April 2017

Developing Predictive Models of Driver Behaviour for the Design of Advanced Driving Assistance Systems

Seyed Mohsen Zabihi
The University of Western Ontario

Supervisor
Dr. Steven Beauchemin
The University of Western Ontario

Graduate Program in Computer Science

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

World-wide injuries in vehicle accidents have been on the rise in recent years, mainly due to driver error. The main objective of this research is to develop a predictive system for driving maneuvers by analyzing the cognitive behavior (cephalo-ocular) and the driving behavior of the driver (how the vehicle is being driven). Advanced Driving Assistance Systems (ADAS) include different driving functions, such as vehicle parking, lane departure warning, blind spot detection, and so on. While much research has been performed on developing automated co-driver systems, little attention has been paid to the fact that the driver plays an important role in driving events. Therefore, it is crucial to monitor events and factors that directly concern the driver. As a goal, we perform a quantitative and qualitative analysis of driver behavior to find its relationship with driver intentionality and driving-related actions. We have designed and developed an instrumented vehicle (RoadLAB) that is able to record several synchronized streams of data, including the surrounding environment of the driver, vehicle functions and driver cephalo-ocular behavior, such as gaze/head information. We subsequently analyze and study the behavior of several drivers to find out if there is a meaningful relation between driver behavior and the next driving maneuver.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Steven Beauchemin, for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Robert Laganiere, Dr. Ken McIsaac, Dr. John Barron, and Dr. Marc Moreno Maza, for their encouragement, insightful comments, and hard questions.

I would like to thank my fellow labmates in RoadLAB: Jamal and Junaed, for the stimulating discussions, and for all the fun we have had in the last few years.

I would also like to express my gratitude to my parents, for supporting me spiritually throughout my life.

Finally, I must acknowledge my wife and best friend, Azadeh, without whose love, encouragement and patience, I would not have finished this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Safety in urban and rural driving is becoming an issue of growing concern since the number of vehicles and also vehicle accidents are increasing considerably. To overcome this issue, a variety of Advanced Driver Assistance Systems (ADAS), such as collision avoidance systems, speed control and warning systems, have been developed to assist in performing the complex driving task. These systems should pay attention to the important role of the human driver as a main element in most driving events and take into consideration all the factors that may concern the driver of the vehicle. Undoubtedly, improving these safety systems is a very important goal for the vehicle industry and governments in providing safer and more comfortable driving. These improvements can reduce the amount and severity of vehicle accidence as well as improve vehicle manufacturers' reputations with respect to vehicle safety. We describe the current literature in the areas of driver modeling and Driver maneuver prediction.

1.1 Literature Survey

A variety of driving assistance systems have been proposed in recent years and could profit from information related to the driver of the vehicle. Some systems analyze driver behavior to realize the driver's intentions in diverse driving situations. Especially in the area of safe driving, the driver and his behavior can demonstrate the driver's intentions. Most vehicle collisions are due to human errors caused by the driver (or drivers). An in-depth study of vehicle collisions carried out in the US in 1999 shows that distracted or inattentive drivers are the cause of more than 70% of the rear-end collisions recorded each year in the US [67]. This finding is backed by results from the 'Naturalistic Driving Study', a large field study conducted in the USA [41], which states that about 80% of collisions are due to distraction. It is obvious that distractions caused by recent in-vehicle devices, such as GPS maps, entertainment systems and mobile telephones, increase a driver's accident risk [57].

Almost every driver has experienced a warning from a passenger, perhaps alerting him or her to prevent an accident with another vehicle or a pedestrian. An intelligent ADAS (i-ADAS) is also trying to work as a co-driver to prevent these accidents by warning the driver or even taking control of the vehicle. It is actually a computer system inside a vehicle that can collaborate with the driver to manage the driving tasks. Detecting the state of the driver (his behavior) and the vehicle (environment), could hence provide valuable information for the system in assisting the driver effectively in different situations.

The main questions in this area are how and what kind of information about the driver could be used in the optimization of driver assistance systems. The field of research engaged in describing the driver and his driving behavior is called driver modeling.

1.1.1 Driver modeling

Driver modeling, as the name suggests, tries to model driver behavior in driving situations. Driver modeling is a very broad field and much work has been done in a variety of research areas for the past 60 years. Plöchl and Edelmann [45] provide an overview of different driving models. They divided the models into four categories, which to some extent can also be seen as a temporal development of the field of driver modeling, fueled by increasing vehicle complexity and computational power:

1. **Focus on the vehicle.** the vehicle is the main goal of these models. In this case, the driver model usually serves as part of the closed loop testing of vehicle performance under various driving conditions.

2. **Focus on the driver.** In this case the driver is the target. This includes efforts to model aspects, such as driving style or psychological states while driving, such as stress or distraction.

3. **Focus on the vehicle/driver combination.** It is a combination of the two previous areas i.e. the interaction between driver and vehicle.

4. **Focus on the environment/traffic.** These models simulate traffic conditions. The main goal is to understand how certain traffic situations emerge and how undesirable ones can be prevented.

In this work, we concentrate on the driver and his behavior to develop a predictive model of driver's intentionality and maneuvers. For this reason, the next section is devoted to the area focusing on the driver.

1.1.2 Driver models focusing on the driver

The goal of this section is describing driver modeling in the context of this work. Plöchl and Edelmann [45] divide the research area interested in the understanding of human driving behavior into two groups:

- Understanding driver and (individual) driver behavior
- Path and speed planning, optimized driver/driving behavior

The first group is related to modeling how we drive a vehicle. The main goal is to figure out how a driver does maneuvers. Driver behavior can be predicted by knowing how a driver acts in certain driving tasks, such as turning, lane changing and overtaking a vehicle. Chandler et al. [12] implemented a vehicle-following model and used data from experiments with real vehicles to set the model's parameters. Ioannou [21] implemented an Adaptive Cruise Control (ACC) system and compared it to three different human driver models. The results showed that the ACC system was able to provide a safer driving. A hybrid three-layered ACC system based on fuzzy logic and neural network has been implemented by Germann in 1995 [14].

Some other maneuvers have been also studied: emergency braking ([53],[52]), lane change ([38], [30]), and turning at intersections ([13],[11]). Predicting the driver's intention could provide valuable information in the context of ADAS, as will be discussed later.

A number of methods have been implemented for assessing the driver's sleepiness. Most of them utilize some features in the driver's face through video cameras and use eye [10],[42],[24], eyelid [10],[51] or head movements [25], [10],[24] to infer sleepiness. A few indirect algorithms, usually using steering behavior in the context of lane keeping [50], [51], have also been developed.

It is shown that distraction is one of the main factors in driving accidents. Accordingly, detecting that a driver is distracted can be very helpful and prevent some dangerous accidents. As mentioned before, the 'Naturalistic Driving Study' concluded that about 80% of collisions were caused by distraction. Some work has been done on detecting the driver's distraction by a video camera capturing the driver's upper body [17],[56],[31]. Cell-phone usage [49] and the on-board information systems, such as radio or GPS [60] may cause distraction.

The second group is related to more complicated maneuvers on a larger timescale than single maneuvers. At this level, driver model is formed based on a series of single maneuvers. For example, overtaking a vehicle consisting not only changing lanes to pass a preceding vehicle, but joining the original lane afterwards. Some methods have been presented in this context to predict the driver's intent [29],[55].

1.1.3 Advanced driver assistance systems

Advanced driver assistance systems (ADAS), which is the main target of driver modeling, are proposed to provide safety for drivers in different driving situations. We give an overview of ADAS systems in this section in order to understand the relationship between driver modeling and these systems.

These systems may use a variety of information, including the vehicle itself (the vehicles velocity, signals, steering angle, etc.), a navigation system or sensors that detect obstacles in the environment. For a systematic overview, this large number of systems can divide to four groups [7]:

- Driver information systems
- Comfort systems

- Semiautonomous systems
- Autonomous systems

These groups will be explained in more detail, describing the current available systems in each group.

Driver information systems

Driver information systems present required information to the driver in driving scenes by auditory and/or visual equipment. Some systems, such as Lane Departure Warning (LDW) or Collision Warning (CW) fall into this group. LDW systems warns the driver if the vehicle is changing the current lane. These systems work based on the detection of the lane, which can be performed by evaluating data from video or infrared sensors monitoring the front view of the vehicle. Nowadays, plenty of vehicle manufacturers offer a LDW system or the more complex Lane Keeping Assist (LKA) system. The goal of CW systems is to inform the driver in the matter of an approaching accident. This driver information is usually the first in a chain of assistance systems, ending up in a automatic operation of the vehicles brake regarding a failure in the response from the driver. Fig. 1.1 depicts a current available case for such a chain published by Mercedes Benz[1] in 2008. Both could profit from models describing the driver and his behavior.

Comfort systems

Comfort systems assist the driver in the case of commonplace driving tasks by taking control of some parts of vehicle. Some systems in this group are

[1] http://www.emercedesbenz.com/Nov08/12_001507_Mercedes_Benz_TecDay_Special_Feature_PRE_SAFE_And_PRE_SAFE_Brake.html

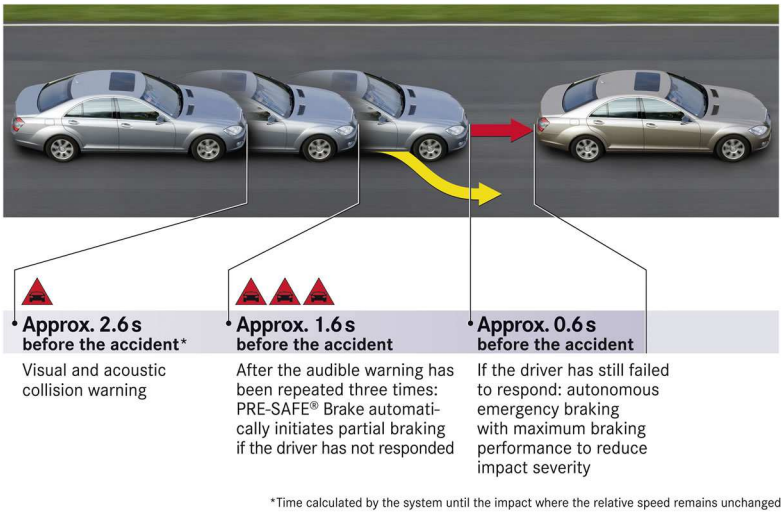


Figure 1.1: *Activation chain of a Collision Warning / Collision Mitigation system.*

ACC, Lane Keeping Assist (LKA) or Active Steering (AS). ACC is well known among the systems which are used in numerous vehicles. This system adjusts the vehicle velocity according to the current traffic and the distance from preceding vehicles. LKA systems employ a small force on the steering wheel to delay or prevent the lane crossing.

Semiautonomous systems

Semi-autonomous systems help the driver in performing complicated maneuvers by controling the vehicle in some cases. Drivers make the last decision and take control of the vehicle in these systems in contrary to the autonomous systems that take control of the system and provide vehicle's safety in some situations. Intelligent Park Assist (IPA) system is an instance for semi-autonomous systems. The system partially takes over control of the steering wheel.

Autonomous systems

Autonomous systems, as the name indicates, take control of the vehicle in specific situations without help from the driver. These systems carry the full responsibility for their actions, so they need to be very specific about the current situation and the state of the vehicle. The goal of autonomous systems is, one day, to design a fully autonomous vehicle and provide safety in all driving situations.

The Antilock Braking System (ABS) was the first assistance system used in commercial vehicles[1]. This system makes the vehicle stay on track during a braking maneuver, even under adverse conditions, such as icy or wet roads. This is done by adjusting the hydraulic brake pressure at each wheel (or front and rear wheels) separately.

1.1.4 Driver maneuver prediction

Predicting driving maneuvers is the main goal of driver modeling in advanced driver assistance systems.

Driving maneuvers

A specific move or series of moves in driving will be called a maneuver. Driving maneuvers can be defined based on traffic and road infrastructure [7]. Reichart [47] and Tolle [62] classified the driving maneuvers, which are listed in Table 1.1. These two classifications are almost similar at some level of granularity with just minor differnces. The proposed maneuvers are enough to do any urban or rural trip as well as driving on highways. The other maneuver classifications are proposed in different contexts are almost similar to the ones

[1]http://www.bosch.com/assets/en/company/innovation/theme03.htm

showed here. For example, Okuno et al. [43] discuss diverse maneuvers taken place on highways.

Table 1.1: Classification of driving maneuvers provided by [47] and [43]

Reichart	**Tölle**
Follow lane	Start
React to obstacle	Follow
Turn at intersection	Approach vehicle
Cross intersection	Overtake vehicle
Turn into street	Cross intersection
Change lane	Change lane
Turn around	Turn at intersection
Drive backwards	Drive backwards
Choose velocity	Park
Follow vehicle	

To model the behavior of a driver accurately, it is necessary to go one step further since people may do the same maneuvers differently in different situations, according to the level of granularity. Therefore, the maneuvers should be partitioned based on, for example, direction (left turn vs. right turn) or region (a lane change in a city vs. one on highways). Other factors, such as weather conditions, traffic flow, number of lanes and etc. can be considered. Therefore, a maneuver is a tactical driving task, such as stopping at a traffic light, changing the lane or turning at an intersection. These maneuvers can be differentiated by the situational factors, such as type of road, speed limit, number of lanes and the existence of other vehicles, pedestrians, traffic signs or traffic lights ahead of the vehicle.

Maneuver prediction models

After defining a driving maneuver, we should determine what a driver model is to be developed to predict the maneuvers. Predicting driving maneuvers involves creating a model of human behavior, so it is a challenging task. A large number of internal and environmental factors can affect human behavior [7]. These factors include emotional factors, such as stress and anger, physical characteristics, environmental conditions, such as bad weather and traffic conditions, distraction, driving skills and so on. Building an effective and complete model based on all these factors is difficult and also impossible in practice. In fact, the proposed methods in the literature have used a subset of above factors. A hybrid driver model has been described in [26] that is able to predict driver behavior. The proposed model is based on visual and vestibular information so that control both the lateral and longitudinal motions of a vehicle.

The driver behavior models can be grouped into two categories [7]. The first is the cognitive driver model, which is based on human psychological behaviors. These models pay attention to, for example, mental and attentional features that a driver shows in performing different maneuvers. The models in the next category which are called behaviorist driver models and they utilize information in the surrounding environment of the driver, including vehicles, pedestrians and the the other objects on road. The most cited methods in this literature are based on one of these two categories and it is clear that each group of these models has its own deficiencies. Evidently, a combination of information in both categories can be more helpful and practical in understanding driver behavior and predicting the most probable next maneuver.

As mentioned above, cognitive driver models deal with human information

processing, such as memory and visual information. The psychological factors can be incorporated in cognitive driver behavior modeling include distraction, reaction time, body strength, vision, impairment (stress, fatigue, alcohol, etc.) and so on [20]. Metari et al. [39] analyzed the cephalo-ocular behavior of drivers in vehicle/road events, such as overtaking and crossing an intersection. This is specially concerned with finding the relationship between vision behavior of older drivers and their actions in road events. They posited the cephalo-ocular information can be related to the driving maneuvers a driver decides to make. Some other researchers have also tested the impact of eye movements on control of actions in everyday events, such as driving [32], [34].

Analysis of driver behavior in a cognitive architecture is very important for determining the motivation behind making a special decision in a driving event [48]. For example, when a driver decides to make a left or right turn, visual information can show that he looks at the mirrors, blind spots, traffic and so on. So, a faster and safer action can be chosen or offered by an ADAS to assist the driver during driving. Baumann and Krems [8] model the situation awareness of a driver in a cognitive architecture.

Despite the fact that cognitive behaviors specifically model visual information, such as gaze tracking, what are important for predicting driver behavior, little research has been done in this domain. The main reason that these models have not received much attentions could be the complexity that the systems have in measuring and providing cognitive behavior information.

On the other hand, behaviorist driver models try to find how the driver interacts with his surrounding environment including the other vehicles and also his vehicle parts, such as the brake and the accelerator pedals, the turn signals and the steering wheel. Modern vehicles are already equipped with some cameras and sensors to measure the internal vehicular information. Radar

systems [1] for detecting distance, lidar systems [36] for obstacle detection, visual systems [66] for road objects detection and vehicle navigation systems, such as GPS [35] have been used in advanced driving assistance systems.

1.2 Research Overview

The main objective in this research is to develop predictive models of driver behavior for the design of advanced automated driving assistance systems for road vehicles. The automated co-driver, as an intelligent system, would be able to alert the driver about an unseen object, such as a vehicle, pedestrian, or any other obstacle, or even take control of the vehicle in some cases when the driver is momentarily incapacitated. The Advanced Driver Assistance System (ADAS) should understand the driver's behavior and intent in different situations sufficiently enough to determine what is the most probable next maneuver and intervene if necessary.

ADAS needs to receive information from the vehicle, road, and environment to put the behavior of the driver in context. Therefore, the system should locate the road ahead of the vehicle, pedestrians and obstacles as well as monitor the behavior of the driver. Developing models of driver intentionality prediction based on such information may lead to advances in providing safety-related vehicular technologies and accident prevention strategies.

1.2.1 Primary Conjecture

Studies of driver behavior models have shown drivers' intentionality can be determined from various perspectives. Advanced driver assistance systems assist the driver in cases such as distraction and tiredness. However, the most

effective solution may be the one that monitors and analyzes driver behavior in order to prevent unsafe maneuvers [16]. Eyes play an important role in detecting driver's intentionality since the driver looks directly at objects and chooses an action based on the information provided by the fixation [6]. According to observations, we believe correlating visual behavior (gaze information, saccades, etc.) and driver environment state (vehicles, pedestrians, traffic signs, etc.) can yield a predictive model of driver behavior.

1.2.2 Hypotheses

It is believed that human error is the primary factor in most vehicle accidents [64]. Consequently, we focus on some hypotheses from which the short term goals of this proposal are derived from the primary conjecture. Towards this aim, we break down the main conjecture into hypotheses which can be investigated effectively and practically:

1. *Driver intentionality and driving actions can be predicted partly by Cephalo-ocular behavior:* Driving actions may be involved with complex cephalo-ocular behaviors. It is clear that ocular behavior cannot directly predict all driving actions since vehicle functions (such as velocity, steering wheel angle, turning signals) and current driving actions should be taken into consideration. Therefore, both current actions and ocular behaviors may be predictive of the next driving action.

2. *Objects detected within the attentional visual area of drivers can reveal a driver's intent:* The attentional visual area of drivers is a central part of safe driving. The probability of unsafe lane departure and vehicle accidents increases when the duration of glances away from the road increases

[33]. As a matter of fact, objects detected in the visual field of a driver may have an impact on driver intent. However, every visual fixation does not mean the driver is paying attention to the event carefully since long fixation may be synonymous with distraction in some cases. Shinar [54] posited that certain ocular patterns, such as fixation with regular saccades (eye movements), demonstrate visual attention which mostly reveals driver intent. Therefore, it is essential to analyze relevant and meaningful cephalo-ocular behaviors of drivers in order to understand driver intent in such a way as to predict it.

3. *Some vehicle accidents may be due to unseen, important information in the visual area the driver does not pay attention to:* It is generally accepted that the human central field of view (the pitch and yaw angles of gaze direction) is $\pm 6.5^{\circ}$ [58]. Consequently, a driver cannot pay attention to the whole surrounding environment and misses some information that affects the choice of a proper action.

Investigation of these hypotheses will extend the current knowledge of cognitive driver models as well as the development of a predictive driving behavior model capable of assisting the driver effectively.

1.2.3 Methodology

To verify the mentioned hypothesis, a modern vehicle with OBD II CAN-bus channels has been equipped with RoadLAB [9]. As Fig. 1.2 shows, this instrumented vehicle can monitor and record the following:

1. front view of the vehicular environment with calibrated stereo cameras;

Figure 1.2: *Vehicular instrumentation configuration in RoadLAB.*

2. the internal network of the vehicle, such as odometry and vehicle functions;

3. the driver cephalo-ocular behavior including head/eye rotation, gaze information, and saccades.

All of these items have been included in sixteen driving sequences for sixteen drivers. These sequences have been recorded with the RoadLAB experimental vehicle in an urban area of London Ontario, Canada and include all the aforementioned information, such as gaze/head information, GPS data and vehicle odometry, and wide angle views of road in real time. Item 1 is addressed with a combination of calibrated, short and long-range stereo cameras and software designed for monitoring driving conditions at the front and sides of the vehicle. Our choice of passive sensors is motivated by the fact that data acquisition is non-invasive and provides information conveyed by visual elements, such as road surface markings and signs, which are critical to the task of driving and yet unavailable to active sensors, such as millimeter-wave radars or laser range finders. Item 2 is addressed with an interface between the CANbus channels of the vehicle and the on-board computer to capture vehi-

cle odometry (independent wheel rotation, acceleration) and partial elements of the driver behavior (operation of steering system, accelerator and brake pedals, turn signals, entertainment devices, electrical windows, and other vehicular equipment). Item 3 is addressed by instrumenting the vehicle with a tracking system measuring relevant cephalo-ocular parameters. We are using FaceLAB, a commercialized gaze and head tracking system, to detect the gaze and head pose. The cross calibration between the image acquisition stereo systems and FaceLAB employs a new algorithm developed as a part of the project.

The vehicular instrumentation provides Driver-Environment-Vehicle (DEV) data streams for which a functional parametrization is required [4]. A behavioral driving agent model relating the cognitive state of the driver (the cephalo-ocular behavior, in particular), the vehicle odometry, and the vehicular environment is inferred in the form of Real-Time Descriptors(RTD) consisting of:

1. a *Contextual Feature Set* (CFS), descriptive of nearby vehicles, lanes and road boundaries, pedestrians, and other commonly encountered elements, as obtained and located with on-board sensors and real-time vision processes;

2. the *Cognitive State of the Driver* (CSD), descriptive of driving maneuvers, usage of vehicle equipment, acknowledged elements within the CFS (by way of intersecting driver 3D gaze direction with elements of the CFS), and level of attention;

3. the *Vehicle State and Odometry* set (VSO), descriptive of current speed, acceleration, position of steering wheel, state of braking system, and other relevant attitudinal information.

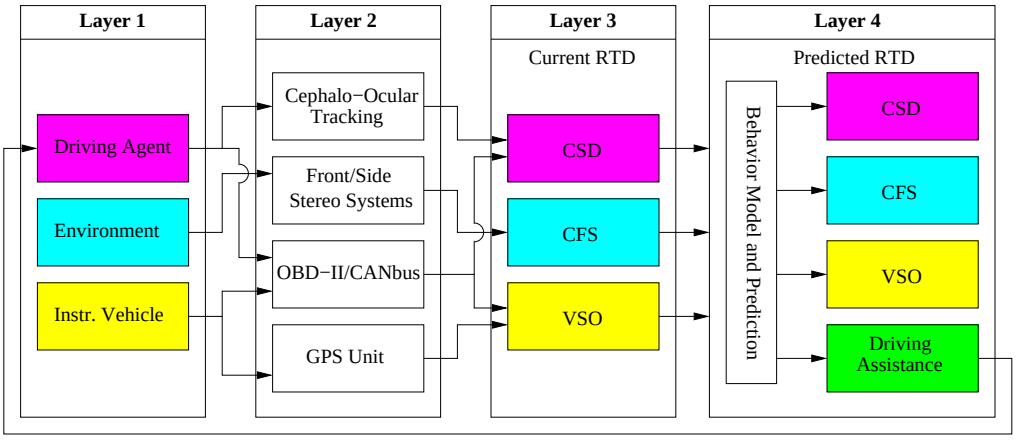


Figure 1.3: *A description of the DEV parametrization within software layers*

We further structure the RTDs in layers of increasing data abstraction (from device drivers to real-time data analysis), and integrate RTDs within a retroactive mechanism (see Figure 1.3) in which both current and predicted RTDs assist in determining the state of the DEV data set, yielding the structure required to test the validity of research hypotheses regarding driver intentionality and action prediction. Research on the parametrization of the DEV set with respect to the creation of effective driver behavior prediction models is currently ongoing.

1.3 Contributions

This thesis is an inherent part of the RoadLAB research program, instigated by Professor Steven Beauchemin, and is entirely concerned with vehicular instrumentation for the purpose of the study of driver intent. Chapters 2, 3, 4, and 5 have been published (or in the process of being) in recognized peer-reviewed venues. In what follows I describe my contributions with regards to each publication within the thesis:

1. Chapter 2: S.M Zabihi, S.S. Beauchemin, G. de Medeiros, and M.A. Bauer, *Frame-Rate Vehicle Detection within the Attentional Visual Area of Drivers, IEEE Intelligent Vehicles Symposium (IV14)*, Dearborn, MI, USA, pp. 146-150, June 8-11 2014.

    - After initial discussions with Professor Beauchemin about the attentional visual area of the driver and the importance of detecting relevant objects located within the attentional visual area of drivers, I developed a technique that is the first of its kind in that it identifies vehicles the driver is most likely to be aware of at any moment while in the act of driving.

2. Chapter 3: S.J. Zabihi, S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer, *Detection and Recognition of Traffic Signs Inside the Attentional Visual Field of Drivers, Submitted to, IEEE Intelligent Vehicles Symposium (IV17)*, Redondo Beach, California, USA, June 11-14 2017.

    - Jamal, researcher at RoadLab, and I decided to extend my vehicle detection approach to detect traffic signs on driving scenes as well. For this purpose, we developed a system to detect and recognize

traffic signs inside the the attentional field of drivers. The system is also able to infer whether the driver was likely to have seen the sign or not based on computing the intersection of the detected bounding box and drivers gaze area.

3. Chapter 4: S.M Zabihi, S.S. Beauchemin, G. de Medeiros, and M.A. Bauer, *Lane-Based Vehicle Localization in Urban Environments, accepted in IEEE International Conference on Vehicular Electronics and Safety (ICVES15)*, Yokohama, Japan, November 5-7 2015.

    - Since GPS is challenged in urban environments where satellite visibility and multipath situations are unavoidable, I proposed a method by which vehicular speed and a map-based lane detection process are called upon to improve the positional accuracy of GPS.

4. Chapter 5: S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer, *Real-Time Driving Manoeuvre Prediction Using IO-HMM and Driver Cephalo-Ocular Behaviour, Submitted to , IEEE Intelligent Vehicles Symposium (IV17)*, Redondo Beach, California, USA, June 11-14 2017.

    - For the purpose of developing an Advanced Driver Assistance System, I developed a model of driver behaviour for turn manoeuvres that I then apply to anticipate the most likely turn manoeuvre a driver will effect a few seconds ahead of time.

1.4 Thesis Organization

The organization of the thesis is as follows, in Chapter 2 and 3, contributions related to vehicles and traffic signs detected within the attentional visual area

of drivers, are presented. In chapter 4, we explain our lane-based vehicle localization method. The approach improves the positional accuracy of GPS. Chapter 5 provides the main contribution to this research, which consists of developing a prediction model to anticipate the most likely manoeuvre a driver will effect a few seconds ahead of time. Finally, chapter 6 offers a conclusion and outlines paths for future research.

Bibliography

[1] Ramzi Abou-Jaoude. Acc radar sensor technology, test requirements, and test solutions. *Intelligent Transportation Systems, IEEE Transactions on*, 4(3):115–122, 2003.

[2] Gabriel Agamennoni, Stewart Worrall, Jon R. Ward, and Eduardo M. Neboty. Automated extraction of driver behaviour primitives using bayesian agglomerative sequence segmentation. In *Intelligent Transportation Systems (ITSC), IEEE 17th International Conference on*, pages 1449–1455. IEEE, 2014.

[3] Hideomi Amata, Chiyomi Miyajima, Takanori Nishino, Norihide Kitaoka, and Kazuya Takeda. Prediction model of driving behavior based on traffic conditions and driver types. In *Intelligent Transportation Systems, ITSC'09. 12th International IEEE Conference on*, pages 1–6. IEEE, 2009.

[4] Angelos Amditis, Katia Pagle, Somya Joshi, and Evangelos Bekiaris. Driver–vehicle–environment monitoring for on-board driver support systems: Lessons learned from design and implementation. *Applied ergonomics*, 41(2):225–235, 2010.

[5] George J. Andersen. Focused attention in three-dimensional space. *Perception & Psychophysics*, 47(2):112–120, 1990.

[6] Dana H. Ballard, Mary M. Hayhoe, and Jeff B. Pelz. Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1):66–80, 1995.

[7] Colin Bauer. *A driver specific maneuver prediction model based on fuzzy logic*. PhD thesis, Freie Universität Berlin, 2012.

[8] Martin RK. Baumann and Josef F. Krems. A comprehension based cognitive model of situation awareness. In *Digital Human Modeling*, pages 192–201. Springer, 2009.

[9] Steven S. Beauchemin, Michael A. Bauer, Taha Kowsari, and Ji Cho. Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality. *Instrumentation and Measurement, IEEE Transactions on*, 61(2):391–401, 2012.

[10] Luis Miguel Bergasa, Jesús Nuevo, Miguel A Sotelo, Rafael Barea, and María Elena Lopez. Real-time system for monitoring driver vigilance. *Intelligent Transportation Systems, IEEE Transactions on*, 7(1):63–77, 2006.

[11] Holger Berndt and Klaus Dietmayer. Driver intention inference with vehicle onboard sensors. In *Vehicular Electronics and Safety (ICVES), 2009 IEEE International Conference on*, pages 102–107. IEEE, 2009.

[12] Robert E. Chandler, Robert Herman, and Elliott W. Montroll. Traffic dynamics: studies in car following. *Operations research*, 6(2):165–184, 1958.

[13] Shinko Yuanhsien Cheng and Mohan M. Trivedi. Turn-intent analysis using body pose for intelligent driver assistance. *Pervasive Computing, IEEE*, 5(4):28–37, 2006.

[14] S. Dermann and R. Isermann. Nonlinear distance and cruise control for passenger cars. In *Proceedings of the American Control Conference*, volume 5, pages 3081–3085. IEEE, 1995.

[15] Ueruen Dogan, Hannes Edelbrunner, and Ioannis Iossifidis. Towards a driver model: Preliminary study of lane change behavior. In *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, pages 931–937. IEEE, 2008.

[16] Birsen Donmez, Linda Ng. Boyle, and John D. Lee. Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis & Prevention*, 39(3):581–590, 2007.

[17] Tiziana DOrazio, Marco Leo, Cataldo Guaragnella, and Arcangelo Distante. A visual approach for driver inattention detection. *Pattern Recognition*, 40(8):2341–2355, 2007.

[18] Sean R. Eddy. Hidden markov models. *Current opinion in structural biology*, 6(3):361–365, 1996.

[19] Yoav Freund and Robert E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

[20] Samer Hamdar. Driver behavior modeling. In *Handbook of Intelligent Vehicles*, pages 537–558. Springer, 2012.

[21] Petros A. Ioannou and Cheng-Chih C. Chien. Autonomous intelligent cruise control. *Vehicular Technology, IEEE Transactions on*, 42(4):657–672, 1993.

[22] Anil K Jain, Jianchang Mao, and K Moidin Mohiuddin. Artificial neural networks: A tutorial. *IEEE computer*, 29(3):31–44, 1996.

[23] Finn V. Jensen. *An introduction to Bayesian networks*, volume 210. UCL press London, 1996.

[24] Qiang Ji, Zhiwei Zhu, and Peilin Lan. Real-time nonintrusive monitoring and prediction of driver fatigue. *Vehicular Technology, IEEE Transactions on*, 53(4):1052–1068, 2004.

[25] Shanshan Jin, So-Youn Park, and Ju-Jang Lee. Driver fatigue detection using a genetic algorithm. *Artificial Life and Robotics*, 11(1):87–90, 2007.

[26] U. Kiencke, R. Majjad, and S. Kramer. Modeling and performance analysis of a hybrid driver model. *Control Engineering Practice*, 7(8):985–991, 1999.

[27] T. Kowsari, S.S. Beauchemin, and J. Cho. Real-time vehicle detection and tracking using stereo vision and multi-view adaboost. In *Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on*, pages 1255–1260. IEEE, 2011.

[28] KF. Kraiss and H. Kuttelwesch. Identification and application of neural operator models in a car driving situation. In *Neural Networks, IJCNN-91-Seattle International Joint Conference on*, volume 2, pages 917–vol. IEEE, 1991.

[29] John Krumm. A markov model for driver turn prediction. *SAE SP*, 2193(1), 2008.

[30] Nobuyuki Kuge, Tomohiro Yamamura, Osamu Shimoyama, and Andrew Liu. A driver behavior recognition method based on a driver model framework. *SAE transactions*, 109(6):469–476, 2000.

[31] MH. Kutila, M. Jokela, T. Mäkinen, J. Viitanen, G. Markkula, and TW. Victor. Driver cognitive distraction detection: Feature estimation and

implementation. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 221(9):1027–1040, 2007.

[32] Michael F. Land. Eye movements and the control of actions in everyday life. *Progress in retinal and eye research*, 25(3):296–324, 2006.

[33] Yi-Ching Lee, John D. Lee, and Linda Ng. Boyle. Visual attention in driving: the effects of cognitive load and visual disruption. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(4):721–733, 2007.

[34] F. Lethaus and J. Rataj. Do eye movements reflect driving manoeuvres. *IET Intelligent Transport Systems*, 1(3):199–204, 2007.

[35] J.P. Löwenau, P.J. Venhovens, and J.H. Bernasch. Advanced vehicle navigation applied in the bmw real time light simulation. *Journal of Navigation*, 53(01):30–41, 2000.

[36] Meng Lu, Kees Wevers, and Rob Van Der Heijden. Technical feasibility of advanced driver assistance systems (adas) for road traffic safety. *Transportation Planning and Technology*, 28(3):167–187, 2005.

[37] Charles C. MacAdam and Gregory E. Johnson. Application of elementary neural networks and preview sensors for representing driver steering control behaviour. *Vehicle System Dynamics*, 25(1):3–30, 1996.

[38] Joel C. McCall, David P. Wipf, Mohan M. Trivedi, and Bhaskar D. Rao. Lane change intent analysis using robust operators and sparse bayesian learning. *Intelligent Transportation Systems, IEEE Transactions on*, 8(3):431–440, 2007.

[39] Samy Metari, Florent Prel, Thierry Moszkowicz, Denis Laurendeau, Normand Teasdale, Steven Beauchemin, and Martin Simoneau. A computer vision framework for the analysis and interpretation of the cephalo-ocular behavior of drivers. *Machine vision and applications*, 24(1):159–173, 2013.

[40] Dejan Mitrovic. Machine learning for car navigation. In *Engineering of Intelligent Systems*, pages 670–675. Springer, 2001.

[41] Vicki L. Neale, Thomas A. Dingus, Sheila G. Klauer, Jeremy Sudweeks, and Michael Goodman. An overview of the 100-car naturalistic study and findings. *National Highway Traffic Safety Administration*, 2005.

[42] Yoshihiro Noguchi, Roongroj Nopsuwanchai, Mieko Ohsuga, and Yoshiyuki Kamakura. Classification of blink waveforms towards the assessment of drivers arousal level-an approach for hmm based classification from blinking video sequence. In *Engineering Psychology and Cognitive Ergonomics*, pages 779–786. Springer, 2007.

[43] Akihiro Okuno, Kenji Fujita, and Atsushi Kutami. Visual navigation of an autonomous on-road vehicle: autonomous cruising on highways. In *Vision-based vehicle guidance*, pages 222–237. Springer, 1992.

[44] Nuria Oliver and Alex P. Pentland. Graphical models for driver behavior recognition in a smartcar. In *Intelligent Vehicles Symposium. Proceedings of the IEEE*, pages 7–12. IEEE, 2000.

[45] Manfred Plöchl and Johannes Edelmann. Driver models in automobile dynamics application. *Vehicle System Dynamics*, 45(7-8):699–741, 2007.

[46] D. Polling, Max Mulder, Marinus Maria van Paassen, and QP. Chu. Inferring the driver's lane change intention using context-based dynamic

bayesian networks. In *Systems, Man and Cybernetics, IEEE International Conference on*, volume 1, pages 853–858. IEEE, 2005.

[47] Günter Reichart. *Menschliche zuverlässigkeit beim führen von kraftfahrzeugen*. VDI-Verlag, 2001.

[48] Dario D. Salvucci. Modeling driver behavior in a cognitive architecture. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2):362–380, 2006.

[49] Dario D. Salvucci and Kristen L. Macuga. Predicting the effects of cellular-phone dialing on driver performance. *Cognitive Systems Research*, 3(1):95–102, 2002.

[50] David Sandberg and Mattias Wahde. Particle swarm optimization of feedforward neural networks for the detection of drowsy driving. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 788–793. IEEE, 2008.

[51] Deng Sanpeng, Xu Xiaoli, Yang Xuecui, and Miao Dehua. Research on the driver fatigue monitoring method based on the dempster-shafer theory. In *Control and Decision Conference (CCDC)*, pages 4176–4179. IEEE, 2010.

[52] J. Schmitt, A. Breu, M. Maurer, and B. Farber. Simulation des bremsverhaltens in gefahrensituationen mittels experimentell validiertem fahrermodell. *VDI BERICHTE*, 2015:75, 2007.

[53] J. Schmitt and B. Färber. Verbesserung von fas durch fahrerabsichtserkennung mit fuzzy logic. *VDI-Berichte*, 2015(1919), 2005.

[54] David Shinar. Looks are (almost) everything: where drivers look to get information. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):380–384, 2008.

[55] Reid Simmons, Brett Browning, Yilu Zhang, and Varsha Sadekar. Learning to predict driver route and destination intent. In *Intelligent Transportation Systems Conferenc. ITSC'06. IEEE*, pages 127–132. IEEE, 2006.

[56] Paul Smith, Mubarak Shah, and Niels da Vitoria Lobo. Determining driver visual attention with one camera. *Intelligent Transportation Systems, IEEE Transactions on*, 4(4):205–218, 2003.

[57] Jane C. Stutts, American Automobile Association, et al. *The role of driver distraction in traffic crashes*. Foundation for Traffic Safety Washington, DC, 2001.

[58] Kenji Takagi, Haruki Kawanaka, Md. Shoaib Bhuiyan, and Koji Oguri. Estimation of a three-dimensional gaze point and the gaze target from the road images. In *Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on*, pages 526–531. IEEE, 2011.

[59] Wataru Takano, Akihiro Matsushita, Keijiro Iwao, and Yoshihiko Nakamura. Recognition of human driving behaviors based on stochastic symbolization of time series signal. In *Intelligent Robots and Systems. IEEE/RSJ International Conference on*, pages 167–172. IEEE, 2008.

[60] Fabio Tango, Luca Minin, Francesco Tesauri, and Roberto Montanari. Field tests and machine learning approaches for refining algorithms and

correlations of drivers model parameters. *Applied ergonomics*, 41(2):211–224, 2010.

[61] Shigeki Tezuka, Hitoshi Soma, and Katsuya Tanifuji. A study of driver behavior inference model at time of lane change using bayesian networks. In *Industrial Technology, 2006. ICIT 2006. IEEE International Conference on*, pages 2308–2313. IEEE, 2006.

[62] W. Tölle. *Ein Fahrmanöverkonzept für einen maschinellen Kopiloten*. PhD thesis, Universität Karlsruhe, 1996.

[63] Jack V. Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11):1225–1231, 1996.

[64] Yoshiyuki Umemura. Driver behavior and active safety (overview). *R&D Review of Toyota CRDL*, 39(2):1–8, 2004.

[65] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition (CVPR). Proceedings of IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

[66] Ljubo Vlacic, Michel Parent, and Fumio Harashima. *Intelligent vehicle technologies: theory and applications*. Butterworth-Heinemann, 2001.

[67] Christopher J. Wiacek and Wassim G. Najm. Driver/vehicle characteristics in rear-end precrash scenarios based on the general estimates system (ges). *SAE Technical Paper*, 1999.

Chapter 2

Vehicle Detection

This Chapter is a reformatted version of the following article:

S.M. Zabihi, S.S. Beauchemin, E.A.M. de Medeiros and M.A. Bauer, *Frame-Rate Vehicle Detection within the Attentional Visual Area of Drivers, IEEE Intelligent Vehicles Symposium (IV14)*, Dearborn, MI, USA, pp. 146-150, June 8-11 2014.

This contribution consists of a frame-rate, vision-based, on-board method which detects vehicles within the attentional visual area of the driver. The method herein uses the 3D absolute gaze point of the driver obtained through the combined use of a front-view stereo imaging system and a non-contact 3D gaze tracker, alongside hypothesis-generation reducing techniques for vehicular detection such as horizon line detection. Trained AdaBoost classifiers are used in the detection process. This technique is the first of its kind in that it identifies vehicles the driver is most likely to be aware of at any moment while in the act of driving.

2.1 Introduction

It has been known for some time that driver ocular movements often provide information on the very next maneuver to be effected [2], such as in the case of negotiating a road bend [4]. Our research program focuses on the development of intelligent, Advanced Driving Assistance Systems (i-ADAS) which include the driver as a pivotal element of driving error prediction and correction.

2.1.1 Context

The underlying principle underpinning this research is to determine to what extent current maneuvers and ocular behavior are predictive of the next maneuver and also whether this predicted maneuver is consonant with the current driving context (other vehicles, pedestrians, signage, and lanes, among others). Our current vision for i-ADAS is that of intervening (by a switch to autonomous driving) only when the next predicted maneuver poses a measurable traffic risk. Vehicular control is then returned once the danger has abated and the driver is capable of correctly controlling the vehicle. Such an i-ADAS, once realized, would certainly significantly reduce vehicle accidents, injuries, and fatalities, without removing the driver from the act of commuting.

The complexities involved in such an endeavor require many stages of research, such as the study of driver behavior as it pertains to ocular movements and current maneuvers, the precise detection and understanding of traffic context surrounding the vehicle, and a correct assessment of objects that are perceived (or not) by the driver. This contribution relates to the latter, as it consists of an attempt at detecting vehicles located in the attentional visual field of the driver.

2.1.2 Background

More than 30,000 people die in vehicle-related accidents in the United States every year [33]. Current Driver Assistance Systems (DAS) such as Electronic Stability Control have reduced these grim numbers over the recent years. However much more remains to be accomplished since driver actions, maneuvers, and ocular behavior as they relate to the environment are usually not included in the retroactive mechanisms composing DAS. Given that most vehicle accidents are rooted in human error, the next steps in ADAS research are clearly laid out. One crucial task is to automatically and reliably detect relevant objects in the environment which the driver may or may not have seen, as a starting point of a driver-aware assistance system.

Many vision-based algorithms have been developed for the purpose of detecting vehicles at frame rate in recent years, but many such techniques lack the required accuracy and speed to be of any practical use. Reducing the search space for Hypothesis Generation (HG) and extracting only the relevant information may assist in reaching higher efficiency. The most relevant information concerning what a driver is looking at comes from the 3D gazing point of the driver in absolute coordinates and what type of objects are located around it [5]. In this contribution our interest focuses on the vehicles the driver is most likely to be aware of at any given time. The experiments were conducted with an improved version of the RoadLAB experimental vehicle described in [19].

The remainder of this contribution is organized as follows: Section II describes the current literature in the area of vehicle detection. The proposed technique is presented in Section III. Results and evaluations are given in Section IV. Section V concludes this work.

2.2 Literature Survey

Vehicle detection has been and continues to be an important area of research in Computer Vision. Statistical methods such as SVM [6, 7], PCA [8] or Neural Networks [9] constituted the main body of early approaches. However, these did not achieve the levels of efficiency required for realistic use.

In 2001, Viola and Jones proposed a powerful method for rapid object detection based on training a sequence of classifiers using Haar-like features [10]. The authors demonstrated high performance rates for their method by applying it to the problem of facial recognition [11]. This approach became widely used in areas including vehicle detection [12], and many researchers have focused on its variants in recent years [13].

In general, any detection problem is formulated in two stages: Hypothesis Generation (HG) and Hypothesis Verification (HV). Various strategies have been used for HG. Some authors used shadows cast on the pavement by the presence of vehicles [14], while others exploited edge maps [15], vehicle symmetry [16], and so on, as cues to finding vehicles in sets of images. Reducing the range of probable vehicle positions in HG is of utmost importance to achieve decent frame rates in the vehicle detection process. In the case of HV, most techniques apply AdaBoost as a strong classifier to verify the existence of vehicles in the the image areas identified as likely candidates by the hypothesis generation mechanism. For instance, Khammari *et al.* used a gradient based algorithm for hypothesis generation and AdaBoost classifier for candidate verification [17]. Alternatively, Song *et al.* generated candidate image regions by use of edge maps and properties of vehicle symmetry followed by texture analysis and AdaBoost for vehicle detection [18]. None of these approaches are attempting to determine which vehicles are in the attentional visual field of the

driver. Rather, they attempt to identify all vehicles within image sequences.

2.3 Proposed Method

For reasons cited above, our interest lies in the detection of vehicles that a driver is most likely to be aware of at any moment. For this reason, we have selected two approaches developed in our laboratories and combined them into the technique we present herein. The first of our techniques, due to Kowsari *et al.* performs vehicle detection by using Hypothesis Generation techniques that significantly reduce the number of candidate image regions while preserving excellent detection rates [23]. The second technique we adopted in this contribution pertains to the identification of the 3D point of gaze in absolute 3D coordinates given in the frame of reference of the vehicle [21]. This technique is unique in that it avoids the parallax errors introduced by the combined usage of 3D gaze trackers with monocular scene cameras. As expected, adjoining these two techniques allows for even more stringent HG strategies since we are only interested in vehicles around the driver point of gaze, thus increasing the nominal frame rate for vehicle detection. Our driving sequences for test purposes were recorded with the RoadLAB experimental vehicle, as depicted in Figure 2.1. Both the forward stereo scene system and the non-contact 3D gaze tracker operate in their own frames of reference. While we define that of the scene stereo system to be that of the vehicle, there is a need to transform the 3D driver gaze, expressed in the coordinates of the tracker, into that of the vehicle. Fortunately, the technique to perform this transformation has been established by Kowsari *et al.* [21] in our laboratories.

Figure 2.1: *Vehicular instrumentation configuration.* **a) (top-left):** *In-vehicle non-contact infra-red binocular gaze tracking system.* **b):** *Forward scene stereo imaging system mounted on the roof of the vehicle.*

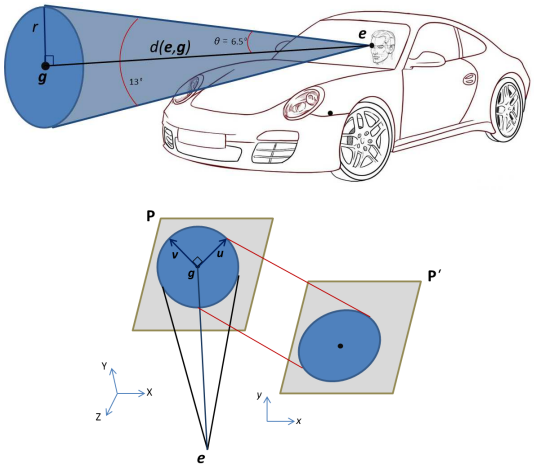


Figure 2.2: **(top):** *A depiction of the driver attentional gaze cone. It is generally accepted that the radius of the cone is approximately* 13° *[1] and* **(bottom):** *reprojection of 3D attentional circle onto the 2D ellipsoid on image plane of the forward stereo scene system.*

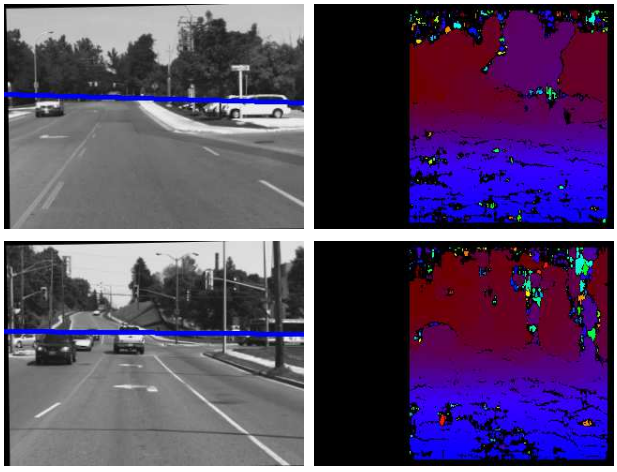


Figure 2.3: **(left):** *Estimated horizon lines and* **(right):** *their corresponding depth maps.*

2.3.1 Hypothesis Generation

The objective of HG is to identify candidate image areas likely to contain vehicles, in order to disregard image areas not likely to contain such objects (such as visible part of the sky, for instance). A considerable number of false positives may be removed by the suppression of image areas in the HG stage for which we know no vehicle can be present. One such powerful constraint applied to the HG process is the fact that vehicles will appear in imagery within a horizontal band centered around the horizon line (if we assume a relatively flat landscape). Furthermore, since our interest is focused on vehicles within an area around the 3D gaze of the driver, then the HG stage need only consider the intersection of the band around the horizon line and the 3D gaze area of the driver. This intersection constitutes a Region of Interest (RoI) within which the HG stage operates. Figure 2.2 depicts the attentional gaze cone of a driver. The radius of the circular gaze area depends on the distance between the fixated object and the driver.

The horizon line is approximated by intersecting the plane parallel to that of the ground plane and passing through the focal point with the image plane of the forward stereo scene system, and converting to image coordinates using stereo calibration parameters. The ground plane estimation is performed by grouping detected 3D ground points in areas of the imagery where no obstacle is known to be present. These points are then used in a robust plane-fitting computation to obtain the ground plane parameters (implementation details for both the ground plane and the horizon line are provided in [23]).

Once provided with an estimate of the horizon line, the first RoI is given by the horizontal band defined as

$$\mathrm{RoI}_1 : f(x) = mx + b \pm \delta \tag{2.1}$$

where m and b specify the horizon line and δ determines the vertical span of the RoI around it. Vehicles are expected to be found within this image region. Figure 2.3 shows the horizon line for two different frames from a test sequence.

The second RoI of interest consists of the circle formed by the intersection of the visual cone of attention with the plane perpendicular to the 3D Line of Gaze (LoG) and containing the 3D Point of Gaze (PoG) of the driver. The hypothesis that visal attention operates as a spotlight within the visual field is corroborated by an number of studies [3]. Additionally, it seems reasonable to equate the span of this spotlight to $\pm 6.5^\circ$ from the pitch and yaw angles of the gaze direction, as it corresponds to the human central field of view [1], resulting in the attentional gaze cone depicted in Figure 2.2.

Once the 3D eye position $\mathbf{e} = (e_x, e_y, e_z)$ from which the LoG is emanating, and the 3D PoG $\mathbf{g} = (g_x, g_y, g_z)$ have been transformed into the frame of reference of the forward stereo scence system (as per [21]), the radius of the

circular gaze area onto the plane perpendicular to the LoG and containing the PoG (plane $\mathbf{P}$) is obtained as

$$r = \tan(\theta) d(\mathbf{e}, \mathbf{g}) \tag{2.2}$$

where $\theta = 6.5°$, and d is the Euclidean distance between $\mathbf{e}$ and $\mathbf{g}$, defined as

$$d(\mathbf{e}, \mathbf{g}) = \sqrt{(e_x - g_x)^2 + (e_y - g_y)^2 + (e_z - g_z)^2} \tag{2.3}$$

At this point, the circle defined by the PoG and radius r, and contained in the 3D plane perpendicular to the LoG is reprojected onto the image plane of the forward stereo scene system and delineates the 2D portion of the scene that falls onto the attentional visual area of the driver. Since the plane perpendicular to the LoG and containing the PoG is generally not parallel to the image plane of the forward stereo scene system, the 3D attentional circle projects to a 2D ellipsoid as illustrated in Figure 2.2. The 3D circle in parametric form can be formulated as

$$\begin{cases} X = g_x + r\cos(\phi)u_x + r\sin(\phi)v_x \\ Y = g_y + r\cos(\phi)u_y + r\sin(\phi)v_y \\ Z = g_z + r\cos(\phi)u_z + r\sin(\phi)v_z \end{cases} \tag{2.4}$$

where $\mathbf{u} = (u_x, u_y, u_z)$ and $\mathbf{v} = (v_x, v_y, v_z)$ are two unit orthogonal vecters in plane $\mathbf{P}$ and ϕ is the radian angle between 0 and 2π.

Using $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ (perspective projection) and applying the intrinsic calibration matrix of the stereo scene system presented in [21] yield 2D ellipsoid on the image plane of the forward stereo scene system. The details of this reprojection are delineated in [21].

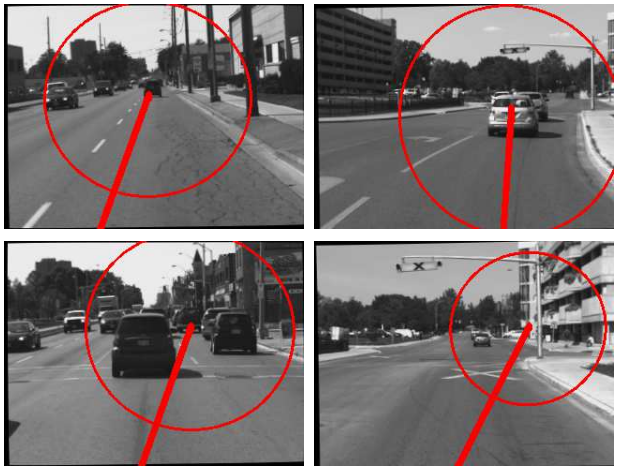


Figure 2.4: *Displays of various LoGs, PoGs, and attentional gaze areas projected onto the forward stereo scene system of the vehicle.*

The second region of interest RoI_2 is thus defined as the image region contained within the 2D circle as it reprojects onto the image plane of the forward stereo scene system. The final RoI of interest for the HG mechanism is consequently defined as the intersection of RoI_1 and RoI_2. Given image region $\mathrm{RoI} = \mathrm{RoI}_1 \cap \mathrm{RoI}_2$, the HG mechanism proceeds according to the algorithm given in [23]. Figure 2.4 displays several Lines of Gaze (LoGs), Points of Gaze (PoG), and attentional visual areas for selected frames.

2.3.2 Hypothesis Verification

AdaBoost, introduced by Freund and Schapire [24], is a method for choosing and combining a set of weak classifiers to build a strong classifier. Adjoining the concept of the integral image as an efficient way of computing Haar-like features and cascaded AdaBoost, Viola and Jones introduced a powerful method for object recognition [10]. We adopted this approach for the HG stage and used four cascaded AdaBoost classifiers to discriminate positive from false-

positive hypotheses. The Haar-training module from OpenCV is used to train our classifiers. Figure 2.5 shows various Haar-like features. We used in excess of two hundred vehicle images as positive samples for each classifier. For the negative examples, a set of more than five hundred images randomly downloaded from the Internet was used.

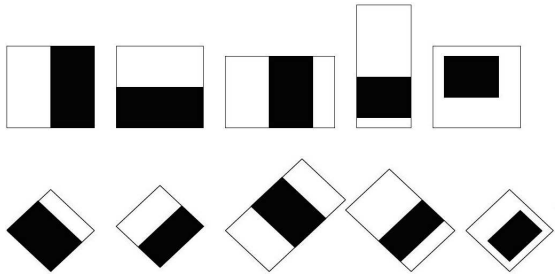


Figure 2.5: *A depiction of Haar-like features.*

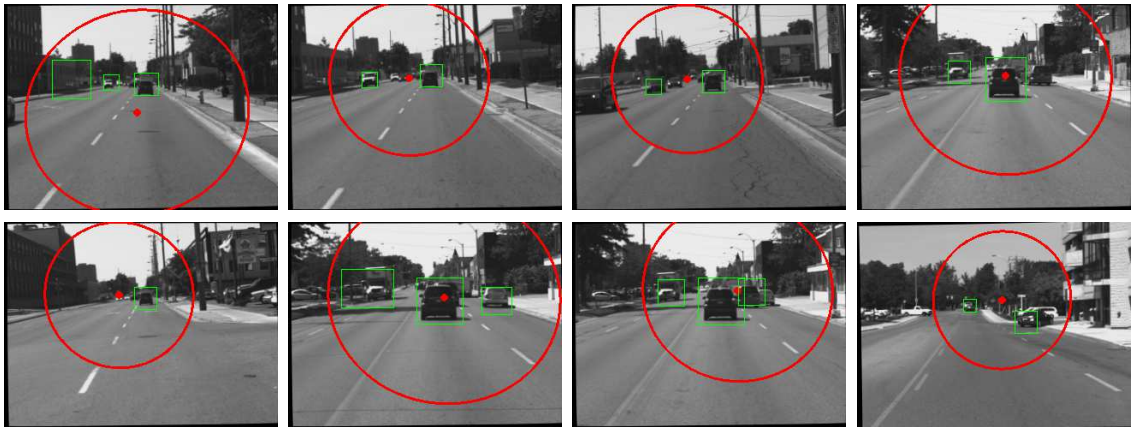


Figure 2.6: *Detection results within the RoI, including false positives and negatives.*

2.4 Experimental Results

The proposed method was tested on 3,326 randomly selected frames, in which 5,751 vehicles in various views and lanes appeared. These were manually an-

Table 2.1: Detection rates and false positives per frame for different views and vehicles

	FV	**RV**	**FSV & BSV**	**SV**	**ALL**
DR	0.9877	0.9890	0.8401	0.8118	0.9867
FP/F	0.25	0.22	0.30	0.31	0.24

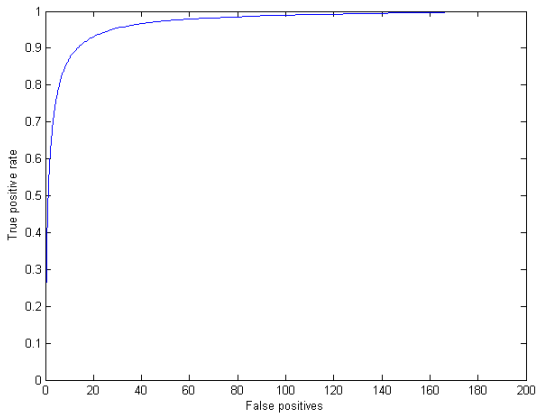


Figure 2.7: *ROC curve obtained from experiments.*

notated for the purpose of evaluating our method. Figure 2.6 shows a small sample of our results where each green rectangular area within the RoI indicates a vehicular detection. Rates of detection and false positives per frame for various vehicle views are given Table 2.1, where **FV**, **RV**, **FSV** & **BSV**, and **SV**, stand for Front View, Rear View, Front-Side and Back Side Views, and Side View, respectively, while **DR** and **FP/F** are Detection Rate and False Positives per Frame.

Figure 2.7 shows the performance of our method measured using a Receiver Operating Characteristics (ROC) curve, characterizes the True Positive Ratio (TPR) versus False Positives (FP). TPR is obtained by dividing True Positives by the number of vehicles. Table 2.2 compares various vehicle detection methods with ours, where **F/S**, **HT**, and **FP** stand for Frames per Second,

Table 2.2: Comparison on frame rates, hit rates, and false positives for various methods

Authors	F/S	HR	FP
Chang and Cho [27]	5	99%	12%
Southall *et al.* [28]	16	99%	1.7%
Bergmiller *et al.* [29]		83.12%	16.7%
Sun *et al.* [30]	10	98.5%	2%
Alonso *et al.* [31]		92.63%	3.63%
Cheng *et al.* [32]	20	90%	10%
Kowsari *et al.* [23]	25	98.6%	13%
Our Method	30	98.9%	11%

Hit Rate, and False Positives, respectively.

2.5 Conclusion

We presented a method for the detection of vehicles located within the attentional visual area of drivers as an initial attempt at identifying which visual stimuli elicit ocular responses from drivers. Our implementation is non-contact and operated on-board an experimental vehicle at frame rate (30Hz). We believe this contribution could easily be extended to include other visual stimuli drivers routinely encounter and attend to, such as pedestrians, cyclists, traffic lights, signs, and more. Ultimately, such an augmented technique will be integrated into i-ADAS as a means to identify objects drivers attend to, and those that they do not, at frame rate.

Bibliography

[1] K. Takagi, H. Kawanaka, S. Bhuiyan, and K. Oguri, *Estimation of a Three-Dimensional Gaze Point and the Gaze Target from Road Images*, IEEE Int.

Conf. Intell. Trans. Sys., pp. 526-531, 2011.

[2] F. Lethaus and J. Rataj, *Do Eye Movements Reflect Driving Manoeuvres?*, Intell. Trans. Sys., Vol. 1, No. 3, pp. 199-204, 2007.

[3] G.J. Andersen, *Focused Attention in Three-Dimensional Space*, Percept. & Psychophys. Vol. 47, pp. 112-120, 1990.

[4] M. Land and J. Horwood, *Which Parts of the Road Guide Steering?*, Nature, Vol. 377, No. 6547, pp. 339-340, 1995.

[5] E. Hartman, *Driver Vision Requirements*, Society of Automotive Engineers, Technical Paper Series, Hillsdale, NJ, 700392, pp. 629-630, 1970.

[6] C. Papageorgiou and T. Poggio, *A Trainable System for Object Detection*, IJCV, Vol. 38, No. 1, pp. 15-33, 2000.

[7] Z. Sun, G. Bebis, and R. Miller, *On-road Vehicle Detection using Gabor Filters and Support Vector Machines*, Int. Conf. on Digit. Sig. Proc., Vol. 2, pp. 1019-1022, 2002.

[8] J. Wu and X. Zhang, *A PCA Classifier and its Application in Vehicle Detection*, Int. Joint Conf. on Neur. Net., Vol. 1, pp. 600-604, 2001.

[9] N. Matthews, P. An, D. Charnley, and C. Harris, *Vehicle Detection and Recognition in Grayscale Imagery*, Control Engineering Practice, Vol. 4, No. 4, pp. 473-479, 1996.

[10] P. Viola and M. Jones, *Rapid Object Detection using a Boosted Cascade of Simple Features*, CVPR, Vol. 1, pp. 511-518, 2001.

[11] P. Viola and M. Jones, *Robust Real-Time Face Detection*, IJCV, Vol. 57, No. 2, pp. 137-154, 2004.

[12] S. Sivaraman and M. Trivedi, *A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking*, IEEE Trans. Intell. Trans. Sys., Vol. 11, No. 2, pp. 267-276, 2010.

[13] C. Caraffi, T. Vojir, J. Trefny, J. Sochman and J. Matas, *A System for Real-Time Detection and Tracking of Vehicles from a Single Car-Mounted Camera*, IEEE Conf. Intell. Trans. Sys., pp. 975-982, 2012.

[14] A. Yoneyama, C. H. Yeh, and C. -C. J. Kuo, *Moving Cast Shadow Elimination for Robust Vehicle Extraction Based on 2D Joint Vehicle/Shadow Models*, IEEE Conf. Adv. Vid. and Sig. Based Surveill., pp. 229-236, 2003.

[15] Z. Sun, R. Miller, G. Bebis and D. DiMeo, *A Real-Time Precrash Vehicle Detection System*, IEEE Workshop Appl. Comp. Vis., pp. 171-176, 2002.

[16] A. Bensrhair, M. Bertozzi, A. Broggi, P. Miche, S. Mousset, and G. Toulminet, *A Cooperative Approach to Vision-Based Vehicle Detection*, IEEE Conf. Intell. Transp. Sys., pp. 207-212, 2001.

[17] A. Khammari, F. Nashashibi, Y. Abramson, and C. Laurgeau, *Vehicle Detection Combining Gradient Analysis and AdaBoost Classification*, IEEE Conf. Intell. Trans. Sys., pp. 66-71, 2005.

[18] G. Y. Song, K. Y. Lee, and J. W. Lee, *Vehicle Detection by Edge-Based Candidate Generation and Appearance-Based Classification*, Intell. Veh. Sym., pp. 428-433, 2008.

[19] S.S. Beauchemin, M.A. Bauer, T. Kowsari, and J. Cho, *Portable and Scalable Vision-Based Vehicular Instrumentation for the Analysis of Driver Intentionality*, IEEE Trans. on Instr. and Meas., Vol. 61, No. 2, pp. 391-401, 2012.

[20] S.S. Beauchemin, M. Bauer, D. Laurendeau, T. Kowsari, J. Cho, M. Hunter, and O. McCarthy, *RoadLab: An In-Vehicle Laboratory for Developing Cognitive Cars*, IEEE Int. Conf. on Comp. Appl. in Ind. and Eng., pp. 7-12, 2010.

[21] T. Kowsari, S.S. Beauchemin, M.A. Bauer, D. Laurendeau, and N. Teasdale, *Multi-Depth Cross-Calibration of Remote Eye Gaze Trackers and Stereoscopic Scene Systems*, submitted, IEEE Intell. Veh. Sym., 2014.

[22] Z. Sun, G. Bebis, and R. Miller, *On-Road Vehicle Detection: A Review*, IEEE Trans. PAMI, Vol. 28, No. 5, pp. 694-711, 2006.

[23] T. Kowsari, S.S. Beauchemin, and J. Cho, *Real-Time Vehicle Detection and Tracking Using Stereo Vision and Multi-View AdaBoost*, IEEE Int. Conf. Intell. Trans. Sys. 2011.

[24] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Computational Learning Theory, Springer, 1995.

[25] N. Corneli, B. Leibe, K. Cornelis, and L. Van Gool, *3D City Modeling Using Cognitive Loops*, Int. Sym. 3D Data Proc., Vis., and Trans., 2006.

[26] B. Leibe, N. Corneli, K. Cornelis, and L. Van Gool, *Dynamic 3D Scene Analysis from a Moving Vehicle*, CVPR, pp. 1-8, 2007.

[27] W.C. Chang and C.W. Cho, *Online Boosting for Vehicle Detection*. IEEE Trans. Sys., Man, and Cyber. B: Cyber. Vol. 40, No. 3, pp. 892-902, 2010.

[28] B. Southall, M. Bansal, and J. Eledath, *Real-Time Vehicle Detection for Highway Driving*, CVPR, pp. 541-548, 2009.

[29] P. Bergmiller, M. Botsch, J. Speth, and U. Hofmann, *Vehicle Rear Detection in Images with Generalized Radial-Basis-Function classifiers*, Intell. Veh. Sym., pp. 226-233, 2008.

[30] Z. Sun, G. Bebis, and R. Miller, *Monocular Precrash Vehicle Detection: Features and Classifiers*, IEEE Trans. Im. Proc., Vol. 15, No. 7, pp. 2019-2034, 2006.

[31] D. Alonso, L. Salgado, and M. Nieto, *Robust Vehicle Detection through Multidimensional Classification for On-Board Video-Based Systems*, ICIP, Vol. 4, pp. 321-324, 2007.

[32] H. Cheng, N. Zheng, C. Sun, and H. van de Wetering, *Vanishing Point and Gabor Feature-Based Multi-Resolution On-Road Vehicle Detection*, Lecture Notes in Computer Science, Vol. 3973, pp. 46-51, 2006.

[33] National Highway Traffic Safety Administration, *Traffic Safety Facts: 2010 Data*, U.S. Department of Transportation, June 2012.

Chapter 3

Traffic Signs Detection and Recognition

This Chapter is a reformatted version of the following article:

S.J. Zabihi, S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer, *Detection and Recognition of Traffic Signs Inside the Attentional Visual Field of Drivers, Submitted to, IEEE Intelligent Vehicles Symposium (IV17)*, Redondo Beach, California, USA, June 11-14 2017.

In this contribution we present a vision-based framework which detects and recognizes traffic signs inside the attentional visual field of drivers. This technique takes advantage of the driver 3D absolute gaze point obtained through the combined use of a front-view stereo imaging system and a non-contact 3D gaze tracker. We used a linear Support Vector Machine as a classifier and a Histogram of Oriented Gradient as features for detection. Recognition is performed by using Scale Invariant Feature Transforms and color information. Our technique detects and recognizes signs which are in the field of view of the driver and also provides indication when one or more signs have been missed by the driver.

3.1 Introduction

Advanced Driver Assistance Systems (ADAS) are rapidly becoming widespread in modern vehicles with the aim to make roads safer for vehicles and pedestrians. Traffic Sign Detection and Recognition (TSDR) techniques are essential components of ADAS.

These methods attempt to make drivers aware of incoming signs on the road and warn them against possible danger. However, there are many factors that can make the process of detection and recognition of traffic signs less successful, such as differences in sign position, in lighting, motion blur, and poor image quality. Additionally, we are interested in the signs a driver may not see, based on the 3D Point of Gaze (PoG) and the attentional visual field of drivers.

TSDR have attracted a great deal of attention in the recent past. While the problem of sign detection may appear solved (in particular for European traffic signs), there is still room for numerous improvements. To the best of our knowledge, our work is the first TSDR method that performs detection and recognition of traffic signs within the attentional visual field of the driver. The usefulness of this approach lies in the capability of informing drivers of road signs that have not intersected their visual field of attention for a certain period of time. The proposed method consists of three different stages:

1. Establishing the attentional field of view of the driver
2. Sign detection performed with a linear Support Vector Machine (SVM) and Histogram of Oriented Gradient (HOG) features
3. Recognition performed using color information and Scale Invariant Feature Transform (SIFT) matching

This contribution is structured as follows. In Section II, we review the related literature in the field of sign detection. Section III discusses the proposed method including detection and recognition. Results and evaluations are given in Section IV. Section V summarizes this paper.

3.2 Literature Survey

Different traffic sign recognition methods have been proposed in the recent past. They usually consist of two sequential processes, namely a detection stage that identifies a Region of Interest (RoI), and a recognition stage that identifies the exact type of sign or rejects the identified RoI.

In many systems, color segmentation is used for detection and recognition. For example, authors in [3] compared the YUV and RGB color spaces and chose the latter as it resulted in lower numbers of false positives and improved computational time. Nevertheless, In [26], the YUV color space was used for similar purposes.

IHLS (an improved version of HLS color space) is another color space that was employed in [11]. A great number of contributors used other color spaces such as HSI [22], [23], [32], and HSV [24], [4]. Soendoro *et al.* [33] performed color filtering using the CIELAB color space coupled with hue and noted its efficiency for localizing traffic signs. According to [12] the color appearance model CIECAM97 performs better than others such as CIELUV, CIELAB, and RGB. Authors in [13] also used the CIECAM97 model for color segmentation. In addition, Liu *et al.* [19] proposed a new color filter which they called Simple Vector Filter (SVF) that is capable of extracting a specific color at high speed and separating objects from backgrounds.

While color based detectors are popular, there are many other approaches

Figure 3.1: *Physical configuration of the experimental vehicle with the external stereo cameras mounted on the roof. The dashboard-mounted eye tracker is displayed in the top-left part of the image.*

based on the shapes of signs. for instance, authors in [30] introduced a colorless method for road sign classification using a Hierarchical Spatial Feature Matching (HSFM) method. Canny's edge detector was used by Garcia *et al.* [14] as a means to extract contours necessary for shape-based traffic sign detectors. Aoyagi and Asakura [1] proposed genetic algorithms and neural networks for the purpose of sign detection. Road sign symmetry properties were used in [28] for detecting candidate image regions. Loy and Barnes [21] suggested a technique for detecting triangular, square, and octagonal road signs using the symmetric nature of these shapes.

Many sign detection strategies include a color segmentation stage followed by some kind of shape extraction. The work of Fang *et al.* [10] includes using hue values as color features and an edge detector method for shape-feature extraction. Contributions from Oh *et al.* [29] and Tsai *et al.* [34] are other examples of integrating color based methods with shape analysis.

In recent years, HOG features [6] were used by many for traffic sign feature extraction such as in Xie *et al.* [35] and in Zaklouta *et al.* [37]. Additionally, Mathias *et al.* [25] used Integral Channel Features (ICF), first established by

Dollar *et al.* [8]. In most of these approaches, the feature extraction stage is followed by a recognition process that ascertains whether detected candidates are actual traffic signs. We have noted that neural networks [16], SVMs [15], and template matching [4],[18] are the most widely used approaches for the traffic sign recognition stage.

3.3 Proposed Method

3.3.1 Gaze Localization

In this work, our main focus lies on detection and recognition of signs within the visual field of the driver. In order to relate the 3D Line-of-Gaze (LoG) of the driver to the depth map obtained by the forward stereo camera system and derive the 3D Point-of-Gaze (PoG), we used a technique proposed in our laboratory [17] to identify the 3D PoG in absolute coordinates expressed in the frame of reference of the vehicle. Figure 3.1 delineates the remote eye tracking system and the stereoscopic vision system. By intersecting the visual cone of attention with the plane perpendicular at the 3D PoG along the 3D LoG of the driver, we are able to form a circle in 3D space which represents our region of interest. Note that this circle becomes a 2D ellipse once projected onto the imaging plane of the stereoscopic sensors. further details and related equations are found in [17] and [36]. Our technique is described by the following procedure:

1. Transformation of eye position $\mathbf{e} = (e_x, e_y, e_z)$ and the 3D PoG $\mathbf{g} = (g_x, g_y, g_z)$ into the frame of reference of the forward scene stereo imaging system as described in [17]

Figure 3.2: *Attentional gaze areas projected onto the forward stereo system of the vehicle.*

2. Calculating the radius of the gaze area based on the Euclidean distance between eye position and point of gaze

3. Re-projecting the defined circle by the PoG and radius and contained in the 3D plane perpendicular to the LoG onto the image plane of the forward stereo scene system.

4. Transformation of the 3D circle into the reference frame of the stereo system

5. Projecting the 3D attentional circle to the corresponding ellipsoid onto the imaging plane of the stereo system

Figure 3.2 depicts attentional gaze areas for a few chosen frames.

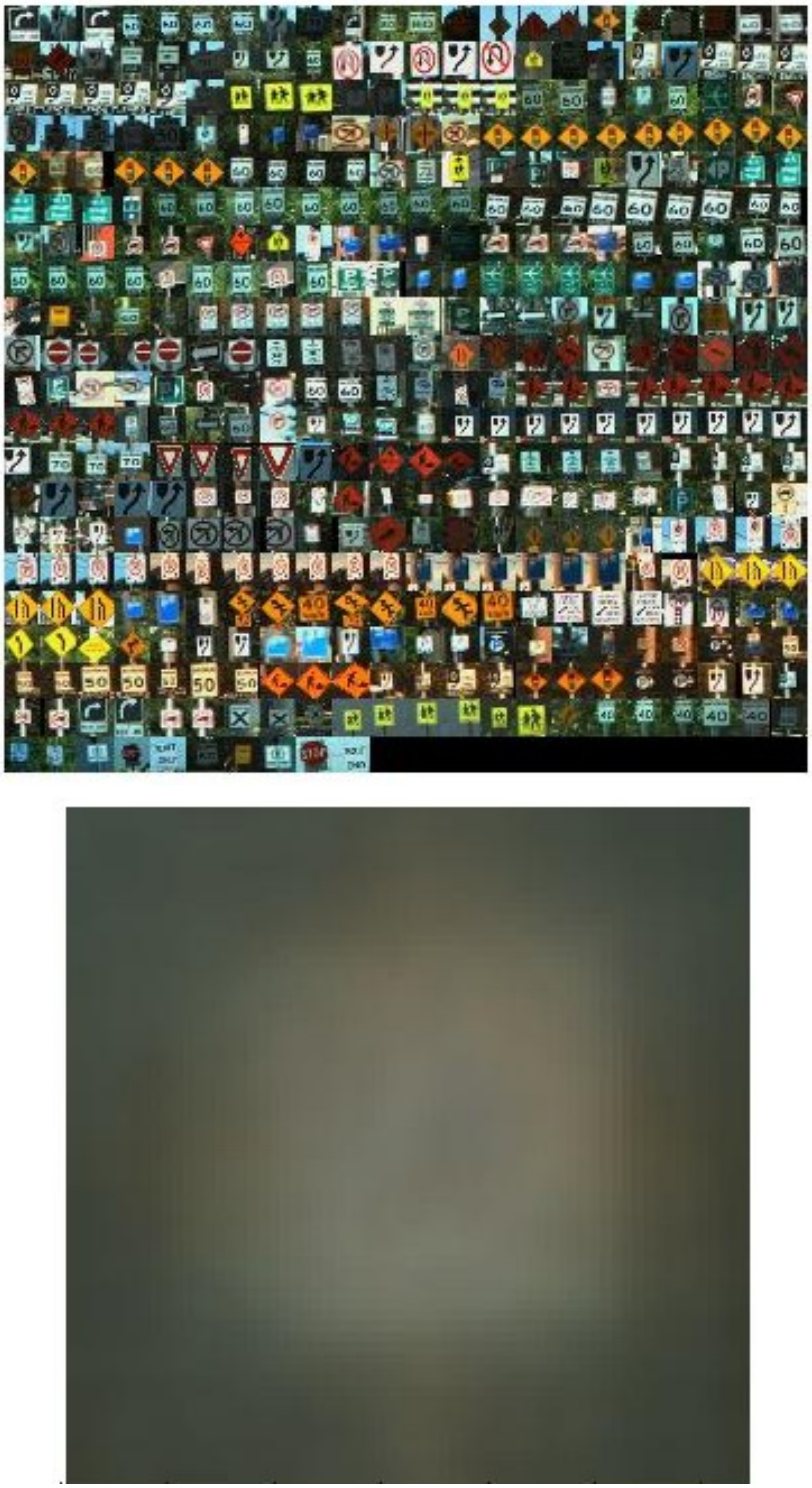


Figure 3.3: **(top)***: Positive samples* **(bottom)***: Average image.*

3.3.2 Detection Phase

It is widely accepted that Histogram of Oriented Gradients (HOG) features are suitable for rigid object detection. These features are based on evaluating well-normalized local histograms of image gradient orientations in a dense grid [6]. Computing HOG features includes estimating first-order image gradients, building the gradient histograms, followed by a block-normalization process resulting in better invariance to illumination, shadowing, and edge contrast. The final stage is gathering the HOG descriptors into a combined feature vector for use in the window classifier. For this system, we chose 9 orientation channels.

After the feature extraction stage, a linear Support Vector Machine (SVM) [5] is used for learning the detector. SVM is a discriminative classifier formally defined by a separating hyper-plane. This classification method is highly accurate and extremely fast which is useful for large amounts of training data. We selected 1000 images as positive training samples. Additionally, we increased the number of positive images by adding the flipped, rotated, and translated versions of the original samples resulting in better detection performance. Figure 3.3 shows the image visualization of the complete list of object training examples and their average. The initial negative samples are selected from the training images with the traffic signs regions cropped out. For boosting the performance of the learned classifier, we use an advanced learning method called Hard Negative Mining (HNM). In essence, every single region that does not contain a traffic sign can be considered as a negative sample. There are too many samples to be used in practice, but we are only looking for key negative samples which can be extracted from the hard negative mining stage. We train the SVM in an iterated procedure, and for each iteration, the detector is

applied to a new image without any traffic signs. Then, we add the resulting false positives (hard negatives) to the training set for the next iteration. We iterated this process 5 times with good results. Finally, the classifier is provided with more key negative samples which makes the detection performance more robust. Figure 3.4 illustrates the extraction of hard negative samples from a traffic sign-free image.

Once the stages of HNM and training are completed, we evaluate the model on test data. We use a sliding window over multiple scales. In order to eliminate redundant detections, a Non Maximum Suppression (NMS) algorithm is used. NMS keeps the highest-scoring detection and removes any other detection whose overlap is greater than a threshold. We used Pascal's overlap score [9] so as to establish the overlap ratio between the two bonding boxes. It is computed as:

$$a_0 = \frac{\text{area}(B_1 \cap B_2)}{\text{area}(B_1 \cup B_2)} \tag{3.1}$$

where a_0 is the overlap ratio. B_1 and B_2 are the bounding boxes.

3.3.3 Recognition Phase

The recognition phase identifies the correct type of road sign candidate. Recognition is performed with SIFT features and color information. The detected candidates are scaled to the same size as our template signs.

We gathered a full set of template traffic signs for use in the recognition stage. Figure 3.5 depicts a few examples of images in the template database. The next step involves the use of color information in order to increase the performance of the SIFT matching. We calculate the color difference between the candidate target and template signs. The HSV color model is used for this purpose since this color space is less sensitive to variable lighting conditions.

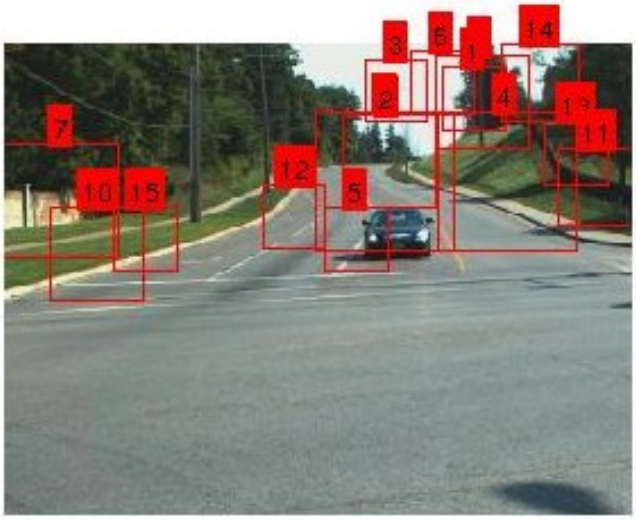


Figure 3.4: *An Illustration of Hard Negative Mining.*

Figure 3.5: *Examples of template signs.*

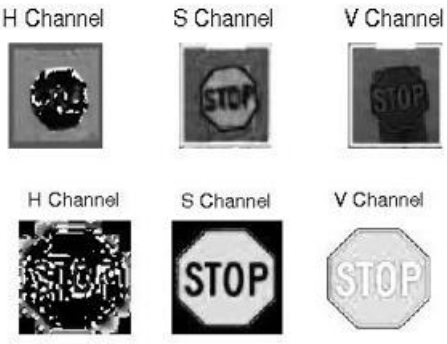


Figure 3.6: *HSV color space images a)* **(top)***: detected sign b)* **(bottom)***: template sign.*

Many researchers considered using this method for traffic sign segmentation. For instance, Paclik *et al.* [31] used this color space due to its similarity to human perception of colors. Figure 3.6 displays the extracted H,S and V values for the template and detected signs. Traffic signs include a wide variety of colors and we differentiate them by using all the components of the HSV color space. We compute the H, S, and V values of the detected candidate and the template signs. Then, we get the average of all values based on a defined mask. This is followed by creating the δ images,

$$\begin{aligned} \delta H &= H\text{channel} - H\text{standard} \\ \delta S &= S\text{channel} - S\text{standard} \\ \delta V &= V\text{channel} - V\text{standard} \end{aligned} \tag{3.2}$$

where Hchannel, Schannel, and Vchannel are the averaged HSV color parameters of the detected candidate, and the other three are averaged HSV color parameters of the template sign. The final value which is the color difference between the two images is obtained as:

$$\delta f = \sqrt{\delta H^2 + \delta S^2 + \delta V^2} \tag{3.3}$$

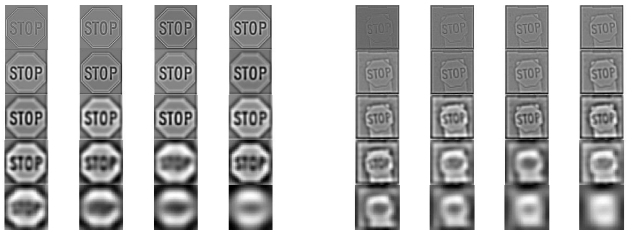


Figure 3.7: *DOG scale space images* **(left)***: template sign* **(right)***: detected sign.*

where δf value is between 0 and 1. If it is close to 1, we can conclude that there is a significant color difference between the candidate traffic sign and the template. We define a threshold, whose value was experimentally obtained, to determine whether the two images are color-wise similar or not. If δf is less than the defined threshold, then we select the corresponding template sign for feature matching. Hence, we perform feature matching only between the candidate image and those images in the template database whose colors are similar to the detected sign. This method removes some of the false matches and significantly improves the performance.

We perform feature matching using the SIFT [20] algorithm. We chose this detector instead of other feature-matching detectors due to its accuracy and speed. Feature extraction is performed with a Difference of Gaussians (DoG) operator within the SIFT descriptor. The features extracted with a DoG are partially invariant to scale, rotation, and position variations. Figure 3.7 illustrates the DoG scale space. Following this, we perform matching on the two sets of descriptors, using a thresholding value. A descriptor d_i is matched to a descriptor d_j if the distance between them multiplied by a threshold is not greater than the distance of d_i to all other descriptors. The value of the threshold we used is 1.5. The RANdom SAmple Consensus (RANSAC) algorithm is also used to discard possible outliers. This is an iterative method

for estimating a mathematical model from a data set that contains outliers. The basic stages of this algorithm are summarized as follows [7]:

- Randomly select the minimum number of points needed to determine the model parameters
- Solve for the parameters of the model
- Determine how many points from the set of selected points fit the model parameters with a predefined tolerance ϵ.
- Re-estimate the model parameters if the fraction of inliers over the total number points in the set is lower than a predefined threshold τ.
- Repeat these steps until an adequate confidence level for the estimated model parameters is attained.

In order to filter out outliers during training, a small set of samples has been used to train a homography model. Then the samples which are within the error tolerance of the homography model are determined. These samples are considered as inliers. If the number of inliers is the largest found so far, we keep the current inlier set. This process is repeated for a number of iterations and returns the model with smallest average error among the generated models.

Finally, the image in the template database which gives the maximum number of matches with the candidate image is considered as a recognized sign. However, if the maximum number of matches between the two images is less than a minimum value, we discard that candidate image.

3.4 Experimental Results

The driving sequences recorded with our experimental vehicle [2] have been used. The proposed method was tested on 3500 frames. Among these frames, 1806 traffic signs appeared which were manually annotated. The size of the recorded images is 320 by 240. While our main focus is on detection and recognition of signs within the visual field of the driver, we also performed sign detection outside this area in order to provide the driver with a response about a possibly unseen traffic sign. Figure 3.8 displays small portions of our results. According to this figure, if the four coordinates of the bounding box are inside the drivers field of view, we can conclude that the driver has seen the sign. Otherwise, the driver has missed the sign. Both *SEEN* and *MISSED* feedbacks are given to the driver right after the detection and recognition of signs.

For assessing the accuracy of sign detection, we report on both the Detection Rate (DR) and the number of False Positives Per Frame ($FPPF$), defined as follows:

$$DR = \frac{TP}{TP + FN} \tag{3.4}$$

$$FPPF = \frac{FP}{F} \tag{3.5}$$

where TP is the number of correct detections, FN is the number of false negatives, FP is the number of false positives, and F is the number of all frames. Tables 5.2 and 4.2 report the performance of the detection and recognition of traffic signs. Figure 3.9 displays the performance of our detector computed using a Receiver Operating Characteristics (ROC) curve, marks the True Positive Rate (TPR) versus False Positive Rate (FPR). The threshold used for the curve is the scoring value of the detected bounding boxes. This value

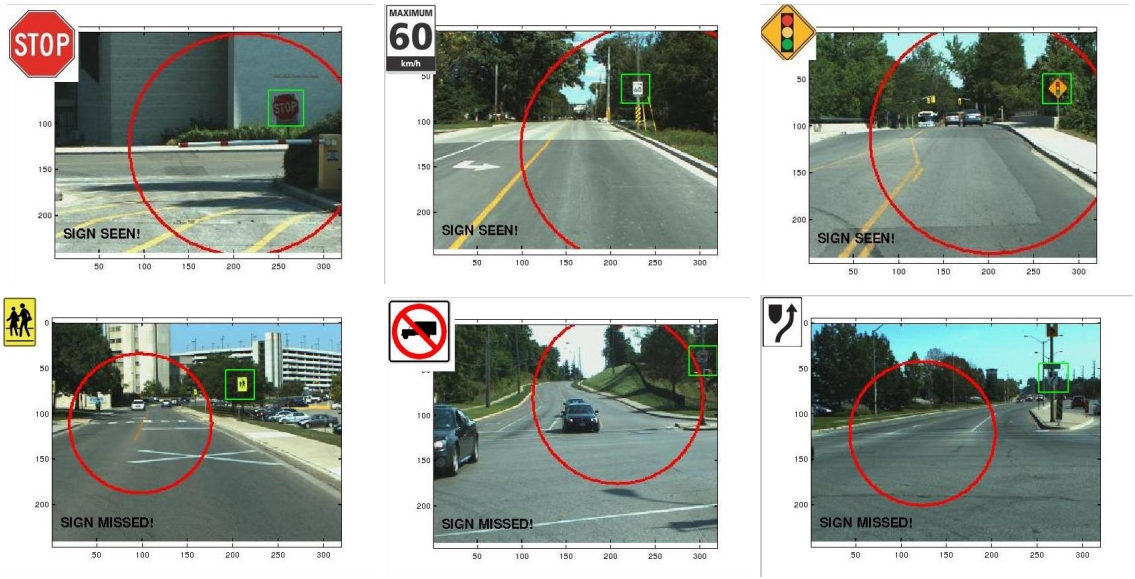


Figure 3.8: *Detection and recognition results* **a) (top)***: SEEN SIGNS* **b) (bottom)***: MISSED SIGNS.*

varies from 0 (definitely negative) to 1 (definitely positive). Another standard evaluation method is the confusion matrix, displayed in Figure 3.10.

Table 3.1: Detection Rate and False Positive Rate per Frame

	DR	*FPPF*
Proposed Method	0.84	0.04

Table 3.2: Accuracy Rate of Recognition Phase

	Image Samples	**Corr. Recog.**	**Accuracy**
Proposed Method	1517	1348	88.9

Not many traffic sign detection methods have been tested on North American sings, as most of the techniques in the current literature rely on existing databases of European road signs. Consequently, we are not able to provide a meaningful comparison between our proposed algorithm and other algorithms

based on North American signs. For instance authors in [27] proposed a detector for the stop, warning, and speed limit signs only and provided separate accuracies for each. In contrast, our detector considers a multitude of traffic signs including warning, temporary conditions, information, direction, and regulatory signs. Such variability precludes the production of a meaningful comparative study at this time.

3.5 Conclusion

We presented an efficient method for the detection and recognition of traffic signs within the attentional visual field of drivers. By using an in-vehicle, non-contact infra-red binocular gaze tracking system installed in our experimental vehicle, we were able to identify the exact attentional visual area of the driver into the depth map provided by the forward stereoscopic system. We were also able to infer whether the driver was likely to have seen the sign or not based on computing the intersection of the detected bounding box and drivers gaze area. While most of the other methods are simply concerned with the detection and recognition of signs within image sequences, we designed an algorithm for inferring the cognitive behaviour of the driver in terms of the 3D localization of gaze in absolute coordinates and determined the approximate size of the driver's attentional visual area. The development of an Advanced Driving Assistance System that increases safety in relation to the 3D localization of the driver's gaze in the environment may be realized by extending our approach to include other important features commonly encountered on the road.

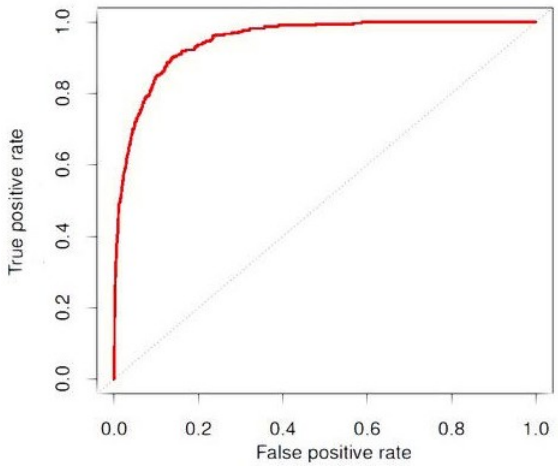


Figure 3.9: *The ROC curve resulting from our experiments.*

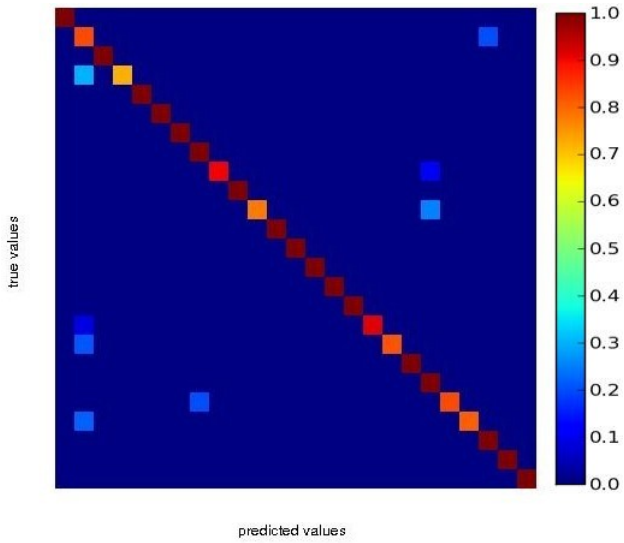


Figure 3.10: *The Confusion Matrix resulting from our experiments*

Bibliography

[1] Yuji Aoyagi and Toshiyuki Asakura. A study on traffic sign recognition in scene image using genetic algorithms and neural networks. In *22nd IEEE International Conference on Industrial Electronics, Control, and Instrumentation (IECON)*, volume 3, pages 1838–1843, 1996.

[2] S.S. Beauchemin, M.A. Bauer, D. Laurendeau, T. Kowsari, J. Cho, M. Hunter, and O. McCarthy. Roadlab: An in-vehicle laboratory for developing cognitive cars. In *23rd International Conference on Computer Applications in Industry and Engineering (CAINE)*, pages 7–12, 2010.

[3] Alberto Broggi, Pietro Cerri, Paolo Medici, Pier Paolo Porta, and Guido Ghisio. Real time road signs recognition. In *IEEE Conference on Intelligent Vehicles Symposium*, pages 981–986, 2007.

[4] Long Chen, Qingquan Li, Ming Li, and Qingzhou Mao. Traffic sign detection and recognition for intelligent vehicle. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 908–913, 2011.

[5] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[6] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, 2005.

[7] Konstantinos G Derpanis. Overview of the ransac algorithm. *Image Rochester NY*, 4(1):2–3, 2010.

[8] Piotr Dollár, Zhuowen Tu, Pietro Perona, and Serge Belongie. Integral channel features. BMVC Press, 2009.

[9] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[10] Chiung-Yao Fang, Sei-Wang Chen, and Chiou-Shann Fuh. Road-sign detection and tracking. *IEEE transactions on vehicular technology*, 52(5):1329–1341, 2003.

[11] Hasan Fleyeh. Color detection and segmentation for road and traffic signs. In *IEEE Conference on Cybernetics and Intelligent Systems*, volume 2, pages 809–814, 2004.

[12] Xiaohong Gao, Kunbin Hong, Peter Passmore, Lubov Podladchikova, and Dmitry Shaposhnikov. Colour vision model-based approach for segmentation of traffic signs. *EURASIP Journal on image and video processing*, 2008(1):1–7, 2007.

[13] Xiaohong W Gao, Lubov Podladchikova, Dmitry Shaposhnikov, Kunbin Hong, and Natalia Shevtsova. Recognition of traffic signs based on their colour and shape features extracted using human vision models. *Journal of Visual Communication and Image Representation*, 17(4):675–685, 2006.

[14] Miguel Ángel García-Garrido, Miguel Ángel Sotelo, and Ernesto Martín-Gorostiza. Fast road sign detection using hough transform for assisted driving of road vehicles. In *International Conference on Computer Aided Systems Theory*, pages 543–548. Springer, 2005.

[15] Jack Greenhalgh and Majid Mirmehdi. Real-time detection and recognition of road traffic signs. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1498–1506, 2012.

[16] M Sajjad Hossain, M Mahmudul Hasan, Mohammad Ameer Ali, Humayun Kabir, and Shawkat Ali. *Automatic detection and recognition of traffic signs*. IEEE, 2010.

[17] T Kowsari, SS Beauchemin, MA Bauer, D Teasdale, and N Laurendeau. Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems. 2014.

[18] Merve Can Kus, Muhittin Gokmen, and Sima Etaner-Uyar. Traffic sign recognition using scale invariant feature transform and color classification. In *23rd International Symposium on Computer and Information Sciences (ISCIS)*, pages 1–6, 2008.

[19] Han Liu, Ding Liu, and Jing Xin. Real-time recognition of road traffic sign in motion image based on genetic algorithm. In *International Conference on Machine Learning and Cybernetics Proceedings*, volume 1, pages 83–86, 2002.

[20] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[21] Gareth Loy and Nick Barnes. Fast shape-based road sign detection for a driver assistance system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 1, pages 70–75, 2004.

[22] S Maldonado-Bascon, S Lafuente-Arroyo, P Siegmann, H Gomez-Moreno,

and FJ Acevedo-Rodriguez. Traffic sign recognition system for inventory purposes. In *IEEE Intelligent Vehicles Symposium*, pages 590–595, 2008.

[23] Saturnino Maldonado-Bascon, Sergio Lafuente-Arroyo, Pedro Gil-Jimenez, Hilario Gomez-Moreno, and Francisco López-Ferreras. Road-sign detection and recognition based on support vector machines. *IEEE transactions on intelligent transportation systems*, 8(2):264–278, 2007.

[24] Rabia Malik, Javaid Khurshid, and Sana Nazir Ahmad. Road sign detection and recognition using colour segmentation, shape analysis and template matching. In *International Conference on Machine Learning and Cybernetics*, volume 6, pages 3556–3560, 2007.

[25] Markus Mathias, Radu Timofte, Rodrigo Benenson, and Luc Van Gool. Traffic sign recognitionhow far are we from the solution? In *The International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2013.

[26] Jun Miura, Tsuyoshi Kanda, Shusaku Nakatani, and Yoshiaki Shirai. An active vision system for on-line traffic sign recognition. *IEICE TRANSACTIONS on Information and Systems*, 85(11):1784–1792, 2002.

[27] Andreas Møgelmose, Dongran Liu, and Mohan M Trivedi. Traffic sign detection for us roads: Remaining challenges and a case for tracking. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1394–1399, 2014.

[28] Christian Nunn, Anton Kummert, and Stefan Muller-Schneiders. A two stage detection module for traffic signs. In *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pages 248–252, 2008.

[29] Jun-Taek Oh, Hyun-Wook Kwak, Young-Ho Sohn, and Wook-Hyun Kim. Segmentation and recognition of traffic signs using shape information. In *International Symposium on Visual Computing*, pages 519–526. Springer, 2005.

[30] Pavel Paclık and Jana Novovicova. Road sign classification without color information. In *Proc. of the 6th Annual Conference of the Advanced School for Computing and Imaging*, 2000.

[31] Pavel Paclık, J Novovičová, Pavel Pudil, and Petr Somol. Road sign classification using laplace kernel classifier. *Pattern Recognition Letters*, 21(13):1165–1173, 2000.

[32] Xu Qingsong, Su Juan, and Liu Tiantian. A detection and recognition method for prohibition traffic signs. In *International Conference on Image Analysis and Signal Processing*, pages 583–586, 2010.

[33] David Soendoro and Iping Supriana. Traffic sign recognition with color-based method, shape-arc estimation and svm. In *International Conference on Electrical Engineering and Informatics (ICEEI)*, pages 1–6, 2011.

[34] L-W Tsai, J-W Hsieh, C-H Chuang, Y-J Tseng, K-C Fan, and C-C Lee. Road sign detection using eigen colour. *IET computer vision*, 2(3):164–177, 2008.

[35] Yuan Xie, Li-feng Liu, Cui-hua Li, and Yan-yun Qu. Unifying visual saliency with hog feature learning for traffic sign detection. In *IEEE Intelligent Vehicles Symposium*, pages 24–29, 2009.

[36] SM Zabihi, Steven S Beauchemin, EAM de Medeiros, and Michael A Bauer. Frame-rate vehicle detection within the attentional visual area

of drivers. In *IEEE Intelligent Vehicles Symposium Proceedings*, pages 146–150, 2014.

[37] Fatin Zaklouta and Bogdan Stanciulescu. Warning traffic sign recognition using a hog-based kd tree. In *IEEE Intelligent Vehicles Symposium*, pages 1019–1024, 2011.

Chapter 4

Vehicle Localization

This Chapter is a reformatted version of the following article:

S.M. Zabihi, S.S. Beauchemin, E.A.M. de Medeiros and M.A. Bauer, *Lane-Based Vehicle Localization in Urban Environments, IEEE International Conference on Vehicular Electronics and Safety (ICVES15)*, Yokohama, Japan, November 5-7 2015.

In this Chapter, we propose a method by which vehicular speed and a map-based lane detection process are called upon to improve the positional accuracy of GPS. Experimental results with urban driving sequences demonstrate that our approach significantly improves the accuracy of positioning the vehicle as compared with systems solely relying on GPS.

4.1 Introduction

In recent years driver assistance systems have contributed remarkably to the mitigation of traffic accidents and their consequences. Vehicle localization is an essential component of these systems and thus their precision and robustness being of great value. Current vehicle localization methods are commonly

based on satellite positioning technology, of which GPS is the most established. However, GPS technology is known to produce inaccurate position estimates in certain conditions and to be reliable only up to a range of several meters.

GPS is a passive satellite-based and easy-to-use positioning system, which was established by the U.S. Department of Defence (DoD) [20]. This system is able to pinpoint the absolute longitude and latitude coordinates of an object on the globe. A GPS system consists of a number of satellites orbiting around Earth. Each satellite frequently sends messages that include the time, the message was transferred, and the satellite location. The messages are received on the ground via a GPS unit and, comparing the time at which the message was received (on its internal clock) against the time which the message was sent, gives distance between the unit and any of these satellites. Some of the common factors influencing GPS accuracy include [5]:

1. **Atmospheric effect:** Both the ionosphere and troposphere impact the speed of GPS radio signals.

2. **Multipath errors:** This occurs when signals are reflected or bounced by coming in contact with surrounding hills, lakes, buildings or any radio wave reflective object before it reaches the receiver. This delay in signal travel time introduces positional errors.

3. **Clock errors:** The internal clocks of both the satellite and receiver have limited accuracy, and they are not precisely synchronized. Since position calculations depend on accurate time, small clock errors can cause significant imprecision in position estimation.

4. **Satellite geometry:** Localization precision is optimal when satellites are located at wide angles from each other from the perspective of the

receiver. Conversely, poor geometry occurs when satellites form a line or find themselves in a tight grouping, resulting in Dilution of Precision (DoP).

Errors due to multipath and reduced satellite visibility are the most difficult to minimize. Others such as atmospheric errors can be compensated for by differential means, including Differential GPS (DGPS) or with the Wide Area Augmentation System (WAAS). Techniques known as dead reckoning and map matching are generally applied to atone for satellite visibility and multipath issues. Dead reckoning employs measurements of the vehicle's motion from on-board sensors such as accelerometers and gyroscopes to extrapolate from the last known vehicle location [21], [24]. Dead reckoning is ineffective when GPS position estimates are unavailable for a long period of time. Map matching methods apply a map of the road environment to narrow the vehicle position as the correct road can be seen. In this case, the vehicle position can be modified to lie on the road, permitting a partial rectification of GPS estimates [27].

Landmark detection provides the ability to sense the surrounding environment of the vehicle to alleviate many of the localization issues identified above. By positioning the vehicle with respect to objects in the environment, it becomes possible to reduce errors in GPS estimates. Furthermore, it should be easier to identify which road the vehicle is on as well as correct the vehicle location by observing objects in the surrounding environment.

In this contribution we propose a method by which vehicular speed and a map-based lane detection process are called upon to improve the positional accuracy of GPS. This contribution is structured as follows: related work is reviewed in section II. Section III describes our proposed method in detail. Results and experiments are presented in section IV. Section V summarizes

our results.

4.2 Literature Survey

Several methods have been proposed for improving the accuracy of GPS. Among them, we find Differential GPS (DGPS) [9] methods. DGPS employs one mobile and one or more stationary GPS receiver stations nearby in order to minimize errors introduced by atmospheric effects. The fixed GPS receiver is in a known position and acts as a reference station, calculating and broadcasting the difference between its known location and that estimated by GPS. This information is applied to the moving GPS receiver in order to correct its position. This technique relies on the assumption that GPS errors are identical for nearby GPS receivers [14].

Another correction technique presented by Lin [18] uses differential correction and Genetic Programming (GP) [16], [18]. In this method, GP generates a correction function from NMEA[1] information derived from the GPS receiver at the known location and the GPS receiver which requires correction. It then uses the generated function to modify its location information.

Other techniques utilize vision or other means of sensing the environment to refine vehicle position estimation. For instance, Georgiev and Allen and Kais et al. apply computer vision methods as a complement to GPS and dead reckoning for positioning a robot in urban environments where satellite visibility may be poor [10], [11]. Both approaches use straight line features that are detectable and abundant in urban environments, such as building edges, doors, windows, trees, poles, traffic signs and lane boundaries. Features used by Kais et al. for localization are mapped within a Geographic Information

[1]NationalMarine Electronics Association (NMEA), http://www.nmea.org

System (GIS) database [11]. The locations of GIS features are transformed to the camera coordinate frame to specify search regions for these features in the acquired image. Conversely, Georgiev and Allen model buildings by their straight line features [10]. The transformation needed to align the extracted features with the model yields the location of the robot relative to the building.

In addition, Barthet al. presented a method for localization of a vehicle's position and orientation with respect to stop lines at intersections based on video sequences and mapped data [3]. Brenner provided a landmark map consisting of extracted poles obtained using a mobile mapping van equipped with LIDAR [6]. Poles are detected from the sensor data and provided as input to a landmark matching algorithm which estimates the vehicle position.

Diverse Visual Odometry (VO) techniques have arisen in recent years, improving vehicle localization performance. The methods in this category, such as [13] and [1] extract 3D features, followed by feature matching, and reconstruct a 3D point cloud used to estimate vehicle pose.

Several recent approaches employ the principles of Simultaneous Localization and Mapping (SLAM) [17], [19], [22], [2], [8], [23]. These types of techniques attempt to build a map as a robotic vehicle navigates through an unknown area while localizing the vehicle within the map simultaneously. Early attempts apply extended Kalman filter (EKF), where the filter state includes the locations of landmarks and robot poses [7], [8]. These methods face covariance complexity problems and hence cannot be used for mapping large environments. To overcome this problem, hierarchical visual SLAM strategies are used to divide an initial map into smaller submaps [26]. For instance, Schleicher et al. presented a real-time hierarchical SLAM system which generates a number of local submaps, each composed of several visual landmarks that are then employed by a standard EKF [22].

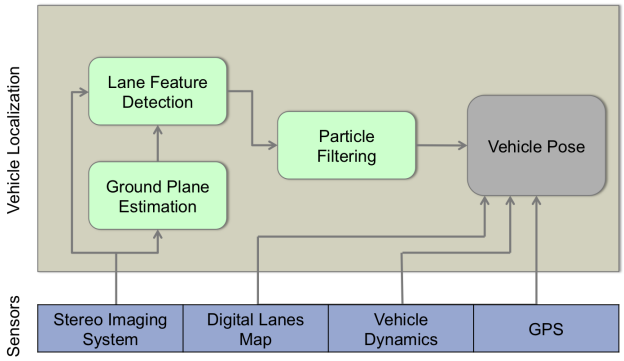


Figure 4.1: *Overview of the proposed vehicle localization approach*

In this contribution, we extend our previous work in lane detection using lane-annotated maps, stereo depth maps and particle filtering [15] to estimate the precise position and orientation of the vehicle by fitting lane features in stereo imagery with lane maps obtained from Google Earth satellite images.

4.3 Proposed Method

We describe a map-based localization approach that has the capability to enhance GPS-based localization by using lanes as landmarks on images obtained by a front-view stereo imaging system. The map-based framework extracts lane boundary features and attempts to fit the lane features with a pre-loaded digital lane map by discovering the best relative modification in the position and orientation from the GPS module of the vehicle. Figure 4.1 depicts an overview of the our localization strategy.

4.3.1 Modelling Observed Road Lanes

In general, features observed in driving environments are not sufficiently unique by themselves to indicate vehicle location unambiguously. This motivates the

use of environment maps containing the position and identity of features (lanes in world coordinates, in our case). We apply a model based on splines that can address the observed lane shapes and cover the entire map of the region of interest [15]. The model contains a number of splines where each spline is a lane marker and consists of a set of control points with known GPS coordinates. Candidate visible splines for each stereo pair are determined based on the current vehicle's position and orientation, and the front stereo system viewing angle.

We use Google Earth satellite images to produce a map that includes all the lanes in a path that was travelled by the experimental vehicle within the city of London, Ontario. The lanes obtained by this method are not occluded by objects such as other vehicles or buildings. These images can also be addressed directly by longitude and latitude which is desirable since we use GPS coordinates to locate the vehicle on the map and extract hypothetically visible lanes from the stereo images.

4.3.2 Localization by Particle Filtering

Particle filtering is used to estimate the vehicle position by integrating measurements from GPS, visual sensors, and context from the map. Information from visual sensors installed on our experimental vehicle provide stereo images and depth maps which are used for detecting lane features in driving scenes. We then attempt to fit our projected spline lane markers onto the image plane of the stereo sensor with the lane features detected in the left stereo image using GPS coordinates as a seed estimate for vehicle position.

Ground Plane Estimation

The ground plane parameters needed for projecting the lanes onto the image can be computed from the depth map obtained from the stereo system. With rectified stereo images, finding disparities and hence depth map merely consists of a 1-D search with a block matching algorithm (our implementation uses the stereo routines from Version 2.4 of OpenCV) Assuming that the ground plane equation is of the form

$$ax + by + cz = d \tag{4.1}$$

where $\vec{n} = (a, b, c)$ is the unit normal vector to the plane, we pose

$$d = \frac{1}{\sqrt{a'^2 + b'^2 + c'^2}} \tag{4.2}$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = d \begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} \tag{4.3}$$

With the coordinates of 3D points in the reference system of the left camera

$$(X_i, Y_i, Z_i) \tag{4.4}$$

we can write

$$\mathbf{Ax} = \mathbf{B} \tag{4.5}$$

and solve for $\mathbf{x}$ in the least-squares sense as

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{B} \tag{4.6}$$

where

$$\mathbf{A} = \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} a' \\ b' \\ c' \end{bmatrix}$$

Often times the ground surface leads to inordinate amounts of outliers, due in part to a lack of texture from the pavement or other drivable surfaces. With the sensitivity of least-squares to outliers being known, we resort to the use of RANSAC in selecting the inliers and obtain a robust estimation of the ground plane coefficients, in the following way:

1. randomly select three points from the 3D points believed to be representative of the ground plane
2. compute the coefficients of the plane defined by the randomly selected points using (4.5)
3. count the points whose distance to the plane is less than a threshold ϵ
4. repeat these steps n times where n is sufficiently large[1]
5. among the n fits choose the largest inlier set which respect to ϵ and compute the coefficients of the ground plane this time using least-squares as in (4.6)

The plane parameters are averaged over a short period of time in order to stabilize them further. The coefficients of the plane are recomputed at each new stereo frame arrival. However, in cases when the number of depth values

[1]Choosing $n > 20$ does not significantly improve the number of inliers with respect to ϵ.

is low (poor texture, etc.) or other vision modules indicate the presence of a near obstacle, the coefficients of the ground plane are not recomputed, the previous parameters are used instead.

Lanes feature detection

We apply a feature detection algorithm to find the boundaries illustrating lanes in the driving environment. The left stereo image and its depth map have been used to create a Gaussian smoothed lane boundary feature image. The lane feature detection algorithm is outlined Algorithm in 1. Constants found in the algorithm are α and β, used for computing the width expectation of the lane markings $L_{\max}$, factored by their distance from the vehicle. Constants NL and LD indicate the state of the lane edge search. NL represents the state in which no lanes are detected, while LD is its complement. Threshold τ_h represents the minimum gradient value required for a transition from NL to LD. Constant O_h is the minimum variation in height from the ground plane for a pixel to be considered part of an obstacle. O_h and τ_h depend on imagery and are experimentally determined.

Figure 4.2a shows the lane features detected in a driving scene and lanes splines of the same scene which are projected to the left stereo camera image can be observed in Figure 4.2b.

GPS Correction

At this stage, we will apply a matching method to fit the lane model projected on the image coordinate system with the lane features acquired from the previous step by finding the best changes in the position and orientation of the vehicle provided by GPS unit. The position and orientation of the ve-

Algorithm 1 *Lane Feature Detection Algorithm*

$G \leftarrow$ 1D Gaussian row smoothing of I with $\sigma = 0.5$
$G \leftarrow$ horizontal gradient of G using 3-point central differences
Remove the values corresponding to obstacles from G using threshold O_h
State $\leftarrow NL$
F initialized to 0
for all rows i in I starting from the image bottom **do**
  $L_{\max} \leftarrow \beta - i\alpha$
  Count $\leftarrow 0$
  for all column j in I **do**
    if $(G_{i,j} > \tau_h \wedge (\text{State} = \text{NL} \vee \text{Count} > L_{\max}))$ **then**
      State $\leftarrow$ LD
    end if
    if $(\text{State} = \text{LD}) \wedge (G_{i,j} < -\tau_h)$ **then**
      for $k = j - \text{Count} \rightarrow j$ **do**
        $F_{i,k} \leftarrow 1$
      end for
      State $\leftarrow$ NL
      Count $\leftarrow 0$
    end if
  end for
end for
$F \leftarrow$ 1D Gaussian row smoothing of I with $\sigma = 0.5$

Figure 4.2: **a) (left):** *Lane features detected by Algorithm 1* **b) (right):** *Projected lane splines on the image*

hicle obtained by unreliable GPS measurements are used only as seed points to find the visible parts of the lane map and start an optimization algorithm to locate the accurate position of the vehicle. The optimization methods produce two parameters δX and $\delta\theta$ which correct the position and orientation of the vehicle. To estimate the best fit parameters between projected lane-marking splines and the detected lane features in the left stereo image, the below likelihood function is defined:

$$\mathcal{L}(z|\mathbf{x}) \tag{4.7}$$

where z is a particular parameter fit, and $\mathbf{x} = (\delta x, \delta\theta)$. With the lane feature image F and the projected, visible lane-marking splines, the likelihood function becomes

$$\mathcal{L}(z|\mathbf{x}) = \sum_{(i,j)\in\mathbf{S}} F(i,j) \tag{4.8}$$

where $\mathbf{S}$ is the set of all projected points of the lane-marking splines.

With the likelihood function, we need to estimate the parameters $\mathbf{x}$ of the fit as:

$$x = \text{argmax}\mathcal{L}(z|\mathbf{x})_{\mathbf{x}} \tag{4.9}$$

Solving this optimization problem is not easily achievable by regular hill-climbing methods due to the non-concavity of the function. Since the search space is large, an exhaustive search is prohibitively expensive while the probability of finding the global maximum remains low [25].

A Particle Swarm Optimization (PSO) method may be more appropriate. The particle swarm lane detection algorithm by Zhou [28] is a single image frame method, which we adapt here as a particle filter working on a

sequence of frames[1]. Our approach consists of generating a set of uniformly distributed particles, each representing a set of possible values for parameters $\mathbf{x} = \delta x, \delta\theta, \delta\lambda$. The likelihood of each particle is estimated with (4.9).

At each iteration, each particle is replaced with a number of newly generated, Gaussian position-disturbed particles. The number of generated particles is proportional to the likelihood of the particle they replace. Their likelihood is estimated again with (4.9) and normalized. This ensures that the stronger particles generate more particles in their vicinity than the weaker ones. Particles with normalized likelihoods lower than a certain threshold are removed and, if the number of particles becomes less than a threshold, the process repeats.

These iterations eventually lead to groups of particles concentrated at the most likely answers in the search space and the particle with the maximum likelihood is chosen as the solution. In addition, keeping the particles over time makes the particle filter to act as a tracker for the lane detection mechanism.

The optimum solution gives us the GPS correction parameters which are relative changes in the position and orientation of the vehicle. Therefore, the accurate position and orientation of the vehicle can be calculated as:

$$\mathbf{X} = \mathbf{x} + \delta\mathbf{x} \tag{4.10}$$

$$\phi = \theta + \delta\theta \tag{4.11}$$

where $\mathbf{x}$ and θ are raw GPS measurements of position and orientation of the vehicle, and $\mathbf{X}$ and ϕ represent the corrected position and orientation.

[1]PSO is a population-based stochastic optimization method first proposed by Eberhart and Kennedy [12].

4.4 Experimental Results

To evaluate the performance of our approach, we used several sequences recorded by driving our instrumented vehicle around the city of London, Ontario [4]. There is much opportunity to observe extreme multipath conditions which usually occur in urban environments. Our vehicle is equipped with stereo camera rigs, a built-in GPS module, and On-Board Diagnostic systems (OBD-II) with CANbus protocol[1]. The front stereo rig is mounted outside of the vehicle on top of the roof, providing stereo images with a resolution of 320 by 240. The driving path covered by the vehicle is illustrated in Figure 4.3a). Figure 4.3b) displays lane-marking splines, each containing several control points, for two distinct driving scenes. Google's static API is used to obtain the images and draw the lane-marking splines.

For evaluation purposes, we assume the vehicle is located in the middle of the lane it is driving on, which we consider ground truth. We keep track of lane information such as opening and closing distances of the lanes from the vehicle during the lane detection and model fitting processes. Hence, we can specify the driver's lane by finding the closest lane to the vehicle. The angle of this lane is considered as the orientation of the vehicle in ground truth.

The localization results modified by the proposed method are compared to the ground truth, where positional and orientational errors are defined as the Euclidean distance between the corrected GPS points and the corresponding points in the ground truth. We have tested our localization method on ten driving sequences, each of them covering the path shown in Figure 4.3 and containing around ninety thousand stereo frames. Figure 4.4 depicts the po-

[1]The CANbus (Controller Area Network bus) provides microcontrollers with the means to communicate with each other within a vehicle.

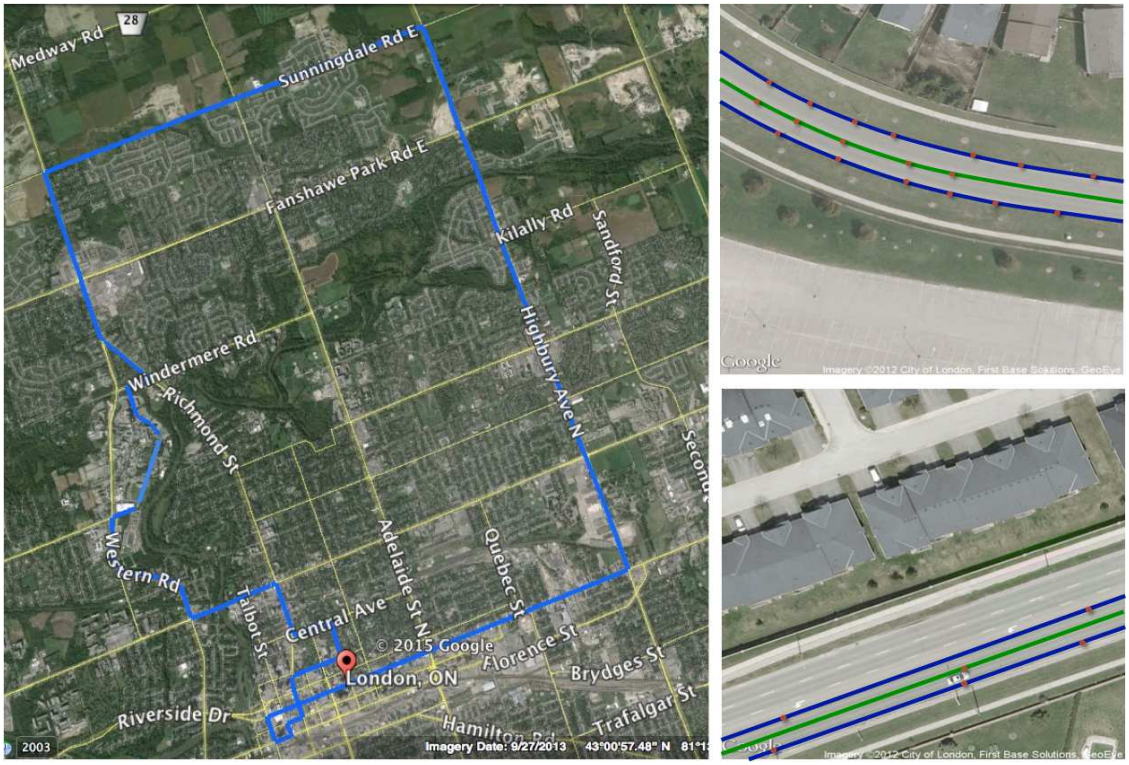


Figure 4.3: **a) (left):** *The path covered by the experimental vehicle.* **b) (right):** *Images obtained by the map building application showing splines as lane markers. The green spline indicates the middle of lanes.*

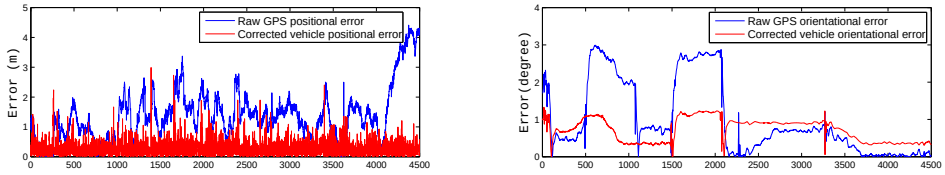


Figure 4.4: *Vehicle positional and orientational errors. The horizontal axis is quantified in frames.*

sitional and orientational errors of raw GPS data and corrected GPS data as compared to ground truth. We observe that the both positional and orientational errors for corrected GPS data are significantly less than those of the raw GPS data. The mean error value and standard deviation of the absolute po-

sitional and orientational errors for these experiments are indicated in Tables 5.2 and 4.2. As the tables show, the average vehicle localization error obtained by using raw GPS is considerably higher than the average error resulting from our proposed technique (where the positional and orientational errors are 0.36 m and 0.72° on average).

Table 4.1: Results of vehicle positioning absolute errors in world coordinate system.

	Mean (m)	**Std (m)**
Raw GPS Data	1.82	1.15
Corrected GPS Data	0.36	0.12

Table 4.2: Results of vehicle orientation absolute errors.

	Mean (°)	**Std (°)**
Raw GPS Data	1.02	0.67
Corrected GPS Data	0.72	0.31

4.5 Conclusion

It is nowadays possible to specify an absolute position anywhere on the globe with GPS. Although GPS works adequately in open environments with no overhead obstructions, it is subject to considerable errors when reception from some of the satellites is blocked. This occurs frequently in urban environments and renders accurate vehicle localization problematic. This contribution proposed a novel approach to improve vehicle localization accuracy by estimating vehicle position and orientation which that minimize the observed difference between detected lane features and projected lane-marking splines using a particle filter.

Bibliography

[1] Pablo Fernández Alcantarilla, José J. Yebes, Javier Almazán, and Luis Miguel Bergasa. On combining visual slam and dense scene flow to increase the robustness of localization and mapping in dynamic environments. In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 1290–1297, 2012.

[2] Tim Bailey and Hugh Durrant-Whyte. Simultaneous localization and mapping (slam): Part ii. In *IEEE Robotics & Automation Magazine*, volume 13, pages 108–117, 2006.

[3] Alexander Barth, Jan Siegemund, and Julian Schwehr. Fast and precise localization at stop intersections. In *Intelligent Vehicles Symposium Workshops (IV Workshops), IEEE*, pages 75–80, 2013.

[4] S.S. Beauchemin, M.A. Bauer, D. Laurendeau, T. Kowsari, J. Cho, M. Hunter, and O. McCarthy. Roadlab: An in-vehicle laboratory for developing cognitive cars. In *23rd International Conference on Computer Applications in Industry and Engineering (CAINE 10)*, pages 7–12, 2010.

[5] Parkinson W. Bradford, J. Spilker, and P. Enge. Global positioning system: theory and applications. In *AIAA Washington DC*, volume 109, 1996.

[6] Claus Brenner. Vehicle localization using landmarks obtained by a lidar mobile mapping system. In *Int. Arch. Photogramm. Remote Sens*, volume 38, pages 139–144, 2010.

[7] Andrew J. Davison. Real-time simultaneous localisation and mapping

with a single camera. In *Computer Vision. Proceedings. Ninth IEEE International Conference on*, pages 1403–1410, 2003.

[8] Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, volume 29, pages 1052–1067, 2007.

[9] Per K. Enge, Rudolph M. Kalafus, and Michel F. Ruane. Differential operation of the global positioning system. In *Communications Magazine, IEEE*, volume 26, pages 48–60, 1988.

[10] Atanas Georgiev and Peter K. Allen. Localization methods for a mobile robot in urban environments. In *Robotics, IEEE Transactions on*, volume 20, pages 851–864, 2004.

[11] Mikael Kais, Stéphane Dauvillier, Arnaud De La Fortelle, Ichiro Masaki, and Christian Laugier. Towards outdoor localization using gis, vision system andstochastic error propagation. In *ICARA International Conference on Autonomous Robots and Agents.*, pages 198–205, 2004.

[12] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Neural Networks, Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948, 1995.

[13] Bernd Kitt, Andreas Geiger, and Henning Lategahn. Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. In *Intelligent Vehicles Symposium (IV), IEEE*, pages 486–492, 2010.

[14] Kazuyuki Kobayashi, Ka C. Cheok, Kajiro Watanabe, and Fumio Munekata. Accurate differential global positioning system via fuzzy logic kalman filter sensor fusion technique. In *Industrial Electronics, IEEE Transactions on*, volume 45, pages 510–518, 1998.

[15] T. Kowsari, S.S. Beauchemin, and M.A. Bauer. Map-based lane and obstacle-free area detection. In *9th International Conference on Computer Vision Theory and Applications (VISAPP14)*, volume 3, pages 523–530, 2014.

[16] John R. Koza. *Genetic programming: on the programming of computers by means of natural selection*, volume 1. MIT press, 1992.

[17] Henning Lategahn, Andreas Geiger, and Bernd Kitt. Visual slam for autonomous ground vehicles. In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 1732–1737, 2011.

[18] Jung Yi Lin. Using evolutionary computation on gps position correction. In *The Scientific World Journal*, volume 2014. Hindawi Publishing Corporation, 2014.

[19] Christopher Mei, Gabe Sibley, Mark Cummins, Paul Newman, and Ian Reid. Rslam: A system for large-scale mapping in constant-time using stereo. In *International journal of computer vision*, volume 94, pages 198–214. Springer, 2011.

[20] M.R. Mosavi. Gps receivers timing data processing using neural networks: Optimal estimation and errors modeling. In *International journal of neural systems*, volume 17, pages 383–393. World Scientific, 2007.

[21] Honghui Qi and John B. Moore. Direct kalman filtering approach for gps/ins integration. In *Aerospace and Electronic Systems, IEEE Transactions on*, volume 38, pages 687–693, 2002.

[22] David Schleicher, Luis Miguel Bergasa, Manuel Ocaña, Rafael Barea, and María Elena López. Real-time hierarchical outdoor slam based on stereovision and gps fusion. In *Intelligent Transportation Systems, IEEE Transactions on*, volume 10, pages 440–452, 2009.

[23] Gabe Sibley, Christopher Mei, Ian Reid, and Paul Newman. Vast-scale outdoor navigation using adaptive relative bundle adjustment. In *The International Journal of Robotics Research*. SAGE Publications, 2010.

[24] Salah Sukkarieh, Eduardo Mario Nebot, and Hugh F. Durrant-Whyte. A high integrity imu/gps navigation loop for autonomous land vehicle applications. In *Robotics and Automation, IEEE Transactions on*, volume 15, pages 572–578, 1999.

[25] E-G Talbi and Traian Muntean. Hill-climbing, simulated annealing and genetic algorithms: a comparative study and application to the mapping problem. In *System Sciences, 1993, Proceeding of the Twenty-Sixth Hawaii International Conference on*, volume 2, pages 565–573, 1993.

[26] Juan D. Tardós, José Neira, Paul M. Newman, and John J. Leonard. Robust mapping and localization in indoor environments using sonar data. In *The International Journal of Robotics Research*, volume 21, pages 311–330, 2002.

[27] George Taylor and Geoff Blewitt. *Intelligent positioning: GIS-GPS unification*. Wiley, 2006.

[28] Yong Zhou, Xiaofeng Hu, and Qingtai Ye. A robust lane detection approach based on map estimate and particle swarm optimization. In *Computational Intelligence and Security*, pages 804–811. Springer, 2005.

Chapter 5

Driving Manoeuvre Prediction

This Chapter is a reformatted version of the following article:

S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer, *Real-Time Driving Manoeuvre Prediction Using IO-HMM and Driver Cephalo-Ocular Behaviour, Submitted to, IEEE Intelligent Vehicles Symposium (IV17)*, Redondo Beach, California, USA, June 11-14 2017.

Driving Assistance Systems increase safety and provide a more enjoyable driving experience. Among the objectives motivating these technologies rests the idea of predicting driver intent within the next few seconds, in order to avoid potentially dangerous manoeuvres. In this work, we develop a model of driver behaviour for turn manoeuvres that we then apply to anticipate the most likely turn manoeuvre a driver will effect a few seconds ahead of time. We demonstrate that cephalo-ocular behaviour such as variations in gaze direction and head pose play an important role in the prediction of driver-initiated manoeuvres. We tested our approach on a diverse driving data set recorded with an instrumented vehicle in the urban area of London, ON, Canada. Experiments show that our approach predicts turn manoeuvres 3.8 seconds before they occur with an accuracy over 80% in real-time.

5.1 Introduction

World-wide injuries in vehicle accidents have increased in recent years, mainly due to driver error. According to a large field study conducted in the USA, around 80% of collisions are due to distraction [21]. It is obvious that distractions caused by recent in-vehicle devices, such as GPS, entertainment systems, and cellphones, increase a driver's accident risk [27]. Over the past few years, several Advanced Driving Assistance Systems (ADAS) have been developed in an attempt to diminish the number of vehicular accidents. Instances of ADAS include Adaptive Cruise Control (ACC), Emergency Braking Systems (EBS), and Collision Avoidance Systems (CAS), among others. These systems pay attention to the role the driver plays in most driving events. Undoubtedly, improving these safety systems is an important objective as they may reduce the frequency and severity of injuries and fatalities.

An intelligent ADAS (i-ADAS) is a sensory and computer system inside a vehicle that collaborates with the driver to manage driving tasks. Detecting a driver's behaviour in conjunction with the dynamics of the vehicle and its surroundings provides valuable information in effectively assisting the driver in different situations. We aim to model driver behaviour and predict the next manoeuvre a driver will most likely perform in the next few seconds (Figure 5.1).

Predicting driving manoeuvres is still a challenging task since a model of human behaviour is required. A large number of factors such as emotional state, physical condition, and driving skills can affect driver behaviour, which ideally should be included in the model in order to provide a faithful representation. A model that incorporates all of these factors would effectively constitute a computational representation of a human being. This is of course

Figure 5.1: **(top):** *Our Vehicular instrumentation configuration.* **(bottom):** *Anticipating the probabilities of different driving manoeuvres using vehicle dynamics and driver cephalo-ocular behaviour.*

highly complex and not yet feasible in practice. Consequently, the presented methods in the literature focus on manageable subsets of these factors.

We have developed a prediction model using an Input-Output Hidden Markov Model (IO-HMM). IO-HMMs deal better with long-term dependencies compared to the standard HMMs and perform sequence production and prediction efficiently [4]. We learn the model parameters from natural driving data including vehicle dynamics and gaze information, then the system outputs the probability of each manoeuvre (left turn, right turn, and going straight) during inference.

This contribution is structured as follows: related work is reviewed in Section II. Section III describes our proposed method in detail. Results and experiments are presented in Section IV. Section V summarizes our results.

5.2 Literature Survey

Driver behaviour prediction models attempt to anticipate actions by observing how a driver interacts with the vehicle and its environment. In the case of

manoeuvre prediction, the models infer the driver's intent by mapping his interaction with control elements available in the vehicle, such as steering wheel, accelerator or turn signal, to the manoeuvre being modelled. Many different techniques have been proposed to model driver behaviour and predict the action a driver will perform next.

Neural networks are powerful for learning sample input/output relationships. The simplest forms of neural networks (single or multi-layer perceptrons) construct a mapping between input and output data by adjusting the weights of neural connections in a learning phase. After the network is trained, it is able to compute output values from input data not used in the training process. In the case of predicting driving maneuvers, the input of a neural network could be behavioral data such as steering wheel angle and speed, and the output of the model a prediction value for a maneuver. For example, authors in [15] implemented a neural network that learns how to execute overtaking manoeuvres from primitive manoeuvres.

Several prediction systems have been developed with the aim of anticipating future human action in various contexts [9, 13, 12]. Similar research is conducted in the context of vehicle driving. For instance, Frolich *et al.* [8], Kumar *et al.*, [17], and Morris *et al.* [20], attempt to predict future lane change manoeuvers. Authors in [8] use turn signals of other vehicles to predict the intent of other drivers before they change lanes or turn. A predictive model for lane changes is developed in [17] where authors use Support Vector Machines (SVM) and Bayesian filtering. Morris *et al.* [20] have developed a real-time prediction system capable of anticipating a driver's intent to change lanes a few seconds before it occurs.

MacAdam and Johnson [18] modelled driver steering behaviour using neural networks, while Dogan *et al.* [7] developed a prediction system using feed

forward and recurrent neural networks to model lane changes on curved roads and compare lane changing to lane keeping scenarios. Mitrovic [19] built a model to predict lateral and longitudinal vehicle acceleration by way of training a neural network.

Even though neural networks are powerful learning mechanisms, their main drawback is that they are very difficult to analyse since the information they encode is not easily interpretable. Another disadvantage is that most neural networks are not able to handle a temporal sequence of data points, but only compute the output for one data vector at a time. In the domain of driver behaviour modelling, especially for the prediction of driving manoeuvres, the data usually consists of sequences of individual phases and including this temporal information is essential.

Bayesian networks (BN) are probabilistic models which provide the possibility to define a structure of causal dependencies between variables in a directed acyclic graph, where the directions of these links imply a causal relation. A variety of driver models have been developed using Bayesian networks. As an example, Tezuka *et al.* [28] implemented a model that discriminates between lane keeping, lane change and emergency lane change using steering wheel angle as input information. Amata *et al.* [1] predicted stopping behaviour at intersections based on environmental conditions (traffic signs, pedestrian crossing, and leading vehicle) and a calculated driver type. Since Bayesian networks are probabilistic models, they are well suited at dealing with uncertainties. This is an advantage for driver modelling, where uncertainty plays a significant role. A drawback that simple Bayesian networks share with neural networks is the difficulty of including temporal information. A more complex form of Bayesian networks, dynamic Bayesian networks, can be used to model changes in a variable over time. However, this makes the construction and

analysis of a network more complex.

Fuzzy logic is a form of logic used for approximate, rather than exact reasoning. Since driving a vehicle is largely a reasoning process, it is intuitive to use fuzzy logic to model driving behaviour, especially in the context of driving manoeuvres. A prediction system based on fuzzy logic was presented in [26] to distinguish emergency braking from merely strong braking behaviour. Ohashi *et al.* also predicted left and right turns using fuzzy logic and case-based learning [22]. Khodayari *et al.* et al. implemented a car-following model based on fuzzy logic aimed at predicting the driver's car-following behavior [11].

Hidden Markov models (HMM) are another type of probabilistic networks, a special case of Bayesian networks. The purpose of an HMM is to estimate a Markov chain. An example of a driver model using HMMs was proposed by Kuge *et al.* [16]. The model consisted of three HMMs, one for each of these manoeuvres: emergency lane change, ordinary lane change and lane keeping. Authors in [10] developed several driving models based on HMM and its extensions to predict turns and lane changes. Other driver models that predict lane changes based on HMM have been proposed in [6], [23], [24]. Since the strength of HMMs lies in sequential pattern recognition, they are suited for the prediction of driving manoeuvres, given that these are sequential in nature.

5.3 Proposed Method

Several factors such as the driver, the vehicle, and surrounding context influence driver intent and manoeuvres. In order to predict driving manoeuvres, we need to model the driver's intent and driving context jointly. A driver

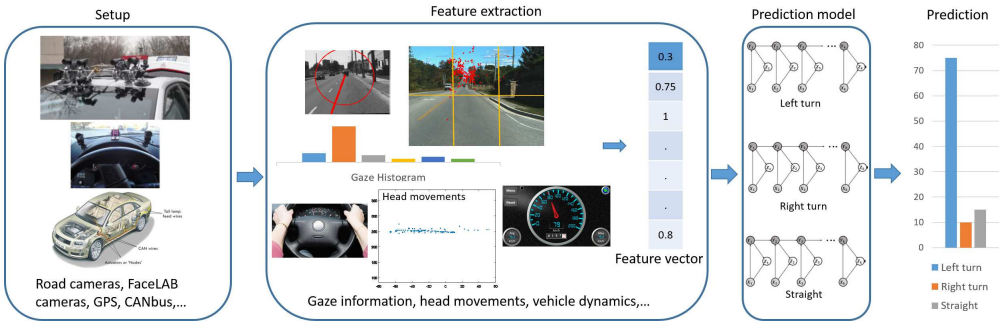


Figure 5.2: *Overview of the proposed approach for predicting driving manoeuvres.*

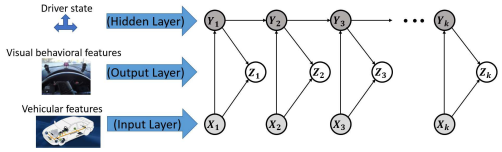


Figure 5.3: **IO-HMM Layers:** *The model includes a Hidden Layer, representing the driver's state; an Output Layer, representing features related to driver cephalo-ocular behaviour; and an Input Layer representing features related to vehicle dynamics.*

behaviour model requires the ability to accept data time series concerning the driving context. Hence, a model such as HMM is better suited than discriminative models such as SVM or others that do not consider these temporal aspects. We chose an Input Output Hidden Markov Model (IO-HMM, an extension of HMM), to build a driving manoeuvre prediction model. Figure 5.2 shows an overview of our system.

5.3.1 Modelling Driver Manoeuvres Using IO-HMM

A Hidden Markov Model (HMM) is a probabilistic model of two sets of random variables, states and outputs [25]. States result from a stochastic process and their evolution cannot be observed over time, but only through substantiation

Figure 5.4: *Gaze points are shown on driving images 5 seconds before a left turn, going straight, and a right turn. Images are divided into six regions.*

of the outputs. IO-HMM is an extension of an HMM in which the distribution of both states and outputs are influenced by a set of input variables [25]. Input variables are associated to the observed sequences in a classification problem where the output variables are considered as classes. IO-HMM presents similarities to HMM, but utilizes the same process as recurrent neural networks to map input sequences to output sequences. The training structure for IO-HMMs is more discriminant than for HMMs and takes advantage of EM algorithms [5]. An IO-HMM graphical model for driving manoeuvres prediction is depicted in Figure 5.3.

5.3.2 Cephalo-Ocular Behaviour and Vehicle Dynamics Features

Eyes play an important role in detecting driver intent since the driver looks directly at objects prior to taking action based on the information provided by

the fixation [2]. We believe that combining cephalo-ocular behaviour (3D Line of Gaze (LoG) and head pose) and vehicle dynamics can yield a predictive model of driver manoeuvres.

We consider two features related to the cephalo-ocular behaviour of the driver, namely the 3D Point of Gaze (PoG) in absolute coordinates, and the horizontal (left-right) head motion extracted from the head pose data. In order to relate the 3D LoG of the driver to a 3D PoG, a cross-calibration technique due to Kowsari *et al.* is used to transform the 3D LoG expressed in the coordinates of the eye-tracker into that of the forward stereo camera system of the experimental vehicle [14]. The 3D PoG is obtained by intersecting the projected 3D LoG onto the imaging plane of the stereo scene with a valid depth estimate. We divide the image scene to six non-overlapping rectangular regions (as shown in Figure 5.4) and construct a histogram of 3D PoGs falling into these regions. Figure 5.4 shows the PoGs over the last 5 seconds before a manoeuvre (left turn, right turn or going straight) occurs. As it is clear in the picture, drivers pay attention to different parts of the scene when deciding to perform different manoeuvres. We also calculate horizontal driver's head movements and build a 20-bin histogram to track the driver's head movement prior to a manoeuvre.

The feature vector forming the input layer of our model includes the data captured from the vehicle's CANbus network. Contemporary vehicles equipped with on-board diagnostic systems (OBD-II) allow sensors to report on current status and constitute the interface through which odometry is made available in real-time. Since 2008, the CANbus protocol has become mandatory for OBD-II. This standardization simplifies the real-time capture of vehicular data. We encode steering wheel angle and speed of the vehicle as features. We make a histogram of steering wheel angles over the last 5 seconds before

Figure 5.5: *Our data set.*

a manoeuvre occurs. We also calculate the minimum, maximum and average speed of the vehicle over this time.

Table 5.1: Description of Driving Sequences Used For Experiments.

Seq. #	Date Of Capture	Age	Gender	Temperature	Weather Condition
Seq. 8	Sep. 12, 2012	21	M	27°C	Sunny
Seq. 9	Sep. 17, 2012	21	F	24°C	Partially Cloudy
Seq. 10	Sep. 19, 2012	20	M	8°C	Sunny
Seq. 11	Sep. 19, 2012	22	F	12°C	Sunny
Seq. 13	Sep. 21, 2012	23	M	19°C	Partially Sunny
Seq. 14	Sep. 24, 2012	47	F	7°C	Sunny
Seq. 15	Sep. 24, 2012	44	F	13°C	Partially Sunny

5.4 Experimental Setup

5.4.1 Driving Sequences

To evaluate our prediction model, we used several driving sequences recorded by driving our instrumented vehicle in the urban area of London Ontario, Canada [3]. These sequences consist of natural driving sequences[1] with the sum of the aforementioned information, such as gaze, head pose, GPS data, vehicle speed, and steering wheel angle. Figure 5.5 depicts a few samples from our 3TB data set. Table 5.1 describes the driving sequences we used to perform the experiments. We annotated the driving videos with 220 events including 65 left turns, 75 right turns, and 80 randomly sampled instances of driving straight. Each turn annotation marks the start time of the manoeuvre before the vehicle starts to yaw.

5.4.2 Prediction Procedure

Algorithm 2 depicts the complete procedure of our prediction model using **IO-HMM**. As it was shown in Figure 5.3, the hidden layer Y, which represents the driver's state depends on the input layer X. The input layer represents features obtained by the vehicle's CANbus data network[2] (vehicle dynamics). The output layer Z describes cephalo-ocular behavioural features (driver gaze and head pose). $\mathbf{P}_M$ is the probability of manoeuvre M.

[1] The stereo cameras operate at 30 frames/sec.

[2] A Controller Area Network (CANbus) is a vehicle bus standard designed to allow micro controllers and devices to communicate with each other in applications without a host computer.

Algorithm 2 *Predicting Driving Maneuvers Using IO-HMM*

Input: Cephalo-Ocular Behaviour and Vehicle Dynamics Features
Output: Probability of each maneuver
  while *driving* **do**
    Extract features $\mathbf{Z}_1^k$ and $\mathbf{X}_1^k$ (Sec. III. B)
    Calculate probabilities of manoeuvres $\mathbf{P}_M = P(M|\mathbf{Z}_1^k, \mathbf{X}_1^k)$
    Choose a manoeuvre if any of $\mathbf{P}_M > 0.7$
  end while

5.5 Experimental Results

To evaluate the accuracy of our method, we needed to calculate how correctly our algorithm anticipates future manoeuvres. We anticipate manoeuvres every 20 frames (0.67 seconds) at which times the algorithm performs a series of processes on the recent driving data. The prediction system produces three probabilities for left turn, right turn, and driving straight events which together sum to 1. After prediction, the system chooses one of the manoeuvres based on these probabilities. If any of the probabilities is above a threshold[1], the system picks that manoeuvre, and chooses otherwise to make no prediction. We rate our algorithm performance using precision (P_r) and recall (R_e) scores:

$$P_r = \frac{t_p}{t_p + f_p} \tag{5.1}$$

$$R_e = \frac{t_p}{t_p + m_p} \tag{5.2}$$

where t_p is the sum of correctly predicted manoeuvres, f_p the sum of wrongly predicted manoeuvres, and m_p the manoeuvres that were wrongly not predicted (the system does not choose any manoeuvre). The precision P_r mea-

[1]The prediction threshold is set to 0.7 for our experiments.

sures the fraction of predicted manoeuvres that are correct, while recall or sensitivity R_e measures the fraction of manoeuvres that are correctly predicted. We also calculate the average Time-to-Manoeuvre which determines the time between prediction and that of the start of a manoeuvre.

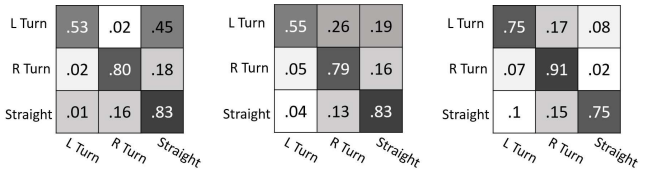
Table 5.2: Results of driving manoeuvres prediction on our data set.

	P_r (%)	R_e (%)	Time-to-Manoeuvre (s)
IO-HMM	74.5	76.6	2.9
IO-HMM G	74.2	78.8	3.2
IO-HMM H	77.9	80.3	3.4
IO-HMM G+H	79.5	83.3	3.8

We use a 5-fold cross validation process to train and test our prediction model. Table 5.2 reports the precision, recall, and Time-to-Manoeuvre for anticipating driving manoeuvres (left turn, right turn, and straight) under three settings:

- **IO-HMM**: the prediction model considers only vehicle dynamics information
- **IO-HMM G**: the prediction model considers only gaze information in the output layer
- **IO-HMM H**: the prediction model considers only head information in the output layer
- **IO-HMM G+H**: the prediction model considers both gaze and head information in the output layer

As Table 5.2 indicates, both precision and recall are higher for **IO-HMM G+H**, where both driver's gaze information and head motion are taken into



(a) ***IO-HMM G*** **(b)** ***IO-HMM H*** **(c)** ***IO-HMM G+H***

Figure 5.6: *Confusion matrices for our prediction model.*

account for building the prediction mechanism. The model predicts manoeuvres 3.8 seconds before they occur, on average. It is interesting to note that while precision is roughly equivalent between **IO-HMM** and **IO-HMM G**, both recall and Time-to-Manoeuvre improve (by 2.2% and 0.3s, respectively). A similar but more pronounced effect occurs when comparing **IO-HMM** with **IO-HMM H**, where all metrics improve. Lastly, **IO-HMM G+H** improves precision by 5%, recall by 6.7% and Time-to-Manoeuvre by 0.9s over **IO-HMM**. These results empirically demonstrate the value of cephalo-ocular behaviour for the prediction of driver-initiated manoeuvres.

Figure 5.6 shows the confusion matrices for joint prediction of all the manoeuvres. Modelling manoeuvre prediction with **IO-HMM** provides a discriminative modelling of the state transition probabilities using features extracted from vehicle dynamics and driver cephalo-ocular behaviour. Figure 5.7 plots the F_1-score changes for different values of the prediction threshold and shows how the threshold can act as a trade-off between precision and recall. The F_1-score is the harmonic mean of precision and recall, defined as

$$F_1 = \frac{2P_rR_e}{P_r + R_e}. \tag{5.3}$$

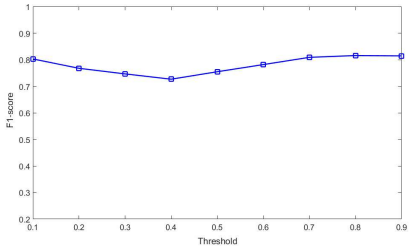


Figure 5.7: *The plot shows the impact of prediction threshold on F_1-score for IO-HMM G+H.*

We observe that the F_1-scores from our prediction algorithm remain relatively stable as the threshold changes. The prediction model anticipates manoeuvres every 0.67 seconds and processes the last 20 frames to current time. The system predicts driving manoeuvres under 3 milliseconds on average on a 3.40GHz Core i7-6700 CPU with Windows 10.

5.6 Conclusion

We developed a prediction model using **IO-HMM** that anticipates a particular type of driving manoeuvres. Our focus was on demonstrating that eye and head movements are predictive of driver-initiated manoeuvres. We utilized features extracted from the cephalo-ocular behaviour of drivers and vehicular dynamics as inputs to our predictive model. Experimental results empirically proved that cephalo-ocular behaviour is at least a partial predictor of driver intent. Future work includes demonstrating this fact for other types of driver manoeuvres.

Bibliography

[1] Hideomi Amata, Chiyomi Miyajima, Takanori Nishino, Norihide Kitaoka, and Kazuya Takeda. Prediction model of driving behavior based on traffic conditions and driver types. In *12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–6, 2009.

[2] Dana H. Ballard, Mary M. Hayhoe, and Jeff B. Pelz. Memory representations in natural tasks. volume 7, pages 66–80. MIT Press, 1995.

[3] S.S. Beauchemin, M.A. Bauer, D. Laurendeau, T. Kowsari, J. Cho, M. Hunter, and O. McCarthy. Roadlab: An in-vehicle laboratory for developing cognitive cars. In *23rd International Conference on Computer Applications in Industry and Engineering (CAINE 10)*, pages 7–12, 2010.

[4] Yoshua Bengio and Paolo Frasconi. An input output hmm architecture. In *Advances in neural information processing systems*, pages 427–434. Morgan Kaufmann Publishers, 1995.

[5] Yoshua Bengio and Paolo Frasconi. Input-output hmms for sequence processing. In *IEEE Transactions on Neural Networks*, volume 7, pages 1231–1249, 1996.

[6] Holger Berndt and Klaus Dietmayer. Driver intention inference with vehicle onboard sensors. In *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pages 102–107, 2009.

[7] Ueruen Dogan, Hannes Edelbrunner, and Ioannis Iossifidis. Towards a driver model: Preliminary study of lane change behavior. In *11th International IEEE Conference on Intelligent Transportation Systems*, pages 931–937, 2008.

[8] Bjorn Frohlich, Markus Enzweiler, and Uwe Franke. Will this car change the lane?-turn signal recognition in the frequency domain. In *Intelligent Vehicles Symposium Proceedings*, pages 37–42. IEEE, 2014.

[9] Kelsey P. Hawkins, Sunny Bansal, Nam N. Vo, and Aaron F. Bobick. Anticipating human actions for collaboration in the presence of task and sensor uncertainty. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2215–2222, 2014.

[10] Ashesh Jain, Hema S. Koppula, Bharad Raghavan, Shane Soh, and Ashutosh Saxena. Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3182–3190, 2015.

[11] Alireza Khodayari, Reza Kazemi, Ali Ghaffari, and Reinhard Braunstingl. Design of an improved fuzzy logic based model for prediction of car following behavior. In *IEEE International Conference on Mechatronics (ICM)*, pages 200–205, 2011.

[12] Kris Kitani, Brian Ziebart, James Bagnell, and Martial Hebert. Activity forecasting. In *Computer Vision–ECCV*, pages 201–214. Springer, 2012.

[13] Hema S. Koppula and Ashutosh Saxena. Anticipating human activities using object affordances for reactive robotic response. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 38, pages 14–29, 2016.

[14] T. Kowsari, S.S. Beauchemin, M.A. Bauer, N. Laurendeau, and D. Teasdale. Multi-depth cross-calibration of remote eye gaze trackers and stereo-

scopic scene systems. In *IEEE Intelligent Vehicles Symposium Proceedings*, pages 146–150, 2014.

[15] K.P. Kraiss and H. Kuttelwesch. Identification and application of neural operator models in a car driving situation. In *IFAC Symposia Series*, volume 5, pages 121–126, 1993.

[16] Nobuyuki Kuge, Tomohiro Yamamura, Osamu Shimoyama, and Andrew Liu. A driver behavior recognition method based on a driver model framework. Technical report, SAE Technical Paper, 2000.

[17] Pranaw Kumar, Mathias Perrollaz, Stéphanie Lefevre, and Christian Laugier. Learning-based approach for online lane change intention prediction. In *Intelligent Vehicles Symposium (IV)*, pages 797–802. IEEE, 2013.

[18] Charles C. MacAdam and Gregory E. Johnson. Application of elementary neural networks and preview sensors for representing driver steering control behaviour. In *Vehicle System Dynamics*, volume 25, pages 3–30. Taylor & Francis, 1996.

[19] Dejan Mitrovic. Machine learning for car navigation. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 670–675. Springer, 2001.

[20] Brendan Morris, Anup Doshi, and Mohan Trivedi. Lane change intent prediction for driver assistance: On-road design and evaluation. In *Intelligent Vehicles Symposium (IV)*, pages 895–901. IEEE, 2011.

[21] Vicki L. Neale, Thomas A. Dingus, Sheila G. Klauer, Jeremy Sudweeks, and Michael Goodman. An overview of the 100-car naturalistic study

and findings. In *National Highway Traffic Safety Administration*, number 05-0400, 2005.

[22] Kunihiro Ohashi, Toru YAMAGUCHI, and Ikuo Tamai. Humane automotive system using driver intention recognition. In *SICE annual conference*, pages 1164–1167, 2004.

[23] Nuria Oliver and Alex P. Pentland. Graphical models for driver behavior recognition in a smartcar. In *Proceedings of the IEEE Intelligent Vehicles Symposium(IV)*, pages 7–12, 2000.

[24] D. Polling, Max Mulder, Marinus Maria van Paassen, and QP. Chu. Inferring the driver's lane change intention using context-based dynamic bayesian networks. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 1, pages 853–858, 2005.

[25] Lawrence Rabiner and B. Juang. An introduction to hidden markov models. In *IEEE ASSP magazine*, volume 3, pages 4–16, 1986.

[26] J. Schmitt and B. Färber. Verbesserung von fas durch fahrerabsichtserkennung mit fuzzy logic. In *VDI-Berichte*, number 1919, 2005.

[27] Jane C. Stutts, American Automobile Association, et al. The role of driver distraction in traffic crashes. In *Foundation for Traffic Safety Washington, DC*, 2001.

[28] Shigeki Tezuka, Hitoshi Soma, and Katsuya Tanifuji. A study of driver behavior inference model at time of lane change using bayesian networks. In *IEEE International Conference on Industrial Technology (ICIT)*, pages 2308–2313, 2006.

Chapter 6

Conclusion and Future Work

Nowadays, driving is a daily, fun and yet intricate task which is composed of several critical subtasks. To explore drivers' intentions in different situations during driving, we must model and analyze drivers' behaviors. Researchers are trying to understand and model driver behavior to predict the most probable next maneuver and assist the driver to make a good decision. In chapter 5, we utilized the driver cephalo-ocular behavior to specify the relationship between this information (gaze position, head position, ...), current driving maneuvers, vehicular attitude and driving behavior of the driver in order to predict the next maneuver. We developed a prediction model using IO-HMM that anticipates turning maneuvers. Experimental results proved that cephalo-ocular behavior and visual search patterns of drivers can be demonstrative of driver intentionality.

In chapter 4, we proposed a novel approach to improve vehicle localization accuracy by estimating vehicle position and orientation which that minimize the observed difference between detected lane features and projected lane-marking splines using a particle filter.

Consequently, we proposed algorithms to detect surrounding objects such as vehicles and traffic signs within the attentional visual area of drivers to investigate their impact on predicting driving maneuvers. Our experiments showed that our method robustly detects vehicles and signs in driving scenes.

Our contribution to this research can be summarized as follows:

1. Constructing a real-time system to detect vehicles within the attentional visual area of drivers.

2. Presenting an approach to detect and recognize traffic signs inside the attentional field of drivers.

3. Proposing a novel approach to improve vehicle localization accuracy.

4. Developing a prediction method to anticipate driver intent.

5. Annotating the sequences. (turning maneuvers, signs, vehicles, ...)

6.1 Future Work

Research on driver intent and advanced driving assistance systems are relatively recent with the potential for significant results and applications in the near future. Here are a few possible research areas that may be undertaken immediately:

1. Objects detected within the attentional visual area of drivers can be used in the driving maneuver prediction model.

2. We believe similar detection systems could easily be developed to identify other objects drivers routinely encounter and attend to, such as pedestrians, cyclists, traffic lights, and more.

3. While the instrumentation represents a successful proof of concept, it was noted that wider viewing angles for the stereo cameras and eye-trackers using more than two cameras (to compensate for head rotations) would allow us to track the 3D driver gaze into the surroundings in a more comprehensive manner.

4. The physical limitations of the instrumentation prevented its use at night and in adverse weather conditions. Such limitations could be removed entirely by a judicious choice of hardware, enabling the study of driver intent in diverse conditions.

Vita

NAME:	Seyed Mohsen Zabihi
POST-SECONDARY EDUCATION AND DEGREES:	Western University London, Canada 2013-2017 PhD in Computer Science
	Ferdowsi University Mashhad, Iran 2004-2009 B.Sc. in Computer Eng.
	Ferdowsi University Mashhad, Iran 2009-2012 M.Sc. in Computer Eng.
AWARDS:	WGRS scholarship for 2013-2017, University of Western Ontario
RELATED WORK EXPERIENCE:	Application Developer Klick Inc. 2017 - present
	Teaching Assistant University of Western Ontario 2013 - 2017

Publications:

S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer, Real-Time Driving Manoeuvre Prediction Using IO-HMM and Driver Cephalo-Ocular Behaviour, IEEE Intelligent Vehicles Symposium (IV17), Redondo Beach, California, USA, June 11-14 2017.

S.J. Zabihi, S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer, Detection and Recognition of Traffic Signs Inside the Attentional Visual Field of Drivers, IEEE Intelligent Vehicles Symposium (IV17), Redondo Beach, California, USA, June 11-14 2017.

Zabihi S.M.; Beauchemin S.S.; de Medeiros G.; and Bauer M.A., Lane-based vehicle localization in urban environments, IEEE International Conference on Vehicular Electronics and Safety (ICVES), Yokohama, Japan 2015.

Zabihi S.M.; Beauchemin S.S.; de Medeiros G.; and Bauer M.A., Frame-Rate Vehicle Detection within the Attentional Visual Area of Drivers, IEEE Intelligent Vehicles Symposium (IV14), Dearborn, MI, USA, June 8-11 2014.

Ghadiri, F.; Zabihi, S.M.; Pourreza, H.R.; and Banaee, T., A Novel Method for Vessel Detection Using Contourlet Transform, National Conference on Communications, IEEE Communication Society, India, 2012.
Tabatabaee, H.; and Zabihi, S.M., Design of Braille alphabet Display for Persian texts using Thermal cells with capability of connectivity to computer, International Review on Computers and Software (IRECOS), 2012.
Zabihi, S.M.; and Akbarzadeh-T, M.R, Generalized Fuzzy C-Means Clustering with Improved Fuzzy Partitions and Shadowed sets, Journal of ISRN Artificial Intelligence, 2011.
Zabihi, S.M.; Pourreza, H.R.; and Banaee, T., Vessel Extraction of Conjunctival Images Using LBPs and ANFIS, Journal of ISRN Machine Vision, 2011.