is applied to predict its position in the next frame. This method gives an accuracy higher than 96%.

In [93], vehicle tracking at urban intersections has been tackled by using Spatial-Temporal Markov Random Field, modeled as a graph. The input image is reduced to smaller blocks which represent nodes in the graph. A solution for the object map in the current image, is found based on the previous frames and the previous object map. The result is used in a Hidden Markov Model (HMM) to detect events like vehicle collisions. In [192], graph correspondence was also proposed to track objects segmented by Gaussian Mixture Model. The objects are classified based on the main orientation of their bounding box. Example images for different weather conditions are shown without quantitative results.

In [170], the proposed approach was based on Markov Chain Monte Carlo sampling within a Bayesian framework. First, a foreground map is computed using background subtraction. A proposal map is computed from the foreground map indicating likely vehicle centroids. The distance of points from the boundary of the foreground map indicates the likelihood in the proposal map. A Bayesian problem is formulated for the vehicle positions. The proposal eliminates overlapping vehicles in 3D space and is evaluated by the match between foreground map and projection silhouettes of the 3D models. Tracking between frames is performed by a Viterbi Optimisation algorithm which finds the optimal track through the set of solutions for every frame.

In [121], the system uses explicit 3D modeling to track vehicles at intersections. The 3D models are used to initialize an object list for every fifth frame based on the convex hull overlap of model projection and motion map, with the camera calibration information. A feature tracker follows the detected objects along some frames before a new initialization takes place. The tracker is used to speed up operation, as the 3D operation would not be fast enough to operate on every frame in real time. The performance was evaluated on 45 minutes of video from two different sites, with 91.5% reliability of the vehicle counter on test data.

In [149], interest points are tracked independently at urban intersections by the mean of KLT algorithm. This provides robustness against errors in the background estimation and can deal with changing viewing




a) b) c)

**Fig. 2.14:** a) Vehicle detection from a visible camera - b) effects of headlights in night time vision - c) example use of vehicle detection with infrared cameras by night using the hypothesis of wheels temperatures [117] (2005)

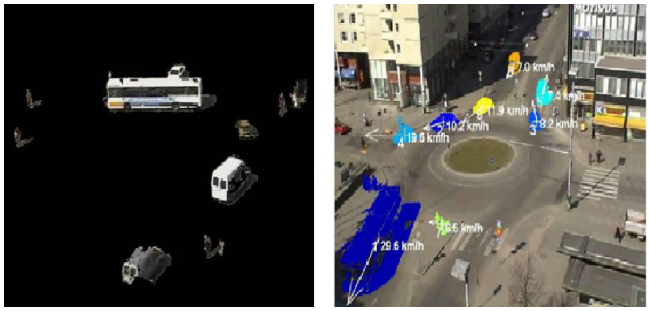


**Fig. 2.15:** Vehicle detection and tracking at intersections by 3D-connected components analysis [123] (2005)

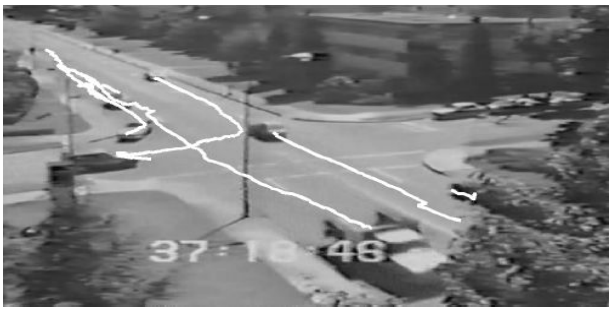


**Fig. 2.16:** A feature-based vehicle tracking at intersections [149] (2006)

**Fig. 2.17:** A multi-camera tracking system at intersections [181] (2013)

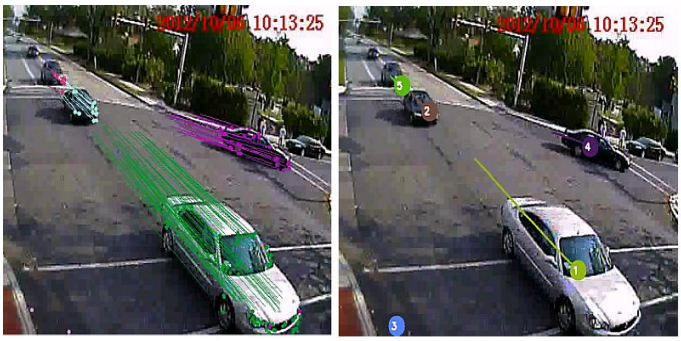



**Fig. 2.18:** Intersection monitoring with large perspective deformation [60] (2014)

angle, as no prior assumption to the constellation of feature points is made. The tracking performance is reported between 85% and 94%.

#### 2.3.3.2 In-vehicle intersection monitoring systems

For in-vehicle systems it appears that Bayesian frameworks are more common. In [3] [15], both motion and depth information are combined to estimate the pose and motion parameters of an oncoming vehicle,

| Roadside works | Sensor setup | Method | Description |
|---|---|---|---|
| (2000) Kamijo et al. [93] | Monocular vision | Frame differencing by Spatio-Temporal Markov Random Field | The algorithm was evaluated on real traffic images, with no assumption of any physical models like shape or textures. They performed multiple vehicle detection and tracking at intersections with occlusion and clutter effects at the success rate of 93%–96% |
| (2004) Messelodi et al. [121] | Monocular vision | Frame differencing | Motion segmentation by estimating the moving map. The performance of detection and classification is over 90 % in good weather. The system does not work during night. |
| (2005) Arvind M. [117] [2] | Monocular vision (visible and infrared) | Adaptive background subtraction | Region-based segmentation by iterative thresholding, for real-time vehicle trajectory estimation at rural intersections. At night, infrared cameras are used to detect vehicle wheels. |
| (2005) Heikki et al. [123] | Monocular vision | Adaptive background subtraction | Velocity profiles of each vehicle is computed from a series of consecutive images. The system cannot deal with cast shadows or occlusions. |
| (2006) Saunier et al. [149] | Monocular vision | Motion-based segmentation (feature vertices) | Points tracks (temporal series of coordinates) are grouped together to generate vehicle hypotheses. Authors reported 88.4% of occluded vehicles detected. They explained that most error are introduced by feature over-grouping |
| (2014) Furuya et al. [60] | Monocular vision | Background substraction + Motion segmentation | Motion segmentation by KLT and corner feature grouping based on similar velocity profiles. Vehicle detection with 56% accuracy. The large error is attributed to the large perspective deformation (Figure 2.18) |

**Tab. 2.1:** Selected representative roadside systems for traffic monitoring at intersections

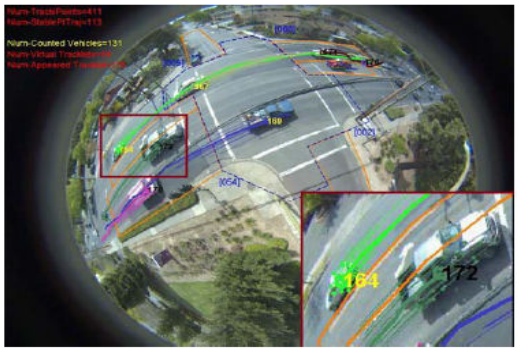


Fig. 2.19: Real time multi-vehicle tracking at Intersections from a fisheye camera [197] (2015)

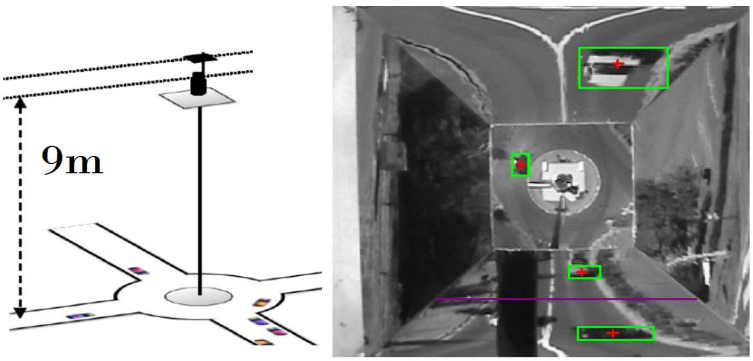
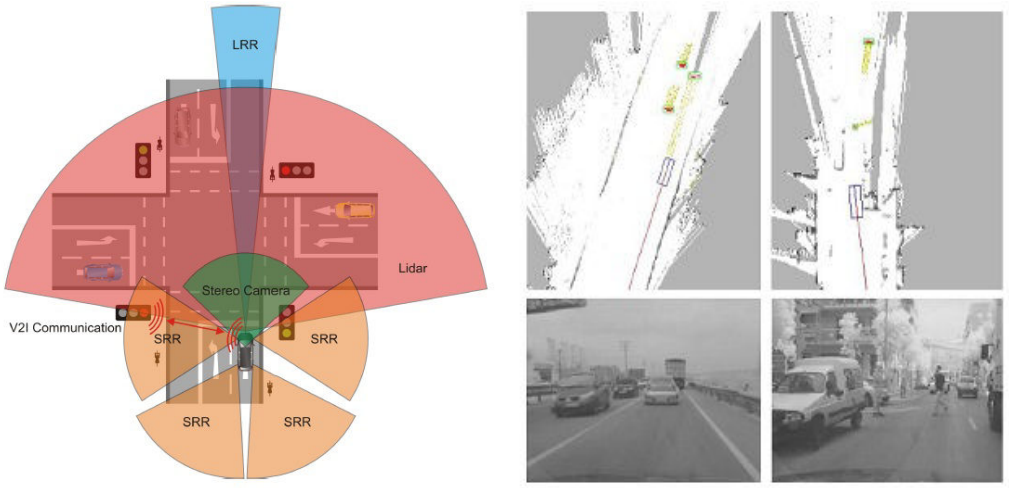



Fig. 2.20: Vehicle detection with an optimized catadioptric cameras. A major drawback is the unnatural image as well as the height of the sensor [65] (2009)

| Omni-vision systems | Description of the methods |
|---|---|
| (2006) Lee et al. [104] | Tracking in image domain by particle filter and motion dynamics – Important failure caused by appearance changes. Not robust to various weather conditions. |
| (2009) Ghorayeb et al. [65] | Pole-mounted optimized catadioptric camera providing 90% of useful image area – Detection by background subtraction |
| (2011) Gee et al. [88] | Pole-mounted at variable elevated positions – Detection by background subtraction – 3D-model based tracking |
| (2015) Wang et al. [197] | Fisheye camera at variable elevated positions – Commercial application – Detection by background subtraction and motion-based verification – Tracking by KLT algorithm with concept of grafting and identity-appearance under constrained motion. |

Tab. 2.2: Omni-vision based systems for intersection monitoring

including the yaw rate, by means of Kalman filters. In order to cover the dynamic range of a vehicle, an Interacting Multiple Model (IMM) filtering is proposed. It is able to automatically choose the right motion model for typical urban scenarios. Moreover, a gauge consistency criteria, as well as a robust outlier detection method allowing for dealing with sudden accelerations and self-occlusions during turn maneuvers is introduced.




a) b)

**Fig. 2.21:** Intersection safety platform developed in the project INTERSAFE a) sensor setup - b) Mapping and moving objects detection results [8] (2011)

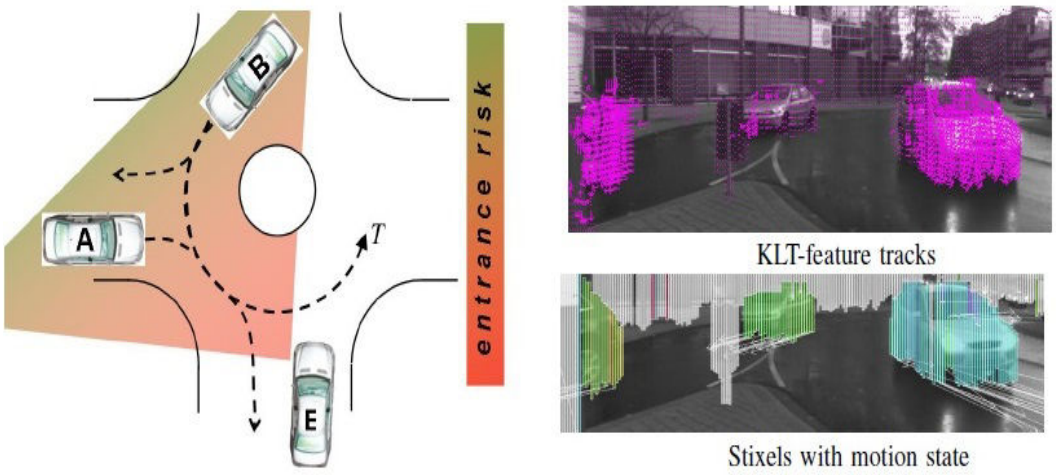



**Fig. 2.22:** Vehicle tracking by dynamic stixels and time-to-contact computation (typical situation where the ego-vehicle E has to consider both cars A and B) [131] (2012)

In [8], the authors deal with tracking multiple objects in an intersection like-scenario from a movable platform. A data association step is first performed in order to assign new objects to the existing tracks, and then optimized thanks to information provided by stereo vision. In fact, at an intersection there may be many objects moving in different directions, vehicles may be crossing or waiting to cross in a direction perpendicular to other oncoming vehicles. Tracks are validated or deleted using the outputs from the data association step. In addition, an on-line adapting version of Interacting Multiple Models (IMM) filtering technique is adopted: four Kalman filters are used to handle four motion models (constant velocity,

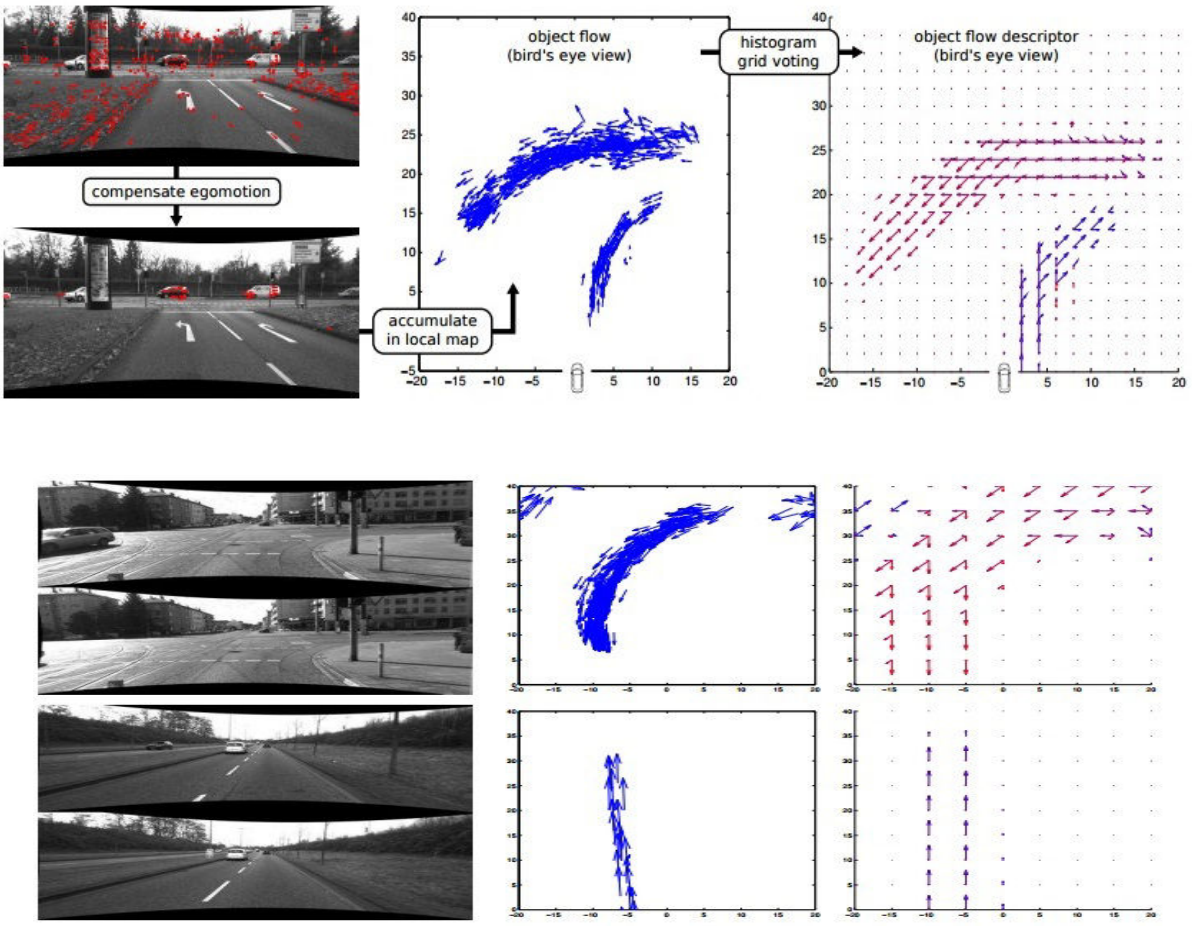



**Fig. 2.23:** Vision-based infrastructure detection with in-vehicle systems. Illustration of a stereo-based motion descriptor that registers flow vectors overtime by visual odometry (top row). Then votes cast by flow vectors are accumulated into an histogram to detect when the vehicle reaches an intersection detect whether a vehicle reaches an intersection (bottom row) [62] (2011)

constant acceleration, left turn and right turn). The output of the tracking process consists of: position and velocity information of the ego vehicle along with a list of moving object with their respective position, orientation, velocity and classification information as well as a reference to their instance in the previous frames.

In [3], a generic system for vehicle tracking, in which objects are modeled as rigid 3D point clouds moving along circular paths [131] was proposed. An extended Kalman filter is used for estimating the pose and motion parameters, including velocity, acceleration, and rotational velocity in terms of yaw rate. This feature-based approach also includes geometrical constraints that require the estimated object pose to be consistent with object silhouettes derived from dense stereo data. The Kalman filter allows for modeling a particular expected dynamic behavior of a tracked

| In-vehicle Works | Sensor setup | Method | Description |
|---|---|---|---|
| (2009) Barth et al. [3] | Stereo vision | Optical flow - Motion cues for clustering 6D points | Vehicles are represented as rigid 3D points clouds, which are grouped based on common motion. Real-time dense stereo disparity maps provide compact stixel world representation. |
| (2011) Aycard et al. [8] | Stereo vision | Sparse optical flow of corners features | A two-level architecture providing 6D-point (3D motion and 3D position) information for obstacle detection |
| (2012) Muffert et al. [131] | Stereo vision | Dynamic Stixels. | They are extracted from a stereo image using SGM along with motion data (optical flow or 6D-vision estimations) |

**Tab. 2.3:** Selected representative in-vehicle systems for traffic monitoring at intersections

instance at intersections, where there are typically two options: straight motion or turn movement. The former is best modeled by a stationary process of constant velocity linear motion, while turning vehicles require higher-order motion models.

### 2.3.4 Vehicle behavior analysis at intersections

Behavior analysis and assessment are major applications of intersection monitoring systems. Generally speaking, safety assessment is a process of finding hazardous locations, detecting accidents and underlying causes using several data sources such as crash report and conflict studies. Traditional methods which rely on manual data collection - survey forms, human field observations or user interviews - have been often used. Manual data collection, while providing rich behavioral cues, has a high monitoring cost since people must be hired for field observations. The data collection method requires significant time to record useful information through forms, interviews, surveys, and accident reports which limit the number of locations that can be analyzed. Further, while human analysis is often used as a ground truth, this may only be true over limited time scales since there is a possibility of miscalculation or missing an event due

to fatigue or other human factors. Hence, automated analysis is better suited for long-term analysis but generally provides less detailed data measurements with less semantic meaning.

Several vision-based intersections monitoring systems are intended to count traffic participants classified in different categories [197]. However, the goal of most intersection monitoring systems is to analyze the interactions between vehicles at intersections in order to detect abnormal events and prevent accidents. Thus, a robust event detection algorithm must be independent of geometric factors, such as geometry of the intersection, angle of video camera, and position where the accident occurred [93].

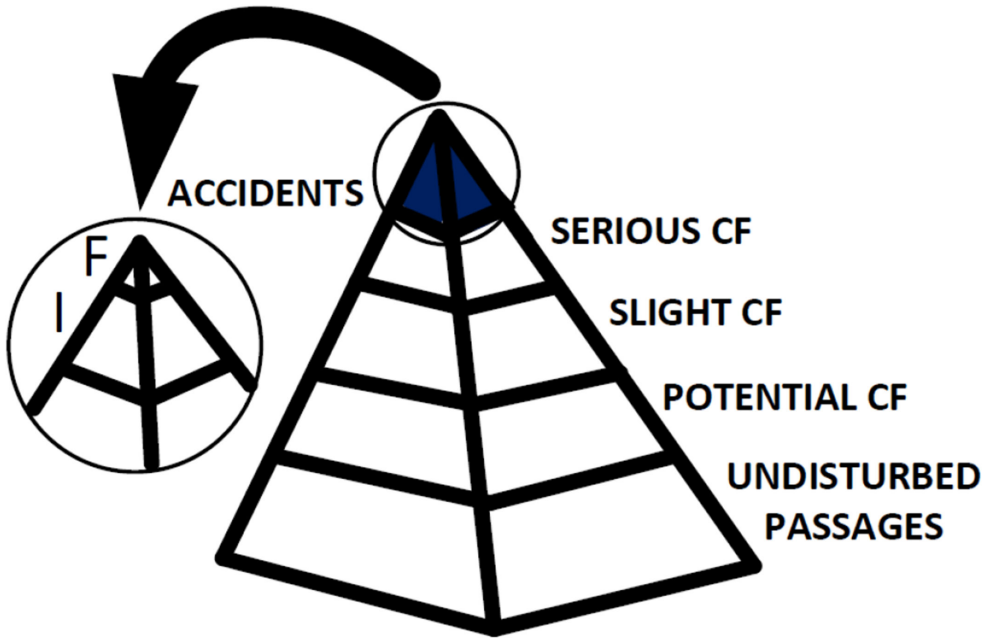



**Fig. 2.24:** Pyramidal traffic safety hierarchy [180]

Conflict and accident analysis are the most frequent addressed topics for vision-based behavior analysis and safety assessment at intersections. Generally a potential accident or collision is defined as an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remained unchanged [180]. A hierarchical concept of safety analysis based on critical observations has been discussed in [180] [180] (Figure 2.24). A recent review of safety analysis at intersections [163], shows that the most popular safety measurements indicators used in intersections related studies [163] are: Time To Collision (TTC, the time for two vehicles to collide if they continue at their present speeds on their paths), Distance To Intersection (DTI, the distance until a vehicle reaches

to stop bar with current speed on a minor road), Time To Intersection (TTI, the time remaining until a vehicle reaches the stop bar with its current speed on a minor road). These indicators are often computed from vehicle trajectories, along with speed and acceleration parameters, for both roadside and in-vehicle systems. In fact, trajectories provide consistent information about vehicle behavior at intersections.

Roadside vision-based vehicle behavior analysis required exploits mostly vehicle trajectories to detect conflicts, but may require manual initialization [162]. In [159], a typical framework is developed to facilitate manual analyses from the video recordings by only providing detection files, typical paths, distance and conflict points. After tracking and recognizing paths, pedestrian and vehicle trajectories are extracted and their counting, behavior and safety information are estimated. The evaluation is based on estimated speed profile, turning movement count, waiting time, TTI and TTC (Figure 2.25). The proposed system is semi-automatic but allows a comprehensive solution for video based behavior, safety and counting analyses at intersections with high accuracy. In [146], vehicle to vehicle conflicts are assessed by calculating TTC (Figure 2.27) from trajectories. Besides, probabilistic trajectory learning is a common technique to detect abnormal situations. In [151], a roadside system which relies on two databases was presented. First, a trajectory database, where the results of the vehicle tracking module are stored, with the generation of a distribution over possible future positions given previous positions for each road user at any instant. Second, an interaction database is created, where trajectories relations between road users are considered with several indicators. In addition, the knowledge about regular road user motion patterns are used to reduce the possibilities, in order to propose more realistic and accurate motion prediction (Figure 2.26).

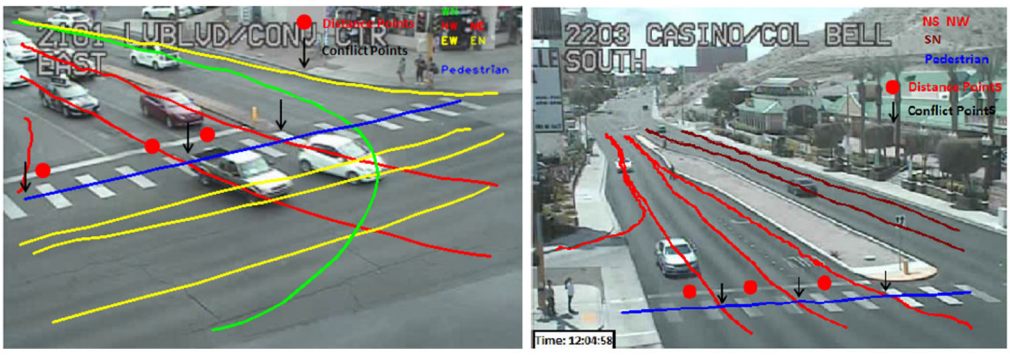



**Fig. 2.25:** Illustration of trajectories and traffic conflict points detection with a typical video-based framework [159]

For in-vehicles systems, trajectory prediction can be as important as scene understanding [105]. In [63], a conditionally independent prob-

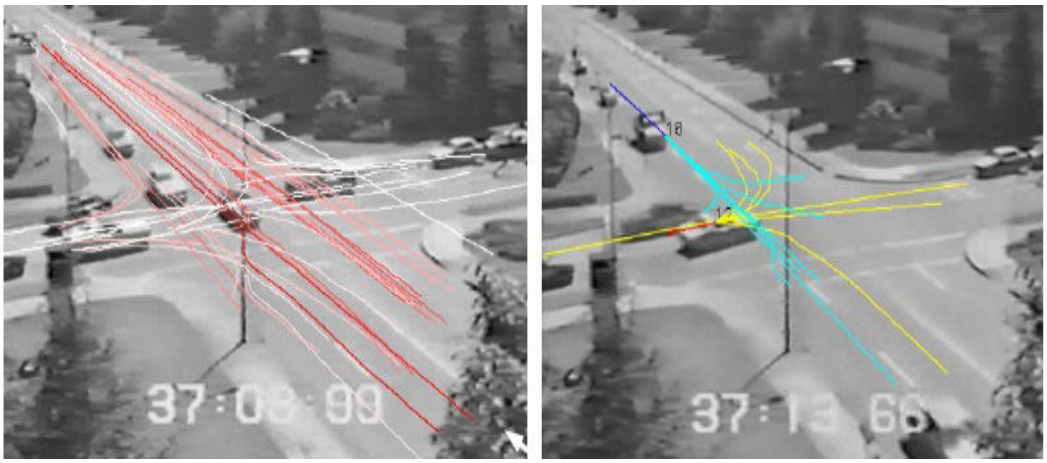


**Fig. 2.26:** Probabilistic collision prediction for roadside intersection monitoring. left: training of traffic conflicts using prototype trajectories – right: an example of movement prediction, the vehicle trajectories are red and blue, with a dot marking their position, and the future positions are respectively cyan and yellow. [151]

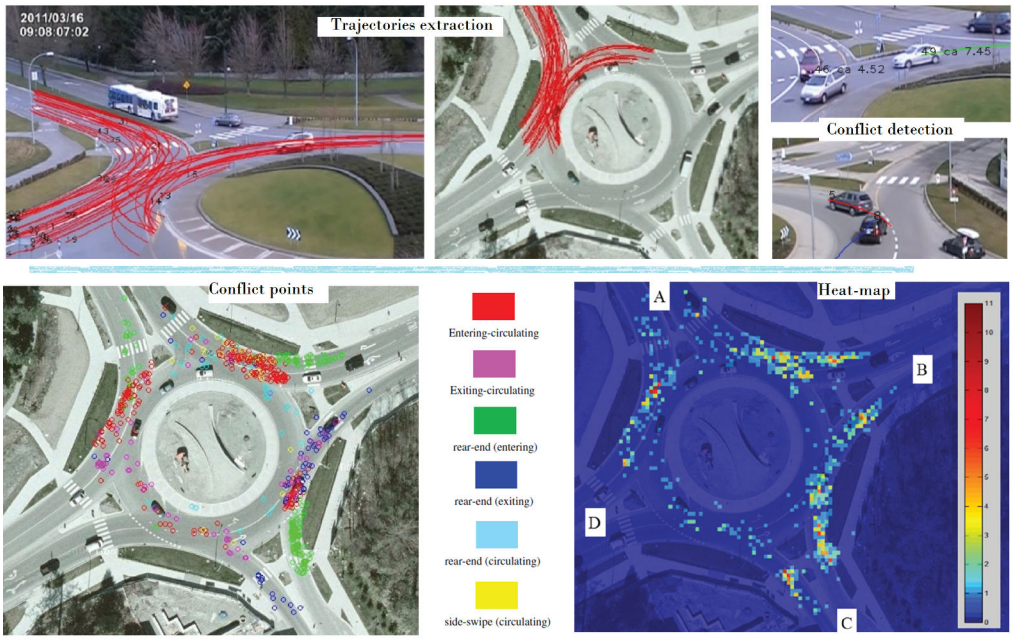



**Fig. 2.27:** Automated roundabout safety analysis by traffic conflict estimation (TTC). Vehicle trajectories are computed used to determine different classes of conflicts (top row). The (Color) Conflicts points are estimated and analyzed, per type or in a heat map [146]

abilistic model that integrates a fusion of multiple cues was proposed (Figure 2.28). On the one hand it allows autonomous vehicles to estimate the layout of urban intersections based on in-vehicle stereo vision. On the other hand, vehicle motions are learned using maximum likelihood and contrastive divergence in order to infer driving directions. The estimation of time-to-contact (TTC) is a good indicator of a potential conflic, and

generally compared with the commonly used total reaction time of 2 seconds. In [131], the goal was to predict whether a safe entrance into the roundabout is possible or not, in front of the ego-vehicle. A least square mechanism is used to fit a cluster of vehicle trajectories over time. The TTC is obtained as the ratio of the length of the circular fitted arc by the mean velocity (Figure 2.30). In [8], a dynamic circle-based strategy has been used for frontal and lateral collision prediction. In this project, the host vehicle and dynamic objects are represented as circles. A potential collision is detected when the host vehicle circle intersects at least one circle of the dynamic objects at the same time (Figure 2.29).

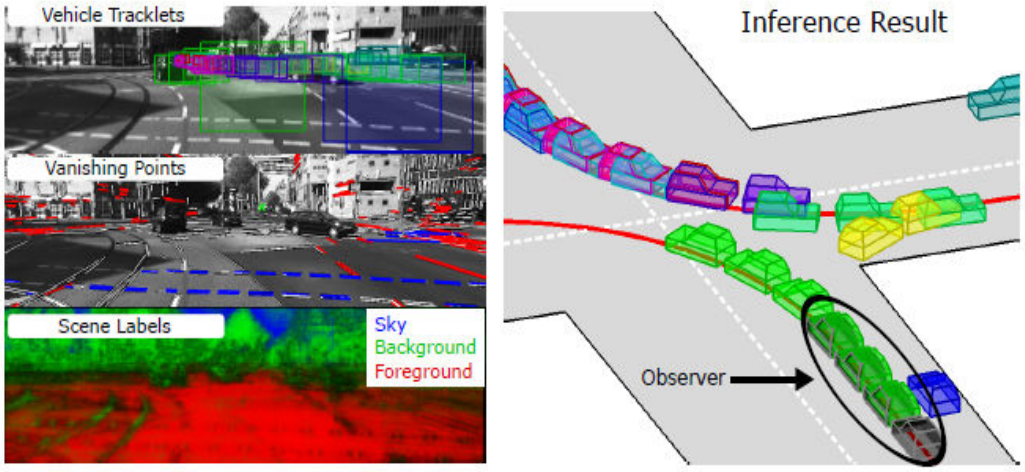



**Fig. 2.28:** Use of appearance and motion feature cues to infer the road layout and the location of traffic participants in the scene from short video sequences. [63]

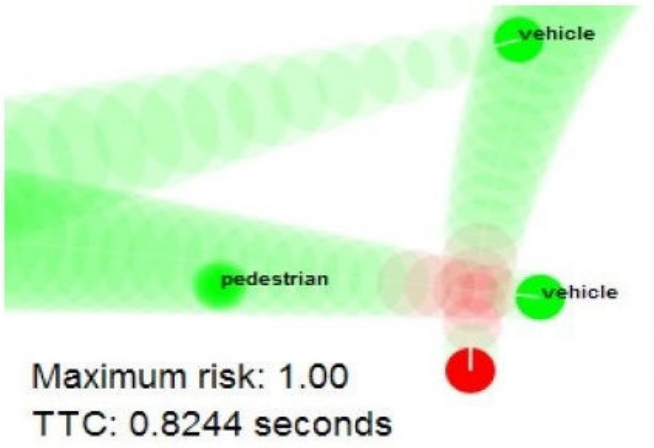



**Fig. 2.29:** Example of a potential collision between the host vehicle (red circle on the bottom) and another vehicle (green circle on the bottom right) [8]

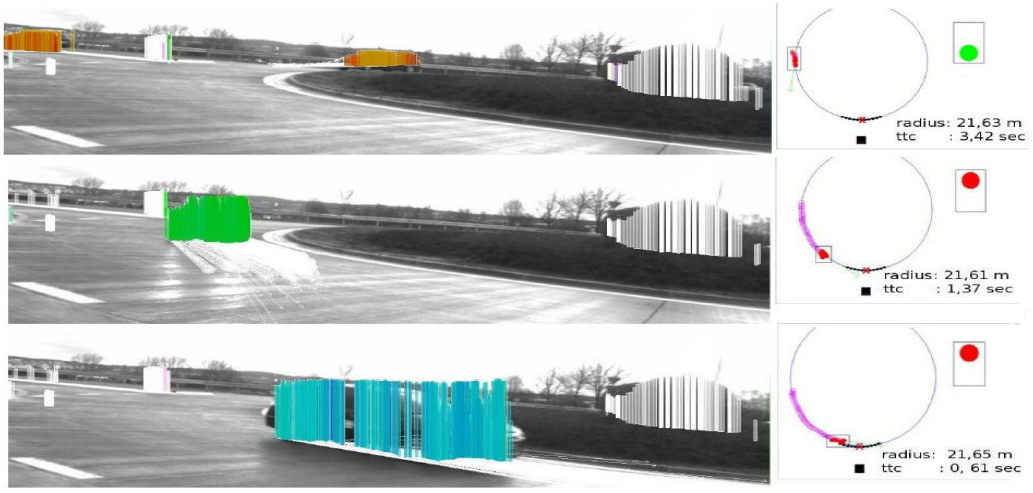



**Fig. 2.30:** Extract of a sequence: typical scene with an incoming vehicle at a roundabout. On the right side: the results of the clustering process and the TTC computation is shown for each scene [131]

## 2.4 Summary and discussion of future trends of traffic monitoring at intersections

The current state-of-the art of vehicle detection and tracking at intersections can be broken into two categories: roadside and in-vehicle systems. There have been several advances in both areas and many papers have reported excellent statistics, however the monitoring performance is still affected in many cases by the environment (illumination, shadows, night time), scene structure (large baseline, different intersections types) and particularly by the vehicle motions at intersections (occlusion, change in appearance, abrupt motions).

### 2.4.1 Roadside systems vs In-vehicle systems

#### 2.4.1.1 System setup

Roadside systems consist generally in pole-mounted cameras fixed on masts, sometimes in the corner or in the center of the intersections. Most of existing systems are not designed to be flexible to different types of intersections and do not offer a wide field of view at the intersection. In the recent literature, omnidirectional cameras have been used. Furthermore, in most studies related to roadside intersection monitoring, a user is required to execute several preprocessing tasks. Therefore, the accuracy of the detection methods highly depends on the parameters set by the user. A challenge would be to develop automatic calibration methods and a general pipeline for traffic monitoring at intersection that would require the least human intervention. Roadside monitoring systems are intended

to work for many hours, therefore data storage and transmission must be optimized to fulfill this requirements [121] [212].

In-vehicle systems for intersection monitoring are developed for Advanced driver Assistance Systems (ADAS). Vision-based systems in this category use stereo setups in order to monitor traffic in front of the vehicle when it reaches the intersection. Cooperative sensor fusion with Lidar and Radar, has been commonly used in this area lately. Also, these systems are generally intended for research and are hardly immediately available for general public use.

#### 2.4.1.2 Vehicle detection and tracking methods

For roadside applications, most real-world vehicle detection systems are based on variations of background subtraction algorithms. This is because methods such as model-based detection have a high computational complexity and barely meets the real time requirements [197]. However, background subtraction can result in important segmentation errors for cluttered scenes and occluded vehicles. Therefore, in the recent literature of roadside systems, vehicle tracking are mainly based on sparse feature matching with the popular KLT algorithm. Additional motion, model cues and prior knowledge of the scene are also necessary to accurately group features and correctly classify traffic participants.

For in-vehicle applications, the first objective is the detection and tracking of upcoming or turning frontal vehicles. Motion cues have been mostly used with different adaptations of the optical flow. Once vehicles are detected, they are generally tracked using generic Bayesian algorithms, especially Kalman and particle filters. The optical flow has also been used to detect when vehicles arrive at an intersection (Figure 2.23).

#### 2.4.1.3 Worthiness of omnidirectional vision for intersection monitoring

Omnidirectional vision allows to monitor an entire intersection, but introduces several challenges because of the large amount of visual information obtained in a single image. There are few omnidirectional vision-based systems for intersections monitoring. An optimized catadioptric camera and fish-eye optics have been used within the recent literature.

A network of calibrated omnidirectional cameras could be used to obtain rich 3D/4D reconstruction of intersections; which has not been done in any related works. This would allow to improve the accuracy of vehicle monitoring at intersections and can be generalized to any type of intersections. It would also provide an excellent realistic augmented map

of the traffic at the intersection. Finally, the system would also need to be flexible to different types of intersections, robust to difficult weather, without requiring important civil engineering, and regular maintenance. That is to say that the design of a real-time omnidirectional-vision based vehicle detection and tracking system, flexible to intersections geometries and easily reconfigurable, remains a research subject worth exploring for intelligent transportation systems.

## 2.5 Conclusion

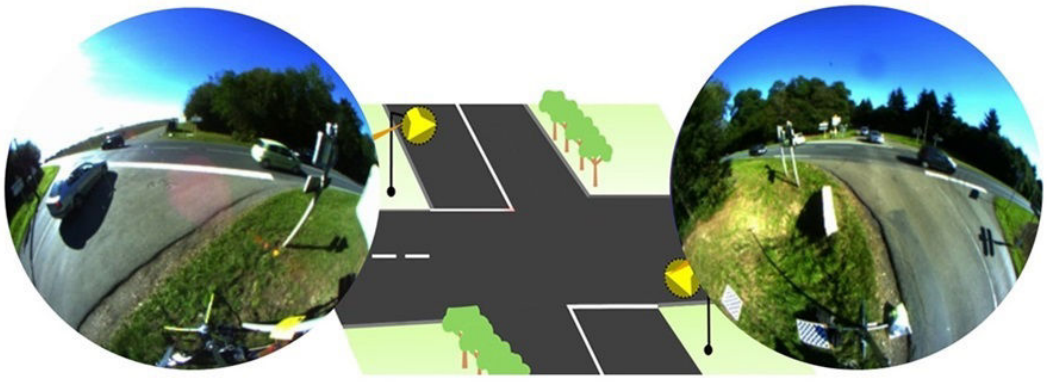
In this chapter, we have presented a review of vision-based vehicle detection and tracking systems for intersection monitoring. There has been an increase of studies focused on intersections over the last decade, with both in-vehicle and roadside systems. The former are often based on stereo-vision, whereas the latter use generally monocular vision and more recently pole-mounted omnidirectional cameras. Omnidirectional cameras provide a wide field of view and can be used to monitor all the roads of an intersection in a single image. However there are several computer vision challenges that arise when using omnidirectional cameras at intersections. In particular, it is necessary to process a large amount of visual information in real time and to monitor objects that have variable scales in the same image and can have different motion patterns. At the same time, it is also necessary to take care of classical vehicle detection and tracking issues caused by occlusion, presence of shadows, abrupt changes in the environment or the scene. There are major possible contributions directions, especially for stationary systems: the automation of the initialization steps such as camera calibration by directly exploiting vehicles motions; the development of a framework for general fusion of active and passive sensors for night time vision and difficult weather; the development of a plug-and-play tools for fast diagnosis. Besides, the use of multi-cameras systems has been hardly investigated for intersections, certainly due to the large baselines. Multiple omnidirectional cameras could be used in a calibrated network, to facilitate the detection of occluded vehicles and improve their evidence from several views. Moreover multiple-view geometry could be studied to improve the accuracy of the monitoring, using 3D and 4D reconstruction at the intersection.

Based on this analysis, the main contribution of this thesis involves the development of an omnidirectional stereo-vision system for roadside intersection monitoring, with the goal to provide as output vehicle trajectories and speeds. In this context, the problem of extrinsic calibration arises as a complex challenge. The next chapter is dedicated to the presentation of the monitoring system proposed in this thesis. We formulate

the extrinsic calibration problem, decoupled into the estimation of a pure rotation and a translation at scale between the cameras, and we describe the different steps for its resolution.

# Part II

## Wide-baseline Fisheye-Stereo for Road Intersection Monitoring



# 3 Our Monitoring System: setup, modeling and extrinsic auto-calibration

> *We view things not only from different sides, but with different eyes; we have no wish to find them alike.*
>
> **— Blaise Pascal**

Compared to human binocular vision limited to 120° in average, some animals, both predators and prays, have developed wide-angle vision, nearly up to 360° for some birds. Indeed, a large and correct perception is essential, to better analyze and understand the environment, in order to detect and react to potential danger that may come from any direction. Early concepts of wide-angle image formation appeared at the end of the 18th century with the first panoramic painting. But it was not before the middle of the 19th century, with the burgeoning field of photography, that the first original idea of the omnidirectional vision sensor was introduced [132]. With the field of artificial imaging, wide-angle perception has deeply evolved in various areas. There are several ways to obtain an omnidirectional image, which can be classified into the following three popular categories:

- **mosaicing from multiple images**: it is possible to generate a wide-angle image, by stitching a sequence of images, acquired either from a rotating camera or a multi-camera system. The first strategy has become a standard functionality available on smartphones. For the second strategy, a well-known example of panoramic vision sensor is the PointGrey Ladybug which reconstructs an image from six single lenses (Figure 3.1-a). One advantage of multi-lens panoramic imaging is the possibility to achieve high spatial resolution. However, the quality is highly constrained by several factors, such as the complexity of large-scale perspective image stitching or issues related to multi-camera cross-calibration.
- **dioptric cameras with wide angle lenses**: the most popular are the fisheye cameras which allow to obtain an omnidirectional image in a single shot. They reproduce a circular image with a field of view close to 180 degree (Figure 3.1-b, top row). A single fisheye camera does not provide a complete panoramic view of the environment. In recent years, a company named Giroptic, has developed imaging

a) Stitching-Panoramic (Ladybug)

b) Fisheye (dioptric)

c) Catadioptric

**Fig. 3.1:** Example of omnidirectional imaging systems: a) panoramic stitched image obtained with a Ladybug camera; b) top row illustrates a typical fisheye camera with resulting image (simulated [54]) - bottom row presents a Giroptic iO full 360 degree clip-on video camera for smartphones; c) a catadiopric camera and the output image [154]

plug-and-play devices, which connect to smarthphones and produce full 360 degree images, based on multi-fisheye image stitching (Figure 3.1-b, bottom row). This is testament to the compactness of fisheye cameras, which are also perfectly suitable for roadside traffic monitoring.

- **catadioptric cameras**: these types are popular for mobile robotics applications [46][154]. They are a combination of a perspective camera and a mirror, and provide nearly 360 degree field of view with a single image (Figure 3.1-c).

Omnidirectional cameras have gained interest in the field of intelligent transportations. However from the state-of-the-art, it appears roadside intersection monitoring systems are essentially based on monocular perspective vision [164, 161]. Intersections are the area of the road network with the most conflict points, where vehicles can have several motion types and abrupt trajectories. They are extremely dangerous though drivers spend a small proportion of time traveling them. Rural intersections, with higher speed limitations, account for a considerable part of the intersection safety problem. Besides, drivers errors represent a major cause of the crashes that occur at intersections. Thus, there are growing challenges related to non-intrusive traffic monitoring at road intersections, especially with omnidirectional cameras [47].

Omnidirectional cameras allow to capture in one shot more visual information of a scene. With these cameras, it is possible to monitor several roads of an intersection at the same time. However the image is highly distorted and it results in a strong variability of the vehicle appearance in the image (Figure 3.2). This is a major issue for vehicle monitoring at intersections. There are few previous works related to traffic monitoring at intersections with wide angle cameras (catadioptric [66] and dioptric [104] [88] [197] vision sensors). It is also important to point out that in all these works, the camera is pole-mounted at an important height above the ground. These works often achieve vehicle detection by background subtraction [66] [197] [88] and vehicle tracking by Bayesian filtering in the image domain [104] or by 3D-models matching [88]. In [197], the concept of identity-appearance is introduced under-constrained motion to improve the tracking and data-association between frames. However these works barely provide results at metric scale. Besides, in general a single camera might not be enough to deal with static or dynamic occlusion. But to the best of our knowledge, multi-cameras or stereo-vision systems at intersections have been hardly studied [181].

In this thesis, we present our flexible monitoring system composed of two synchronized fisheye cameras installed in the corners of the intersection, not higher than two and a half meter. This is one of the main

**Fig. 3.2:** Vehicle appearance variability in fisheye images (synthetic data from [54])

differences compared to some existing works, where the vision sensor is generally required to be placed on high masts, sometimes up to ten meters. Three main criteria were to be considered in the proposal of our fisheye-stereo sensor: the easiness of installation especially for non-specialists, the adaptability to the infrastructure allowing to fit different intersections, and the monitoring quality. The complete system requires no drastic civil engineering and can be installed without disrupting the traffic. However, this configuration brings up a problem: the large baseline between the cameras which sometimes can exceed twenty meters. And in this context, the question of calibration arises as major challenge.

Camera calibration is an important task for roadside vision-based traffic monitoring. It is often divided into two closely related parts: the intrinsic calibration which gives the camera model parameters; and the extrinsic calibration. For a single camera, the extrinsic calibration generally refers to recovering the orientation of the sensor with respect to the road. Whereas in case of multi-camera systems, like in our case, it refers to recovering the transformation that relates cameras to one another. While the intrinsic calibration parameters can be estimated prior deployment, the extrinsic calibration needs to be estimated once the cameras are installed. For traffic camera calibration, many solutions require user inputs and prior knowledge of the scene geometry to achieve accurate calibration, in order to derive traffic parameters [94] [169]. The quality of the calibration highly depends on the accuracy of these initializations. Thus, there is a need for semi or fully automatic systems [53]. The more automated the better, in order to ease the work of traffic managers.

As defined, the extrinsic calibration task is not straightforward. The estimation of the extrinsic calibration parameters for stereoscopic sensors, and especially with wide-angle cameras, is a well studied topic in the literature. However the existing popular techniques are mostly applied for short baseline multi-camera rigs, with overlapping field of view. In

practice the calibration can be achieved by taking pictures of a specific calibration object or pattern at different positions near the sensors [59], or by using the infrastructure [75], or the ground plane as reference [97] [96]. For wide-angle stereo-vision, previous works regarding the extrinsic calibration generally involve Structure From Motion (SFM) techniques, using points and lines features [140]. When it comes to traffic monitoring, these methods are hardly feasible and can be bulky. For our system especially, the large baseline does not allow the application of traditional calibration methods which are based on the use of patterns and require supervision. Indeed, it is hardly conceivable to stop the traffic and to install a large-scale chessboard pattern on the road surface for calibration purposes. So generally, the extrinsic calibration for traffic applications is estimated using several vanishing points (VPs). The latter are computed thanks to parallel lines extracted from specific landmarks such as lane marking, or from man-made structures when available [4] [94]. In the absence of extractable parallel sets of lines, the estimation of vanishing points can be performed using vehicle trajectories with the assumption of planar motion [53].

In this chapter we first provide an overview - concise and necessary to understand our work - of omnidirectional image modeling, especially for fisheye camera intrinsic calibration. The rest of the chapter is dedicated to the formulation of the extrinsic calibration problem, in the context of our challenging wide-baseline stereoscopic omnidirectional system for rural intersections monitoring. We present our complete solution to recover automatically the absolute extrinsic parameters of the wide-baseline fisheye stereo.

## 3.1 Spherical camera model

A camera projection model describes the bidirectional mapping between a world point and its corresponding image point. The projection models for omnidirectional cameras are more complex than perspective cameras because of the distortions. It results in a non-linear mapping of a 3D ray and a pixel on the image plane. Distortions are generally categorized into two types: radial and tangential often negligible. The estimation of the projection function including the distorting parameters is referred to as the intrinsic calibration of the camera. It is a necessary step for several computer vision applications and can be done offline. More details on the subject of omnidirectional camera and related projection models can be found in these references [39] [140] [154].

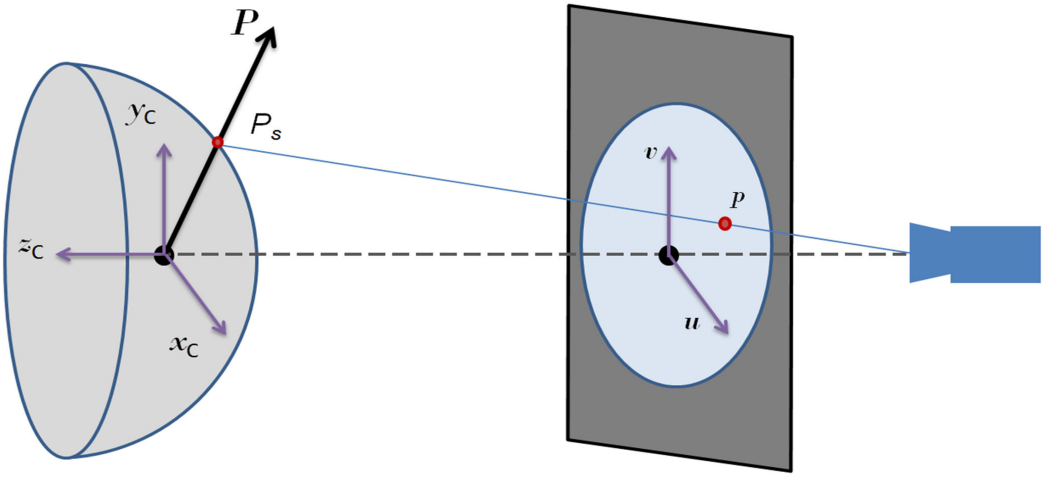



**Fig. 3.3:** Intrinsic calibration: generic spherical projection model

There are several intrinsic calibration models for omnidirectional cameras [140]. In this work we use the spherical camera model. This model was first introduced by Geyer and Daniilidis [64], and refined for central catadioptric cameras [14]. Later it has been demonstrated that the spherical mapping can also be extended to fisheye cameras [206] [39].

To estimate the intrinsic camera parameters, we use the extended generic polynomial model proposed in [153]. This generic model approximates the projection model with a parametric function instead of a specific projection function, regardless of the type of mirror or lens in the optics. Thus, the selected model allows our approach to be easily generalized to perspective, catadioptric, and dioptric sensors. As illustrated in Figure 3.3, the direction of a 3D ray $P = (x, y, z)$ projected into a fisheye image point $(u, v)$ is given as follows:

$$P = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ f(u,v) \end{bmatrix} \tag{3.1}$$

Then the point $P_s$ is obtained on a unitary sphere by normalizing the vector $P = (x, y, z)$ as follows:

$$P_s = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \frac{P}{||P||} = \begin{bmatrix} \frac{u}{\sqrt{u^2+v^2+f(u,v)^2}} \\ \frac{v}{\sqrt{u^2+v^2+f(u,v)^2}} \\ \frac{f(u,v)}{\sqrt{u^2+v^2+f(u,v)^2}} \end{bmatrix} \tag{3.2}$$

An important assumption of the model is the axial rotational symmetry of the fisheye lens, therefore the polynomial function $f$ can be written as:

$$f(\rho) = \sum_{i=0}^{n} a_{2i} . \rho^{2i} \tag{3.3a}$$

$$\rho = \sqrt{u^2 + v^2} \tag{3.3b}$$

The polynomial order is set to 4 for the calibration. For a complete model, the back-projection of a 3D-point onto the fisheye image is obtained by the inverse function $f^{-1}$ such as:

$$f^{-1}(\theta) = \rho(\theta) \tag{3.4a}$$

$$\theta = \arctan \frac{z_s}{\sqrt{x_s^2 + y_s^2}} \tag{3.4b}$$

Thus the projection of a 3D point on the spherical camera toward the fisheye image is obtained as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x_s \frac{\rho(\theta)}{\sqrt{x_s^2 + y_s^2}} \\ y_s \frac{\rho(\theta)}{\sqrt{x_s^2 + y_s^2}} \end{bmatrix} \tag{3.5}$$

In the model, the fisheye image coordinates $(u, v)$ are defined from the image center $(u_0, v_0)$. The authors proposed an affine transformation to compensate for errors in the camera setting, misalignment between lens and image plane axis and digitalising artefacts. Pixels coordinates $(u, v)$ can be transformed into coordinates $(u', v')$ defined from the upper left corner of the image. The inverse affine transformation is also defined.

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} c & d \\ e & 1 \end{bmatrix} . \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \tag{3.6}$$

The intrinsic calibration of the fisheye camera is done by estimating the coefficients of the projection function $f$ and its inverse $\rho$, as well as the parameters of the affine transformation $c, d, e$. It is done by taking several pictures of a pattern at different random positions (Figure 3.4). The toolbox OcamCalib [152] computes all the parameters by minimizing the reprojection of corners between black and white squares using Levenberg-Marquardt algorithm (Figure 3.4). Provided the spherical camera parameters, the circular fisheye image can be undistorted into a panoramic image. However in our work, we prefer to work using the spherical representation (Figure 3.5).

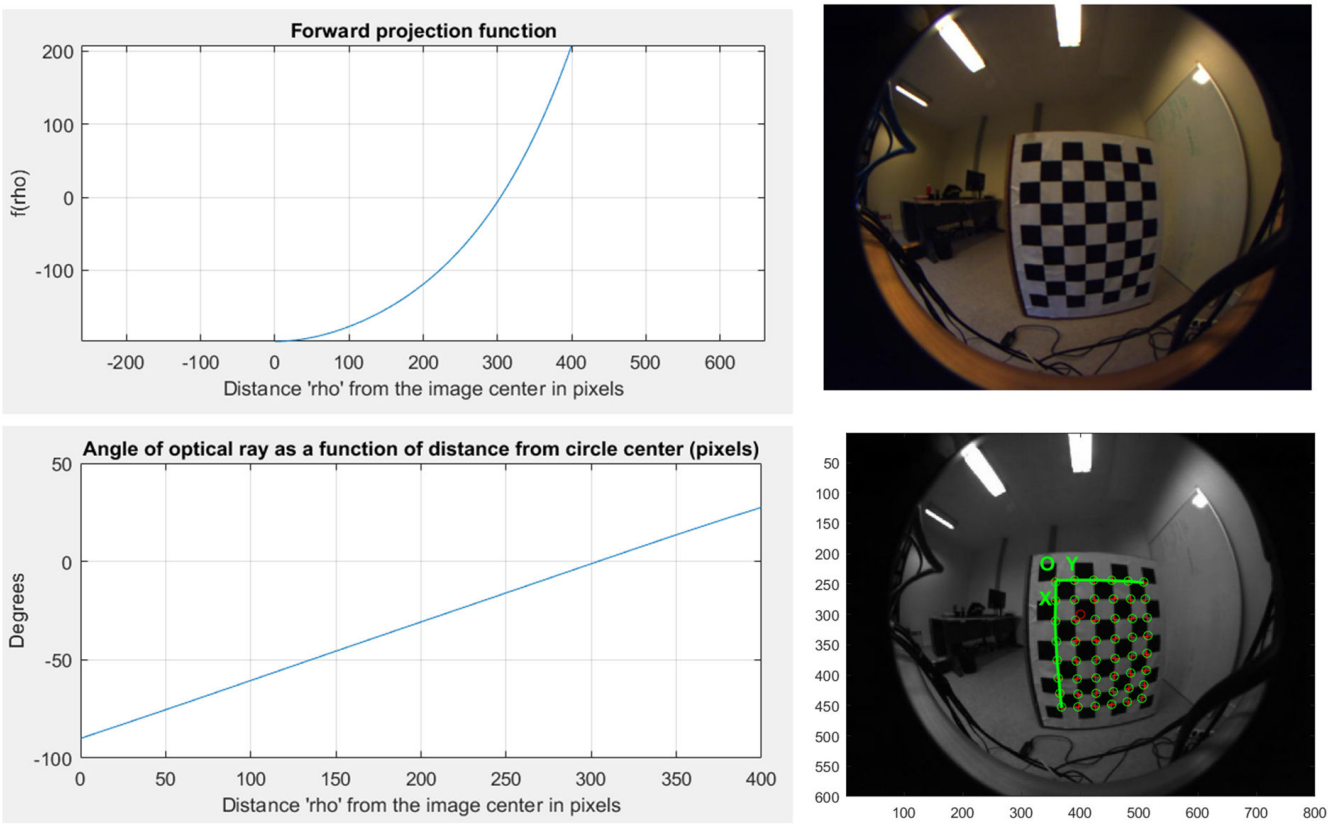



**Fig. 3.4:** Intrinsic calibration: Projection function and angle (in the example mean of the reprojection error computed over all checkerboards is 0.58; sum of squared errors 66.41) - illustration of the corner reprojection

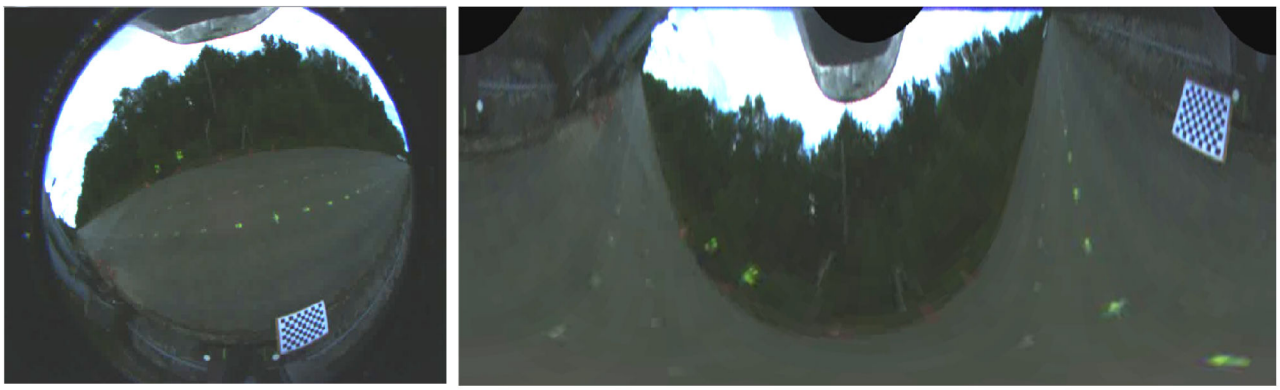


**Fig. 3.5:** Intrinsic calibration: qualitative verification by undistorting the image into a panoramic image

## 3.2 Problem Formulation: large-baseline fisheye-stereo extrinsic calibration

Examples of images acquired by the proposed monitoring system are presented in Figures 3.6 and 3.7. The monitoring sensor is composed of two synchronized fisheye cameras - *(GigE cameras, CMOS, 35.6 fps,* $1600 \times 1200$*, combined with a high resolution fish-eye lens up to 5 megapixel that offers 360* $\times$*185 degrees of viewing)* - installed in each corner of the intersection and connected to a central processing computer over WiFi. The cameras are placed on tripods, or directly adapted into the infrastructure when possible, not higher than two and a half meters. This setup allows rapid and flexible deployment for different scenarios. The monitoring system was first tested in the lab condition where several experiments were achieved (Figure 3.6). The system has been used to collect traffic data at risky intersections in Normandy (Figure 3.7). The multi-view perception of the scene offered by the system allows to strengthen vehicles evidence even in presence of occlusion, in difficult environmental conditions. However because of the large-baseline and the orientation of the cameras, the extrinsic calibration arises as a main challenge. In fact, as explained earlier in the introduction, using a calibration pattern is not possible in our context. Instead vehicles themselves will be considered as dynamic calibration objects. Thus, we have introduced an alternative novel extrinsic calibration method that is efficient, and which uses a joint analysis of vehicle motion and appearance. For generalization purposes we apply the spherical camera model. In this section, we formulate the extrinsic calibration problem. First of all we present the following coordinate systems and parameters as is depicted in Figure 3.8:

**Fig. 3.6:** Introducing the fisheye-stereo monitoring system - Lab dataset: note how the occluded black car in the left image is fully visible in the right image; multi-view data association is also a major issue

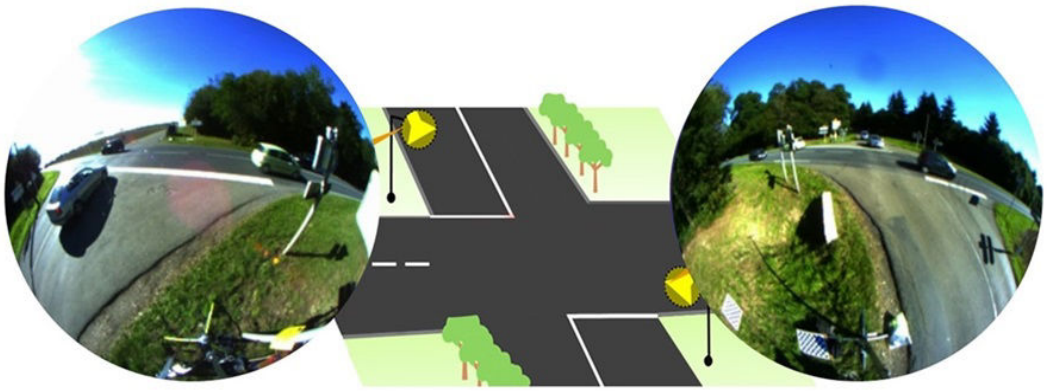


**Fig. 3.7:** Introducing the fisheye-stereo monitoring system - Rural dataset: several roads of the intersection are visible in a single image. Note also how there are important useless regions in the image covered by vegetation.

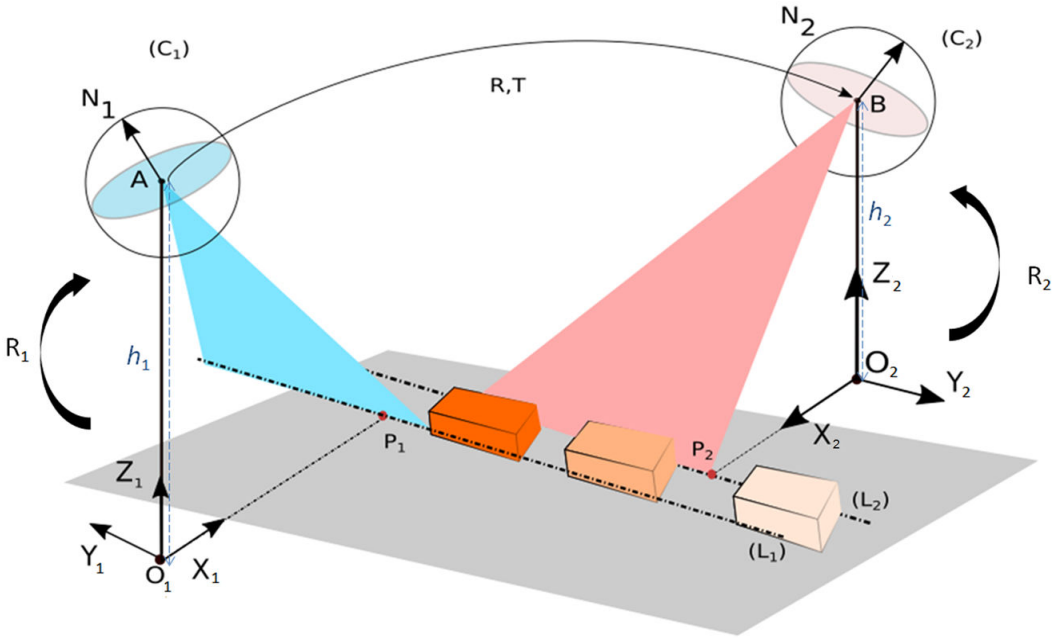



**Fig. 3.8:** Extrinsic calibration: problem formulation

• The cameras ($\mathbf{C}_1$) and ($\mathbf{C}_2$) are placed respectively at heights $h_1$ and $h_2$. The height is the precise distance between the optical center and the ground plane. As opposite to previous works [197], in which cameras should be placed high above the ground up to 10 m, we do not exceed a maximum height of 2.5 m. This allows to have a good compromise between easiness of deployment, the adaptability to the infrastructure and the monitoring quality.

• Two local references are defined vertically bellow cameras ($\mathbf{C}_1$) and ($\mathbf{C}_2$), respectively $\mathscr{R}_1=(O_1,X_1,Y_1,Z_1)$ and $\mathscr{R}_2=(O_2,X_2,Y_2,Z_2)$. Note that the world reference is placed at $\mathscr{R}_1$, and the following relations are verified:

$$X_2 = -X_1; Y_2 = -Y_1; Z_2 = Z_1 \tag{3.7}$$

• The transformation from the cameras ($\mathbf{C}_1$) and ($\mathbf{C}_2$) with their respective local mast reference $(O_1, X_1, Y_1, Z_1)$ and $(O_2, X_2, Y_2, Z_2)$ are: $[\mathbf{R}_1|\mathbf{T}_1]$ and $[\mathbf{R}_2|\mathbf{T}_2]$, where $\mathbf{R}_1$ and $\mathbf{R}_2$ are 3×3 rotation matrices, $\mathbf{T}_1 = [0, 0, h_1]^T$ and $\mathbf{T}_2 = [0, 0, h_2]^T$ are 3×1 translation vectors.

• The extrinsic calibration between the cameras ($\mathbf{C}_1$) and ($\mathbf{C}_2$) is given by the transformation $[\mathbf{R}|\mathbf{T}]$, where $\mathbf{T}$ is the extrinsic translation between the cameras at scale ($\overrightarrow{AB}$). We have the following relations:

$$\mathbf{T} = [\mathbf{T}_X, \mathbf{T}_Y, \mathbf{T}_Z]^T \tag{3.8a}$$

$$\mathbf{R} = \mathbf{R}_A^B = (\mathbf{R}_1)^{-1} \cdot \mathbf{R}_2 \cdot \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.8b}$$

Once the cameras are installed, raw measurements of the camera heights and the extrinsic translation can be obtained. These are approximate measurements that can provide an initial guess with an uncertainty window. The extrinsic calibration of the monitoring system is completely defined once we estimate and refine in the world coordinates: the extrinsic rotation matrix $\mathbf{R}$, the extrinsic translation at scale $\mathbf{T}$, the camera heights $h_1$,$h_2$. Next, we will present our approach for the extrinsic calibration.

## 3.3 Our approach: joint motion-appearance based extrinsic auto-calibration

We propose a robust approach to estimate the extrinsic calibration, by considering vehicles on the main road as dynamic calibration objects undergoing planar motion [165]. We also assume that most vehicles have a constant velocity or very negligible acceleration when they travel the intersection on the priority road. The general flowchart is presented in Figure 3.9. Our method is based on the estimation of key features which are vanishing points. They are estimated thanks to a joint analysis of the motion and the appearance of vehicles on the main road.

### 3.3.1 Vanishing Points Geometry on the Sphere

A set of parallel lines observed under perspective intersect at infinity in a vanishing point. Vanishing points (VPs) are invariant to translation and have been widely used to study the geometry of urban environments

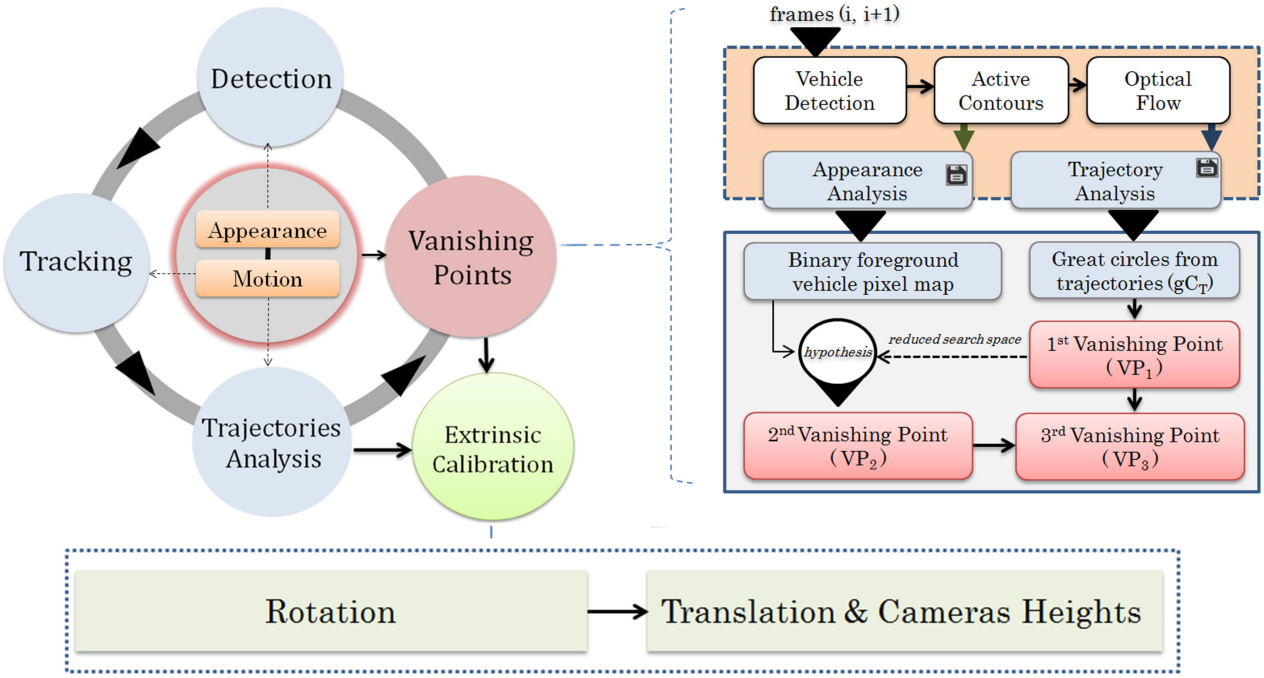



**Fig. 3.9:** Flowchart of our approach

[5] [207] as well as traffic scenes [94]. On the spherical camera, the lines project to great circles and the vanishing points result in antipodal points on the sphere [103] [100]. This has a great advantage as it allows to determine all vanishing points as finite elements on the unitary sphere. A great circle can be simply represented by its normal. Lines and vanishing point detection in omnidirectional cameras have been well studied. The Hough transform is the most popular technique to extract lines, and it has been naturally extended to omnidirectional cameras. However there are some problems regarding this method which are mainly the difficulty to find a good threshold, and the non-homogeneous resolution of the Hough parameters space on the sphere [137]. A different approach is proposed in [16], where lines are constructed from chains of connected edge pixels projected onto the sphere. A split-merge procedure is applied to cluster pixels and refine the lines estimates. Besides, there are several methods to compute vanishing points given a set of lines normals on the unitary sphere. RANSAC-based methods are efficient and can deal with large proportion of outliers. For instance in [16], the authors proposed a real-time RANSAC-based algorithm that estimates three vanishing points on Google Street View images with buildings, by inherently enforcing their orthogonality constraint. However, the estimation of vanishing points becomes a big issue when lines are unavailable in the environment or simply hard to extract. Therefore we propose an alternative solution.

### 3.3.2 Orthogonal trihedron of vanishing points

We propose to estimate three orthogonal directions by considering **vehicles moving on the main road of the intersection** as dynamic calibration objects (Figure 3.10). The method works as a joint motion-appearance closed-loop feedback framework which refines the estimations iteratively. The global context of the scene also provides important relevant information. The first vanishing point represents the first direction of the road plane and traffic stream, it coincides with the intersection at infinity of parallel trajectories seen under perspective. The second vanishing point gives the second direction of the road plane, it coincides with the intersection of parallel edges support lines seen under perspective orthogonally to the traffic motion. The third vanishing point represents the normal of the road plane.

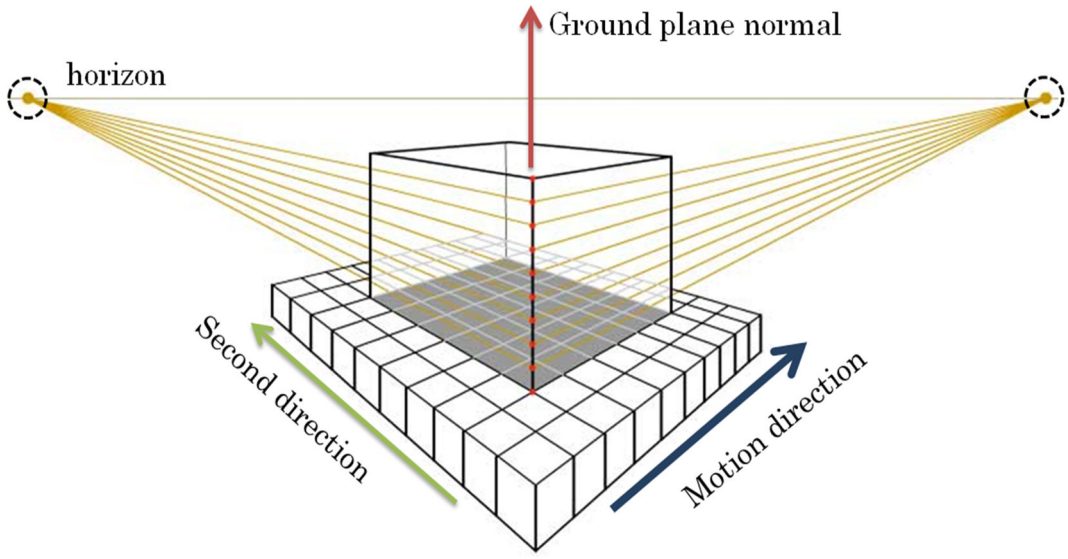
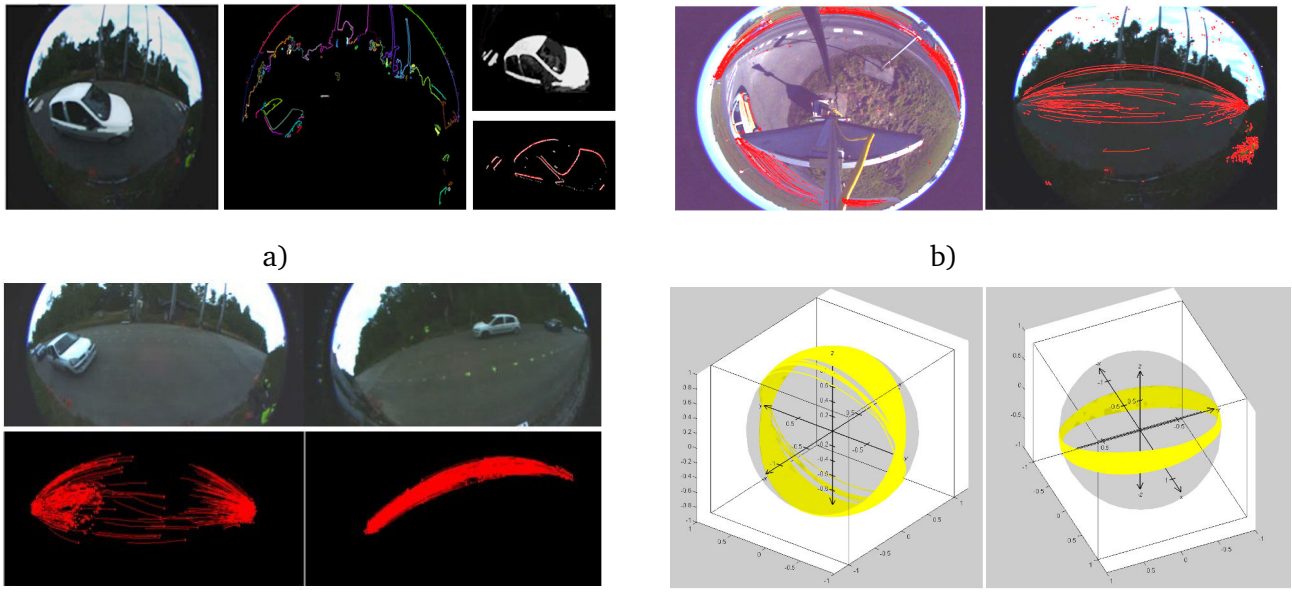



**Fig. 3.10:** Vehicle-related orthogonal vanishing direction

#### 3.3.2.1 First Vanishing Point $(VP_1)$

The first vanishing point represents the direction of the traffic motion. For traffic monitoring applications, vehicles trajectories can be used to generate virtual lines [53]. Figure 3.11 illustrates the process. Vehicles are first detected using ViBe algorithm, which incorporates a memoryless update policy and is resilient to noisy data [13] or any other robust background subtraction method [20]. Then an active contour map is computed from the foreground objects using Canny algorithm (Figure 3.11-a). Then we apply the popular KLT algorithm to track Harris corner features. The KLT tracker assumes the inter-frame motion is small, similar for neighboring pixels, and can be modeled by an affine transformation.



a) b)

c) d) Mapping on the unitary sphere

**Fig. 3.11:** Multi-view trajectory analysis framework: (a) Vehicle detection and moving edge extraction; (b) Raw vehicle tracking (noisy); (c) Multi-view vehicle trajectories beams extraction (filtering); (d) Mapping on the unitary sphere.

In other words the pixel intensities of an object will not change between consecutive frames, allowing the movement between two consecutive images to bounded withing a small patch of size $(2w_x+1)\times(2w_y+1)$, where $w_x$ and $w_y$ are integers. Thus, given a pixel $(u,v)$ in the current frame, the algorithm finds in the next frame the value of the pixel defined as $(u',v')=(u'=u+d_u, v'=v+d_v)$ that minimizes the residual intensity function $e_I$ between the two positions:

$$e_I(d_u,d_v)=\sum_{u-w_x}^{u+w_x}\sum_{v-w_x}^{v+w_x}\Big(I(u,v)-I(u+d_u,v+d_v)\Big)^2 \tag{3.9}$$

The pyramidal implementation of the KLT tracker [19] is used directly on the omnidirectional image and gives good results, without the need to undistort the image. However, the raw tracker might be noisy due to the illumination or moving clouds, or fail for few frames due to occlusion (Figure 3.11-b). In order to improve the quality of the optical flow, we perform a bilateral error verification and regular update [60] (Figure 3.11-c). The tracking results in a collection trajectories beams: $\{traj_k\}_{k:1....n}=\{(u_i,v_i,t)_k\}_{k:1....s}$, tracked corners positions over time. These trajectories beams are parallel in the world reference frame, as vehicles move straight on the main direction of the intersection. The trajectory density can also be used as a cue to detect the dominant traffic

stream. We map all the trajectories beams on the sphere using the intrinsic parameters. The following step allows to estimate the great circles that best fit each spherical trajectory. This is done by linear regression and singular value decomposition (Figure 3.11-d).

A great circle is defined as the intersection of the spherical camera and a plane passing by the center of the sphere. Therefore, for a single corner trajectory, we want in the first place to find the plane that is as close as possible to the set of $n$ consecutive corner positions expressed on the unit sphere $(P_1, ..., P_n)$. The proximity is measured by the square sum of the orthogonal distances between the plane and the corner trajectory points. Let the position of the plane be represented by a point $P_c$ belonging to the plane and let the unit vector **N** be the normal to the plane determining its direction. The goal is to minimize the orthogonal distance between a corner trajectory point $(P_i)$ and the plane ($P_c$,**N**) such as:

$$\min_{P_c, ||\mathbf{N}||=1} \left( \sum_{i=1}^{n} (P_i - P_c)^T \cdot \mathbf{N} \right) \tag{3.10}$$

A simple solution to the previous equation for $P_c$ is given by:

$$P_c = \frac{1}{n} \sum_{i=1}^{n} P_i \tag{3.11}$$

In order to estimate the normal of the plane **N**, Equation 3.10 is re-formulated as follows, by introducing the $3 \times n$ matrix **A** :

$$\min_{||\mathbf{N}||=1} \left\| A^T \cdot \mathbf{N} \right\|_2^2 \tag{3.12a}$$

$$\mathbf{A} = [P_1 - P_c, P_2 - P_c, ..., P_n - P_c] \tag{3.12b}$$

The computation of the singular decomposition of **A** gives the spherical normal **N** of the corner trajectory plane such as [7]:

$$\mathbf{A} = USV^T \tag{3.13a}$$

$$\mathbf{N} = U(:, 3); \text{given as the third column of } U \tag{3.13b}$$

The spherical normal **N** corresponds directly to the normal of the great circle from a corner trajectory. Therefore, we obtain a list of great circles from trajectories on the sphere. Under the assumption of planar motion, these trajectories are parallel and intersect at infinity in a vanishing point.

Given a group of great circles from trajectories of the dominant traffic stream, we want to determine the corresponding vanishing point $(VP_1)$. For this we propose the spherical VP-RANSAC algorithm (Algorithm 1). RANSAC [56] [34] was developed from within computer vision community and allows to deal with a large proportion of outliers in the input data. It is a resampling technique that generates candidate solutions by using the minimum number observations required (here two great circles) to estimate the underlying model parameters (the desired vanishing point $VP_1$). The parameters are defined as follows:

• $m = 2$, the random sample size; as minimum of two great circles is necessary to compute a vanishing point hypothesis

• $p$, the probability that at least one of the sets of random samples does not include an outlier.

• $u$, the probability that any great circle is an inlier, given a vanishing point hypothesis.

• $n_{max} = \frac{\log(1-p)}{\log(1-u^m)}$, the maximum number of iterations.

• $\epsilon$, the tolerance, geodesic maximum inlier error.

• $\tau$, the threshold, the minimum ratio of inliers required for a good hypothesis selection.

Two great circles are randomly selected and their associate vanishing point estimated as the cross-product of their normals (VP-guess). Then we count the number of great circles that fit the guess within a geodesic tolerance (inliers). If the ratio of inliers exceeds a predefined threshold, we keep this VP-guess as a possible good model (hypothesis). The previous steps are repeated for a defined number of iterations. At the end, the hypothesis that shows the highest consensus is selected as solution. Finally the vanishing point is re-estimated with great circles normals inliers. In order to improve the accuracy, we proceed to the re-estimation by weighing the normals of great circles, by the length of corresponding trajectories. Thus, longer and more stable trajectories contribute the most in the estimation. The estimation of the first vanishing point is accurate and fast.

Because the RANSAC algorithm is non deterministic, a preliminary analysis has been achieved in order to find the suitable initialization parameters. The probability $p$ that at least one of the sets of random samples does not include an outlier is set to 99%, while the probability of having an inlier $u$ is set to 80%. Besides, the RANSAC tolerance $\epsilon$ and threshold $\tau$ are the parameters which can mostly affect the estimation (Algorithm 1). At initialization the tolerance $\epsilon$ is set to $1^\circ$ and the threshold $\tau$ is set to $95\%$.

**Algorithm 1:** Spherical VP-RANSAC algorithm

**Result:** Vanishing Point $VP_1$

**Input:**

$\rightarrow G_{VP_1}=\{g_j\}_{j:1....s}$, $N_{VP_1}=\{n_j\}_{j:1....s}$ input great circles and normals

$\rightarrow$ RANSAC parameters: $[m=2;\ p;\ u;\ n_{max};\ \epsilon;\ \tau]$

initialization: $ii \leftarrow 0$, number of hypothesis ;

**for** $k \leftarrow 1$ **to** $n_{max}$ **do**
- Select randomly m=2 great circles with normals:$\{n_1,n_2\}$;
- Compute k-th vanishing point guess: $\Upsilon_k = n_1 \times n_2$ ;

Count the number of great circles inliers ($I_k$) that best fit such as:

**for** $j \leftarrow 1$ **to** $s$ **do**
- **if** $|\arccos(\Upsilon_k \cdot n_j)| < \epsilon$ **then**
  - $I_k \leftarrow I_k + 1$
  - keep this normal: $\Xi_k\{j\} \leftarrow n_j$
- **end**

**end**

**if** $\frac{I_k}{s} < \tau$ **then**
- $ii \leftarrow ii+1$
- keep this hypothesis: $hyp_k \leftarrow <\Upsilon_k|\Xi_k>$

**end**

**end**

Select the best hypothesis: $<\Upsilon_B|\Xi_B>$ at index B, such as $I_B=\max\{I_k\}_{k:1....s}$

Re-estimate the model: $VP_i^{cam}= \frac{1}{w}\sum_{i=1}^{r} w_i.(\Gamma_i)$ with $w=\sum_{i=1}^{r} w_i$

- $r=\binom{\alpha}{2}$, is the number of 2-subsets $\{SubG_i\}_{i:1....r}$ of $\Xi_B$
- $\{\Gamma_i = SubG_i(1) \times SubG_i(2)\}_{i:1....r}$ is the set of new VP-hypothesis
- $w_i$ is the weight of the i-th great circle inlier *(length of the trajectory or number of pixels of edges)*

In most cases the pair $(\tau = 95\%, \epsilon = 1^\circ)$ appears to be over-constrained and the algorithm fails to find a good vanishing point model after several iterations. Thus, the algorithm has been implemented to automatically tune these parameters if no VP-estimation has been performed after the maximum number of iterations ($n_{max} = 16$). Offline learning revealed that the suitable values should be taken as follows:

$$\begin{cases} 1^\circ < \epsilon < 10^\circ \\ \\ 60\% < \tau < 95\% \end{cases} \tag{3.14}$$

An example result of the traffic motion vanishing points is illustrated in Figure 3.14. In this case, both antipodal spherical vanishing point can be back-projected on the fisheye image, because the circular fisheye camera has an horizontal field of view about $185°$.

The great circle defined by the normal $(VP_1)$ contains all the possible orthogonal vanishing points to $(VP_1)$ (Figure 3.13-a). Thus the first vanishing point $(VP_1)$ is used to generate a list of hypothesis for $(VP_2)$ and $(VP_3)$, using the orthogonality constraint. This allows to generate a reduced search space for $(VP_2)$ and $(VP_3)$.

#### 3.3.2.2 Second Vanishing Point $(VP_2)$

The second vanishing point is estimated using directly vehicles' active contour map. A possible solution [49] consists in the extraction of moving vehicle edges [67] which are accumulated in parallel to the tracking. In a similar way to trajectories, the extracted edges can be mapped on the sphere and clustered. By modeling a vehicle as a rigid body with three dominant orthogonal classes of edges, the second and third vanishing points can be estimated. The estimation of this second vanishing point requires however an important amount of extracted edges segments to converge to an accurate solution. This step was particularly challenging because of the strong distortion and variable vehicle appearance. Thus, we propose to work directly on the binary moving edge map, without the need to extract and cluster edges segments.

We introduce a direct scale-invariant pixel-wise estimation method (SIP-VP, Algorithm 2). Each $VP_2$-hypothesis ($VP_{2_{hyp}}$) is tested for all pixels in the active edge map. The SIP-VP algorithm is applied directly on the sphere. The algorithm finds its roots based on the cross-product relation (Figure 3.13-b) between a pixel $(P)$ projected on a great circle, the normal $(N)$ of a great circle that contains the pixel, and a vanishing point corresponding to this great circle (here $VP_{2_{hyp}}$) :

$$P \times (VP_{2_{hyp}}) = N \tag{3.15}$$

The proposed SIP-VP algorithm (Algorithm 2) is a top-down approach. The number of possible second vanishing points hypothesis generated is controlled with the parameter $(a_p)$. For each $VP_{2_{hyp}}$-hypothesis, we compute all possible imaginary normals and store the result in a lookup table to speed up the processing. Then, the idea based on equation 3.15 (Figure 3.13-b), is to count the pixels $(P)$, in the vehicle binary edge map at the current frame, that contribute to imaginary great circles of normal $(imN)$ linked to a given vanishing point hypothesis $VP_{2_{hyp}}$ itself (Figure 3.12). The best $VP_{2_{hyp}}$-candidate gives the maximum pixel counts. The algorithm can perform well with just one frame where the appearance of

**Algorithm 2:** Scale-invariant pixel-wise VP algorithm

**Result:** Vanishing Point $VP_2$

**Input:** $\rightarrow$ SIP parameters: [$VP_1$; $\epsilon_p = 1°$; $a_p = 0.5°$ ]

Generate $VP_2$ hypothesis, with an angular step $a_p$ ;

k $\leftarrow 1$ **for** $i \leftarrow 0$ **to** $\left(\frac{90°}{a_p}\right)$ **do**

  $VP_2\_hyps(n)$=$gen\_hyp$( $VP_1$, $a_p * i$ )  $k++$;

**end**

For ($n_{frames}$) binary edge images, the pixels are projected onto the unitary sphere: $\rightarrow BinMap(i)$=$\{P_j\}_{j:1....c_k}$; where $c_k$ is the number of vehicle pixels in the current frame. For each $VP_2$-hypothesis, we compute all possible imaginary great circles normals defined by pixels in $BinMap(k)$.

**for** $ii \leftarrow 1$ **to** $k$ **do**

  **for** $j \leftarrow 1$ **to** $c_k$ **do**

    normals_vp2_hyp_LUT(ii, j) $\leftarrow P_j \times VP_2\_hyps(j)$

  **end**

**end**

- For each vanishing point hypothesis and for all corresponding imaginary great circle normals from the Lookup table ($normals\_vp2\_hyp\_LUT$), use an accumulator to count the pixels that belong to the line within the geodesic error $\epsilon_p$.

**for** $ii \leftarrow 1$ **to** $k$ **do**

  **for** $j \leftarrow 1$ **to** $c_k$ **do**

    $pixCount(ii) \leftarrow 0$

    **if** $|\arccos$(normals_vp2_hyp_LUT(ii, j) $\cdot P_j)| < \epsilon$ **then**

      count the number of pixels that contribute:
      $pixCount(ii) \leftarrow pixCount(ii) + 1$

    **end**

  **end**

**end**

How to select the best candidate ? The best candidate for $VP_2$ gives the maximum pixels count summed up for all imaginary great circles:

$VP_2_candidate \leftarrow <VP_2_hyps(b)|$normals_vp2_hyp_LUT(b, :)$>$ at index b, such as: $pixCount(b)$=$\max\{pixCount_k(1), pixCount_k(2), ...pixCount_k(n)\}$

These steps can be applied for several frames, and the most repeated $VP_2\_candidate$ is chosen as the final solution

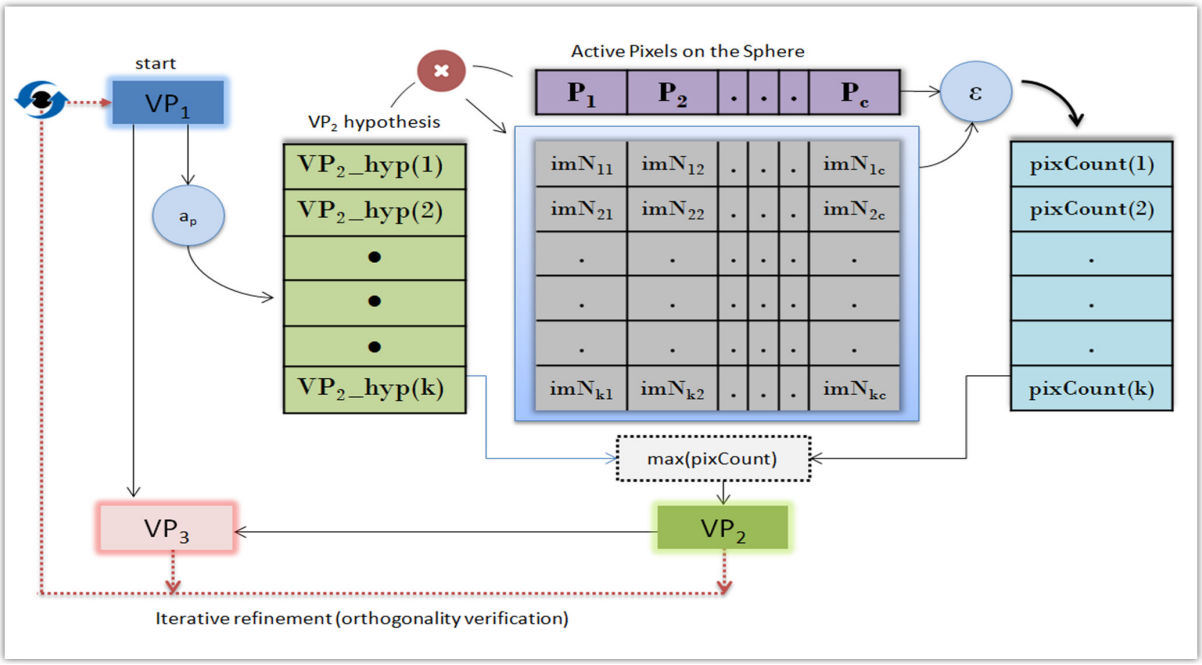
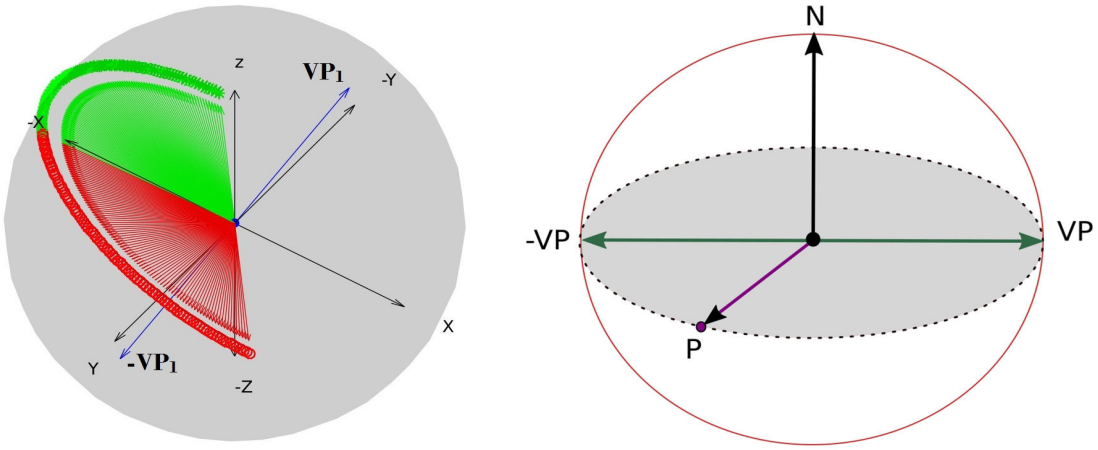


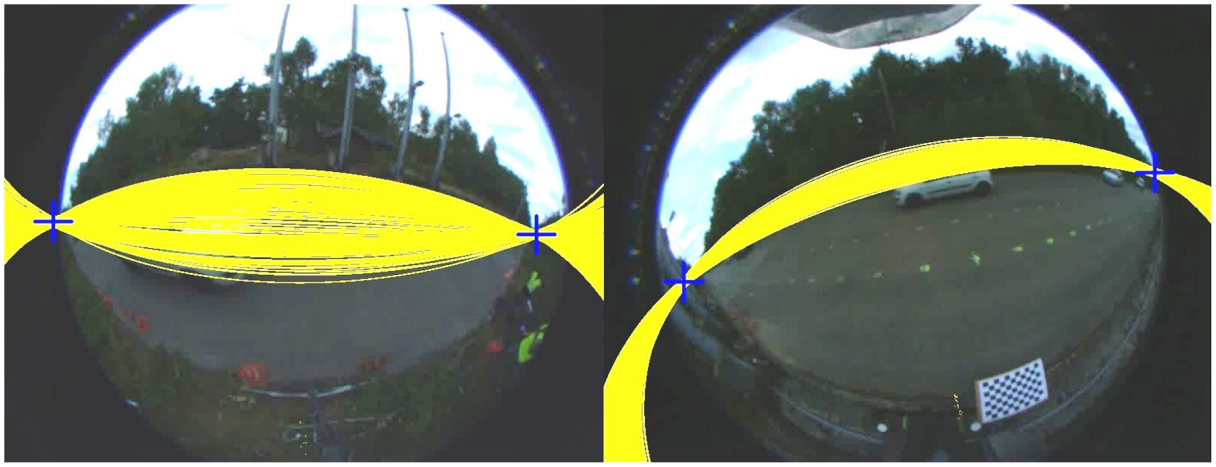

**Fig. 3.12:** Detailed illustration of algorithms for the estimation of the orthogonal trihedron of VPs




a) b)

**Fig. 3.13:** Spherical relations: a) the great circle of normal $(VP_1)$ contains all the possible orthogonal vanishing points to $(VP_1)$. The search space is divided into two subspaces of $(90°)$ using the antipodality property, in order to formulate hypothesis for $(VP_2)$ (green) and $(VP_3)$ (red). b) relation between a pixel $(P)$ projected on a great circle, the normal $(N)$ of a great circle that contains the pixel, and a vanishing point corresponding to this great circle (Algorithm 2)

vehicles is clear. However to improve the accuracy of the algorithm we repeat the main steps for a certain number of frames. Doing so we inherently take into consideration different vehicles appearances and scales, and converge to a stable global maximum.



a)

b)

c)

**Fig. 3.14:** Orthoganal trihedron of vanishing point estimation approach. a) great circles normals from trajectories are retro-projected in the fisheye image (yellow), the reprojection of $VP_1$ is represented in blue. b) location of vanishing points hypothesis, $VP_2$ (green) and $VP_3$ (red). c) the complete solution ($VP_2$ - green ; $VP_1$ - blue ; $VP_3$ - red).

#### 3.3.2.3 Third Vanishing Point $(VP_3)$

The third vanishing point gives the normal of the road plane in the spherical camera. The estimated normal can be used to extract the road plane [17]. It is computed by a simple cross product:

$$VP_3 = VP_1 \times VP_2 \tag{3.16}$$

The complete thrihedron can be refined iteratively or after a certain number of frames in a closed-loop feedback framework (Figure 3.12). The estimation is allowed to stop at any time once the orthogonality constraint is verified with a maximum error of $1^\circ$. A qualitative observation of the different steps is provided in figure 3.14.

Once the vanishing points are estimated, we can compute the extrinsic calibration with six degree of freedom. First of all the extrinsic rotation is obtained by matching vanishing points. Then we propose two methods to estimate the extrinsic translation at scale: a semi automatic approach (Figure 3.15) which requires manual inputs provided by an operator; and a fully automatic approach (Figure 3.16) which only requires the dimensions of a calibration vehicle.

### 3.3.3 Extrinsic Rotation from Vanishing Points

Each triplet of vanishing points $[VP_2\ VP_1\ VP_3]$ encapsulates the rotation matrix with respect to the local mast reference ($\mathbf{R}_1$ and $\mathbf{R}_2$ for respective cameras). Thus, the trihedrons of orthogonal vanishing points can be directly matched, with respect to the world coordinates (Figure 3.8). This leads to a direct estimation of the extrinsic rotation using a closed-form solution similarly to [114][45]:

$$[VP_2^{(\mathbf{C}_2)}\, VP_1^{(\mathbf{C}_2)}\, VP_3^{(\mathbf{C}_2)}] = \mathbf{R}_1^2.[VP_2^{(\mathbf{C}_1)}\, VP_1^{(\mathbf{C}_1)}\, VP_3^{(\mathbf{C}_1)}] \tag{3.17}$$

### 3.3.4 Extrinsic Translation from Virtual Lines

After the extrinsic rotation is estimated, we proceed to the estimation of the extrinsic translation along with the camera heights. In our early works, we presented a semi-automatic solution (Figure 3.15). For the estimation of camera heights required to select one control point for each camera on the road surface, with known distance to the camera mast [49]. The same way, front wheel positions were hand-selected in each camera as input for the extrinsic translation estimation. In order to deal with larger baselines and variable vehicle appearances, we propose to further robustify these steps. Instead of selecting manually iterative

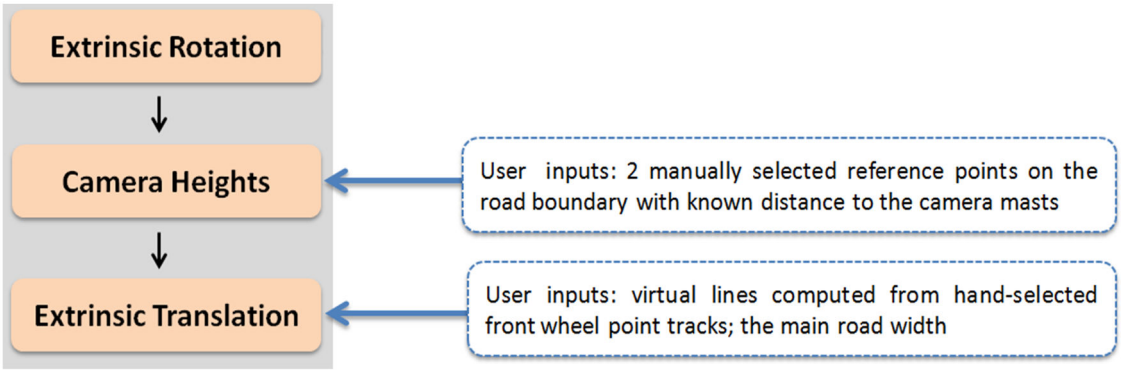



**Fig. 3.15:** Pipeline of the extrinsic semi-automatic calibration

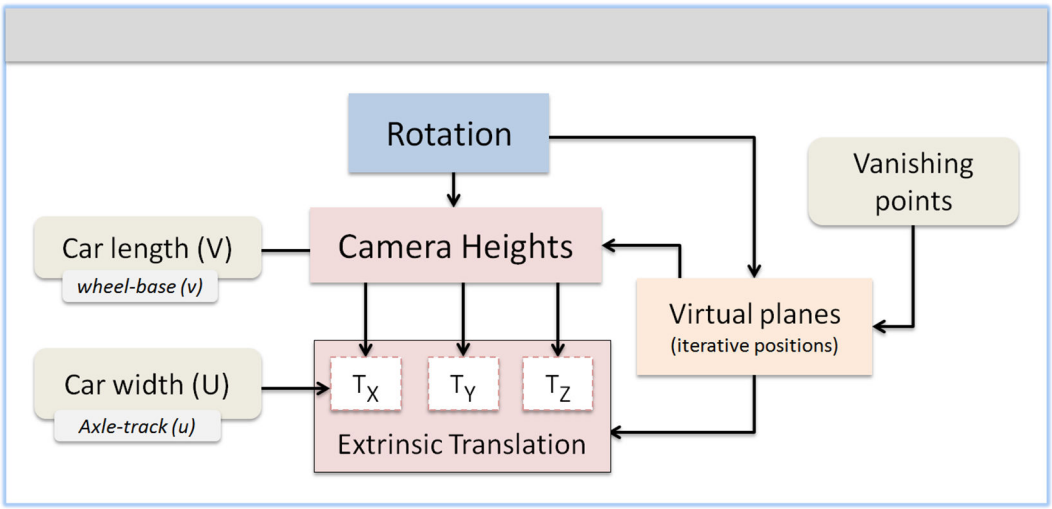



**Fig. 3.16:** Pipeline of the extrinsic automatic calibration

feature tracks, we propose an automatic method which is based on virtual planes detections. These virtual planes are tangent to a moving vehicle's front and back bumper orthogonally to the road surface.

The novel automatic framework is illustrated in Figure 3.16. The philosophy of our method is based upon the fact that a single car, moving on the main road, can be used as a dynamic calibration object. Therefore it is important to know the dimensions of the car used in the process. We can drive by with our own car, or we identify a car from the video. The dimensions required as the only inputs to our algorithm are: **the width (U) and the length (V)** (Figure 3.17, top). In practice the dimensions are obtained from automobile manufacturers specifications. It might also be possible to infer automatically these parameters using statistical knowledge of the car dimensions as a function of its blob appearance in the image [53].

In this section, we start by presenting the method used to detect the virtual planes as introduced earlier. Then we explain how to compute the

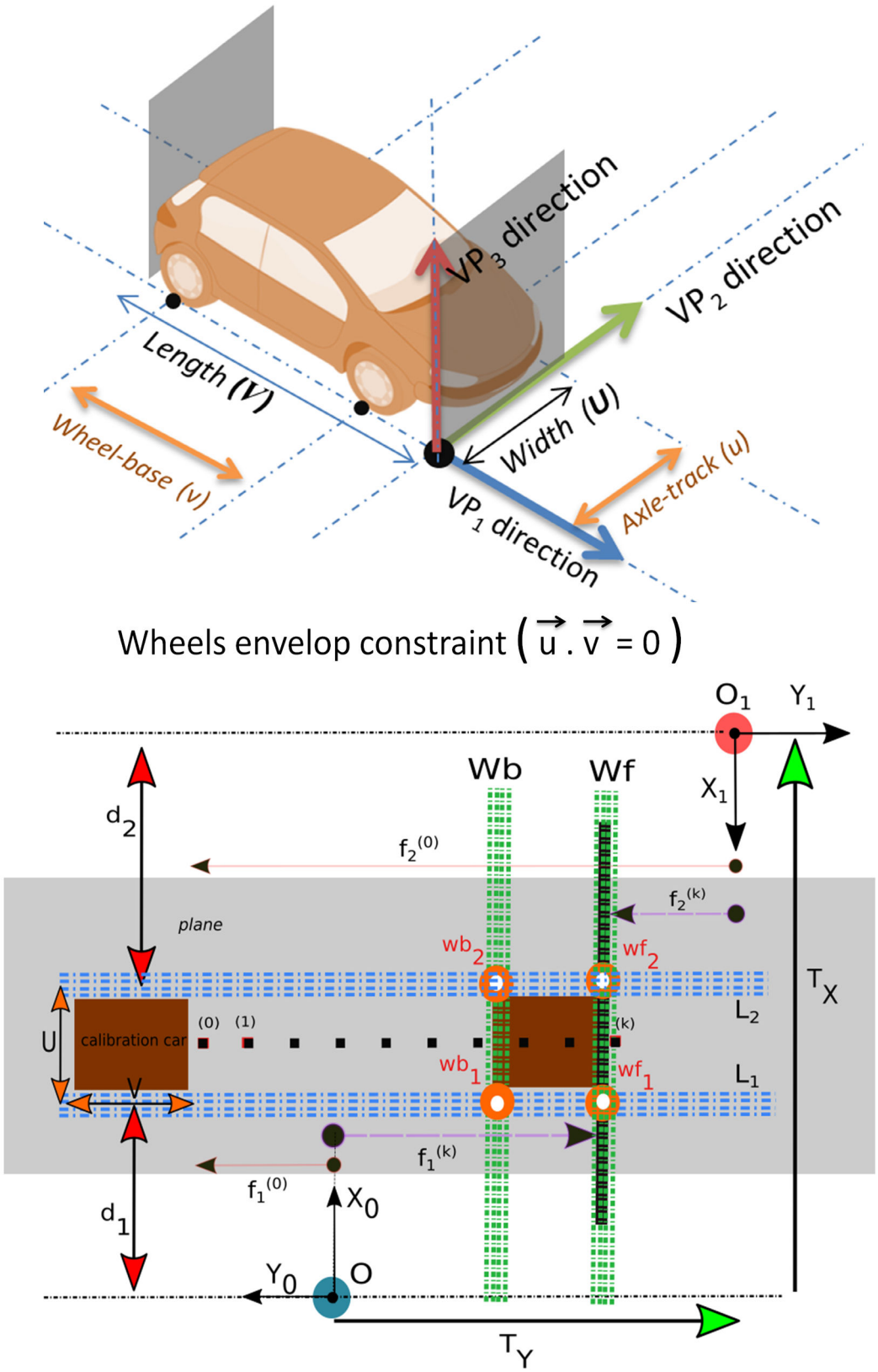



**Fig. 3.17:** Extrinsic translation estimation. Top) definition of the virtual planes used for the extrinsic translation and camera heights estimation. Bottom) top-view geometric description of the iterative virtual planes detection and representation of different parameters

camera heights and the extrinsic translation between the cameras, using the dimensions of the calibration vehicle and the virtual planes computed iteratively for few frames.

#### 3.3.4.1 Iterative Virtual plane detection

We define two virtual planes considered tangent to the calibration vehicle's front and back bumpers orthogonally to the road surface. The virtual planes detection strategy is used to provide inputs for the remaining calibration steps. The concept is to detect in each frame the position of the vehicle, defined by both virtual planes, and to feed these iterative 3D positions into our calibration pipeline.

The cameras are oriented in such a way they observe the sides of the calibration car most of time as it travels through the intersection on the main road. The detection of the virtual planes in each frame gives a good localization of the car. Besides, the virtual planes normals in the world reference coincide with the first vanishing point direction. Therefore, the origin of the planes on the main road can be defined by the intersection of two lines lying on the road plane as follows (see Figure 3.17):

• The front virtual plane is located by the pair of lines $W_f$ and $L_1$ as seen from Camera 1; $W_f$ and $L_2$ as observed from camera 2

• The back virtual plane is located by the pair of lines $W_b$ and $L_1$ as seen from Camera 1; $W_b$ and $L_2$ as observed from camera 2

The lines $W_b$ and $W_f$ match with the direction of the second vanishing point along X-axis, whereas lines $\{L_1$ or $L_2\}$ verify the direction of the first vanishing point $VP_1$ along Y-axis. In order to detect these lines and estimate their parameters, we perform a direct processing of the fisheye image similar to a bounding-box estimation as follows:

• the bounding omnidirectional line toward $VP_1$, tangent to the down-most pixel of the calibration vehicle blob is selected as the line $L_i$ (i=1,2 to denote the camera);

• the bounding omnidirectional lines toward $VP_2$, tangent to the front-bumper and back-bumper, are selected respectively as the lines $W_f$ and $W_b$.

The accuracy of the virtual planes detection obviously depends on the quality of the calibration vehicle blob segmentation in the fisheye image. While abnormal segmentation is likely to entail imperfections in the virtual plane detection, we consider that the possible errors are negligible regarding the large baseline between the cameras.

For each camera, the normals of the lines $W_f$, $W_b$ and $L_i$ are derived afterward on the spherical image. Then we want to compute

the intersection points of the lines $W_f$ and $L_i$ on the one hand, and the lines $W_b$ and $L_i$ on the other hand (as illustrated in Figure 3.17). One advantage of the spherical model is that we can estimate the intersection by a simple cross-product between the lines' normals [49] such as:

• the point $wf_i$ is the intersection of the line $W_f$ and $L_i$, and indicates the position of the front-bumper virtual plane as seen from camera i. On the spherical camera the point $wf_i$ is mapped to $sf_i$

• the point $wb_i$ is the intersection of the line $W_b$ and $L_i$, and indicates the position of the back-bumper virtual plane as seen from camera i. On the spherical camera the point $wb_i$ is mapped to $sb_i$.

The operation is repeated for a sequence of pair-wise synchronized frames. We obtain consecutive estimations for the intersection points expressed on the spherical image for each camera.

$$\left\{ SF_i = sf_i^{(1)}, sf_i^{(2)} ... sf_i^{(k)} \right\}; \left\{ SB_i = sb_i^{(1)}, sb_i^{(2)} ... sb_i^{(k)}, ... \right\}$$

#### 3.3.4.2 Estimation of the camera heights

Any point $(\mathbf{wg}_i)$ on the road plane, can be expressed in 3D-coordinates with respect to the local reference $\mathscr{R}_i=(O_i, X_i, Y_i, Z_i)$. The transformation is computed as the intersection of the ground plane and the projective ray $\left(\mathbf{C}_i, \mathbf{sg}_i\right)$. The point $(\mathbf{sg}_i)$ corresponds to the image of $(\mathbf{wg}_i)$ on the unitary sphere rectified with the estimated camera orientation. We have the following equations which allow to establish the relation between the 3D point $(\mathbf{wg}_i)$ and its image on the sphere $(\mathbf{sg}_i)$ .

$$wg_i(h_i) = \underbrace{\left( -\frac{Z_i \cdot [0,0,h_i]^T}{Z_i \cdot (\mathbf{R}_i^{-1} \cdot \mathbf{sg}_i)} \right)}_{\text{(projection scale)}} \cdot \overbrace{(\mathbf{R}_i^{-1} \cdot \mathbf{sg}_i)}^{\text{(rectified projective ray)}} + \begin{bmatrix} 0 \\ 0 \\ h_i \end{bmatrix} \tag{3.18}$$

$$\alpha_g = \arccos\left( \mathbf{R}_i^{-1}(\mathbf{sg}_i) \cdot [0,0,1]^T \right) \tag{3.19}$$

In order to estimate the height of each camera $h_i$, we can apply the previous relations (Equations 3.18,3.19) to the set of points $SF_i$ and $SB_i$ (Figure 3.18). A valid height estimation must allow to verify, at any instant k, the correctness of the vehicle length (the distance between points $wf_i^k$ and $wb_i^k$ on the ground). Given that condition, we compute the camera heights by minimizing a cost function **(E)** which depends on the previous estimations of the 3D-points $wf_i^k$ and $wb_i^k$ (function of $h_i$). We use several pairs of frames of the car at different positions in the intersection, in order

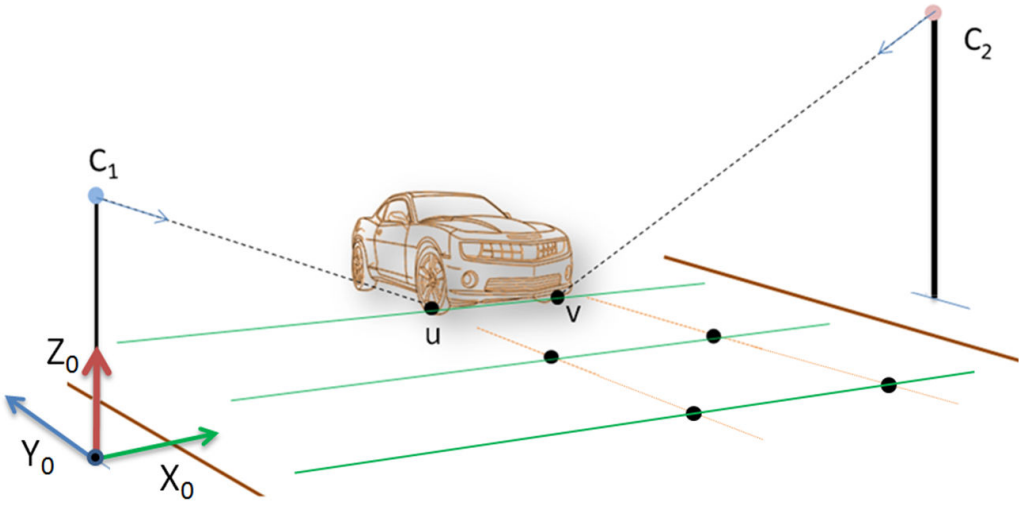



**Fig. 3.18:** Ground points lifting and 3D ray estimation in each camera frame (Illustration for the front-bumper virtual plane keypoints; the same applies for the back-bumper virtual plane keypoints)

to take into account the imperfect flatness of the road surface, as well as the strong variability of the car appearance as it travels. The cost function is given as follows **(E)** (Figure 3.18):

$$\mathbf{E}(h_i) = \sum_{k=1}^{n} \left( \left\| wf_i(h_i) - wb_i(h_i) \right\|_2 - \mathbf{V} \right) \tag{3.20}$$

$$\mathbf{h}_i = \arg\min_{h_i} \left( \mathbf{E} \right) \tag{3.21}$$

Once the cameras' heights computed, we move to the extrinsic translation. The estimation of the extrinsic translation for multi-omnidirectional cameras can be solved using multi-view geometry properties on the sphere [185] [116]. However existing methods deal with short baseline stereo-rigs generally for robotics applications [128] [129] [140]. Line features are often preferred and used in the process because they are more stable than points features [115]. Matching lines is particularly challenging with wide angle cameras in the context of intersections, not only because of the large baseline between the cameras as depicted earlier, but also because they might not be extractable and relevant lines.

In order to tackle the absence of lines, we propose to use joint motion and appearance cues. Our idea is to match the virtual planes previously estimated. In order to compute the extrinsic translation only the sets of points $SF_i$ (i=1,2) that define the iterative positions of the front-bumper virtual plane will be exploited.

#### 3.3.4.3 Translation along the X-axis ($\mathbf{T}_X$)

It can be easily noticed that, the translation along the X-axis can be computed if the distances between the side lines ($L_1$, $L_2$) and the respective cameras masts are known, however up to an uncertainty. The uncertainty is related to the vehicle width or axle-track dimension (**U**).

$$\mathbf{T}_X = d_1 + d_2 + \mathbf{U} \tag{3.22}$$

The distances $d_1$ and $d_2$ represent the lateral position of the calibration vehicle with respect to the cameras' masts [49]. In fact the distances $d_i$ ($i = \{1,2\}$) represent the X-coordinates of the points which lie on the lines $L_i$. Therefore $d_i$ can be computed from points $(wf_i^k)$ after a certain number of frames (s>1) such as:

$$\left(\frac{1}{s}\sum_{k=1}^{s}(wf_i^k)_{\mathscr{R}_i}\cdot X_i\right) - \mathbf{d}_i = 0 \tag{3.23}$$

#### 3.3.4.4 Translation along the Y-axis ($\mathbf{T}_Y$)

We define the vectors $\overrightarrow{f_1^{(k)}}$ and $\overrightarrow{f_2^{(k)}}$, which represent the location of the virtual plane with respect to each camera mast (Figure 3.17). Thus, it can be easily noticed that the difference between these vectors must have constant norm in time ($\lambda > 0$) regardless of the position of the car in time (frame k) [49], as follows:

$$||\overrightarrow{f_2^{(0)}} - \overrightarrow{f_1^{(0)}}|| = ||\overrightarrow{f_2^{(1)}} - \overrightarrow{f_1^{(1)}}|| = \cdots = ||\overrightarrow{f_2^{(k)}} - \overrightarrow{f_1^{(k)}}|| = \lambda \tag{3.24}$$

The vector $\overrightarrow{f_i^{(k)}}$ (camera $i = \{1,2\}$, frame $k$) is innately controlled through the formulation of the virtual plane. As illustrated, we observe that it actually corresponds to the component of the points $(wf_i^k)$ along the Y-axis in the associated camera frame. Therefore we have:

$$\overrightarrow{f_i^{(k)}} = (wf_i^k)_{\mathscr{R}_i}\cdot\overrightarrow{Y_i} \quad (i = \{1,2\}) \tag{3.25}$$

The constant value $\lambda$ which actually represents the norm of the extrinsic translation along Y-axis ($\mathbf{T}_Y$) can be computed after a certain number of frames (s>1) such as:

$$\left(\frac{1}{s}\sum_{k=1}^{s}||\overrightarrow{f_2^{(k)}} - \overrightarrow{f_1^{(k)}}||\right) - \mathbf{T}_Y = 0 \tag{3.26}$$

#### 3.3.4.5 Translation along the Z-axis ($\mathbf{T}_Z$)

The translation component ($\mathbf{T}_Z$) can be naturally computed as the difference of the cameras heights. It is intrinsically refined through minimization defined above.

$$\mathbf{T}_Z = \left( h_1 - h_2 \right) \tag{3.27}$$

## 3.4 Conclusion

In this chapter we have presented our main contribution. In summary, we have introduced a method which allows to estimate the extrinsic calibration of a large baseline fisheye-stereo at road intersections. Vehicle motion and appearance cues on the main road are used to estimate vanishing points in each camera frame, which in turn are matched to estimate the extrinsic rotation. Then a single vehicle with known axle-track and wheelbase dimensions can be taken as a dynamic calibration object to compute the extrinsic translation at scale. Our approach is suitable for difficult environment such as rural scenes where the absence of lines makes the calibration challenging. As formulated, the method can be applied to estimate the pose of traffic cameras for different applications, provided straight planar motion on one main direction.

In the next chapter we present experiments carried out in the lab, as well as and on real intersections, in order to discuss the performance of our extrinsic auto-calibration approach. The monitoring system is tested for distance measurement on the road surface, vehicle lateral position estimation. We also introduce and evaluate a vehicle trajectory estimation framework, which allows vehicle speed estimation.

# 4 Experimental Results: Extrinsic calibration and Trajectory Analysis

> *Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning*
>
> **— Albert Einstein**

This chapter presents results of extensive experiments performed in order to evaluate and discuss the performance of the wide-baseline fisheye-stereoscopic system. It is important to point out, that to the best of our knowledge there were no roadside traffic dataset acquired with fisheye cameras in rural environments, along with ground truth intrinsic and extrinsic calibration parameters. The only dataset that we came across recently [54] appeared to be unsuitable for the evaluation of our method. It contains only one short sequence recorded with a fisheye camera installed high above the ground-level, not in a rural environment, neither an intersection. Because of this lack, we have carried out extensive datasets acquisition, both in the lab and at rural intersections.

## 4.1 Datasets and Evaluation Protocol

We present and discuss experiments carried out in real conditions, in the lab (Figure 4.1) as well as at rural intersections (Normandy-France, Figure 4.2). All the sequences are acquired in sunny or cloudy weather, with various lighting conditions, and with presence of vegetation in the environment. However we did not perform any experiment in hazardous weather conditions. The complete evaluation datasets contain three wide-baseline fisheye stereo sequences in the lab (LA, LB, LC), and the same number for rural intersections (SA, SB, SC). Our evaluation involves the following aspects: vanishing points, extrinsic calibration, trajectory analysis.

—**Vanishing Points Error**: we discuss the performance of our algorithms for orthogonal vanishing points estimation [datasets LA, LB, SA, SC]. The performance is discussed according to the geodesic Line-VP error. The latter measures the distance on the unit sphere between a set of parallel ground truth lines and the corresponding estimated vanishing point. Ground truth lines are either hand-selected from purposely drawn pat-

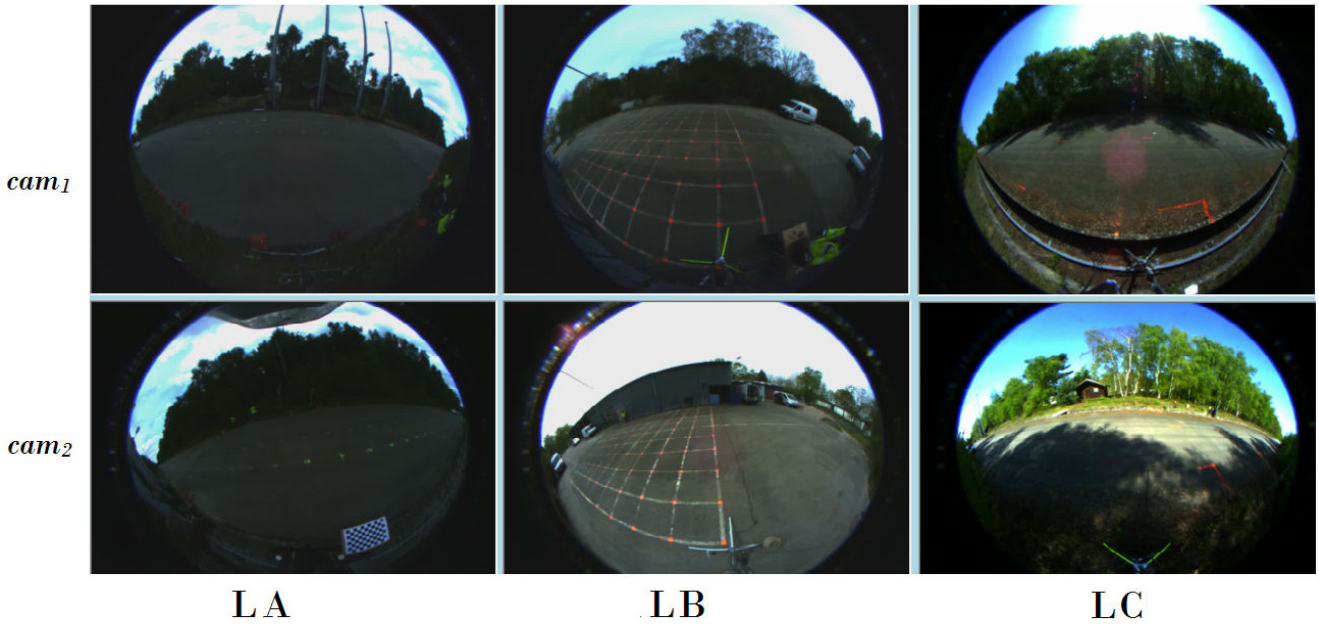



**Fig. 4.1:** Wide-baseline fisheye-stereo dataset acquired in the lab, in various lighting conditions

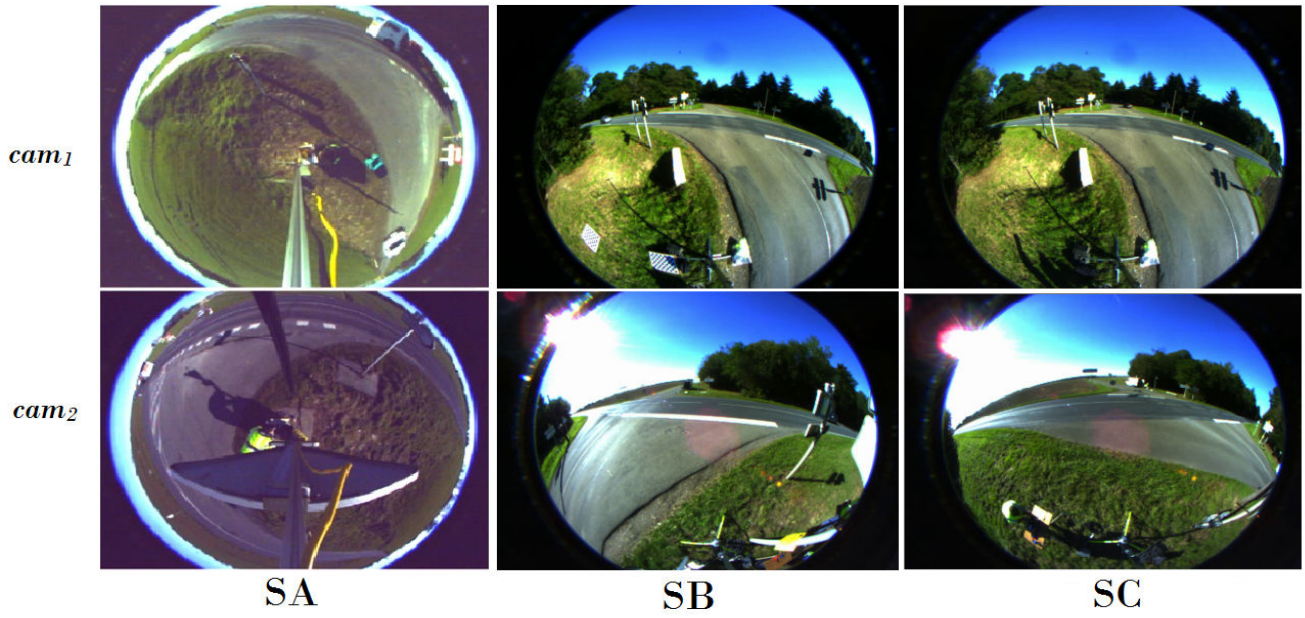



**Fig. 4.2:** Wide-baseline fisheye-stereo dataset acquired at rural intersections in Normandy, France, in various lighting conditions (Note that Sequences SB and SC are acquired on the same intersection, but with a different position for the second camera)

terns (only possible in the lab), either from available structures in the scene, or generated from extended vehicle edgelets. Especially for the third vanishing point, which defines the road plane normal, vertical poles known to be perfectly orthogonal to the road surface are selected. All three vanishing points are evaluated on datasets LA (first camera) and LB (both camera), whereas we focus on the third vanishing point for datasets SA (both cameras) and SC (both cameras).

—**Extrinsic Calibration Error**: we present results obtained for the extrinsic rotation and the extrinsic translation at scale [only datasets LB (baseline larger than 13 m) and SB (baseline larger than 22 m)]. Complete ground truth references for extrinsic calibration parameters were only available for the lab dataset LB recorded on a high-speed test drive field. To acquire the ground truth a $10\ m^2$ square calibration grid is painted on the road surface. The ground truth poses of the cameras and heights are obtained by solving a plane induced homography on the equivalent sphere [17] [152]. We also make a comparison between the semi-automatic [49] and the full-automatic approaches. However no ground truth extrinsic calibration was available for the rural intersection dataset SB. Thus we propose a different evaluation approach, which give excellent insights of the extrinsic calibration accuracy, even in extreme conditions. In this case our evaluation simply quantifies the reprojection and localization error of vehicle wheels in the image. We compute vehicles dimensions (wheel-base and axle-track) on the main road of the intersection, and compare the results to ground truth retrieved form car manufacturers specifications.Then, we evaluate the orthogonality between the wheel-base and the axle-track.

—**Trajectory Analysis**: we present results regarding vehicle trajectory reconstruction and speed measurement [datasets LC and SB]. Our approach falls within the scope of structure-from-motion (SFM), which is primarily concerned with the 3D reconstruction of a rigidly moving object seen by a static camera [101]. We introduce a method for metric trajectory reconstruction within a Bayesian framework along with speed estimation. In this work, we are mainly interested in vehicle driving with higher speed limitations on the main road of intersections. Thus turning motion analysis is not concerned for this evaluation. We first evaluate vehicle trajectory with dataset LC, and we discuss particularly the average lateral position within a wide visibility range, with comparison to ground truth data provided by a LIDAR. Then we evaluate the performance of average speed estimation with dataset SB on a rural intersection, where ground speeds are computed by estimating the travel time between virtual gates.

## 4.2 Vanishing Points Estimation

Vanishing points are essential features in our approach. The performance of the extrinsic calibration highly depends on the accuracy of the complete triplet of orthogonal vanishing points. The first vanishing point $VP_1$ is particularly important and computed from corner trajectories (see Algorithm 1, parameters: tolerance $\epsilon = 2^\circ$ and threshold $\tau = 90\%$). A short sample of extracted corners trajectories is presented in Figure 4.3, with back-projected trajectories (yellow conics) and corresponding vanishing point ($VP_1$, blue) in the fisheye image. For this example, the maximum orthogonality error between the computed $VP_1$ and any of few randomly selected inliers corner trajectories is about 0.3° (both cameras, Figure 4.4). In fact, regardless of the dataset, the robustness and stability of the first vanishing point $VP_1$ is crucial, because it is used to generate hypothesis for $VP_2$ and $VP_3$ (see Algorithm 2).

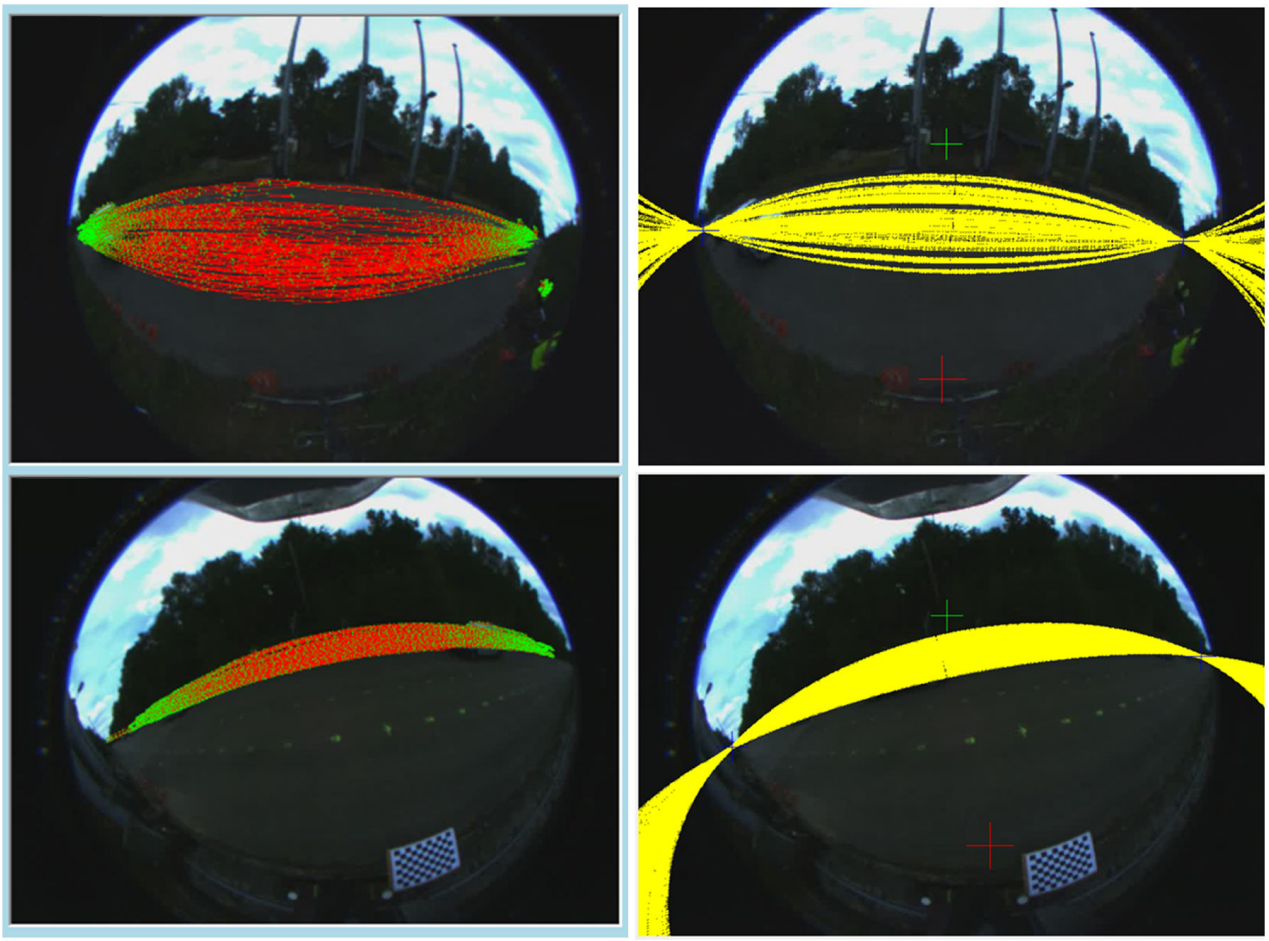


**Fig. 4.3:** Sample of extracted corner trajectories, back-projected inliers conics [LA dataset], and computed vanishing points ($VP_1$ in blue, $VP_2$ in green, $VP_3$ in red)

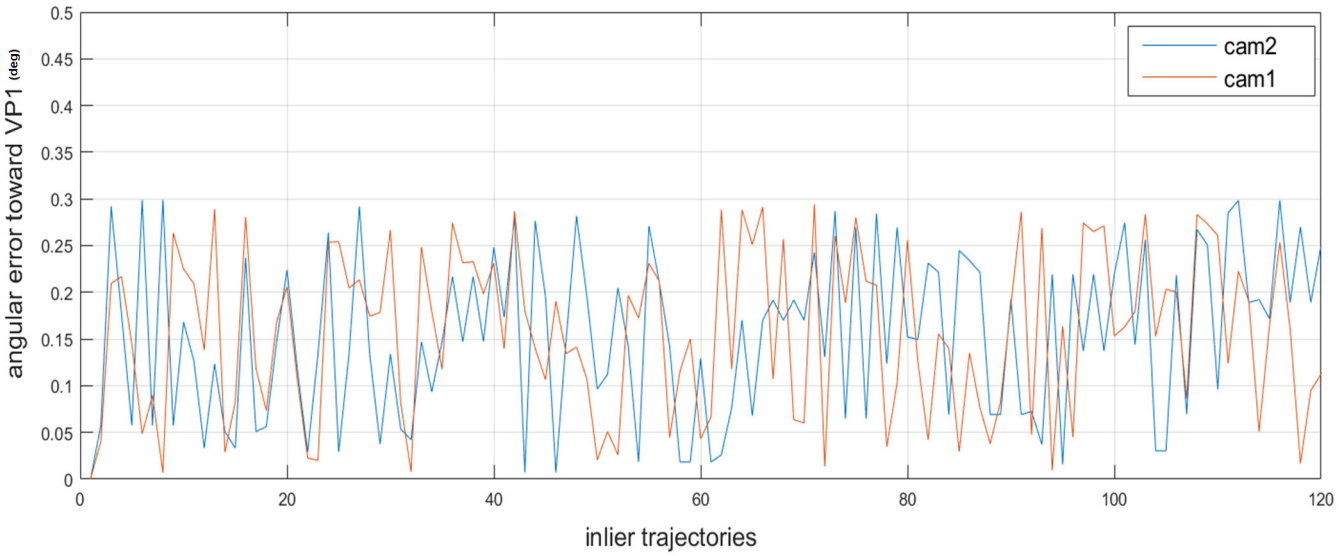



**Fig. 4.4:** Trajectory inliers accuracy with respect to $VP_1$ [LA dataset Camera 1 and 2]: reprojection error for a subset of 120 inlier corner trajectories (out of nearly 1000 in each camera, for about 6 seconds of recording)

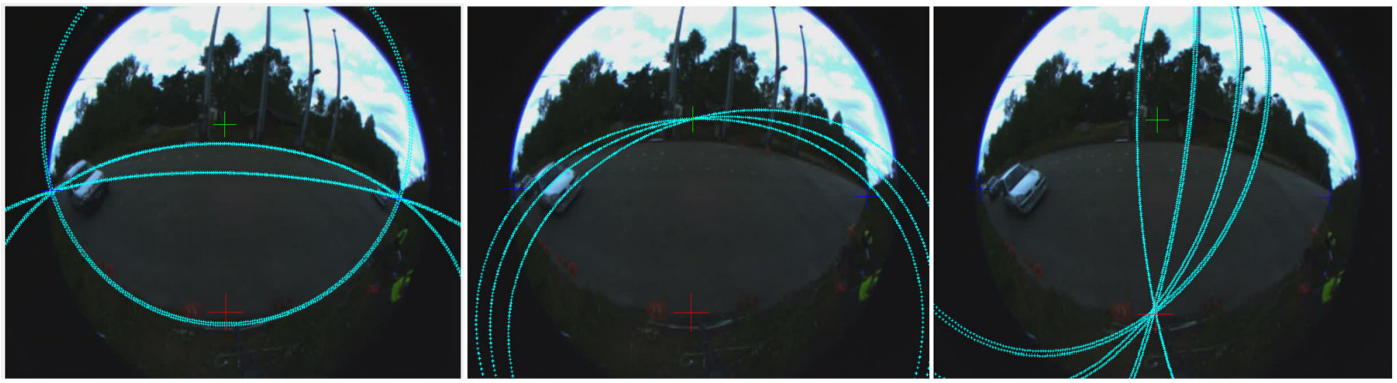


**Fig. 4.5:** Vanishing points evaluation [dataset LA-camera 1], by matching manually generated lines to estimated vanishing points (for $VP_1$ road boundaries selected, for $VP_2$ vehicle edgelets extended to lines, for $VP_3$ lines constructed from vertical poles)

In order to evaluate the accuracy of vanishing points we introduce the Line-VP error ($e_{lv}$) (Equation 4.1). The latter simply measures the geodesic distance between a given vanishing point $(VP)$ and the projection of related parallel lines on the spherical camera. The key idea is to verify the orthogonality on the spherical camera between the normal (N) of any reprojected line with respect the related vanishing point $(VP)$. For a given vanishing point, we compute the mean $e_{lv}$-error from few ground truth lines (Figures 4.5,4.8,4.10,4.12). The experimental results are presented in Table 4.1 (Figure 4.15) and discussed hereby.

$$e_{lv} = \left|90 - \arccos(N \cdot VP)\right| \tag{4.1}$$

We evaluate the complete triplet of orthogonal vanishing points ($VP_1$, $VP_2$, $VP_3$) in the lab (both cameras for dataset LB, see Figure 4.7). Our approach proved to be very accurate, as we get $e_{lv}$-errors between **0**° and **2**° in the worst case. The errors are close to zero for dataset LA. For dataset LB, the mean $e_{lv^1}$-error for the first vanishing point is equal to **0.53**° (camera 1) and **0.95**° (camera 2). Sample of extracted trajectories and back-projected inliers, computed vanishing points are illustrated in Figures 4.3, 4.6, 4.9 and 4.11. Besides, for dataset LB, the mean $e_{lv^2}$-error for the second vanishing point is equal to **0.88**° (camera 1) and **1.95**° (camera 2). This small increase of the error can be associated to the imperfect flatness of the road surface over a long distance. Illustration of the voting space distribution for the second vanishing point is illustrated in Figures 4.13 (dataset LB) and 4.14 (dataset SC). For the third vanishing point $VP_3$, we also achieve a high accuracy, with negligible mean $e_{lv^3}$-error, both in the lab and at a the rural intersection dataset. The results obtained for these experiments have demonstrated the robustness and the accuracy of the complete orthogonal triplet of vanishing points. Now we will quantify the extrinsic rotation and translation errors.

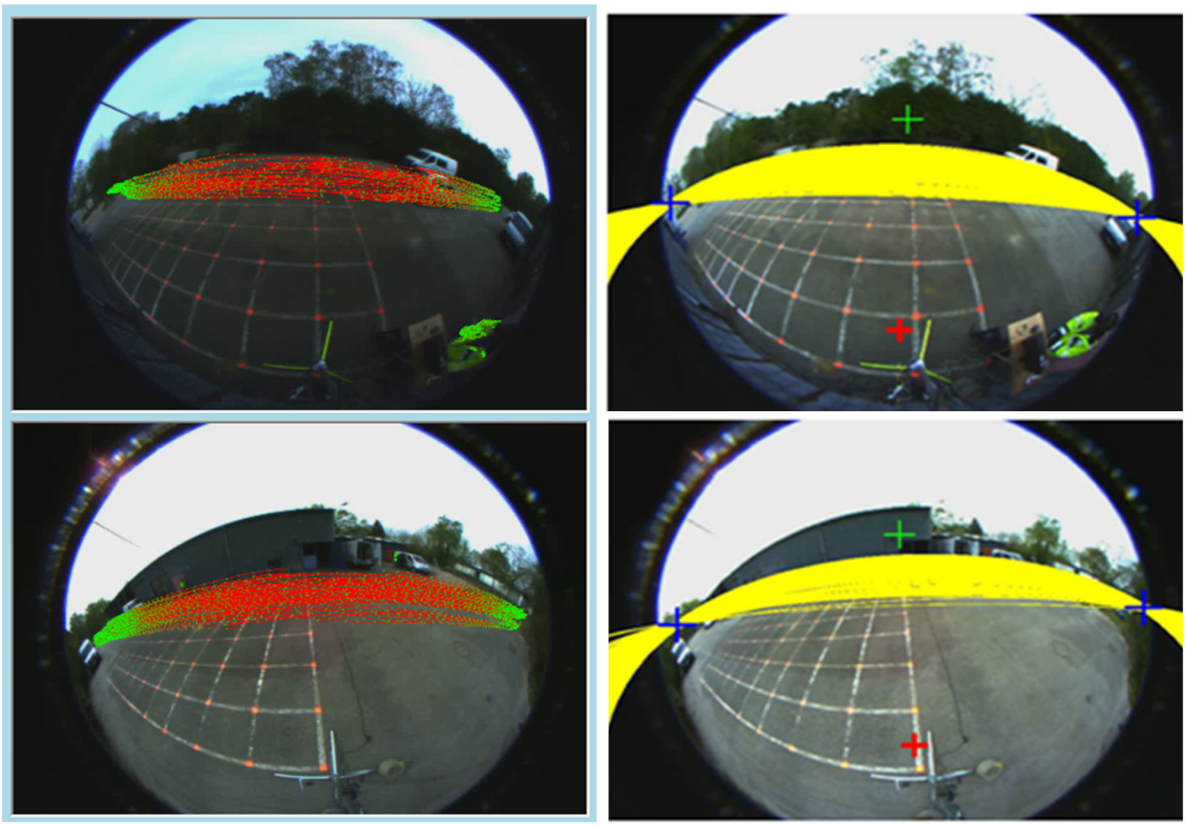


**Fig. 4.6:** Sample of extracted corner trajectories, back-projected inliers conics, and vanishing points [LB dataset]

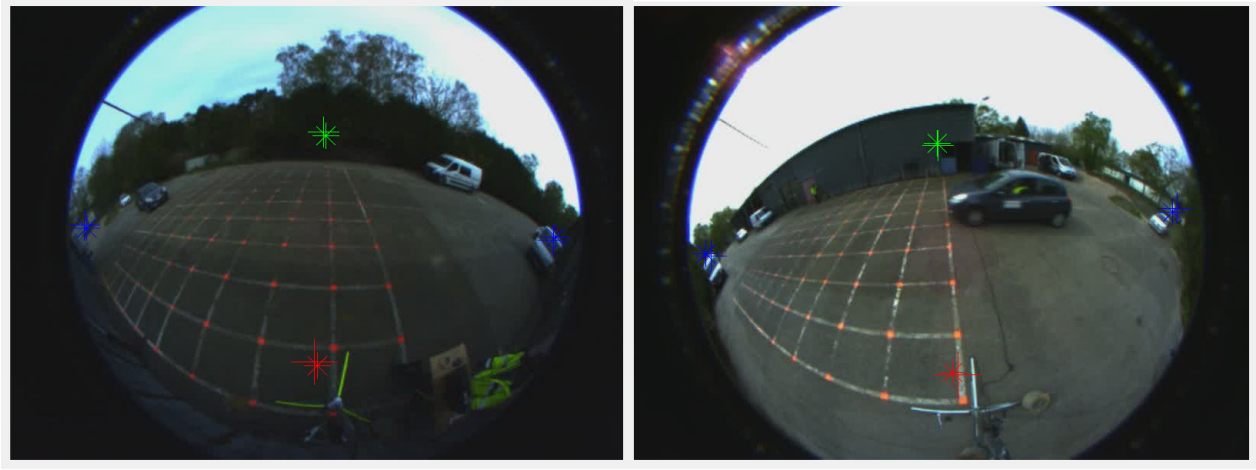
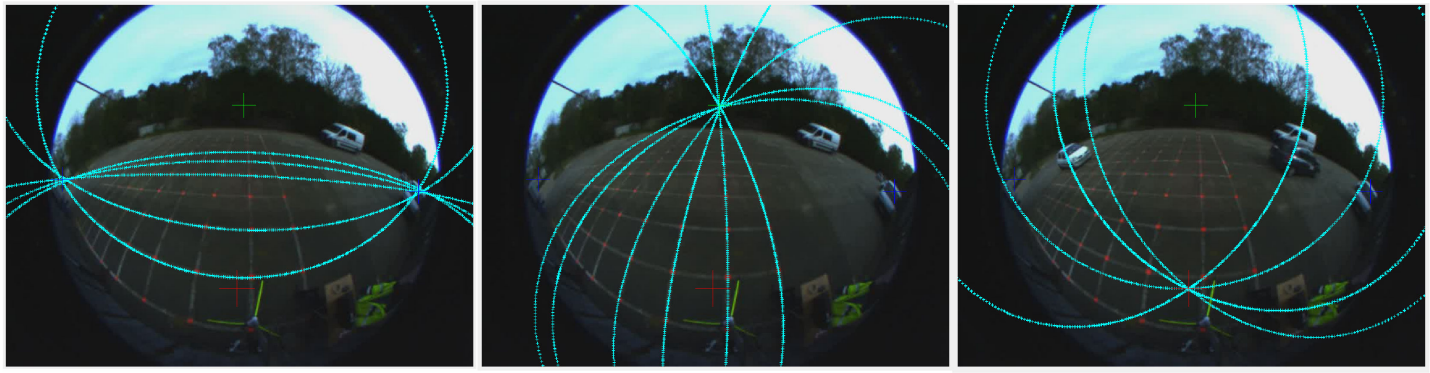
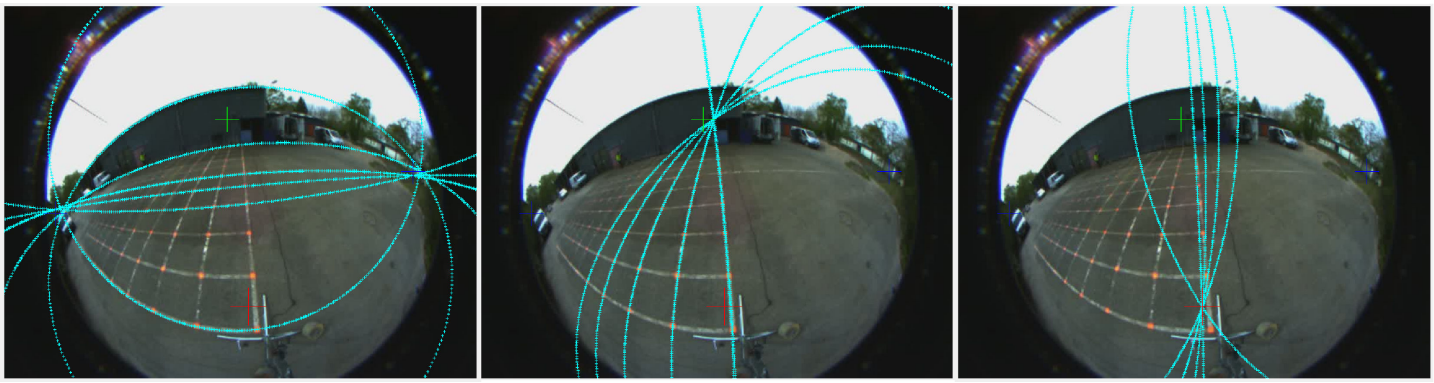


**Fig. 4.7:** Vanishing Points Comparison [dataset LB]: estimated (+) ground truth (*)



camera 1



camera 2

**Fig. 4.8:** Vanishing points evaluation [dataset LB-both cameras], by matching manually generated lines to estimated vanishing points (for $VP_1$ drawn lines on the road surface and the infrastructure are used, for $VP_2$ drawn lines on the road surface are used, for $VP_3$ lines are constructed from vertical structures in the image orthogonal to the road)

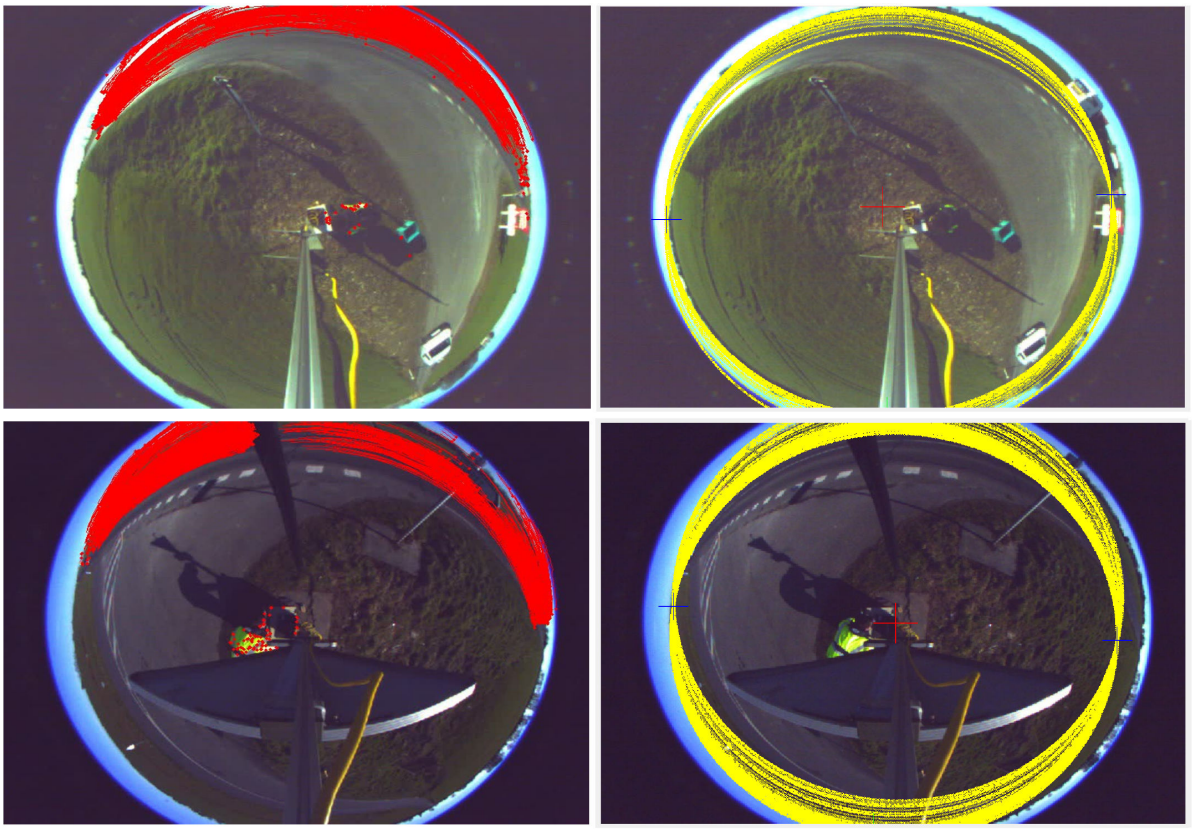


**Fig. 4.9:** Sample of extracted corner trajectories, back-projected inliers conics, and vanishing points [SA dataset]

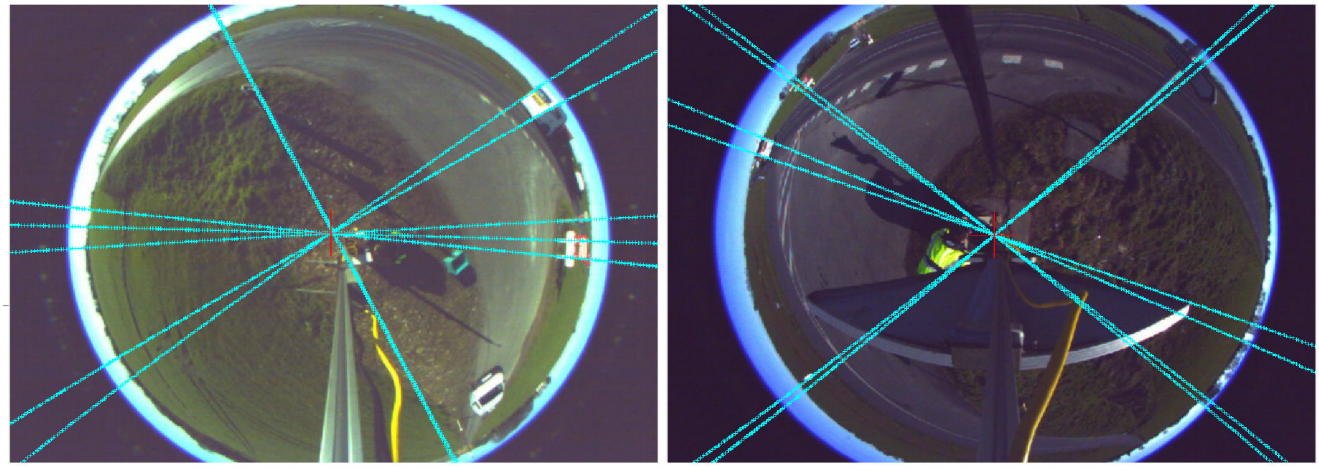


**Fig. 4.10:** Vanishing points evaluation [dataset SA-both cameras], by matching manually generated lines to the third vanishing point $VP_3$. Ground truth lines are constructed from vertical poles or structures in the image orthogonally to the road)

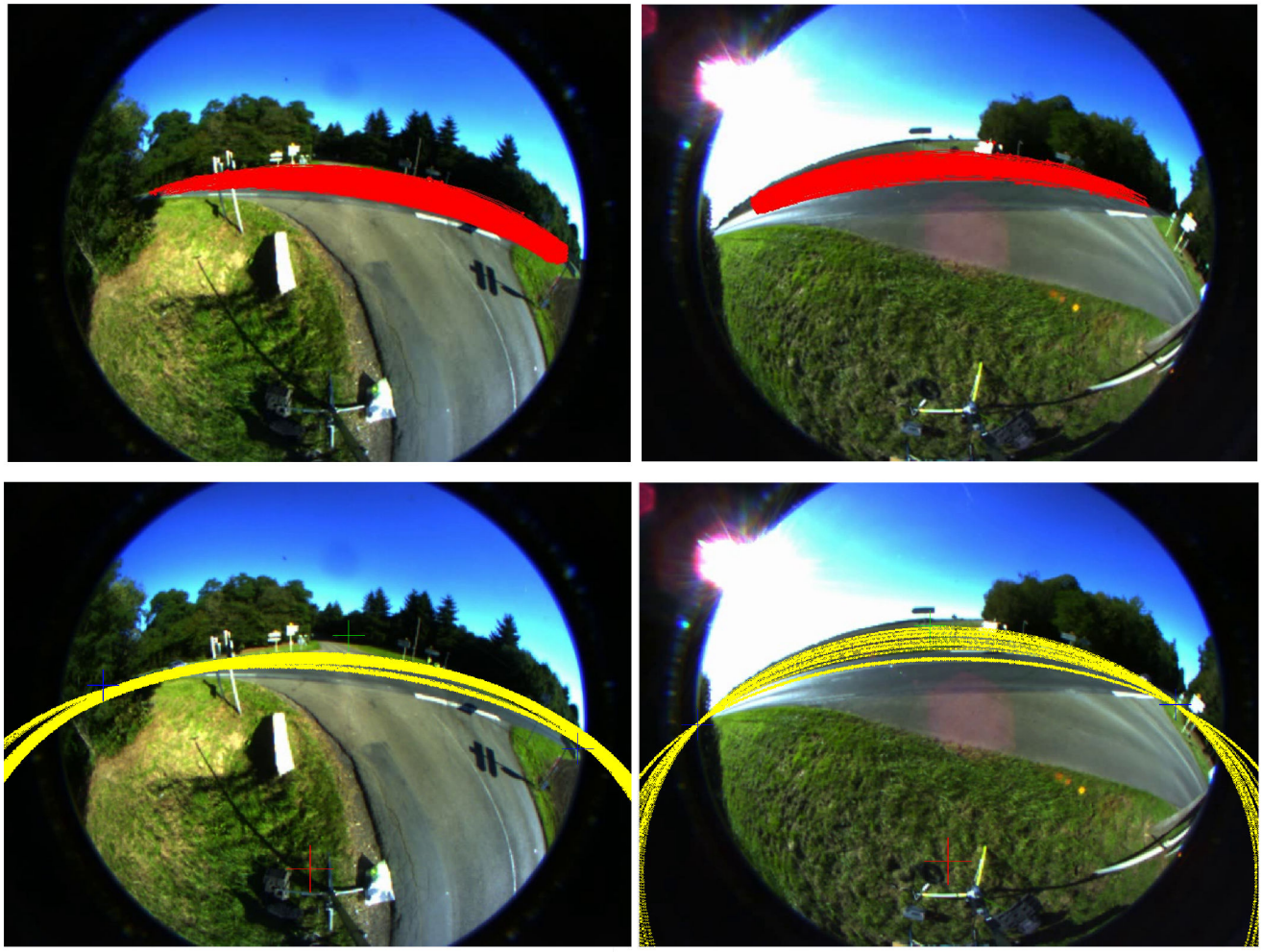


**Fig. 4.11:** Sample of extracted corner trajectories, back-projected inliers conics, and vanishing points [SC dataset]

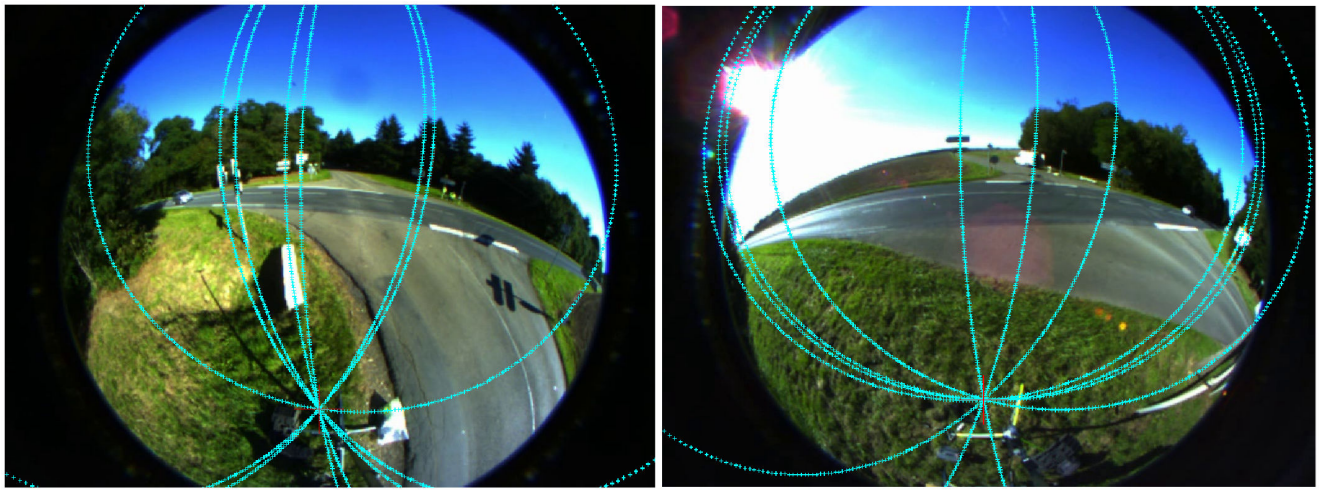


**Fig. 4.12:** Vanishing points evaluation [dataset SC-both cameras], by matching manually generated lines to the third vanishing point $VP_3$. Ground truth lines are constructed from vertical poles or structures in the image orthogonally to the road)

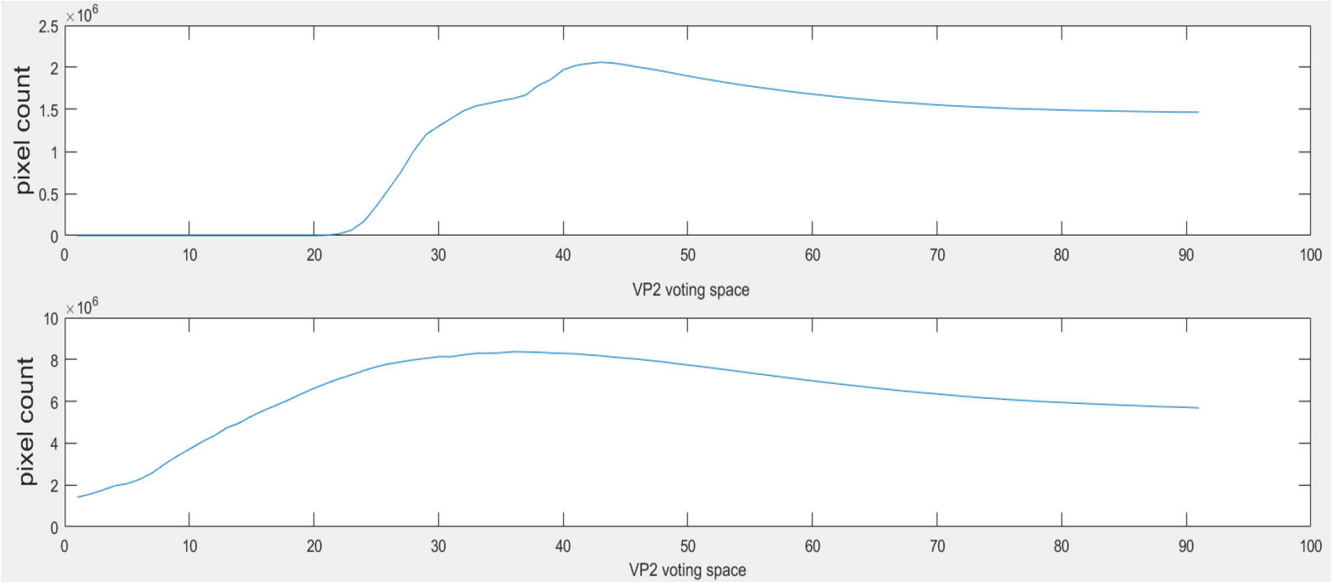



**Fig. 4.13:** Illustration of VP2 accumulator space (Algorithm 2) for both cameras (top-camera 1, bottom-camera 2, dataset LB)

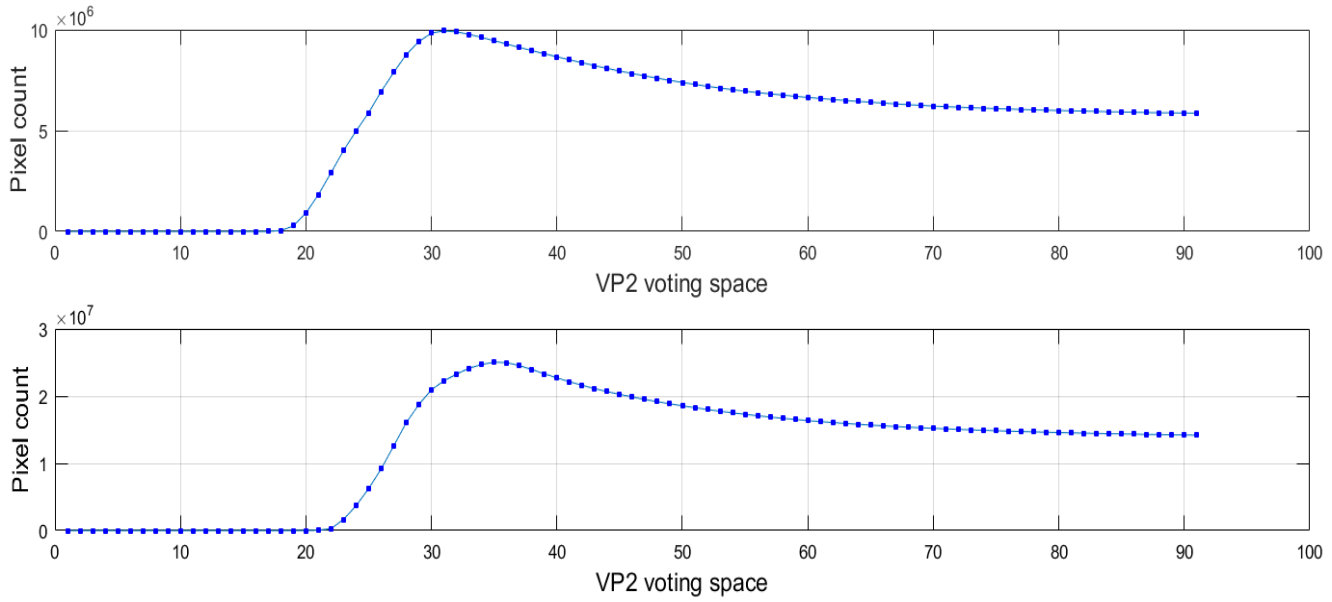



**Fig. 4.14:** Illustration of VP2 accumulator space (Algorithm 2) for both cameras (top-camera 1, bottom-camera 2, dataset SC)

| Dataset LA | |
|---|---|
| — Lines-$VP_1^{cam1}$ = **0.17**° | |
| — Lines-$VP_2^{cam1}$ = **0.08**° | |
| — Lines-$VP_3^{cam1}$ = **0.02**° | |
| **Dataset LB** | **Dataset LB** |
| — Lines-$VP_1^{cam1}$ = **0.53**° | — Lines-$VP_1^{cam2}$ = **0.95**° |
| — Lines-$VP_2^{cam1}$ = **0.88**° | — Lines-$VP_2^{cam2}$ = **1.95**° |
| — Lines-$VP_3^{cam1}$ = **0.06**° | — Lines-$VP_3^{cam2}$ = **0.04**° |
| **Dataset SA** | **Dataset SA** |
| — Lines-$VP_3^{cam1}$ = **0.01**° | — Lines-$VP_3^{cam2}$ = **0.03**° |
| **Dataset SC** | **Dataset SC** |
| — Lines-$VP_3^{cam1}$ = **0.06**° | — Lines-$VP_3^{cam2}$ = **0.05**° |

**Tab. 4.1:** Vanishing points accuracy: mean geodesic Lines-VP error for few datasets

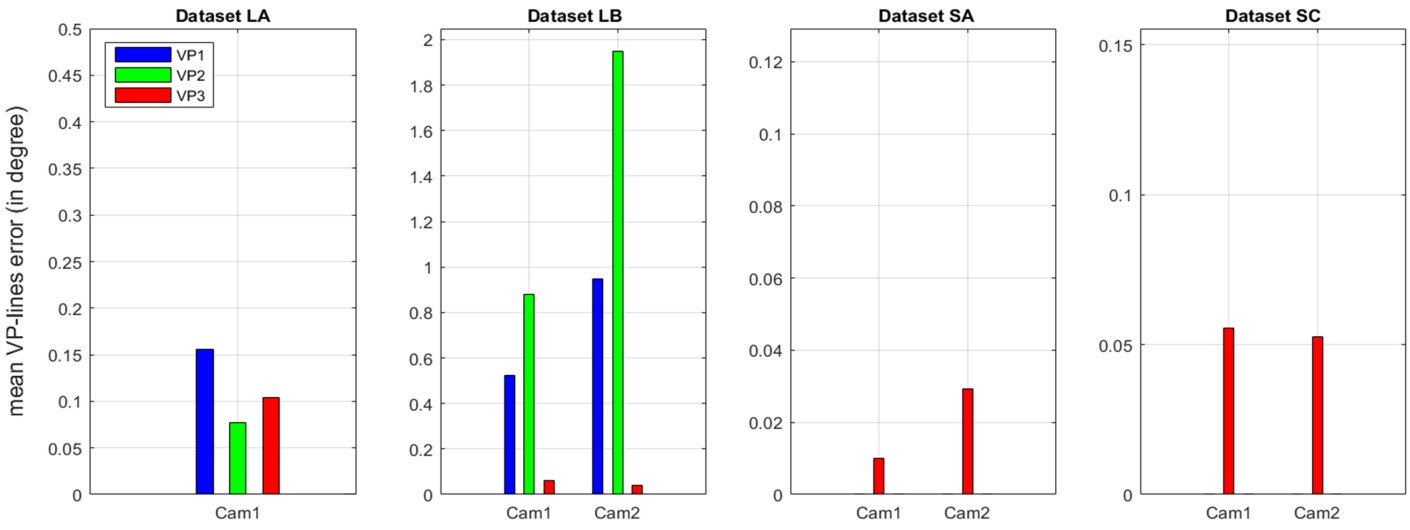
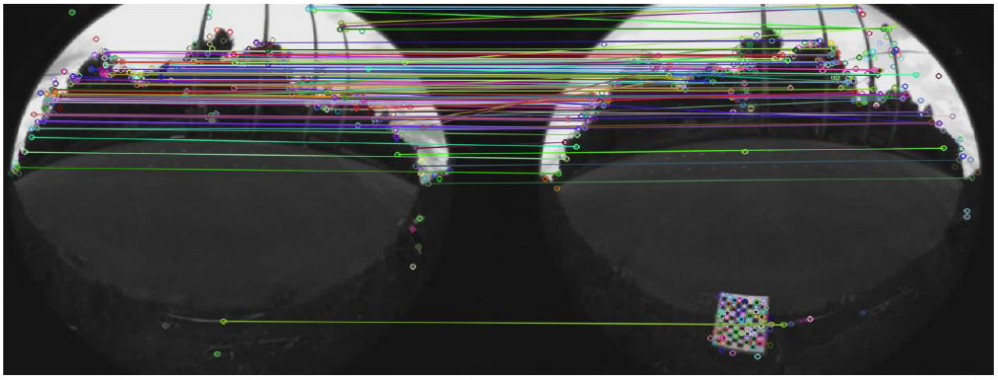


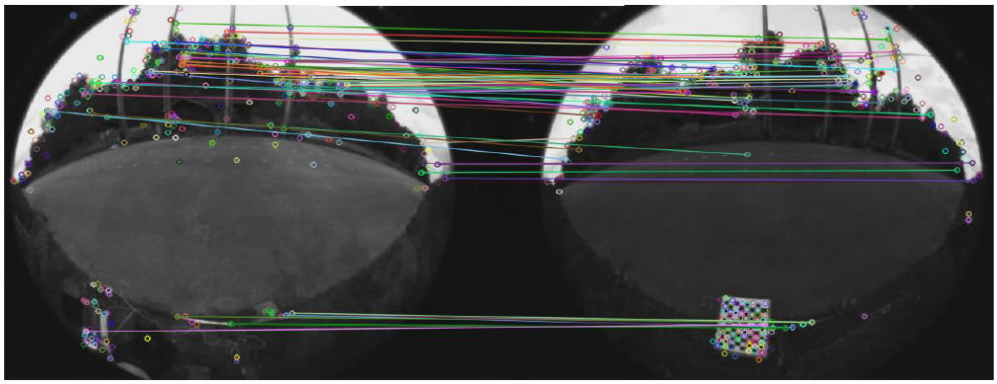

**Fig. 4.15:** Vanishing Points Accuracy

## 4.3 Extrinsic Calibration Evaluation

For classical stereo-vision (small baseline and orientation between the cameras), the estimation of the relative pose is feasible by computing the fundamental or essential matrix from points features correspondences. Here, because of the strong view difference and the vegetation, it is quite difficult to detect and match efficiently points features.



a) 2m



a) 8m

**Fig. 4.16:** SIFT features matching failure (cameras on the same side)

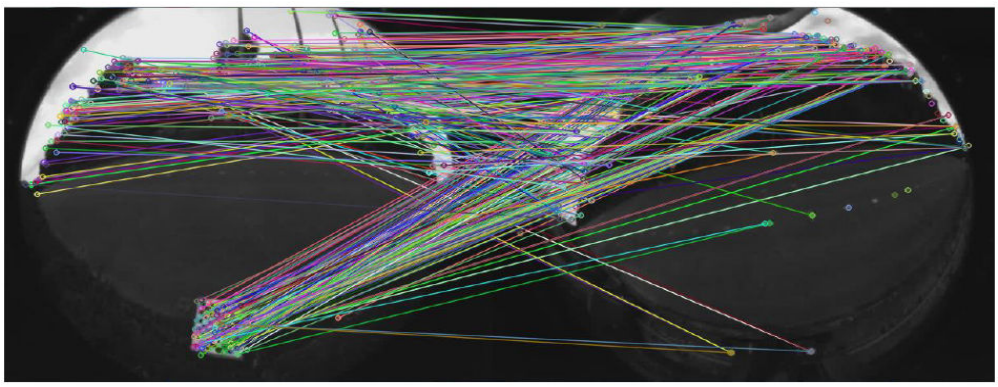


**Fig. 4.17:** Fisheye-stereo SIFT features matching failure

Figure 4.16 illustrates an example of SIFT features matchings with two fisheye cameras placed on the same side of a road: first with a baseline of $2m$ and then $8m$. With the shorter baseline (Figure 4.16-a), there are numerous correct features matchings, but most are unstable and present in the foliage or vertical poles. Besides, this configuration scenario is not convenient for our wide-baseline fisheye stereo. For the second case with a baseline of $8m$ (Figure 4.16-b), the matching clearly fails and the number of false SIFT points correspondences increases. In both previous cases, we notice however stable matching at infinity. The difficulty of wide-baseline correspondence is even more obvious in a different experiment where both cameras are placed in front of each other with a baseline close to $8m$, but with a strong extrinsic rotation near to $180°$ (Figure 4.17). These are some of the issues which have initially motivated our point-correspondence-free extrinsic calibration approach. In this section we present extrinsic calibration results based on our approach, in which vehicles are used as dynamic calibration objects. The experiments have been carried out both in the lab and at a rural intersection.

**—Experiment in the lab (dataset LB)**

We first evaluate the calibration approach in the lab environment with a baseline equal to 13.45 meters [49]. We use dataset LB, recorded on a test drive field, describing the traffic on the main road of an intersection. A given car, of which ground truth required dimensions are known, is used for the calibration. In order to acquire ground truth extrinsic calibration, a $10\ m^2$ square grid is painted on the road surface (only possible in the lab). The ground truth poses of the cameras, the extrinsic rotation and extrinsic translation at scale are obtained by solving a plane induced homography on the equivalent sphere [17]. We compare the extrinsic parameters obtained by the fully-automatic approach to the semi-automatic [49]. To recall, the latter requires hand-selected wheel points tracks as well as control points on the road surface. In addition, in the semi-automatic approach, the translation $\mathbf{T}_X$ along X-axis is assumed known at installation. Whereas in the fully-automatic approach the only prior knowledge to compute the full translation at scale is the dimension of a calibration vehicle as described in the previous chapter. The experimental results obtained are presented below in the Table 4.2, and described hereby:

• We observe that the rotation estimation is slightly improved (see Table 4.2) in the fully automatic approach compared to the semi-automatic method [49]. We obtain an absolute extrinsic orientation error lower than **2.4**$°$ in the worst case (the rotation error with respect to the X-axis, which is related to the imperfect flatness of the road surface).

| Extrinsic calibration | Ground truth | Estimation (automatic method) | Error (automatic) | Error (semi-automatic [49]) |
|---|---|---|---|---|
| **R** | $\begin{bmatrix} -1 & -0.05 & -0.06 \\ 0.06 & -1 & -0.08 \\ -0.06 & -0.08 & 1 \end{bmatrix}$ | $\begin{bmatrix} -1 & -0.09 & -0.07 \\ 0.09 & -1 & -0.03 \\ -0.07 & -0.04 & 1 \end{bmatrix}$ | $X_1 - axis$ : 2.3455° $Y_1 - axis$ : 0.3344° $Z_1 - axis$ : 2.2351° | $X_1 - axis$ : 2.8937° $Y_1 - axis$ : 0.0566° $Z_1 - axis$ : 2.9184° |
| $\mathbf{h}_1$ (m) | 2.4 ± 0.1 | 2.33 | 0.07 | 0.07 |
| $\mathbf{h}_2$ (m) | 2.3 ± 0.1 | 2.21 | 0.09 | 0.12 |
| $\mathbf{T}_X$ (m) | 10 ± 0.1 | 10.23 | 0.23 | — |
| $\mathbf{T}_Y$ (m) | 9 ± 0.1 | 9.13 | 0.13 | 0.11 |
| $\mathbf{T}_Z$ (m) | 0.1 ± 0.1 | 0.12 | 0.02 | 0.05 |

**Tab. 4.2:** Evaluation of the extrinsic calibration (lab experiment): comparison between the semi-automatic [49] and the full automatic methods in similar setups. The performance is similar, but with several improvements: the rotation error is reduced, besides all the components of the translation are computed automatically (in [49], $\mathbf{T}_X$ was assumed known up to an uncertainty)

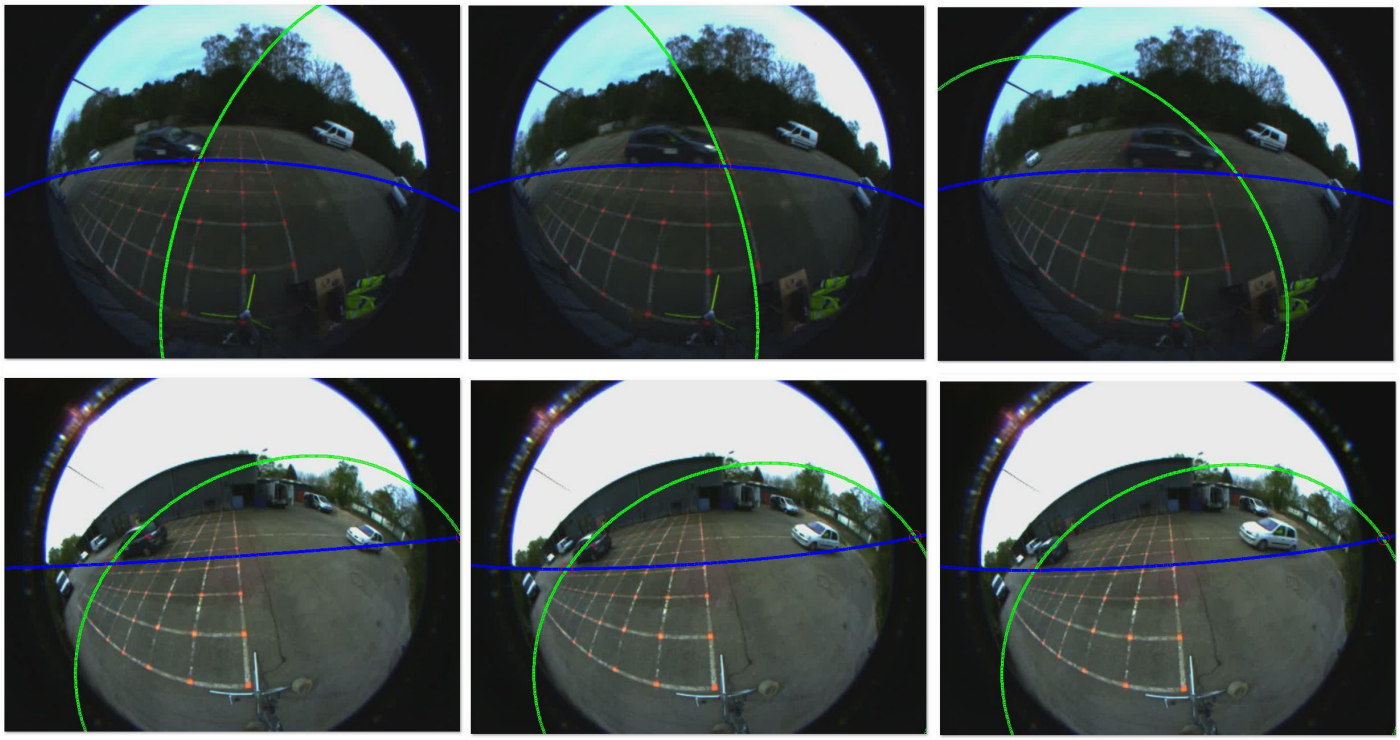


**Fig. 4.18:** Illustration of the iterative front-bumper virtual plane detections by virtual lines (lab dataset)

• The translation computed with the automatic approach is close to the result of the semi-automatic approach (see Table 4.2). It confirms the precision of the iterative virtual lines-planes detection (Figure 4.18),

despite eventual errors related to vehicle blob segmentation. The overall baseline error ($e_T$) is about **0.26 cm**, and expressed in percentage as follows:

$$e_T = \frac{\left| \sqrt[2]{\mathbf{T}_X^2 + T_Y^2 + T_Z^2} - ||\mathbf{T}_{truth}|| \right|}{||\mathbf{T}_{truth}||} = \mathbf{1.93\ \%} \tag{4.2}$$

The extrinsic calibration results demonstrate the good performance of our approach (see Table 4.2). The results obtained are quite promising regarding the very wide baseline between the cameras, and given that a single vehicle with rectilinear and planar motion can efficiently be used as a calibration object for traffic monitoring applications. Provided the extrinsic calibration, a good 3D localization of vehicles can be achieved (Figure 4.19). Based on these results, we have moved toward the evaluation of the extrinsic auto-calibration verification in real traffic scenes at a rural intersection.

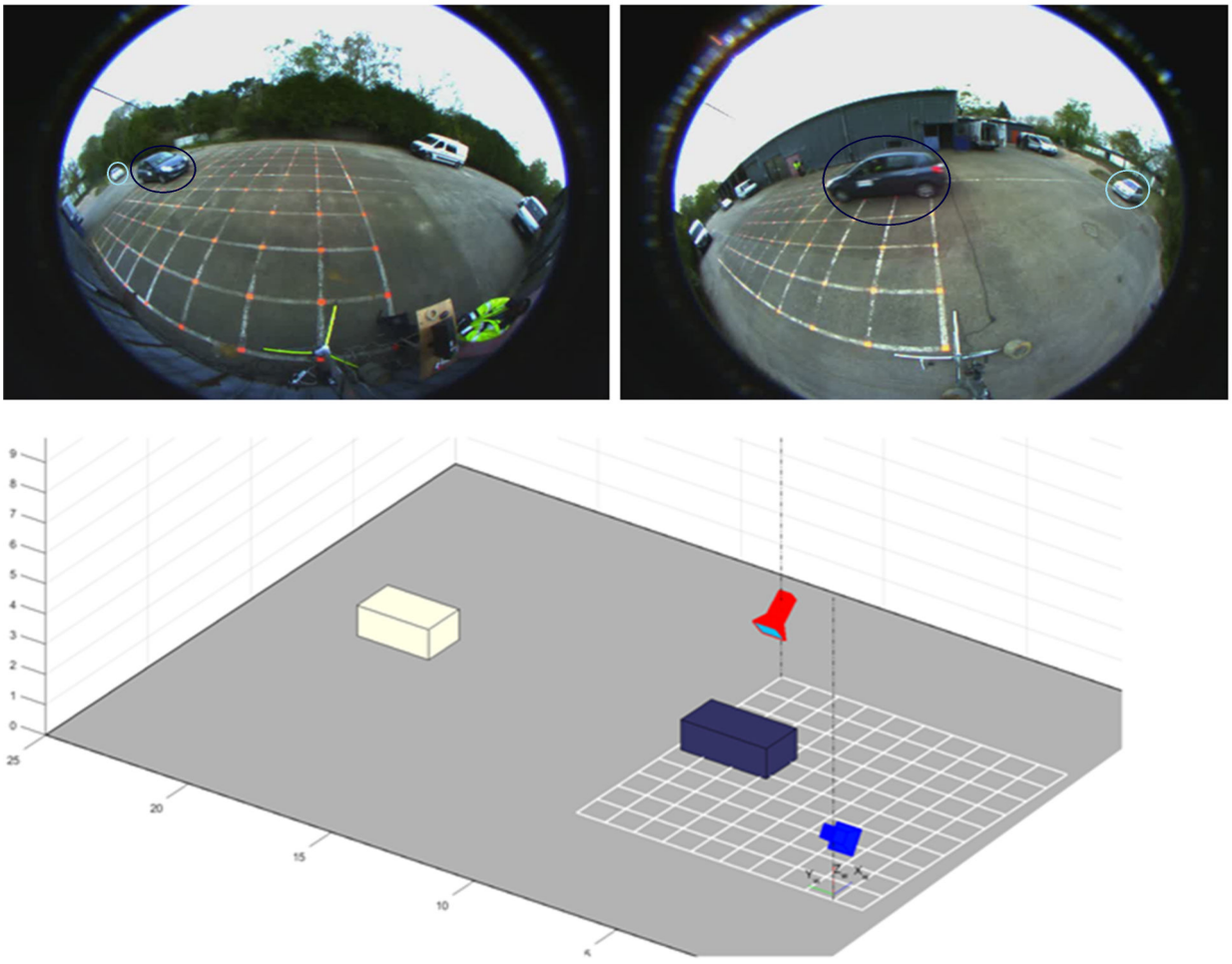


**Fig. 4.19:** 3D Localization and Reconstruction of vehicle modeled as 3D rigid bodies on the road plane (highlighted in blue and white)

**—Experiments on a real intersection (dataset SB)**

We run the automatic pipeline in order to estimate the extrinsic calibration of the large baseline fisheye stereoscopic vision system installed at a rural intersection in Normandy (Figure 4.20, top view of the system). We could not however obtain accurate ground truth of the extrinsic rotation and extrinsic translation, and therefore we proposed an indirect approach. The online evaluation method for the extrinsic translation and camera heights is not straightforward, rather indirect, but it gives good insights of the accuracy of the complete calibration. After running the extrinsic calibration, we compute the wheel-base (**v**) and axle-track (**u**) dimensions for several vehicles traveling on the intersection. The estimated vehicles dimensions are compared with ground truth data obtained from manufacturers specifications, and we analyze the error. We evaluate the orthogonality constraint of the four wheels envelop (defined by the wheel-base and the axle-track) projected back on the road surface. In other words, our evaluation quantifies the reprojection error of few vehicles wheels on the road surface.

The complete calibration results obtained are presented in Table 4.3. For the rural intersection dataset, we have added additional constraints for the virtual planes detection (both front and back). In order to deal with the presence of noise and shadows, we robustify the planes localization by finding the vertical bounding omnidirectional line directed toward $VP_3$, tangent to front and back vehicle bumpers. The results are illustrated in figure 4.38. For each vehicle used in the evaluation we hand-select the wheels; and given the intrinsic and extrinsic calibration obtained, we estimate the 3D-coordinates of the wheels on the road surface, and in the world reference. Now, with the 3D-positions of the wheels expressed

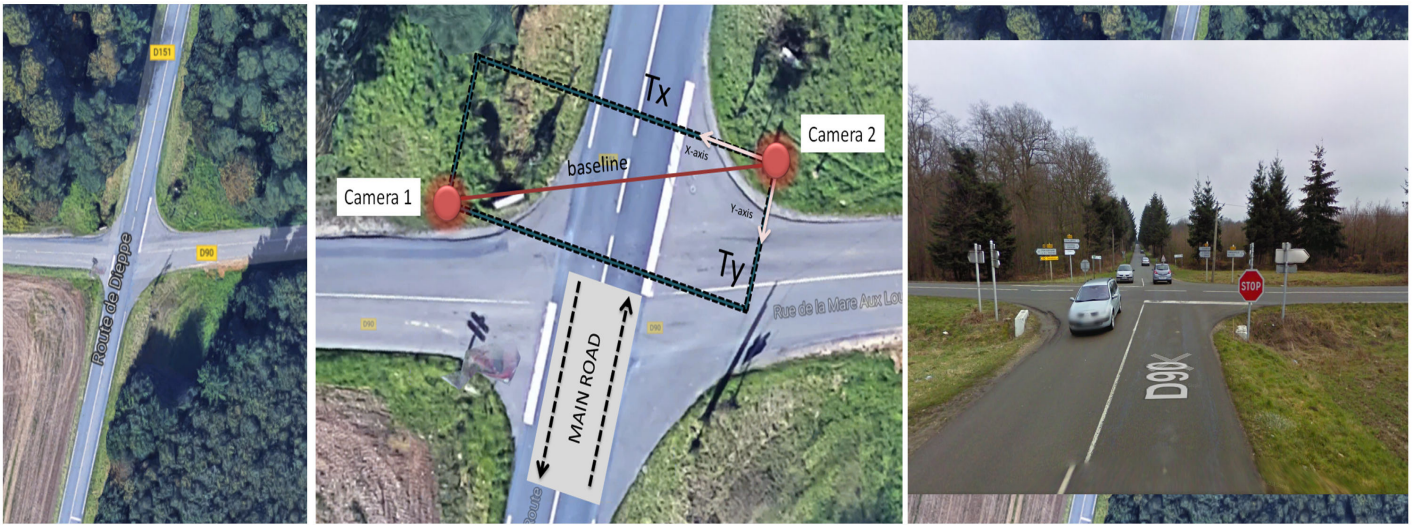



**Fig. 4.20:** Wide-baseline fisheye-stereo setup in Dataset SB (rural intersection, Normandy (D151:D90), France

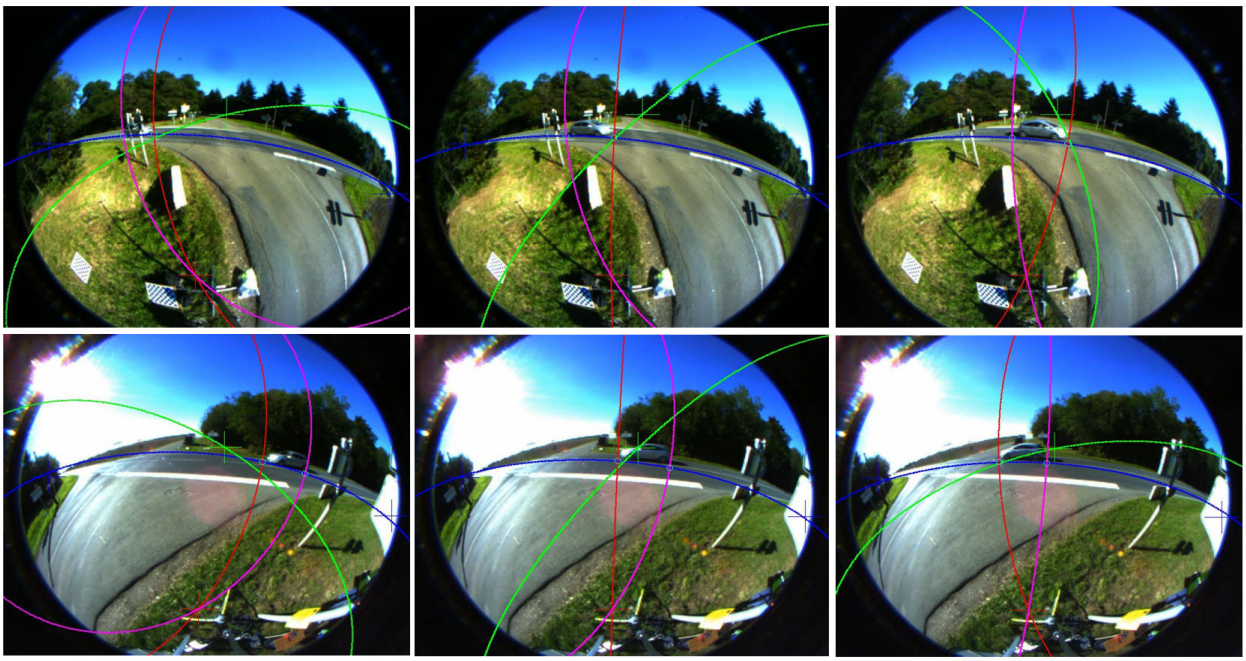


Fig. 4.21: Illustration of the virtual plane detection by virtual lines in few frames at a rural intersection dataset

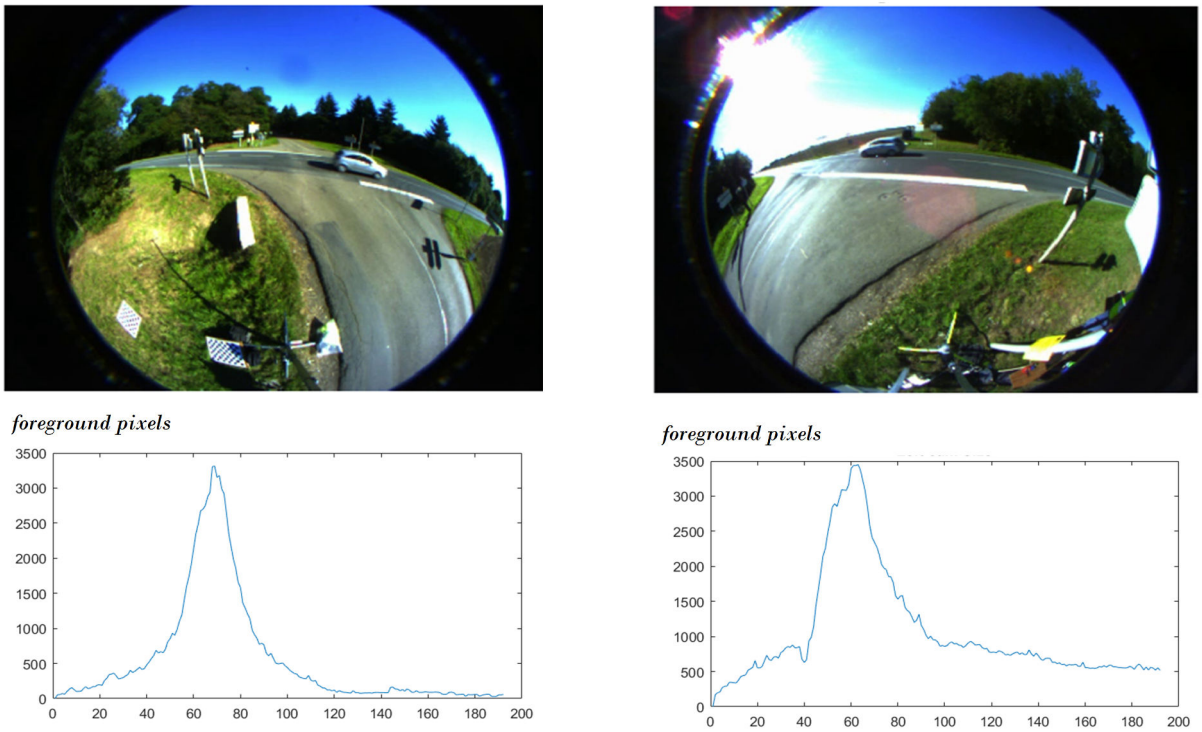



Fig. 4.22: Example of distribution of a single vehicle size traveling on the main road of the intersection

in world coordinates, we can estimate the wheel-base ($\mathbf{v}$) and axle-track ($\mathbf{u}$). The orthogonality between the wheel-base and the axle-track is also verified. We repeat this operation in several frames for each vehicle in the

| Extrinsic Rotation | | |
|---|---|---|
| R1 | R2 | R |
| $\begin{bmatrix} 0.475 & -0.199 & -0.857 \\ -0.15 & -0.978 & -0.144 \\ -0.867 & -0.06 & -0.495 \end{bmatrix}$ | $\begin{bmatrix} -0.436 & 0.180 & 0.882 \\ 0.117 & 0.983 & 0.143 \\ -0.892 & 0.041 & 0.449 \end{bmatrix}$ | $\begin{bmatrix} -0.998 & -0.026 & -0.051 \\ 0.026 & -1 & -0.009 \\ -0.051 & -0.008 & 0.999 \end{bmatrix}$ |

| Extrinsic Translation | | | Camera Heights | |
|---|---|---|---|---|
| TX | TY | TZ | h1 | h2 |
| 20.468 | 10.638 | 0.092 | 1.890 | 1.798 |

**Tab. 4.3:** Extrinsic calibration parameters estimated for dataset SB (rotation, translation, height)

evaluation, in order to take into account the scale variation throughout the intersection (Figure 4.22). We consider a short sequence of the video recorded by the cameras for the evaluation at a rural intersection with very low traffic during the experiments. The evaluation dataset is restrained to 20 vehicles, which are clearly identifiable in the images, from variables types and different manufacturers, and with precise ground truth dimensions. For each vehicle we compute the proposed metrics. The results obtained are hereby presented and analyzed:

• The results for the wheel-base (**v**) are detailed in Table 4.4. It allows to validate distance measurements toward the first vanishing point ($VP_1$). The absolute error for this dimension ranges from **0 cm** [0 %] to **33 cm** [13.4%] (Figure 4.23,4.24). Besides, the mean and median errors for (**v**) are respectively **9 cm** and **7 cm**, and the 95th percentile is equal to **15 cm** (Table 4.6) (Figure 4.25).

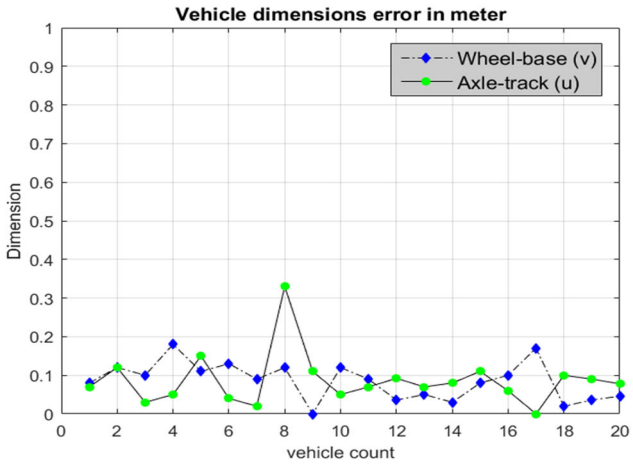



**Fig. 4.23:** Vehicle dimensions error in meter

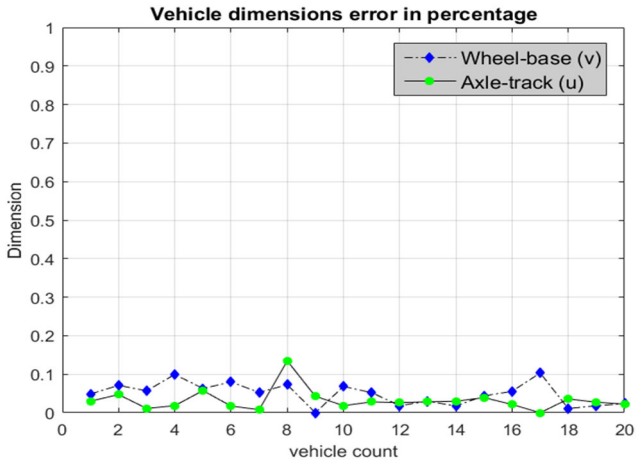



**Fig. 4.24:** Vehicle dimensions error in percentage

| Vehicles wheel-base (v) | | | |
|---|---|---|---|
| A Second Head | Estimated Distance | Error (m) | Error (%) |
| 2.34 | 2.27 | 0.07 | 2.99 |
| 2.54 | 2.66 | 0.12 | 4.72 |
| 2.70 | 2.67 | 0.03 | 1.11 |
| 2.72 | 2.77 | 0.05 | 1.84 |
| 2.61 | 2.76 | 0.15 | 5.75 |
| 2.30 | 2.26 | 0.04 | 1.74 |
| 2.51 | 2.53 | 0.02 | 0.80 |
| 2.45 | 2.78 | 0.33 | 13.47 |
| 2.58 | 2.47 | 0.11 | 4.26 |
| 2.71 | 2.76 | 0.05 | 1.85 |
| 2.49 | 2.56 | 0.07 | 2.81 |
| 3.50 | 3.59 | 0.09 | 2.63 |
| 2.47 | 2.54 | 0.07 | 2.83 |
| 2.69 | 2.61 | 0.08 | 2.97 |
| 2.80 | 2.69 | 0.11 | 3.93 |
| 2.80 | 2.86 | 0.06 | 2.14 |
| 2.45 | 2.45 | 0.00 | 0.00 |
| 2.80 | 2.70 | 0.10 | 3.57 |
| 3.30 | 3.21 | 0.09 | 2.73 |
| 3.50 | 3.42 | 0.08 | 2.23 |

**Tab. 4.4:** Wheel-base error (v)

| Vehicles axle-track (u) | | | |
|---|---|---|---|
| A Second Head | Estimated Distance | Error (m) | Error (%) |
| 1.64 | 1.72 | 0.08 | 4.88 |
| 1.69 | 1.81 | 0.12 | 7.10 |
| 1.76 | 1.66 | 0.10 | 5.68 |
| 1.81 | 1.63 | 0.18 | 9.94 |
| 1.75 | 1.64 | 0.11 | 6.29 |
| 1.63 | 1.76 | 0.13 | 7.98 |
| 1.71 | 1.80 | 0.09 | 5.26 |
| 1.65 | 1.53 | 0.12 | 7.27 |
| 1.70 | 1.70 | 0.00 | 0.00 |
| 1.76 | 1.88 | 0.12 | 6.82 |
| 1.72 | 1.81 | 0.09 | 5.23 |
| 1.96 | 1.92 | 0.04 | 1.84 |
| 1.67 | 1.62 | 0.05 | 2.99 |
| 1.71 | 1.68 | 0.03 | 1.75 |
| 1.83 | 1.91 | 0.08 | 4.37 |
| 1.80 | 1.90 | 0.10 | 5.56 |
| 1.65 | 1.48 | 0.17 | 10.30 |
| 1.83 | 1.81 | 0.02 | 1.09 |
| 1.99 | 1.95 | 0.04 | 1.81 |
| 1.96 | 1.91 | 0.05 | 2.35 |

**Tab. 4.5:** Axle-track error (u)

| Computed vehicle geometry error (m) | | | |
|---|---|---|---|
| Dimension | Mean | Median | 95th percentile |
| **Axle-track (u)** | **0.08** | **0.09** | **0.17** |
| **Wheel-base (v)** | **0.09** | **0.07** | **0.15** |
| **Average** | **0.085** | **0.08** | **0.16** |

**Tab. 4.6:** Average absolute vehicle dimension

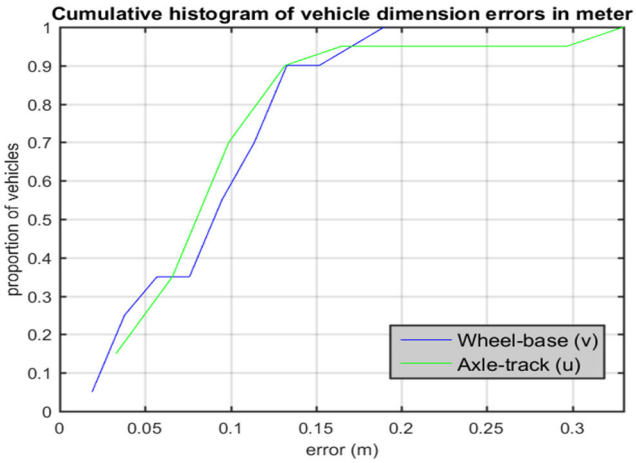



**Fig. 4.25:** Cumulative error (meter)

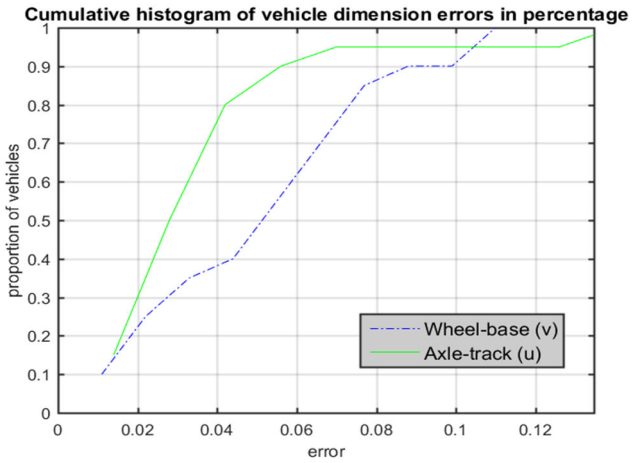



**Fig. 4.26:** Cumulative error (percentage)

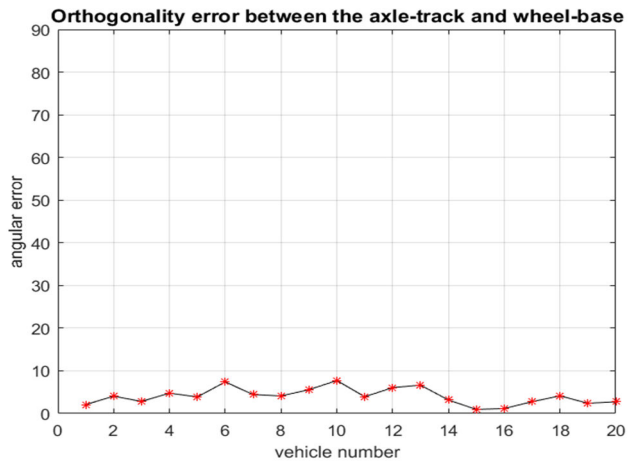



**Fig. 4.27:** Orthogonality error between the axle-track and wheel-base

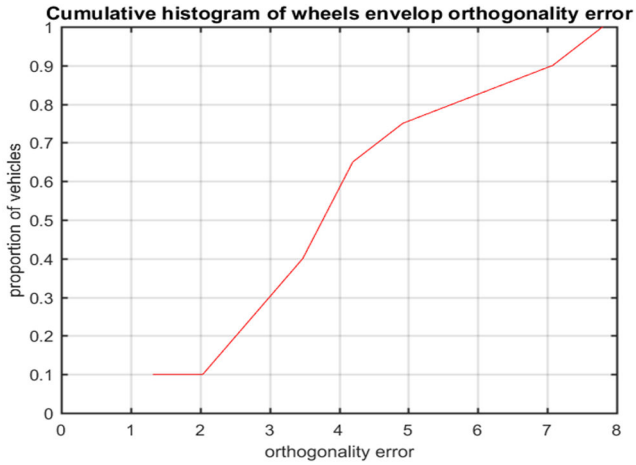



**Fig. 4.28:** Cumulative histogram of wheels envelop orthogonality error

• The results for the axle-track (**u**) are detailed in Table 4.5. It allows to validate distance measurements toward the second vanishing point ($VP_2$). The absolute error for this dimension ranges from **0 cm** [0 %] to **18 cm** [10.3%] (Figure 4.23,4.24). Besides, the mean and median errors for (**v**) are respectively **8 cm** and **9 cm**, and the 95th percentile is equal to **17 cm** (Table 4.6) (Figure 4.25).

• In average, with our extrinsic calibration at scale (Table 4.6) we obtain errors below **16 cm**. The cumulative histograms of absolute dimensions errors also validates the accuracy of our approach (Figure 4.25). In terms of percentage, we observe that distance measurement toward the first vanishing point corresponding to the traffic stream is more accurate than distance measurement toward the second vanishing point orthogonally to the traffic stream (Figure 4.26).

• Along with the dimensions evaluation, the orthogonality between the axle track and wheel-base directions is also evaluated. The results shows that the angular error of the wheels envelop projected onto the road surface ranges from **0.96**° to **7.73**°, with a mean value of **4.05**° (Figure 4.27,4.28).

• Finally, the calibrated network of camera is used to measure the lengths of stop marking lines on the minor road (Fig. 4.22) with known ground truth. In both cases, the absolute error of the stop line error is negligible, in average **2.5 cm** (Table 4.7).

| Ground truth (stop marking line length, see fig. 4.22) | Estimation | Absolute error |
|---|---|---|
| $m_1$ = 6.69 m | 6.67 m | 0.02 m |
| $m_2$ = 11.89 m | 11.86 m | 0.03 m |

**Tab. 4.7:** Evaluation of the ground plane distance estimation based on the extrinsic calibration (rural intersection dataset)

The experiments carried out validate our complete auto-extrinsic calibration method. The system also allows a reliable estimation of vehicle position on the main road of the intersection. In the next section we evaluate distance measurement on the road surface.

## 4.4 3D-Trajectory Estimation

Vehicle trajectory estimation is a complex task with the proposed wide-baseline fisheye-stereo. In fact, the strong change in appearance of vehicles in the fisheye image as they move throughout the intersection, along with the view difference between the cameras, makes it very difficult to achieve a dense and complete reconstruction of vehicles trajectories. Besides, it remains complex to keep track of vehicles in a busy intersection and when interactions between vehicles on severals roads are important. For these reasons, we propose a trajectory reconstruction approach which takes advantage of the virtual plane formulation proposed to recover the extrinsic calibration. In our approach we defined the position of the vehicle by the virtual plane. Therefore provided the extrinsic calibration estimated previously, the position of the vehicle in the 3D scene can be can be directly computed from 2D-vehicle information. Because of vehicle segmentation error, presence of shadows or vehicle appearance change, the identification of the front virtual plane may not be perfectly correct during the entire trajectory estimation. In result the 3D estimated position

of the vehicle will be noisy. In order to handle noise, we propose to use Bayesian filtering strategy. On the first hand we deal directly with noise in the image domain, by applying a blob tracker in order to ensure consistency of vehicle segmentation and re-identification. Then we deal with noise in the world domain, by applying a Kalman filter to refine the trajectory reconstruction of any vehicle traveling on the main road.

The Kalman filter offers an effective framework to manage measurement and model noises as well as estimation of state vectors. As the position of any vehicle is defined by the front virtual plane, we simply consider tracking references points $wf_i$ expressed in the world frame. When the vehicle is clearly visible only in the image acquired from either $\mathbf{C}_1$ or $\mathbf{C}_2$, its 3D-position is given in the world reference by either point $wf_1^k$ or $wf_2^k$. When the vehicle is clearly visible in both cameras, then its 3D-position is adjusted by the fisheye-stereo as the mean of the points $wf_1^k$ and $wf_2^k$. The dynamic trajectory can be modeled as a discrete time system ($d_t$=1), which consist of the position and the velocity parameters along the road plane axes such as:

$$\mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \epsilon_t \tag{4.3}$$

- $\epsilon_t$ is the process noise and indicates the uncertainty in the model. It is here assumed to be Gaussian and defined with a 4×4 square covariance matrix $\mathbf{D}_t$ with all values set to 0.2 m (high confidence in the model).
- $\mathbf{F}_t$ is the state transition model which applies the effect of each system state parameter at time $t-1$ on the system state at time t. We assume rectilinear planar motion at constant velocity (x, y, z=0), thus negligible acceleration for vehicles on the main road of the intersection, which lead to the following transition matrix:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.4}$$

At each iteration, a measurement is made and the new position of the front virtual plane is computed. Thus the lateral and longitudinal position are always observed to handle abrupt motion change and noise. The measurement model of the true vector state is given as follows:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \delta_t \tag{4.5}$$

- $\delta_t$ is the measurement noise, assumed to be Gaussian and defined with a 4×4 square covariance matrix $\mathbf{Q}_t$ with all values set to 2 m, which takes into account vehicle segmentation errors in the foreground.
- $\mathbf{H}_t$ is the transformation matrix that maps the state vector parameters into the measurement domain. Since we only measure position parameters at each iteration, the observation matrix $\mathbf{H}$ is defined as follows:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{4.6}$$

The Kalman filter process is divided into two steps which are the prediction and the update. The prediction step, uses a previously estimated state and the linear model to predict the value of the next state as well as the state estimate covariance:

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_{t-1|t-1}\mathbf{x}_t \tag{4.7}$$

$$\hat{\mathbf{P}}_{t|t-1} = \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^T + \mathbf{D}_t \tag{4.8}$$

- where $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{P}}_{t|t-1}$, are respectively the estimated state vector and a priori covariance matrix.

The update step of the Kalman filter uses the current measurement of the output together with the statistical properties of the model, to correct the state estimate. The values calculated is the innovation covariance, the Kalman gain resulting in the updated state estimate and state estimate covariance:

$$\hat{y}_t = \mathbf{z}_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1} \tag{4.9}$$

$$\mathbf{K}_t = \hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T(\mathbf{H}_t\hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T + \mathbf{Q}_t)^{-1} \tag{4.10}$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\hat{y}_t \tag{4.11}$$

$$\hat{\mathbf{P}}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\hat{\mathbf{P}}_{t|t-1} \tag{4.12}$$

where $\hat{\mathbf{x}}_{t|t}$ and $\hat{\mathbf{P}}_{t|t}$ are the estimated a posteriori state vector and covariance matrix respectively.

**—Experiment in the lab (dataset LC)**

A first set of experiments was performed in a controlled driving field environment (dataset LC), but with difficult processing conditions: very sunny, strong vegetation, leaves shadows. We aim to evaluate particularly vehicles lateral position with respect to the world frame ($lp_1$ and $lp_2$, see Figures 4.30 and 4.29). Ground truth is acquired with a LIDAR (Sick LMS511 laser scanner) at a frequency of 50Hz, and synchronized with the wide-baseline fisheye-stereo. The transformation between the LIDAR frame and the world frame is estimated at installation. Then, during the experiment we apply the extrinsic auto-calibration between the cameras. We consider the scenario of crossing-encountering vehicles (a black car driving on a lane close to the camera-1, and a white van driving on a lane near to the second camera) with dynamic occlusion. The results of average lateral position evaluation are summarized in Table 4.8 and illustrated in Figure 4.31. We achieve in average a lateral vehicle position estimation lower than **20 cm**, which confirms the robustness of our approach.

**—Experiments on a real intersection (dataset SB)**

In order to validate our approach in a real intersection, we propose to evaluate the velocity of vehicles [169], and thus the performance of the Kalman-based tracker. A preliminary study with a radar on over 1300 vehicles at the rural junction revealed that the speed of vehicles follows a normal distribution (Figure 4.33), with a mean of 82.29 km/h (with speed limited to 90 km/h). This gives us a good basis to initialize the tracker. For the experiments, a virtual gate is defined on the road surface. Then the travel time of vehicles through the gate entry and exit is retrieved to compute the actual average ground truth speed. We perform the speed evaluation on a subset of 10 vehicles, for which we compute the trajectory and motion parameters by our approach. A blob tracker can be applied together with our method to ensure correct vehicle identification and inter-frame data-association (Figure 4.37). Our method performs well (Figure 4.34) and achieves robust trajectory reconstruction: with the largest speed error below **5.53 km/h** and a mean error of **3.25 km/h**.

| Driving Experiments | Lateral position | Ground Truth LIDAR (m) | Estimation lp (m) | Absolute error (m) |
|---|---|---|---|---|
| Test 1 | car (black) | 3.88 | 4.06 | 0.18 |
| | van (white) | 6.81 | 6.96 | 0.15 |
| Test 2 | car (black) | 3.66 | 3.89 | 0.23 |
| | van (white) | 7.5 | 7.7 | 0.2 |

**Tab. 4.8:** Vehicle lateral position error

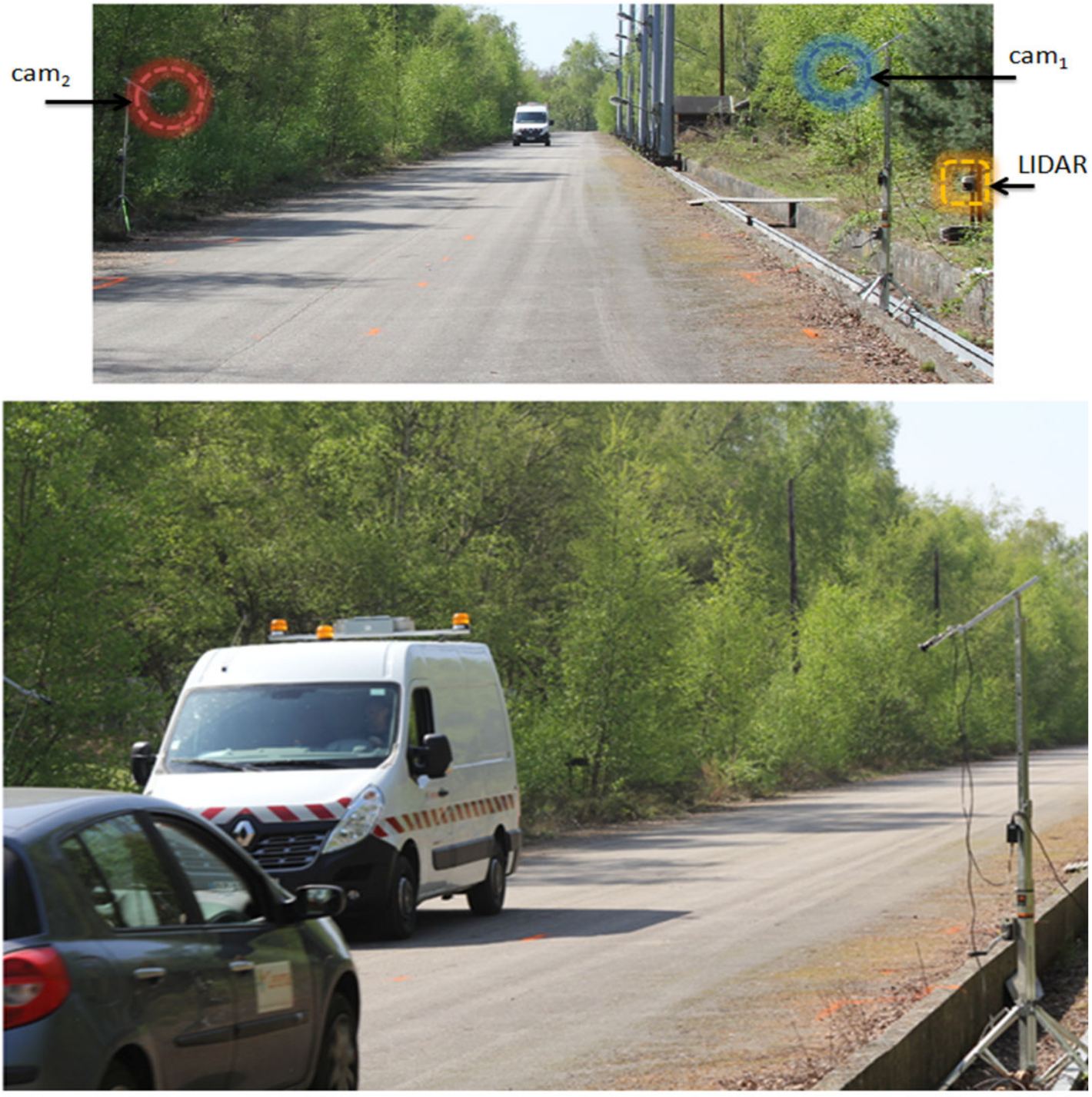



Fig. 4.29: View of the experimental setup

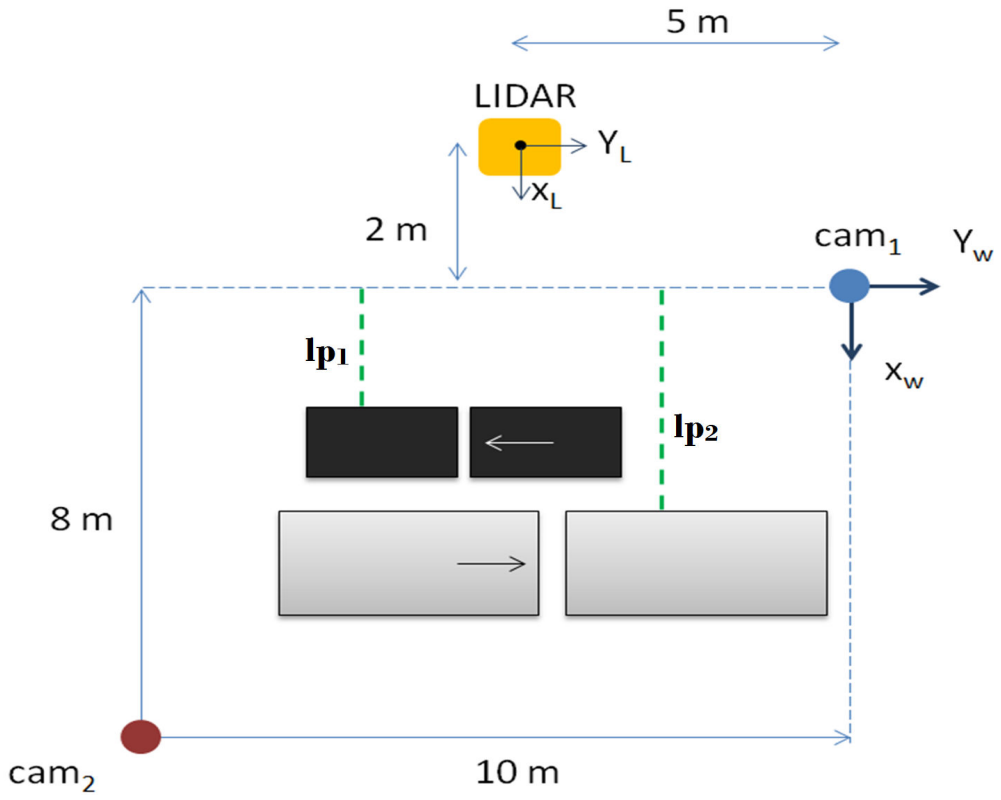



**Fig. 4.30:** Experimental setup for lateral position evaluation: two vehicles move straight in opposite directions. The world frame is defined vertically below camera-1 as described in the previous chapter 3.2

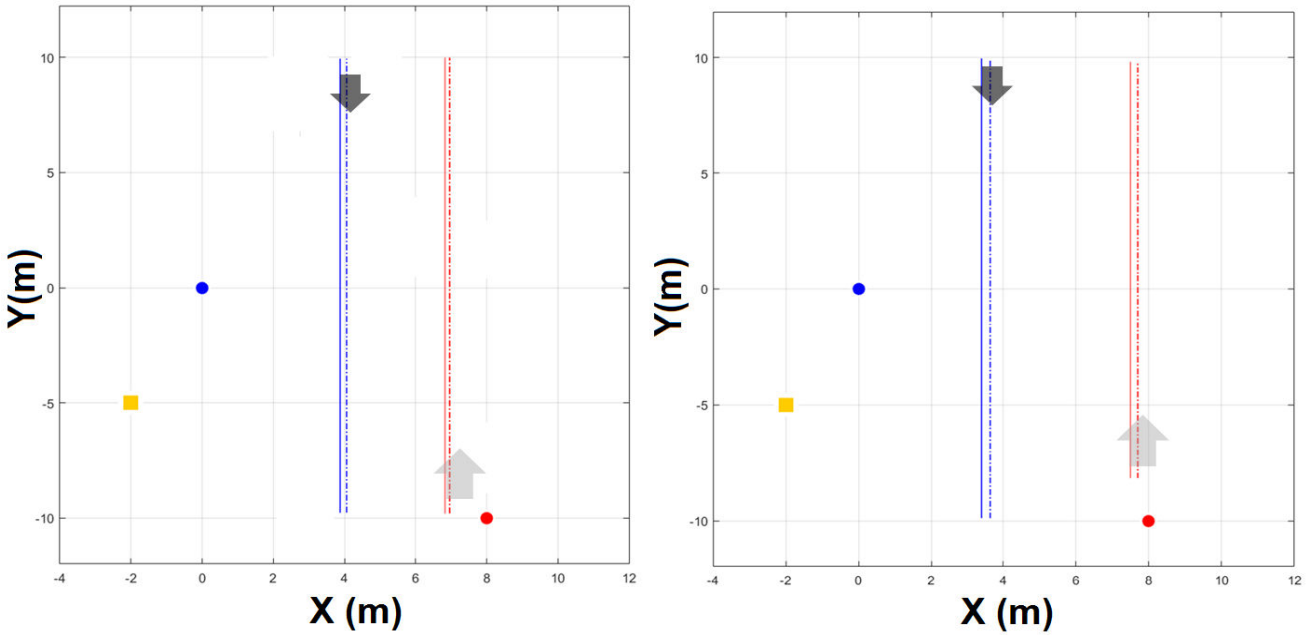
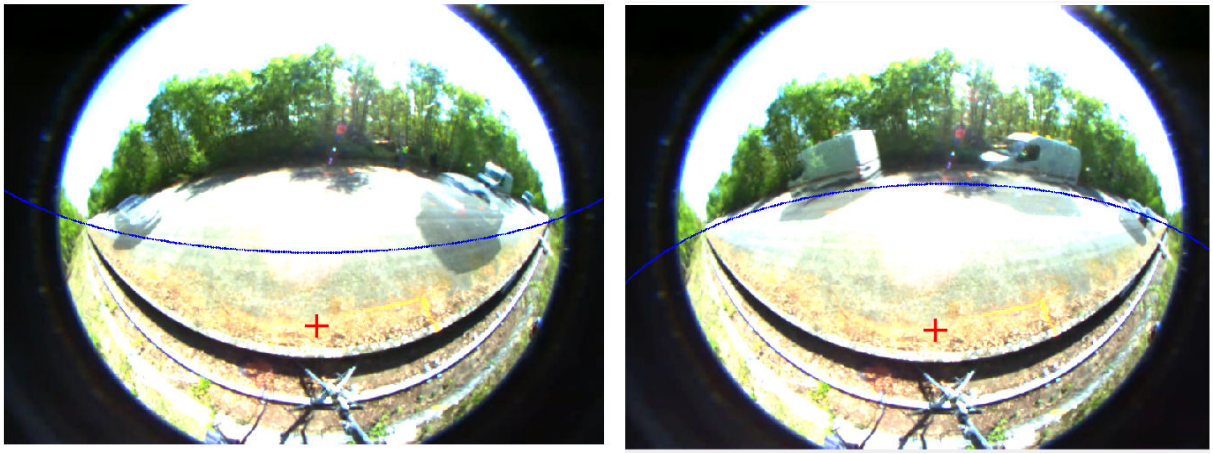



**Fig. 4.31:** Vehicle lateral position evaluation: first (Figure 4.32-a) and second (Figure 4.32-b) experiments, from left to right. Arrows indicate vehicles (black car and white van) opposite driving directions

a) first drive



b) second drive

**Fig. 4.32:** Estimated virtual lines from vehicle average lateral position (as formulated in Figure 4.30). The red cross represent the projected origin of the world frame ($VP_3$). It can be seen that the estimated lateral position is more precise when vehicle are near the cameras. Which can be actually linked to a visibility window between -10m and +10m along X-axis (likewise the LIDAR ground truth in Figure 4.31), where vehicle span a representative size in the fisheye image.

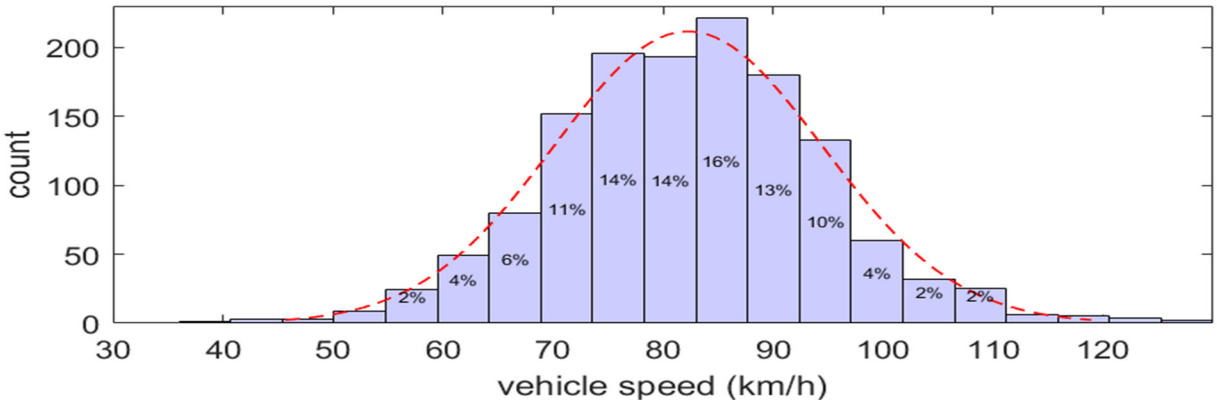



**Fig. 4.33:** Preliminary study: vehicle speed distribution

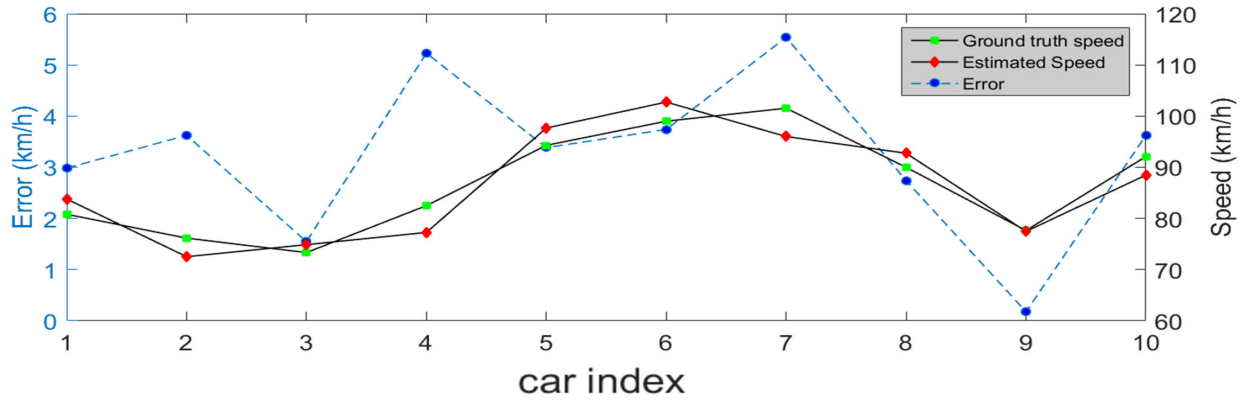



**Fig. 4.34:** Trajectory evaluation: vehicle speed analysis

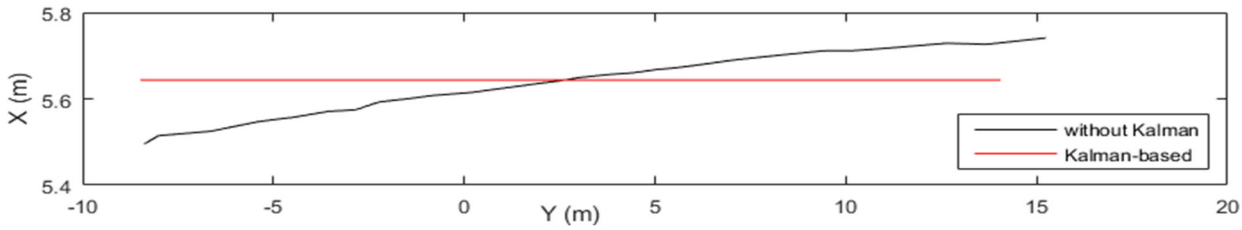



**Fig. 4.35:** Kalman filter performance for the example (over 20m)

In summary, the evaluation confirms the efficiency the proposed auto-calibration method, which leads to accurate vehicle localization (Figure 4.38). An example trajectory of a vehicle moving straight on the main road, is depicted in Figures 4.36,4.35. We can see that the use of the Bayesian filter allows to refine the trajectory and to compensate for possible vehicle segmentation errors over time. Our future works will involve further experiments, and comparison with Lidar-based processing in similar setup. In the next chapter we present a summary of our work and our contributions.

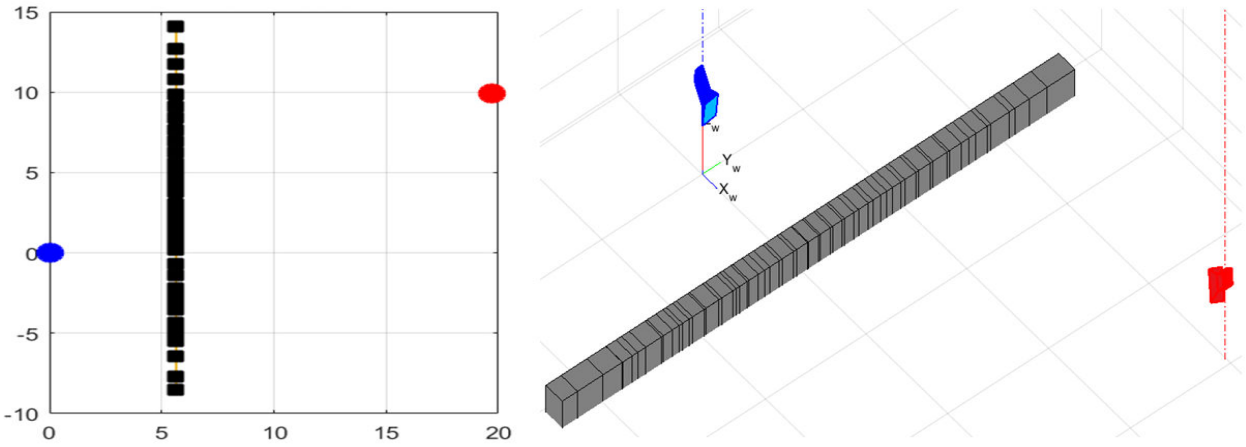



**Fig. 4.36:** Example of reconstructed trajectory (Figure 4.38)

**Fig. 4.37:** Illustration of vehicle re-identification: example of dynamic occlusion; a blob tracker is used along with our method

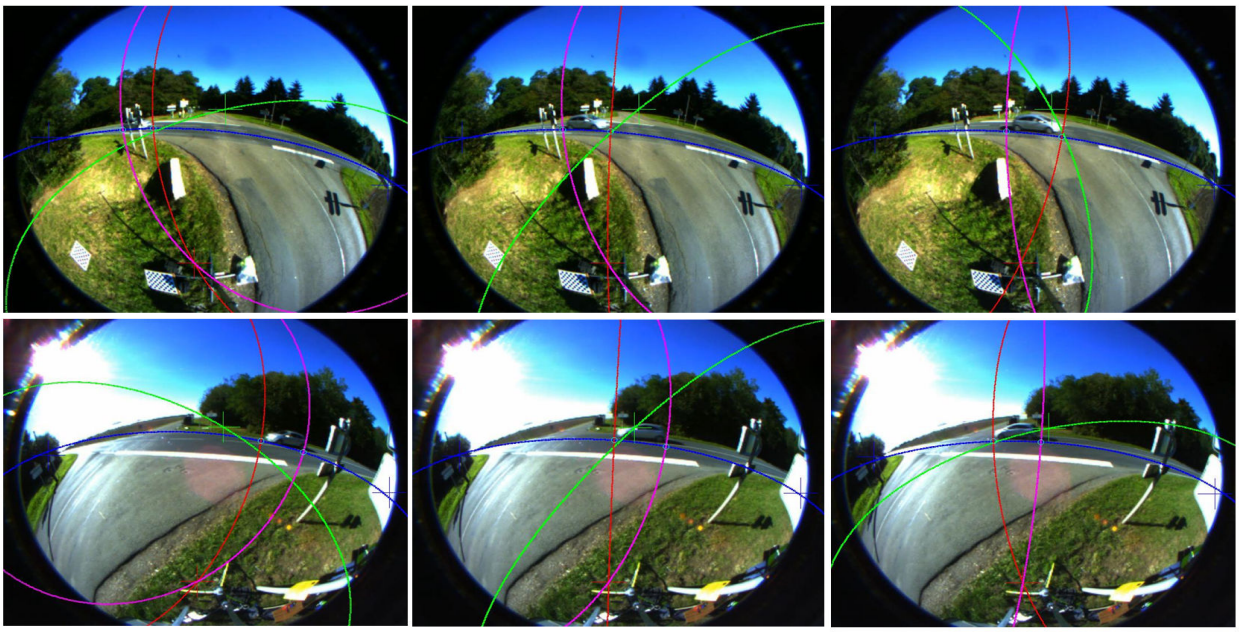


**Fig. 4.38:** Auto-calibration toward trajectory estimation: virtual lines (plane) tracking and association

# Conclusion and Perspectives

# 5

> *The only source of knowledge is experience*
>
> **— Albert Einstein**

The presented manuscript is the summary of my research works, toward safety improvement at road intersections, over the last three years. This thesis has mainly demonstrated the feasibility of a wide-baseline fisheye stereoscopic system for intersection monitoring: self-calibration and 3D localization of vehicles on the main road. In this final chapter, we provide a summary of our key contributions. We also discuss the limitations and introduce future directions. The work and ideas presented in this thesis previously featured in the following main papers (French and International submissions, already accepted or published):

—1) S. René E. DATONDJI, Yohan DUPUIS, Peggy SUBIRATS and Pascal VASSEUR, "Wide-baseline Omni-Stereo at Junctions: Extrinsic Auto-Calibration, Trajectory and Speed Estimation", IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 2017 [Conference]

—2) S. René E. DATONDJI, Yohan DUPUIS, Peggy SUBIRATS et Pascal VASSEUR, "Trajectographie à l'échelle absolue par Stéréovision-Fisheye Large Entraxe aux Carrefours", SAGEO, Rouen, 2017 [Short paper and poster]

—3) S. René E. DATONDJI, Yohan DUPUIS, Peggy SUBIRATS and Pascal VASSEUR, "Rotation and translation estimation for a wide baseline fisheye-stereo at crossroads based on traffic flow analysis", in IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 1534-1539, Rio de Janeiro, 2016 [Conference] [49]

—4) S. René E. DATONDJI, Yohan DUPUIS, Peggy SUBIRATS, and Pascal VASSEUR, "A survey of vision-based traffic monitoring of road intersections", in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 10, pp. 2681-2698, published in October 2016 [Journal, Survey and Analysis] [47]

—5) S. René E. DATONDJI, Yohan DUPUIS, Peggy SUBIRATS, and Pascal VASSEUR, "Calibration d'un dispositif stéréo-fisheye large baseline pour le diagnostic d'intersections routières", Reconnaissance de Formes et Intelligence Artificielle (RFIA)-Journée Transports Intelligents, Juin 2016 [Conference] [48]

In this thesis, we have presented a survey of vision-based intersection monitoring. Then we have proposed a wide baseline fisheye-stereo system for traffic monitoring in this context. We have introduced a method which allows to estimate automatically the extrinsic calibration between the cameras. Our approach is suitable for difficult environments such as rural scenes where the absence of lines makes the calibration especially challenging. In fact, the approach requires neither the knowledge of the scene geometry, nor the use of a calibration pattern. Instead vehicles are used as dynamic calibration objects, by jointly analyzing in the process their motion and appearance cues. The proposed approach allows to obtain metric localization and therefore to compute motion parameters of vehicles, especially on the main road where they travel faster and are visible on a long distance. As formulated, the method can be applied to estimate the pose of traffic cameras for different applications, provided straight planar motion on one main direction. Extensive results in the lab and at rural intersections validate the proposed auto-calibration approach which is extensible to several cameras types thanks to the spherical model.

There are several perspectives to this thesis. Though our approach allows to estimate vehicle positions at scale with good accuracy, trajectory reconstruction remains challenging. The critical aspect of the processing which mostly affect the performance is related to vehicles segmentation. Besides, the proposed wide-baseline fisheye-stereo is yet to be used to its full extent. In fact, data association between the cameras remains a complex task and will require deeper research. Future works also involve direct near-miss multi-view conflicts or accidents detection, for example by trajectory entropy analysis. In addition, extensive experiments need to be carried out on more intersections, in challenging conditions such as by nighttime or in rainy weather.

This thesis project has been a great professional challenge. It was a great experience to contribute to the fields of computer vision and intelligent transportations, through my thesis and while serving as a reviewer (IEEE ITSM and IEEE T-ITS). During this thesis, I have gained considerable communication skills as I took part twice in the Three Minute Thesis Challenge of Normandy and ended up each time in the final. These experiences also led me to win two awards, the best oral presentation at the Ph.D. days and the best poster at the LITIS Lab scientific day. Subsequently and above all, I have gained an important experience in research project management and software development. For my future career, my goal is to keep working toward the development of innovative solutions in order to improve traffic monitoring and road safety.

# Bibliography


[1]Ömer Aköz and M Elif Karsligil. „Traffic event classification at intersections based on the severity of abnormality“. In: *Machine vision and applications* (2014), pp. 1–20 (cit. on p. 16).

[2]Lee Alexander, Pi-Ming Cheng, Alec Gorjestani, et al. „The Minnesota mobile intersection surveillance system“. In: *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*. IEEE. 2006, pp. 139–144 (cit. on pp. 13, 32).

[3]David Pfeiffer Alexander Barth and Uwe Franke. „Vehicle Tracking at Urban Intersections Using Dense Stereo“. In: *3rd Workshop on Behaviour Monitoring and Interpretation, BMI*. Ghent, Belgium, Nov. 2009, pp. 47–58 (cit. on pp. 5, 31, 35, 36).

[4]S Álvarez, David Fernández Llorca, and MA Sotelo. „Hierarchical camera auto-calibration for traffic surveillance systems“. In: *Expert Systems with Applications* 41.4 (2014), pp. 1532–1542 (cit. on p. 50).

[5]Michel Antunes and Joao P Barreto. „A global approach for the detection of vanishing points and mutually orthogonal vanishing directions“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 1336–1343 (cit. on p. 57).

[6]Jon Arróspide and Luis Salgado. „A study of feature combination for vehicle detection based on image processing“. In: *The Scientific World Journal* 2014 (2014) (cit. on p. 20).

[7]K Somani Arun, Thomas S Huang, and Steven D Blostein. „Least-squares fitting of two 3-D point sets“. In: *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), pp. 698–700 (cit. on p. 60).

[8]Olivier Aycard, Qadeer Baig, Silviu Bota, et al. *Intersection Safety using Lidar and Stereo Vision sensors*. University of Grenoble - FRANCE, 2011 (cit. on pp. 5, 13, 14, 29, 34, 36, 40).

[9]Peyman Babaei. „Vehicles tracking and classification using Traffic zones in a hybrid scheme for intersection traffic management by smart cameras“. In: *Signal and Image Processing (ICSIP), 2010 International Conference on*. IEEE. 2010, pp. 49–53 (cit. on p. 17).

[10] Hernán Badino, Uwe Franke, and David Pfeiffer. „The stixel world-a compact medium level representation of the 3d-world“. In: *Pattern Recognition*. Springer, 2009, pp. 51–60 (cit. on p. 20).

[11] Xuegang Jeff Ban and Marco Gruteser. „Towards fine-grained urban traffic knowledge extraction using mobile sensing“. In: *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. ACM. 2012, pp. 111–117 (cit. on p. 16).

[12] François Bardet and Thierry Chateau. „MCMC particle filter for real-time visual tracking of vehicles“. In: *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*. IEEE. 2008, pp. 539–544 (cit. on p. 25).

[13] Olivier Barnich and Marc Van Droogenbroeck. „ViBe: A universal background subtraction algorithm for video sequences“. In: *Image Processing, IEEE Transactions on* 20.6 (2011), pp. 1709–1724 (cit. on pp. 18, 19, 58).

[14] Joao P Barreto and Helder Araujo. „Issues on the geometry of central catadioptric image formation“. In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 2. IEEE. 2001, pp. II–422 (cit. on p. 51).

[15] Alexander Barth. „Vehicle tracking and motion estimation based on stereo vision sequences“. PhD thesis. IGG, 2010 (cit. on pp. 24, 29, 31).

[16] Jean-Charles Bazin and Marc Pollefeys. „3-line RANSAC for orthogonal vanishing point detection“. In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE. 2012, pp. 4282–4287 (cit. on p. 57).

[17] Jean-Charles Bazin, Pierre-Yves Laffont, Inso Kweon, Cédric Demonceaux, and Pascal Vasseur. „An original approach for automatic plane extraction by omnidirectional vision“. In: *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE. 2010, pp. 752–758 (cit. on pp. 67, 77, 87).

[18] N Blanc, B Steux, and T Hinz. „LaRASideCam: A fast and robust vision-based blindspot detection system“. In: *Intelligent Vehicles Symposium, 2007 IEEE*. IEEE. 2007, pp. 480–485 (cit. on p. 20).

[19] Jean-Yves Bouguet. „Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm“. In: *Intel Corporation* 5 (2001), pp. 1–10 (cit. on pp. 20, 59).

[20] Thierry Bouwmans, Fida El Baf, and Bertrand Vachon. „Background modeling using mixture of gaussians for foreground detection-a survey“. In: *Recent Patents on Computer Science* 1.3 (2008), pp. 219–237 (cit. on pp. 19, 58).

[21]Norbert Buch, Sergio A Velastin, and James Orwell. „A review of computer vision techniques for the analysis of urban traffic“. In: *Intelligent Transportation Systems, IEEE Transactions on* 12.3 (2011), pp. 920–939 (cit. on pp. 8, 12, 18).

[22]Norbert Buch, James Orwell, and Sergio A Velastin. „Detection and classification of vehicles for urban traffic scenes“. In: *Visual Information Engineering,* (5th International Conference on. IET, 2008.) (cit. on pp. 18, 20, 21).

[23]Thomas Bucher, Cristobal Curio, Johann Edelbrunner, et al. „Image processing and behavior planning for intelligent vehicles“. In: *Industrial Electronics, IEEE Transactions on* 50.1 (2003), pp. 62–75 (cit. on p. 19).

[24]Xianbin Cao, Jinhe Lan, Pingkun Yan, and Xuelong Li. „Vehicle detection and tracking in airborne videos by multi-motion layer analysis“. In: *Machine Vision and Applications* 23.5 (2012), pp. 921–935 (cit. on p. 25).

[25]Yanpeng Cao, A. Renfrew, and P. Cook. „Novel optical flow optimization using pulse-coupled neural network and smallest univalue segment assimilating nucleus“. In: *Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. International Symposium on*. 2007, pp. 264–267 (cit. on p. 20).

[26]Yanpeng Cao, A. Renfrew, and P. Cook. „Vehicle motion analysis based on a monocular vision system“. In: *Road Transport Information and Control - RTIC 2008 and ITS United Kingdom Members' Conference, IET*. 2008, pp. 1–6 (cit. on p. 20).

[27]Y-M Chan, S-S Huang, L-C Fu, P-Y Hsiao, and M-F Lo. „Vehicle detection and tracking under various lighting conditions using a particle filter“. In: *IET intelligent transport systems* 6.1 (2012), pp. 1–8 (cit. on p. 25).

[28]Yi-Ming Chan, Shih-Shinh Huang, Li-Chen Fu, and Pei-Yung Hsiao. „Vehicle detection under various lighting conditions by incorporating particle filter“. In: *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. IEEE. 2007, pp. 534–539 (cit. on p. 20).

[29]Bo-Hao Chen and Shih-Chia Huang. „Probabilistic neural networks based moving vehicles extraction algorithm for intelligent traffic surveillance systems“. In: *Information Sciences* 299 (2015), pp. 283–295 (cit. on p. 22).

[30]L. Chen and C. Englund. „Cooperative Intersection Management: A Survey“. In: *IEEE Transactions on Intelligent Transportation Systems* 17.2 (2016), pp– 570–586 (cit. on p. 5).

[31]Hong Cheng, Nanning Zheng, and Chong Sun. „Boosted Gabor features applied to vehicle detection“. In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 1. IEEE. 2006, pp. 662–666 (cit. on p. 22).

[32] Hsu Yung Cheng and Jenq Neng Hwang. „Multiple-target tracking for crossroad traffic utilizing modified probabilistic data association“. In: *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*. Vol. 1. IEEE. 2007, pp. I–921 (cit. on p. 30).

[33] Minkyu Cheon, Wonju Lee, Changyong Yoon, and Mignon Park. „Vision-based vehicle detection system with consideration of the detecting location“. In: *Intelligent Transportation Systems, IEEE Transactions on* 13.3 (2012), pp. 1243–1252 (cit. on p. 20).

[34] Sunglok Choi, Taemin Kim, and Wonpil Yu. „Performance evaluation of RANSAC family“. In: *Journal of Computer Vision* 24.3 (1997), pp. 271–300 (cit. on p. 61).

[35] Robert T Collins, Alan Lipton, Takeo Kanade, et al. *A system for video surveillance and monitoring*. Vol. 2. Carnegie Mellon University, the Robotics Institute Pittsburg, 2000 (cit. on p. 2).

[36] European Commission. *Road safety in the European Union: Trends, statistics and Challenges*. http://goo.gl/LqWwl7. Online; accessed 01-October-2015. 2015 (cit. on p. 3).

[37] European Commission. *Traffic Safety Basic Facts 2012: Junctions*. http://goo.gl/YYpqD6. Online; accessed 01-October-2015. 2013 (cit. on p. 3).

[38] European Commission. *Traffic Safety Basic Facts on Junctions*. http://goo.gl/ZkhX15. Last accessed May 2016. 2015 (cit. on p. 3).

[39] Jonathan Courbon, Youcef Mezouar, Laurent Eckt, and Philippe Martinet. „A generic fisheye camera model for robotic applications“. In: *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE. 2007, pp. 1683–1688 (cit. on pp. 50, 51).

[40] Marco Cristani, Michela Farenzena, Domenico Bloisi, and Vittorio Murino. „Background subtraction for automated multisensor surveillance: a comprehensive review“. In: *EURASIP Journal on Advances in signal Processing* 2010 (2010), p. 43 (cit. on p. 19).

[41] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. „Detecting moving objects, ghosts, and shadows in video streams“. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.10 (2003), pp. 1337–1342 (cit. on p. 20).

[42] E Dabbour and S M Easa. „Perceptual framework for a modern left-turn collision warning system“. In: *World Academy of Science, Engineering and Technology* 33 (2009), pp. 640–646 (cit. on pp. 4, 13).

[43]R Danescu, S Nedevschi, MM Meinecke, and T Graf. „Stereovision based vehicle tracking in urban traffic environments“. In: *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. IEEE. 2007, pp. 400–404 (cit. on p. 24).

[44]Radu Danescu, Florin Oniga, and Sergiu Nedevschi. „Modeling and tracking the driving environment with a particle-based occupancy grid“. In: *Intelligent Transportation Systems, IEEE Transactions on* 12.4 (2011), pp. 1331–1342 (cit. on p. 19).

[45]René DATONDJI, Yohan DUPUIS, Peggy SUBIRATS, and Pascal VASSEUR. „Calibration d'un dispositif stéréo-fisheye large baseline pour le diagnostic d'intersections routières“. In: *Reconnaissance de Formes et Intelligence Artificielle (RFIA)-Journée Transports Intelligents*. 2016 (cit. on p. 67).

[46]René Emmanuel Datondji, Nicolas Ragot, Yassine Nasri, Redouane Khemmar, and Rémi Boutteau. „Odométrie visuelle par vision omnidirectionnelle pour la navigation autonome d'une chaise roulante motorisée“. In: *Journées francophones des jeunes chercheurs en vision par ordinateur*. 2015 (cit. on p. 48).

[47]S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur. „A Survey of Vision-Based Traffic Monitoring of Road Intersections“. In: *IEEE Transactions on Intelligent Transportation Systems* 17.10 (2016), pp. 2681–2698 (cit. on pp. 48, 105).

[48]S. René E. Datondji, Yohan Dupuis, Peggy Subirats, and Pascal Vasseur. „Calibration d'un dispositif stéréo-fisheye large baseline pour le diagnostic d'intersections routières“. In: (2016) (cit. on p. 105).

[49]S. René E. Datondji, Yohan Dupuis, Peggy Subirats, and Pascal Vasseur. „Rotation and Translation Estimation for a Wide Baseline Fisheye-Stereo at Crossroads Based on Traffic Flow Analysis“. In: *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC 2016), Rio de Janeiro, Brazil, November 1-4* (2016) (cit. on pp. 63, 67, 71, 73, 77, 87, 88, 105).

[50]Kaushik Deb, Ibrahim Khan, Anik Saha, and Kang-Hyun Jo. „An Efficient Method of Vehicle License Plate Recognition Based on Sliding Concentric Windows and Artificial Neural Network“. In: *Procedia Technology* 4 (2012), pp. 812–819 (cit. on p. 22).

[51]Hai Dinh and Hua Tang. „Simple method for camera calibration of roundabout traffic scenes using a single circle“. In: *IET Intelligent Transport Systems* 8.3 (2013), pp. 175–182 (cit. on p. 26).

[52]Arnaud Doucet and Adam M Johansen. „A tutorial on particle filtering and smoothing: Fifteen years later“. In: *Handbook of Nonlinear Filtering* 12 (2009), pp. 656–704 (cit. on p. 25).

[53] Marketa Dubska, Jakub Sochor, and Adam Herout. „Automatic Camera Calibration for Traffic Understanding“. In: *Proceedings of BMVC 2014* (2014), pp. 1–10 (cit. on pp. 18, 25, 49, 50, 58, 68).

[54] A. Eichenseer and A. Kaup. „A data set providing synthetic and real-world fisheye video sequences“. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016, pp. 1541–1545 (cit. on pp. 47, 49, 75).

[55] Friedrich Erbs, Alexander Barth, and Uwe Franke. „Moving vehicle detection by optimal segmentation of the dynamic stixel world“. In: *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE. 2011, pp. 951–956 (cit. on p. 20).

[56] Martin A Fischler and Robert C Bolles. „Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography“. In: *Communications of the ACM* 24.6 (1981), pp. 381–395 (cit. on p. 61).

[57] Yoav Freund and Robert E Schapire. „A decision-theoretic generalization of on-line learning and an application to boosting“. In: *Journal of computer and system sciences* 55.1 (1997), pp. 119–139 (cit. on p. 22).

[58] Yoav Freund and Robert E Schapire. „A desicion-theoretic generalization of on-line learning and an application to boosting“. In: *Computational learning theory*. Springer. 1995, pp. 23–37 (cit. on p. 22).

[59] Qiang Fu, Quan Quan, and Kai-Yuan Cai. „Calibration of multiple fish-eye cameras using a wand“. In: *IET Computer Vision* 9.3 (2014), pp. 378–389 (cit. on p. 50).

[60] Takashi Furuya and Camillo J. Taylor. „Road Intersection Monitoring from Video with Large Perspective Deformation“. In: *21st World Congress on Intelligent Transport Systems*. 2014 (cit. on pp. 17, 18, 24, 25, 31, 32, 59).

[61] Stefan Gehrig, Clemens Rabe, and Lars Krüger. „6D vision goes fisheye for intersection assistance“. In: *Computer and Robot Vision, 2008. CRV'08. Canadian Conference on*. IEEE. 2008, pp. 34–41 (cit. on p. 5).

[62] Andreas Geiger and Bernd Kitt. „Object flow: A descriptor for classifying traffic motion“. In: *Intelligent Vehicles Symposium (IV), 2010 IEEE*. IEEE. 2010, pp. 287–293 (cit. on pp. 21, 35).

[63] Andreas Geiger, Martin Lauer, Christian Wojek, Christoph Stiller, and Raquel Urtasun. „3d traffic scene understanding from movable platforms“. In: *Pattern Analysis and Machine Intelligence* (IEEE Transactions on 36.5 (2014): 1012-1025.) (cit. on pp. 3, 27, 38, 40).

[64]Christopher Geyer and Kostas Daniilidis. „A unifying theory for central panoramic systems and practical implications“. In: *Computer Vision-ECCV 2000*. Springer, 2000, pp. 445–461 (cit. on p. 51).

[65]Ali Ghorayeb, Alex Potelle, Laure Devendeville, El Mustapha Mouaddib, et al. „Capteur omnidirectionnel Optimal pour le diagnostic de la circulation dans les carrefours urbains“. In: *ORASIS'09-Congrès des jeunes chercheurs en vision par ordinateur*. 2009 (cit. on pp. 4, 14, 17, 28, 33).

[66]Ali Ghorayeb, Alexis Potelle, Laure Devendeville, and El Mustapha Mouaddib. „Optimal omnidirectional sensor for urban traffic diagnosis in crossroads“. In: *Intelligent Vehicles Symposium (IV), 2010 IEEE*. IEEE. 2010, pp. 597–602 (cit. on pp. 26, 48).

[67]Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. „LSD: A fast line segment detector with a false detection control“. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2008) (cit. on p. 63).

[68]Michael Goldhammer, Elias Strigel, Daniel Meissner, et al. „Cooperative multi sensor network for traffic safety applications at intersections“. In: *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE. 2012, pp. 1178–1183 (cit. on p. 5).

[69]Mohinder S Grewal. *Kalman filtering*. Springer, 2011 (cit. on p. 25).

[70]GRIDSMART. *Traffic management systems for intersection monitoring*. https://gridsmart.com. Online; accessed 01-October-2015 (cit. on pp. 5, 29).

[71]Hanqi Guo and Zuchao Wang. „TripVista : Triple Perspective Visual Trajectory Analytics and Its Application on Microscopic Traffic Data at a Road Intersection“. In: *in Visualization Symposium (PacificVis)* (2011) (cit. on p. 3).

[72]Yanlin Guo, Cen Rao, Supun Samarasekera, et al. „Matching vehicles under large pose transformations using approximate 3d models and piecewise mrf model“. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE. 2008, pp. 1–8 (cit. on pp. 19, 21).

[73]Peng Hao, Xuegang Ban, Kristin P Bennett, Qiang Ji, and Zhanbo Sun. „Signal timing estimation using sample intersection travel times“. In: *Intelligent Transportation Systems, IEEE Transactions on* 13.2 (2012), pp. 792–804 (cit. on p. 16).

[74]Marko Heikkila and Matti Pietikainen. „A texture-based method for modeling the background and detecting moving objects“. In: *IEEE transactions on pattern analysis and machine intelligence* 28.4 (2006), pp. 657–662 (cit. on p. 20).

[75] Lionel Heng, Mathias Bürki, Gim Hee Lee, et al. „Infrastructure-based calibration of a multi-camera rig“. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2014, pp. 4912–4919 (cit. on p. 50).

[76] Stefan Hoeglinger et al. *Traffic Safety Basic Facts 2007: Junctions*. http://goo.gl/AE8SYX. Online; accessed 01-October-2015. 2008 (cit. on pp. 2, 3).

[77] Christian Hoffmann. „Fusing multiple 2D visual features for vehicle detection“. In: *Intelligent Vehicles Symposium, 2006 IEEE*. IEEE. 2006, pp. 406–411 (cit. on p. 20).

[78] Timothy Hospedales, Shaogang Gong, and Tao Xiang. „Video behaviour mining using a dynamic topic model“. In: *International journal of computer vision* 98.3 (2012), pp. 303–323 (cit. on pp. 15, 16).

[79] Chao-Yung Hsu, Li-Wei Kang, and Hong-Yuan Mark Liao. „Cross-camera vehicle tracking via affine invariant object matching for video forensics applications“. In: *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE. 2013, pp. 1–6 (cit. on p. 20).

[80] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. „A Survey on Visual Surveillance of Object Motion and Behaviors“. In: *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions* 34.3 (2004), pp. 334–352 (cit. on pp. 3, 20).

[81] Weiming Hu, Xuejuan Xiao, Zhouyu Fu, et al. „A system for learning statistical motion patterns“. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28.9 (2006), pp. 1450–1464 (cit. on pp. 15, 16, 27).

[82] Weiming Hu, Xuejuan Xiao, Dan Xie, Tieniu Tan, and Steve Maybank. „Traffic accident prediction using 3-D model-based vehicle tracking“. In: *Vehicular Technology, IEEE Transactions on* 53.3 (2004), pp. 677–694 (cit. on p. 19).

[83] Zhencheng Hu, Chenhao Wang, and Keiichi Uchimura. „3D Vehicle extraction and tracking from multiple viewpoints for traffic monitoring by using probability fusion map“. In: *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. IEEE. 2007, pp. 30–35 (cit. on p. 4).

[84] Michael Isard and Andrew Blake. „Conditional density propagation for visual tracking“. In: *International journal of computer vision* (1998) (cit. on p. 25).

[85] Karim Ismail, Tarek Sayed, and Nicolas Saunier. „Camera calibration for urban traffic scenes: practical issues and a robust approach“. In: *Transportation Research Board Annual Meeting Compendium of Papers*. 2010 (cit. on p. 18).

[86]*Japan, Intelligent Transportation Systems, Green Safety: public-Private collaborative projects that aims "For a Greener and Safer Society" using Cooperative systems*. http://www.its-jp.org/english/its-green-safety-showcase/. Online; accessed 01-October-2015 (cit. on p. 6).

[87]Amirali Jazayeri, Hongyuan Cai, Jiang Yu Zheng, and Mihran Tuceryan. „Vehicle detection and tracking in car video based on motion model“. In: *Intelligent Transportation Systems, IEEE Transactions on* 12.2 (2011), pp. 583–595 (cit. on p. 20).

[88]Timothy F. Gee Jeffery R. Price. „Omnidirectional imaging and computer vision for transportation applications: from conception to deployment“. In: *Report-AldisCorp* (2011) (cit. on pp. 4, 5, 14, 26–29, 33, 48).

[89]Jean-Philippe Jodoin, Guillaume-Alexandre Bilodeau, and Nicolas Saunier. „Urban Tracker: Multiple object tracking in urban mixed traffic“. In: *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*. IEEE. 2014, pp. 885–892 (cit. on pp. 15, 17).

[90]Björn Johansson, Johan Wiklund, Per-Erik Forssén, and Gösta Granlund. „Combining shadow detection and simulation for estimation of vehicle size and position“. In: *Pattern Recognition Letters* 30.8 (2009), pp. 751–759 (cit. on pp. 20, 21).

[91]Eugen Kafer, Christoph Hermes, C Wohler, Helge Ritter, and Franz Kummert. „Recognition of situation classes at road intersections“. In: *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE. 2010, pp. 3960–3965 (cit. on p. 3).

[92]Rudolph Emil Kalman. „A new approach to linear filtering and prediction problems“. In: *Journal of Fluids Engineering* 82.1 (1960), pp. 35–45 (cit. on p. 25).

[93]S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi. „Traffic monitoring and accident detection at intersections“. In: *Intelligent Transportation Systems, IEEE Transactions on* 1.2 (2000), pp. 108–118 (cit. on pp. 17, 27, 29, 30, 32, 37).

[94]Neeraj K Kanhere and Stanley T Birchfield. „A taxonomy and analysis of camera calibration methods for traffic monitoring applications“. In: *IEEE Transactions on Intelligent Transportation Systems* 11.2 (2010), pp. 441–452 (cit. on pp. 18, 49, 50, 57).

[95]V Kastrinaki, M Zervakis, and Kostas Kalaitzakis. „A survey of video processing techniques for traffic applications“. In: *Image and vision computing* 21.4 (2003), pp. 359–381 (cit. on pp. 3, 8).

[96] Moritz Knorr, Jose Esparza, Wolfgang Niehsen, and Christoph Stiller. „Extrinsic calibration of a fisheye multi-camera setup using overlapping fields of view“. In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 1276–1281 (cit. on p. 50).

[97] Moritz Knorr, Wolfgang Niehsen, and Christoph Stiller. „Online extrinsic multi-camera calibration using ground plane induced homographies“. In: *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE. 2013, pp. 236–241 (cit. on p. 50).

[98] Teuvo Kohonen. „An introduction to neural computing“. In: *Neural networks* 1.1 (1988), pp. 3–16 (cit. on p. 22).

[99] T Kowsari, SS Beauchemin, and J Cho. „Real-time vehicle detection and tracking using stereo vision and multi-view AdaBoost“. In: *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE. 2011, pp. 1255–1260 (cit. on p. 20).

[100] Till Kroeger, Dengxin Dai, and Luc Van Gool. „Joint vanishing point extraction and tracking“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 2449–2457 (cit. on p. 57).

[101] Suryansh Kumar, Yuchao Dai, and Hongdong Li. „Multi-body non-rigid structure-from-motion“. In: *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE. 2016, pp. 148–156 (cit. on p. 77).

[102] Raphael Labayrade, Didier Aubert, and J-P Tarel. „Real time obstacle detection in stereovision on non flat road geometry through" v-disparity" representation“. In: *Intelligent Vehicle Symposium, 2002. IEEE*. Vol. 2. IEEE. 2002, pp. 646–651 (cit. on p. 19).

[103] Jeong-Kyun Lee and Kuk-Jin Yoon. „Real-time Joint Estimation of Camera Orientation and Vanishing Points“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 1866–1874 (cit. on p. 57).

[104] Sang-Mook Lee and Hojong Baik. „Origin-destination (OD) trip table estimation using traffic movement counts from vehicle tracking system at intersection“. In: *IEEE Industrial Electronics, IECON 2006-32nd Annual Conference on*. IEEE. 2006, pp. 3332–3337 (cit. on pp. 4, 5, 26–28, 33, 48).

[105] Stéphanie Lefèvre, Christian Laugier, and Javier Ibañez-Guzmán. „Evaluating risk at road intersections by detecting conflicting intentions“. In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE. 2012, pp. 4841–4846 (cit. on p. 38).

[106] Stéphanie Lefèvre, Christian Laugier, and Javier Ibañez-Guzmán. „Risk assessment at road intersections: Comparing intention and expectation“. In: *Intelligent Vehicles Symposium (IV), 2012 IEEE*. IEEE. 2012, pp. 165–171 (cit. on pp. 2, 3).

[107]Bastian Leibe, Konrad Schindler, Nico Cornelis, and Luc Van Gool. „Coupled object detection and tracking from static cameras and moving vehicles“. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30.10 (2008), pp. 1683–1698 (cit. on p. 25).

[108]Young-Chul Lim, Minho Lee, Chung-Hee Lee, Soon Kwon, and Jong-hun Lee. „Improvement of stereo vision-based position and velocity estimation and tracking using a stripe-based disparity estimation and inverse perspective map-based extended Kalman filter“. In: *Optics and Lasers in Engineering* 48.9 (2010), pp. 859–868 (cit. on p. 19).

[109]Mingxiu Lin and Xinhe Xu. „Multiple vehicle visual tracking from a moving vehicle“. In: *Intelligent Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on*. Vol. 2. IEEE. 2006, pp. 373–378 (cit. on p. 21).

[110]Wei Liu, XueZhi Wen, Bobo Duan, Huai Yuan, and Nan Wang. „Rear vehicle detection and tracking for lane change assist“. In: *Intelligent Vehicles Symposium, 2007 IEEE*. IEEE. 2007, pp. 252–257 (cit. on p. 20).

[111]Jianguang Lou, Tieniu Tan, Weiming Hu, Hao Yang, and Stephen J Maybank. „3-D model-based vehicle tracking“. In: *Image Processing, IEEE Transactions on* 14.10 (2005), pp. 1561–1569 (cit. on pp. 19, 21).

[112]Chen Change Loy, Tao Xiang, and Shaogang Gong. „Detecting and discriminating behavioural anomalies“. In: *Pattern Recognition* 44.1 (2011), pp. 117–132 (cit. on p. 16).

[113]Bruce D Lucas, Takeo Kanade, et al. „An iterative image registration technique with an application to stereo vision.“ In: *IJCAI*. Vol. 81. 1981, pp. 674–679 (cit. on p. 25).

[114]Dieu Sang Ly, Cédric Demonceaux, Pascal Vasseur, and Claude Pégard. „Extrinsic calibration of heterogeneous cameras by line images“. In: *Machine vision and applications* 25.6 (2014), pp. 1601–1614 (cit. on pp. 18, 67).

[115]Dieu-Sang Ly, Cédric Demonceaux, Ralph Seulin, and Yohan Fougerolle. „Scale invariant line matching on the sphere“. In: *IEEE International Conference on Image Processing, ICIP'2013*. 2013 (cit. on p. 72).

[116]Sang Ly, Cedric Demonceaux, and Pascal Vasseur. „Translation estimation for single viewpoint cameras using lines“. In: *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE. 2010, pp. 1928–1933 (cit. on p. 72).

[117]Arvind M. „Sensor Fusion for Real-time Gap Tracking and Vehicle Trajectory Estimation at Rural Intersections“. In: *Doctoral dissertation* (2005) (cit. on pp. 14, 28, 31, 32).

[118]Emilio Maggio and Andrea Cavallaro. *Video tracking: theory and practice*. John Wiley & Sons, 2011 (cit. on p. 24).

[119] Ling Mao, Mei Xie, Yi Huang, and Yuefei Zhang. „Preceding vehicle detection using histograms of oriented gradients“. In: *Communications, Circuits and Systems (ICCCAS), 2010 International Conference on*. IEEE. 2010, pp. 354–358 (cit. on pp. 20, 22).

[120] Xue Mei and Haibin Ling. „Robust visual tracking and vehicle classification via sparse representation“. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33.11 (2011), pp. 2259–2272 (cit. on p. 25).

[121] Stefano Messelodi, CarlaMaria Modena, and Michele Zanin. „A computer vision system for the detection and classification of vehicles at urban road intersections“. English. In: *Pattern Analysis and Applications* 8.1-2 (2005), pp. 17–31 (cit. on pp. 4, 18, 19, 21, 25, 26, 28–30, 32, 42).

[122] O Moisan, P Subirats, O Bisson, et al. „A safer road with no accidents: a case study“. In: *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*. 2014 (cit. on pp. 6, 7).

[123] Matthieu Molinier, Tuomas Häme, and Heikki Ahola. „3D-connected components analysis for traffic monitoring in image sequences acquired from a helicopter“. In: *Image Analysis*. Springer, 2005, pp. 141–150 (cit. on pp. 27, 31, 32).

[124] Brendan Morris and Mohan Trivedi. „Robust classification and tracking of vehicles in traffic video streams“. In: *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*. IEEE. 2006, pp. 1078–1083 (cit. on p. 25).

[125] Brendan Tran Morris and Mohan M Trivedi. „Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach“. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33.11 (2011), pp. 2287–2301 (cit. on pp. 3, 15, 16).

[126] Brendan Tran Morris and Mohan Manubhai Trivedi. „A survey of vision-based trajectory learning and analysis for surveillance“. In: *Circuits and Systems for Video Technology, IEEE Transactions on* 18.8 (2008), pp. 1114–1127 (cit. on pp. 2, 18).

[127] Brendan Tran Morris and Mohan Manubhai Trivedi. „Understanding vehicular traffic behavior from video: a survey of unsupervised approaches“. In: *Journal of Electronic Imaging* 22.4 (2013), pp. 041113–041113 (cit. on p. 2).

[128] Saleh Mosaddegh, David Fofi, and Pascal Vasseur. „Line based motion estimation and reconstruction of piece-wise planar scenes“. In: *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*. IEEE. 2011, pp. 658–663 (cit. on p. 72).

[129] Saleh Mosaddegh, David Fofi, and Pascal Vasseur. „Short baseline line matching for central imaging systems“. In: *Pattern Recognition Letters* 33.16 (2012), pp. 2292–2301 (cit. on p. 72).

[130]Ulrich Muehlmann, Miguel Ribo, Peter Lang, and Axel Pinz. „A new high speed CMOS camera for real-time tracking applications“. In: *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*. Vol. 5. IEEE. 2004, pp. 5195–5200 (cit. on p. 17).

[131]M. Muffert, T. Milbich, D. Pfeiffer, and U. Franke. „May I enter the roundabout? A time-to-contact computation based on stereo-vision“. In: *Intelligent Vehicles Symposium (IV), 2012 IEEE*. 2012, pp. 565–570 (cit. on pp. 29, 34–36, 40, 41).

[132]Vishvjit Singh Nalwa. *Panoramic viewing system with support stand*. US Patent 6,128,143. 2000 (cit. on p. 46).

[133]M Naylor and CI ATTWOOD. „Annotated digital video for intelligent surveillance and optimized retrieval: final report“. In: *ADVISOR connortium* (2003) (cit. on p. 2).

[134]US D.O.T NGSIM. *Next Generation Simulation Systems*. http://goo.gl/zHOX5M. Online; accessed 01-October-2015 (cit. on pp. 15, 16).

[135]Hossein Tehrani Niknejad, Akihiro Takeuchi, Seiichi Mita, and David McAllester. „On-road multivehicle tracking using deformable object model and particle filter with improved likelihood estimation“. In: *Intelligent Transportation Systems, IEEE Transactions on* 13.2 (2012), pp. 748–758 (cit. on p. 25).

[136]Jesus Nuevo, Ignacio Parra, Jonas Sjoberg, and Luis M Bergasa. „Estimating surrounding vehicles' pose using computer vision“. In: *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE. 2010, pp. 1863–1868 (cit. on p. 20).

[137]Seon Ho Oh and Soon Ki Jung. „A Great Circle Arc Detector in Equirectangular Images.“ In: *VISAPP (1)*. 2012, pp. 346–351 (cit. on p. 57).

[138]Artur Ottlik and H-H Nagel. „Initialization of model-based vehicle tracking in video sequences of inner-city intersections“. In: *International Journal of Computer Vision* 80.2 (2008), pp. 211–225 (cit. on pp. 16, 20, 21, 24).

[139]Igor E Paromtchik, Mathias Perrollaz, and Christian Laugier. „Fusion of telemetric and visual data from road scenes with a lexus experimental platform“. In: *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE. 2011, pp. 746–751 (cit. on p. 29).

[140]Luis Puig, Jesús Bermúdez, Peter Sturm, and José Jesús Guerrero. „Calibration of omnidirectional cameras in practice: A comparison of methods“. In: *Computer Vision and Image Understanding* 116.1 (2012), pp. 120–137 (cit. on pp. 50, 51, 72).

[141]Clemens Rabe, Uwe Franke, and Stefan Gehrig. „Fast detection of moving objects in complex scenarios“. In: *Intelligent Vehicles Symposium, 2007 IEEE*. IEEE. 2007, pp. 398–403 (cit. on p. 20).

[142]Richard J Radke. „A survey of distributed computer vision algorithms“. In: *Handbook of Ambient Intelligence and Smart Environments*. Springer, 2010, pp. 35–55 (cit. on p. 14).

[143]Bernhard Rinner and Wayne Wolf. „An introduction to distributed smart cameras“. In: *Proceedings of the IEEE* 96.10 (2008), pp. 1565–1575 (cit. on p. 4).

[144]Lee August Rodegerdts. *Roundabouts: An informational guide*. Vol. 672. Transportation Research Board, 2010 (cit. on p. 3).

[145]Paul E Rybski, Daniel Huber, Daniel D Morris, and Regis Hoffman. „Visual classification of coarse vehicle orientation using histogram of oriented gradients features“. In: *Intelligent Vehicles Symposium (IV), 2010 IEEE*. IEEE. 2010, pp. 921–928 (cit. on pp. 20, 22).

[146]Haytham Sadeq and Tarek Sayed. „Automated roundabout safety analysis: diagnosis and remedy of safety problems“. In: *Journal of Transportation Engineering* 142.12 (2016), p. 04016062 (cit. on pp. 38, 39).

[147]Andres Sanin, Conrad Sanderson, and Brian C Lovell. „Shadow detection: A survey and comparative evaluation of recent methods“. In: *Pattern recognition* 45.4 (2012), pp. 1684–1695 (cit. on p. 18).

[148]Ikuro Sato, Chiharu Yamano, and Hirohiko Yanagawa. „Crossing obstacle detection with a vehicle-mounted camera“. In: *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE. 2011, pp. 60–65 (cit. on p. 20).

[149]N. Saunier and T. Sayed. „A feature-based tracking algorithm for vehicles in intersections“. In: *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*. 2006, pp. 59–59 (cit. on pp. 18, 23, 25, 27, 29–32).

[150]Nicolas Saunier and Tarek Sayed. „Automated analysis of road safety with video data“. In: *Transportation Research Record: Journal of the Transportation Research Board* 2019.1 (2007), pp. 57–64 (cit. on p. 4).

[151]Nicolas Saunier, Tarek Sayed, and Clark Lim. „Probabilistic collision prediction for vision-based automated road safety analysis“. In: *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. IEEE. 2007, pp. 872–878 (cit. on pp. 38, 39).

[152]Davide Scaramuzza, Roland Siegwart, Roland Siegwart, and Roland Siegwart. *A practical toolbox for calibrating omnidirectional cameras*. Swiss Federal Institute of Technology, 2007 (cit. on pp. 52, 77).

[153]Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart. „A toolbox for easily calibrating omnidirectional cameras“. In: *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE. 2006, pp. 5695–5701 (cit. on p. 51).

[154]Miriam Schönbein. *Omnidirectional Stereo Vision for Autonomous Vehicles*. Vol. 32. KIT Scientific Publishing, 2015 (cit. on pp. 47, 48, 50).

[155]Werner von Seelen, Cristóbal Curio, J Gayko, Uwe Handmann, and Thomas Kalinke. „Scene analysis and organization of behavior in driver assistance systems“. In: *Image Processing, 2000. Proceedings. 2000 International Conference on*. Vol. 3. IEEE. 2000, pp. 524–527 (cit. on p. 22).

[156]Jie Sha, Yipu Zhao, Wenda Xu, et al. „Trajectory analysis of moving objects at intersection based on laser-data“. In: *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE. 2011, pp. 289–294 (cit. on p. 14).

[157]Yu Shi and Fábio Dias Real. „Smart cameras: Fundamentals and classification“. In: *Smart Cameras*. Springer, 2010, pp. 19–34 (cit. on p. 4).

[158]Patrick Yuri Shinzato, V Grassi, Fernando Santos Osório, and Denis F Wolf. „Fast visual road recognition and horizon detection using multiple artificial neural networks“. In: *Intelligent Vehicles Symposium (IV), 2012 IEEE*. IEEE. 2012, pp. 1090–1095 (cit. on p. 22).

[159]Mohammad Shokrolah Shirazi and Brendan Morris. „A typical video-based framework for counting, behavior and safety analysis at intersections“. In: *Intelligent Vehicles Symposium (IV), 2015 IEEE*. IEEE. 2015, pp. 1264–1269 (cit. on p. 38).

[160]Mohammad Shokrolah Shirazi and Brendan Morris. „Contextual Combination of Appearance and Motion for Intersection Videos with Vehicles and Pedestrians“. In: *Advances in Visual Computing*. Springer, 2014, pp. 708–717 (cit. on p. 15).

[161]Mohammad Shokrolah Shirazi and Brendan Morris. „Observing behaviors at intersections: a review of recent studies & developments“. In: *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2015, pp. 1258–1263 (cit. on pp. 2, 48).

[162]Mohammad Shokrolah Shirazi and Brendan Morris. „Vision-based turning movement counting at intersections by cooperating zone and trajectory comparison modules“. In: *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE. 2014, pp. 3100–3105 (cit. on p. 38).

[163]Mohammad Shokrolah Shirazi and Brendan Tran Morris. „Looking at Intersections: A Survey of Intersection Monitoring, Behavior and Safety Analysis of Recent Studies“. In: *IEEE Transactions on Intelligent Transportation Systems* 18.1 (2017), pp. 4–24 (cit. on p. 37).

[164] Mohammad Shokrolah Shirazi and Brendan Tran Morris. „Vision-Based Turning Movement Monitoring: Count, Speed & Waiting Time Estimation“. In: *IEEE Intelligent Transportation Systems Magazine* 8.1 (2016), pp. 23–34 (cit. on pp. 2, 48).

[165] Sudipta N Sinha, Marc Pollefeys, and Leonard McMillan. „Camera network calibration from dynamic silhouettes“. In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2004, pp. I–195 (cit. on p. 56).

[166] Sayanan Sivaraman and Mohan M Trivedi. „Active learning for on-road vehicle detection: A comparative study“. In: *Machine vision and applications* 25.3 (2014), pp. 599–611 (cit. on p. 22).

[167] Sayanan Sivaraman and Mohan Manubhai Trivedi. „Looking at Vehicles on the Road : A Survey of Vision-Based Vehicle Detection , Tracking and Behavior Analysis“. In: *Intelligent Transportation Systems* 14.4 (IEEE Transactions on 14.4 (2013): 1773-1795.), pp. 1773–1795 (cit. on pp. 8, 12).

[168] Andrews Sobral and Antoine Vacavant. „A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos“. In: *Computer Vision and Image Understanding* 122 (2014), pp. 4–21 (cit. on pp. 18, 19).

[169] Jakub Sochor, Roman Juránek, Jakub Španhel, et al. „BrnoCompSpeed: Review of Traffic Camera Calibration and A Comprehensive Dataset for Monocular Speed Measurement“. In: *(Under Review, IEEE T-ITS)* (2017) (cit. on pp. 49, 98).

[170] Xuefeng Song and Ram Nevatia. „Detection and tracking of moving vehicles in crowded scenes“. In: *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*. IEEE. 2007, pp. 4–4 (cit. on pp. 25, 30).

[171] Chris Stauffer and W Eric L Grimson. „Adaptive background mixture models for real-time tracking“. In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. Vol. 2. IEEE. 1999 (cit. on p. 19).

[172] Milos Stojmenovic. „Real time machine learning based car detection in images with fast training“. In: *Machine Vision and Applications* 17.3 (2006), pp. 163–172 (cit. on p. 22).

[173] Elias Strigel, Daniel Meissner, Florian Seeliger, Benjamin Wilking, and Klaus Dietmayer. „The Ko-PER intersection laserscanner and video dataset“. In: *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE. 2014, pp. 1900–1901 (cit. on pp. 5, 15, 16).

[174] Peggy Subirats, Yohan Dupuis, Eric Violette, David Doucet, and Guy Dupre. „A new tool to evaluate safety of crossroad“. In: *4th International Symposium on Highway Geometric Design. Valence*. 2010, pp. 2–5 (cit. on p. 3).

[175]Zehang Sun, George Bebis, and Ronald Miller. „Monocular precrash vehicle detection: features and classifiers“. In: *Image Processing, IEEE Transactions on* 15.7 (2006), pp. 2019–2034 (cit. on p. 20).

[176]Zehang Sun, George Bebis, and Ronald Miller. „On-road vehicle detection-A review“. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28.5 (2006), pp. 694–711 (cit. on p. 8).

[177]Zehang Sun, George Bebis, and Ronald Miller. „On-road vehicle detection using evolutionary Gabor filter optimization“. In: *Intelligent Transportation Systems, IEEE Transactions on* 6.2 (2005), pp. 125–137 (cit. on p. 20).

[178]Zehang Sun, George Bebis, and Ronald Miller. „On-road vehicle detection using Gabor filters and support vector machines“. In: *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*. Vol. 2. IEEE. 2002, pp. 1019–1022 (cit. on p. 22).

[179]Zhanbo Sun and Xuegang Jeff Ban. „Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors“. In: *Transportation Research Part C: Emerging Technologies* 36 (2013), pp. 268–283 (cit. on p. 16).

[180]Åse Svensson and Christer Hydén. „Estimating the severity of safety related behaviour“. In: *Accident Analysis & Prevention* 38.2 (2006), pp. 379–385 (cit. on p. 37).

[181]Hua Tang. „Development of a Multiple-Camera Tracking System for Accurate Traffic Performance Measurements at Intersections“. In: *Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota* (2013) (cit. on pp. 4, 28, 31, 48).

[182]Hua Tang and Hai Dinh. „A Tracking-Based Traffic Performance Measurement System for Roundabouts and Intersections“. In: *Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota* (2012) (cit. on pp. 26, 27).

[183]Christian Thiemann, Martin Treiber, and Arne Kesting. „Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data“. In: *Transportation Research Record: Journal of the Transportation Research Board* (2008) (cit. on p. 16).

[184]Carlo Tomasi and Takeo Kanade. *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991 (cit. on p. 25).

[185]Akihiko Torii, Atsushi Imiya, and Naoya Ohnishi. „Two-and three-view geometry for spherical cameras“. In: *Proceedings of the sixth workshop on omnidirectional vision, camera networks and non-classical cameras. Citeseer (cf. p. 81)*. Citeseer. 2005 (cit. on p. 72).

[186] Federal Highway Administration US Department of Transportations. *The National Intersection Safety Problem (FHWA-SA-10-005)*. http://goo.gl/DXC4JZ. Online; accessed 01-October-2015. 2009 (cit. on p. 2).

[187] Mohan M Trivedi, Tarak Gandhi, and Joel McCall. „Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety". In: *Intelligent Transportation Systems, IEEE Transactions on* 8.1 (2007), pp. 108–120 (cit. on p. 4).

[188] Quoc Bao Truong and Byung Ryong Lee. „Vehicle detection algorithm using hypothesis generation and verification". In: *Emerging Intelligent Computing Technology and Applications*. Springer, 2009, pp. 534–543 (cit. on p. 22).

[189] Luo-Wei Tsai, Jun-Wei Hsieh, and Kuo-Chin Fan. „Vehicle detection using normalized color and edge map". In: *Image Processing, IEEE Transactions on* 16.3 (2007), pp. 850–864 (cit. on p. 20).

[190] B. Ulmer. „VITA-an autonomous road vehicle (ARV) for collision avoidance in traffic". In: *Intelligent Vehicles '92 Symposium., Proceedings of the*. 1992, pp. 36–41 (cit. on p. 2).

[191] Vladimir Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2000 (cit. on p. 22).

[192] H. Veeraraghavan, O. Masoud, and N. Papanikolopoulos. „Vision-based monitoring of intersections". In: *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*. 2002, pp. 7–12 (cit. on pp. 18, 30).

[193] Harini Veeraraghavan and Nikolaos Papanikolopoulos. „Combining multiple tracking modalities for vehicle tracking at traffic intersections". In: *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*. Vol. 3. IEEE. 2004, pp. 2303–2308 (cit. on p. 27).

[194] Harini Veeraraghavan, Osama Masoud, and Nikolaos P Papanikolopoulos. „Computer vision algorithms for intersection monitoring". In: *Intelligent Transportation Systems, IEEE Transactions on* 4.2 (2003), pp. 78–89 (cit. on pp. 18, 27).

[195] Paul Viola and Michael Jones. „Rapid object detection using a boosted cascade of simple features". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2001, pp. I–511 (cit. on p. 22).

[196] Jung Ming Wang, Yun-Chung Chung, SC Lin, et al. „Vision-based traffic measurement system". In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 4. IEEE. 2004, pp. 360–363 (cit. on p. 19).

[197]Wei Wang, Tim Gee, Jeff Price, and Hairong Qi. „Real Time Multi-vehicle Tracking and Counting at Intersections from a Fisheye Camera“. In: *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*. IEEE. 2015, pp. 17–24 (cit. on pp. 18, 25, 26, 29, 33, 37, 42, 48, 55).

[198]Xiaogang Wang, Xiaoxu Ma, and W Eric L Grimson. „Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models“. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.3 (2009), pp. 539–555 (cit. on pp. 15, 16).

[199]Xuezhi Wen, Hong Zhao, Nan Wang, and Huai Yuan. „A rear-vehicle detection system for static images based on monocular vision“. In: *Control, Automation, Robotics and Vision, 2006. ICARCV'06. 9th International Conference on*. IEEE. 2006, pp. 1–4 (cit. on p. 21).

[200]KIT Intersection monitoring datasets in various wheather. http://goo.gl/wLlnIN. Online; accessed 01-October-2015 (cit. on pp. 15, 16).

[201]M. Williams. „The PROMETHEUS Programme“. In: *Towards Safer Road Transport - Engineering Solutions, IEE Colloquium on*. 1992, pp. 4/1–4/3 (cit. on p. 2).

[202]Chunpeng Wu, Lijuan Duan, Jun Miao, Faming Fang, and Xuebin Wang. „Detection of front-view vehicle with occlusions using AdaBoost“. In: *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*. IEEE. 2009, pp. 1–4 (cit. on p. 22).

[203]George Yannis, Constantinos Antoniou, and Petros Evgenikos. „Comparative analysis of junction safety in Europe“. In: *Proceedings of the 12th World Conference on Transport Research (WCTR)*. 2010 (cit. on p. 2).

[204]Alper Yilmaz, Omar Javed, and Mubarak Shah. „Object tracking: A survey“. In: *Acm computing surveys (CSUR)* 38.4 (2006), p. 13 (cit. on pp. 18, 23).

[205]Zhaozheng Yin and Robert Collins. „Belief propagation in a 3D spatio-temporal MRF for moving object detection“. In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE. 2007, pp. 1–8 (cit. on p. 20).

[206]Xianghua Ying and Zhanyi Hu. „Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model“. In: *Computer Vision-ECCV 2004*. Springer, 2004, pp. 442–455 (cit. on p. 51).

[207]Menghua Zhai, Scott Workman, and Nathan Jacobs. „Detecting vanishing points using global image context in a non-manhattan world“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 5657–5665 (cit. on p. 57).

[208] Guohui Zhang, Ryan P Avery, and Yinhai Wang. „Video-based vehicle detection and classification system for real-time traffic data collection using uncalibrated video cameras“. In: *Transportation Research Record: Journal of the Transportation Research Board* 1993.1 (2007), pp. 138–147 (cit. on p. 20).

[209] Hongyi Zhang, Andreas Geiger, and Raquel Urtasun. „Understanding high-level semantics by modeling traffic patterns“. In: *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE. 2013, pp. 3056–3063 (cit. on pp. 3, 27).

[210] Xuetao Zhang, Nanning Zheng, Yongjian He, and Fei Wang. „Vehicle detection using an extended hidden random field model“. In: *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE. 2011, pp. 1555–1559 (cit. on p. 20).

[211] Yan Zhang, Stephen J Kiselewich, and William A Bauson. „Legendre and Gabor moments for vehicle recognition in forward collision warning“. In: *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*. IEEE. 2006, pp. 1185–1190 (cit. on p. 22).

[212] Huijing Zhao, Jinshi Cui, Hongbin Zha, et al. „Sensing an intersection using a network of laser scanners and video cameras“. In: *Intelligent Transportation Systems Magazine, IEEE* 1.2 (2009), pp. 31–37 (cit. on pp. 5, 42).

[213] Quanwen Zhu, Long Chen, Qingquan Li, et al. „3d lidar point cloud based intersection recognition for autonomous driving“. In: *Intelligent Vehicles Symposium (IV), 2012 IEEE*. IEEE. 2012, pp. 456–461 (cit. on p. 14).

[214] Ying Zhu, Dorin Comaniciu, Martin Pellkofer, and Thorsten Koehler. „Reliable detection of overtaking vehicles using robust information fusion“. In: *Intelligent Transportation Systems, IEEE Transactions on* 7.4 (2006), pp. 401–414 (cit. on p. 19).

[215] Zoran Zivkovic, Ali Taylan Cemgil, and Ben Krose. „Approximate Bayesian methods for kernel-based object tracking“. In: *Computer Vision and Image Understanding* 113.6 (2009), pp. 743–749 (cit. on p. 23).