PREDICTION OF DRIVER BEHAVIOR

Thesis submitted for the degree of Doktor-Ingenieur
(Dr.-Ing.)

Universität Bielefeld
Technische Fakultät
Universitätsstr. 25
33615 Bielefeld
Germany

Michaël Garcia Ortiz
March 2014

Acknowledgments

I would like to thank people which helped me during the three years of this PhD work. Working on this new and fascinating scientific topic which is driver behavior prediction would not have been possible without the inspirational discussions and pieces of advice I received by researchers in Honda Research Institute Europe GmbH and in the Laboratory for Cognition and Robotics.

First of all, I would like to thank Alexander Gepperth, for giving me the chance to have my first steps in the research world. After a master thesis, a PhD and a Post-doc with him, I am sure he secretly hopes to get rid of me one day. His views on Artificial Intelligence and research in general are an inspiration. I wish to thank him and his wife, Britta, for their friendship and their hospitality. It is always a pleasure to be around them.

I would like to thank the people in Honda Research Institute Europe GmbH. I learned a lot on a wide range of scientific fields thanks to them. I am also thankful for all the nice moments we had, and the friendly atmosphere I encountered there. Especially, I wish to thank Jens Schmüdderich, Jannik Fritsch, and Thomas Weisswange for their support and guidance. For the same reasons, and for accepting to be part of my examination committee, special mention goes to Christian Goerick. I would also like to thank Burkhard Zittel, our dear system administrator, for his never-ending bright mood and his skills in repairing all the mistakes I made with his computers. I am thankful to the management team for allowing me to pursue this thesis in such an environment. Finally, I would like to thank the PhD students I encountered at Honda, Tobias Kühnl, Sarah Bonnin, Mathias Platho, and Cem Karaoguz. I had a lot of fun with them and with the other students, at work or around for a drink.

I wish to thank the people in Bielefeld for their kindness every time I went there. Especially, I would like to thank Prof. Dr.-Ing. Franz Kummert for his help throughout my thesis, and for his precious pieces of advice. I wish to thank Prof. Dr.-Ing. Helge Ritter and Dr. Thies Pfeiffer for accepting to be part of my examination committee on a really short notice.

Speaking of last moments, I would like to once again thank Jens Schmüdderich for his patience at the end of my PhD. He kindly accepted to be my supervisor even if he had no time to spare, and I learned a lot with him, not only on the scientific side, but also on the human side. He taught me a lot about responsibility, and I think I learned one great lesson thanks to him.

I would like to thank my family for their support. They always encouraged me, and I am where I am today thanks to them. I wish to thank my friends from Paris and from Grenoble. Being away was not easy, but knowing that they were always present when I came back helped a lot.

Finally, I would like to thank Justine Roure. I was lucky enough to find this one person who makes everything easy, who understands me, and who is always here to support me. I would not have achieved anything if it was not for her. Merci demoiselle.

Contents

Chapter 1

Introduction and Motivation

1.1 Context of the work

This work took place as a partnership between Honda Research Institute Europe GmbH and The Laboratory for Cognition and Robotics of the Bielefeld University (see Fig. 1.1). Predicting the behavior of a human being, in the context of driving situations, is both a challenging research topic and a source of possible applications with high value for the industry. In this first section we will describe the scientific implications and the motivations of this thesis.

Figure 1.1: Honda Research Institute Europe GmbH and The Laboratory for Cognition and Robotics of the Bielefeld University collaborated to work on the topic of driver behavior prediction.

1.1.1 Reasons for researching intelligent systems for cars

Since several years, the scientific community and the car industry are working on the next generation of vehicles incorporating intelligent functionalities.

Before introducing the field of research, we start with a definition of intelligent systems for vehicles. We define it as a system that observes the driving environment, detects participants and relevant information about the driving scene, interprets what is happening (which requires some degree of understanding) and possibly predicts what is going to happen, and finally takes actions or gives feedback to the driver. The applications of such intelligent systems are the improvement of traffic safety [73, 31], reduction of energy consumption [86, 14] or improvement of the comfort of the driver.

Tremendous innovations in the area of autonomous or semi-autonomous vehicles and driving assistance systems are presented each year by companies or research groups. Google is developing a fleet of autonomous cars since several years [87]. They presented a vehicle which incorporates sensing abilities and the possibility to control the actuators (pedals, signals and steering wheel) using embedded computers. Google released several videos and made multiple demonstration of the autonomous driving capability of their prototype car.

Similarly, Toyota recently presented their semi-autonomous car, developed in order to improve driving safety [38]. Stanford and Volkswagen are collaborating in order to create an autonomous racer [1], which is able to perform maneuvers at high speed on a test track. Examples of prototype cars can be observed in Fig. 1.2.

It is difficult, due to confidentiality of the projects, to obtain information about this kind of prototypes. Of course, the time when we will be able to sit in our car, start the auto-pilot mode, and read the news while our vehicle drives us to work has yet to come, and it is not easy to estimate when these products will be ready for the market, but it might occur sooner than first expected. Autonomous vehicles are the science fiction of yesterday and the reality of tomorrow.

Figure 1.2: Left image: Google driverless car operating on a testing path. Right Image: Junior, a robotic Volkswagen Passat, at Stanford University in October 2009 (Pictures by Steve Jurvetson).

It is not hard to imagine why this area of research is an interesting one, both for the researcher and for the car manufacturer. It allows researchers to work on topics such as computer vision, planning and reasoning, trajectory estimation, or interaction between human and machine, and to apply and see the fruit of their work on a real-world application. Intelligent vehicles hold the promise to one day remove the main reason for accidents from the equation: the human driver. Finally, the implications of autonomous or assisted driving for society are considerable, in terms of possible improvement of safety, or reduction of energy consumption,

Improve driving safety by preventing accidents from happening

Every year, we face a large number of casualties due to traffic accidents. Several studies by the World Health Organization gave an alarm signal regarding the amount of killed and injured in road traffic related accidents (see [70] and [71]). These reports estimated that in 2002, 1.2 million persons died in traffic-related accidents, and that between 20 and 50 million are injured every year. We extracted several relevant numbers about the amount of killed and injured in 2007 in several developed countries, which we presented in Tab. 1.1 as an illustration of the necessity of improving road safety. The data are provided by the countries themselves, so they depend on the local definition of what a traffic-related injury is. This could explain the huge difference between France and Germany observed in Tab. 1.1. Even if the amount of killed is similar, the amount of injured differs greatly.

Still, the numbers are impressive, and the impact on society is enormous. On a cold and pragmatic economic level, these accidents cost a lot of money in destruction of infrastructure, mobilization of rescue teams, and hospital fees. The same reports estimate that traffic accidents each year cost around 518 billion euros worldwide, and on average 1 to 2 % of the Gross Domestic Product. In Western Europe alone, traffic accidents cost each year 180 billion euros. In order to appreciate the amount of money wasted, we can compare it to the budget of the European Union for the period 2007-2013: 864.3 billion euros [90], or to the

Country	Total killed	Total injured	#injured per 10000 persons
Germany	4949	431419	53
France	4620	77007	12
United States	42642	3305237	108
United Kingdom	3298	264288	43
Italy	5669	332995	57
Japan	6639	1034445	81

Table 1.1: Amount of road casualties in representative countries in 2007. Data from World Health Organization (see [70] and [71]).

regular budget of United Nations for 2012-2013: 5.15 billions dollars [66]. On the human side, accidents leave victims injured both physically and psychologically. Of course it also affects the well-being of their family, their productivity at work, they suffer long-term troubles. The real cost of an accident can go well beyond the limits of immediate material destruction, and is impossible to evaluate.

Researchers and car companies are looking for solutions in order to reduce the amount of accidents. They already developed a considerable set of technologies in order to decrease the amount of casualties. Most of them (like airbags, seat-belts, anti-lock braking systems, shock absorbing car bodies) are efficient in decreasing the impact of an accident, and in protecting the passengers of the car. These technologies already saved a lot of lives, but they are rarely able to avoid accidents because they do not anticipate them. Moreover, if they are protecting in many cases the passengers of the car, they do not prevent more vulnerable traffic participants, like pedestrians or bicyclists, from getting injured. The next step for safety systems is to prevent accidents from happening, instead of just reducing their impact. This can only be possible using intelligent systems that can observe the driving environment, reason and decide if there is a danger, determine how to avoid it and act if necessary. As an illustration of situations where such systems could be useful, please refer to Fig. 1.3.

Figure 1.3: Typical situations where safety systems need the perception of the scene in order to handle possibly dangerous situations. Unexpected situations or traffic participants behaving in a dangerous way are not a rare occurrence, and intelligent systems could help drivers avoid accidents.

Reduce energy consumption by optimizing the driving

The reduction of energy consumption is certainly one of the main challenges we have to face for the next decades. Transportation is one of the major factors in fossil energy consumption, and it is also responsible for a large amount of CO_2 pollution. The European Commission estimated that light-duty vehicles emit 15 % of European Union's CO_2 (see, e.g. [27, 28]). Considering the objectives of the Kyoto protocol [65], reducing the emission of ground transportation is an important part of the solution. At the same time, we need to be pragmatic: we live in a society where personal means of transportation are not only considered indispensable, but also part of the culture. It is difficult to ask individuals to voluntarily limit the use of their vehicle if they do not have a strong incentive to do so, especially in regions where vehicles are indispensable to go to work every day. It stands to reason that if it is difficult to decrease the amount of vehicles, part of the solution is to make them more energy-efficient.

Improvements have been made on the engines, which are more and more optimized and need less fuel to operate, and hybrid and electric cars have been developed and are continuously being improved. But we can go beyond these solutions that do not take into account the environment in which a vehicle is driving. A growing number of scientific contributions presented intelligent systems used in order to improve energy efficiency and reduce fuel consumption, based on the optimization of the way vehicles are performing maneuvers, namely providing eco-driving functionalities (see [7, 11]). As an example, some models of hybrid cars can save energy using regenerative braking, by converting the kinetic energy to recharge batteries (see [94]). If a system can anticipate what the driver will do, it can optimize the trajectory in order to limit the fuel consumption and use predicted decelerations in order to recharge the batteries for the electric engine more often and more efficiently.

Improve comfort by anticipating driving maneuvers

Finally, another application for intelligent vehicles is the improvement of driving comfort. Car industry is a very competitive market, and companies are looking for new segments where they can distinguish themselves from competitors. Many potential customers will be attracted by systems improving comfort while driving, so part of the research in intelligent systems for cars focuses on how to improve the driving experience, i.e. make it easier and more enjoyable.

As an example, lane keeping assistant systems are technologies that actively keep the vehicle in the lane on highways if the driver drifts out of it. Automatic speed regulation keeps the car at a certain speed without requiring to touch the gas pedal. This can be really interesting for, e.g., truck drivers that spend a lot of time on highways. But these technologies have a limitation: in the case of automatic speed regulation, this technology can not cope if a vehicle ahead drives slower than the desired speed, or if another vehicle cuts into the lane. This case requires the driver to have a constant focus on the road. In order to provide more comfort, it is better if the system can adapt to changes in its dynamic environment: let the vehicle adapt to the speed of the preceding vehicle, or autonomously change lane when required. Again, this requires knowledge about the environment, detection capabilities, reasoning and action planning. Intelligent systems can be used in order to create more attractive and more enjoyable vehicles.

Figure 1.4: Traffic congestion is a problem for energy consumption as well as for driving comfort (images free of rights). They can be avoided by reducing traffic accidents, and by optimizing the way cars maneuver.

1.1.2 Scientific challenges around intelligent vehicles

Whether it is for safety, eco-driving or comfort driving, we clearly see the objectives and possibilities related to intelligent vehicles. But do we have the technical means to implement such systems? We can identify three families of research challenges corresponding to perception, reasoning and action.

Perception of the environment

In order to understand the scene and possibly warn the driver or perform maneuvers, a Driver Assistance System has to be able to perceive its environment. Two approaches, which are not exclusive, are available: ego-centered sensors and Vehicle to Infrastructure or Vehicle to Vehicle (V2I/V2V) communication.

Ego-centered perception is based on sensors mounted on the vehicle. Numerous technologies exist to perceive the traffic scenes. Recent contributions in the domain of Intelligent Transportation Systems introduced advanced perception systems involving combinations of state-of-the-art sensors (Radar, LIDAR, cameras) and complex algorithms (SLAM [62], texture-based analysis [51], object detection [25]) in order for a vehicle to perceive its environment. Vision sensors such as cameras are used together with pattern recognition algorithms to detect traffic participants [24, 53]. As an example, results from pedestrian detection based on monocular camera images are presented in Fig.1.5. Range

Figure 1.5: Example of pedestrian detection using camera images and pattern recognition. Reproduced with permission from [35].

sensors like radar, lidar or velodyne sensors allow to determine their distance and speed of the equipped vehicle to other traffic participants, or estimate the geometry of the scene [62, 64, 85]. Finally, inertial sensors are used to measure the motion of the vehicle equipped, and a whole array of dedicated sensors built into the car serve as driver monitoring. In addition, direct driver monitoring like head cameras or foot cameras are used to measure his physical behavior inside the car [63]. The combination of these different types of sensors allow high quality detection of relevant driving scene elements: infrastructure and traffic participants.

Communication between traffic participants and infrastructure, namely Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) technologies, is a different approach. Traffic participants can share information with other traffic participants [8] and with the infrastructure [75] using wireless communication, which can lead to great advantages. For example, instead of having expensive vision sensors inside a vehicle, a fixed external sensory system positioned at key locations (a zebra crossing for example) could send information about the scene to traffic participants. Also, traffic participants could share their internal states and their position with other traffic participants and with the infrastructure. Thus, the reliability of the information improves: instead of depending on range sensors, the information comes from the internal sensors of the vehicles themselves, or from the infrastructure, and can come from multiple sources.

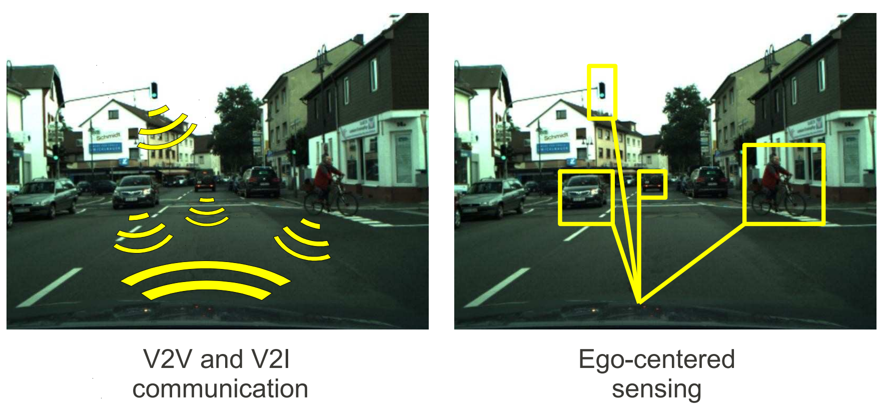

Figure 1.6: V2V and V2I communication allows traffic participants and traffic infrastructure to share their status, whereas ego-centered perception relies on the combination of cameras and range sensors to perceive the surroundings of the vehicle.

The possibility of cars equipped with sensory systems sharing information may improve scene representation by avoiding occlusion issues and having multiple points of views for traffic participants detection. The perception becomes a network of sensors which can be fused [13]. This would once again require a standard communication language between cars of different manufacturers and infrastructures.

So the detection technologies exist and are being developed. Even if they are not as good as human vision system, the recent advances are impressive, and the surrounding environment of the vehicles can be analyzed with ever-growing precision (see [43, 62, 47]).

Reasoning about the driving scene

There is no intelligent system without some form of reasoning. The core functions of an intelligent car are the consolidation of the information provided by a perception system and the understanding of the driving scene (estimate immediate danger, predict present and future traffic participants behavior and trajectories).

Once data is acquired by a perception system, it has to be exploited in order to create an internal representation of the scene [34, 82]. Temporal and semantic consistency may be used in order to make the knowledge about the scene more reliable. Work is being conducted on filtering and tracking algorithms [3, 9], as well as on expectation maps [33]. It is used in order to ensure, e.g., that the system does not suddenly detect a bicycle in the sky, or that a car which was not here before suddenly appears in front of the ego-vehicle (the ego-vehicle is the vehicle where cameras and range sensors are mounted. It is the vehicle equipped with the perception and reasoning system).

The understanding of the scene is the step where the system tries to estimate what the traffic participants are doing, why they are doing it, tries to predict what they are going to do next. Based on present and past observations of the scene, the system has to use acquired or built-in knowledge in order to deduce what constitutes a danger, and whether it is safe to perform certain maneuvers [20, 93]. For example, is it safe to pass a following car on highway given the current configuration of the scene? This is a mandatory step: to compute future trajectories of the ego-vehicle, to anticipate possible dangers from the driving environment, and to identify safe driving maneuvers.

Our work will focus mainly on this part of intelligent systems for cars, and more precisely on the prediction of the behavior of traffic participants.

Influence of the system on the driver

We can classify intelligent systems in vehicles depending on the way they influence the driving process of the human driver. This can go from a small LED display on the dashboard to leaving the entire control to the system.

The first level of influence is to give feedback to the the driver in specific cases (for example in case of danger). This could be a sound ([84]), a vibration of the seat or the steering wheel ([10]), or a visual warning on the dashboard. It could also be a more complex signal, e.g. displaying information about the scene on the windshield, so that data is superimposed onto the visual scene. This technology, called heads-up display, uses augmented reality to provide important information to the driver [17].

The next level of influence is an extension of the previous one, where the driver is assisted in certain maneuvers by a system performing trajectory planning and partly controlling some actuators. Such a system detects when the driver is facing a situation that requires taking control. For example, an Advanced Driver Assistance System could perform an emergency breaking, or avoidance maneuver in case the driver does not react quickly enough to certain kinds of scenes. Most car companies are orienting their research in this direction for various reasons, one of them being that it is considered to be a step toward fully autonomous vehicles. This approach allows to develop specific assistant systems, which are active only when needed.

The last level of influence is autonomous driving. The system completely replaces the driver for both sensory tasks and performance of driving maneuvers. This is a very difficult task to solve, however the possibility of having a car without driver is particularly attractive. Many prototypes have been presented in the past few years, as introduced in Sec. 1.1.1. In [79], the authors present a prototype of autonomous car driving in inner-city in Braunschwieg, Germany. In France, INRIA experimented with a driverless vehicle to carry people in a pedestrian area [2]. Even if the prototypes were demonstrated only in simple scenarios, the results are impressive.

The technologies to improve the driving experience exist in principle. Whether it is by taking partial or total control of the car, or by simply providing valuable information to the driver, they allow intelligent vehicles to actively influence the way humans drive.

1.1.3 Developing Driver Assistance Systems at Honda Research Institute

Honda Research Institute's objective, like many other major car companies, is to develop intelligent systems to improve the driver's experience. Whether it is for safety, eco-driving or comfort, researchers in Honda Research Institute soon realized that some work outside of the classical themes of transportation science was needed in order to develop advanced functionalities. One needs to go beyond classical Driver Assistance Systems (DAS) in order to propose better safety, less fuel consumption and more comfort.

Researchers in Honda Research Institute already produced a great amount of work in the field of object detection and computer vision [21, 82, 60, 59, 41, 61, 46]. The following questions need to be asked: Is it possible to predict the behavior of traffic participants, and if it is, how to predict it? It was realized that understanding what the driver and what other traffic participants are doing will be a crucial information to develop truly advanced systems.

I have a lot of interest for the prediction of the future using observations in the present, for intelligent system design and for real-world application. Understandably, I wanted to participate in this project, with the aim to explore the possibilities of human behavior prediction in the car domain.

1.2 Necessity of behavior prediction for advanced driving applications

As will be shown in this section, predicting behaviors of the different traffic participants is a mandatory milestone toward intelligent vehicles. It is also a key element to improve the perception and interpretation of the environment.

1.2.1 Predicting the current behavior of traffic participants

Knowing why the driver or another traffic participant is doing what he does is a valuable information. Similarly, the difference between how the system expects a traffic participant to behave and how he actually acts holds a lot of information.

The human vision system is impressive, as it allows us to easily detect traffic participants on the road, in a dynamic, constantly changing environment. Despite decades of research in computational vision processing and the use of more and more precise sensors, not a single artificial vision system can reach comparable performance. All drivers have similar capacities when it comes to object recognition, and they usually do not make mistakes in recognizing a car or a pedestrian in a driving scene. Attention plays a major role while driving a car. As drivers, we can not look everywhere at the same time, and sometimes we miss some valuable information that could have prevented an accident, because attention is limited (see [89, 18]). The huge difficulty is to assess if a driver correctly comprehends the situation and if he reacts correctly given the driving scene.

By knowing how and when the driver usually reacts to certain scene elements, we can identify when he does not react properly, or when he might be late to react. For example, if the way a driver stops at a red traffic light, e.g. when he starts decelerating, or how steep is his deceleration, can be learned, then the system can detect when he is not sufficiently decelerating. By measuring the difference between what the system predicts (that is what the driver usually does) and what the system observes (what the driver is actually doing), we can detect unusual reactions to a certain scene, or the complete lack of reaction. If the driver approaches a pedestrian crossing, the system will predict that he has to start braking or decelerating at a certain distance from the crossing, depending on the presence and position of a pedestrian. If he does not, the difference between the prediction and the observed behavior can trigger a warning signal or a planned deceleration.

1.2.2 Predicting the future behavior of traffic participants

Inexperienced drivers are not as good as experienced drivers in anticipating what will happen next in a driving scene. In order to improve, the human brain develops prediction models based on experience (see [30]). Better anticipation of what other traffic participants are going to do in the near future provides more time for planning our own behavior. The same rule applies to intelligent vehicles.

In order to anticipate how the driving scene is going to evolve, we must predict the future behavior of the other traffic participants. Classic systems react to the current scene, but we have to go beyond that and react to a likely future scene. If an intelligent system can predict the future behavior of other traffic participants and anticipate what they are going to do, it can warn the ego-vehicle driver if another vehicle is likely to influence his behavior. Therefore, the driver can react before a dangerous situation can actually occur.

Another application of such anticipation is eco-driving. Anticipation of the ego-vehicle's behavior can lead to even better energy management: if the system can anticipate the future maneuvers of the driver, it can optimize them in order to be efficient and consume less energy. Moreover, in the case of electric and hybrid vehicles, if the system can predict future decelerations, it can directly use them to convert the kinetic energy in order to charge the batteries instead of using the brakes.

Figure 1.7: Illustration of the usefulness of behavior prediction. On the left image, behavior prediction of the ego-vehicle can help anticipate future trajectories and initiate eco-driving strategies. On the right image, predicting what other traffic participants will do on highway is necessary to anticipate danger or to provide advanced knowledge about the scene for autonomous driving.

1.2.3 Behavior prediction to improve perception systems

By predicting the future behavior of a traffic participant, we can anticipate where it is going to be located in the future driving scene. This prediction can be used in order to improve the perception system.

Usually, tracking algorithms use physical models to determine the movement of an object and its future positions in space (see [81, 6, 26]). By doing so, these algorithms filter the data in order to decrease the effect of noise. However most of these models do not take into account possible changes in behavior: they are based on the assumption that the movement continues with a certain regularity. What if a car is going to cut-in in front of the ego-vehicle on highway? Indeed, if a system tracks objects based only on physical trajectories, there will be no indicators of a sudden maneuver like a cut-in. However, if the system expects the car to cut-in, the tracking algorithm can adapt to this information and improve the tracking.

Another motivation for predicting the future behavior of traffic participant is the possible reduction of computing power needed by detection systems. The computational power is a precious resource in embedded systems. Knowing the future behavior of a traffic participant can be used to make the system focus the processing on certain relevant parts of the scene: where we predict the traffic participants will be.

1.3 Scope of this thesis

In this thesis. we will address selected topics related to behavior prediction that we deemed important. We chose to focus our work on the following topics: the application of machine learning for behavior prediction systems, the adaptation of the system to the driver, the handling of the complexity of the scene, the possibility to apply knowledge to other traffic participants, and finally the feasibility of behavior prediction in limited sensory setting.

Machine learning in order to avoid design issues

Considering possible driving behaviors in reaction to the multitude of possible driving scenes makes one speculate: is it only possible to create "by hand" a system taking into account all of this variability? Is it possible to design a system which is able to tackle every specific driving situation? One may assume that the answer is "no".

The limitations of such approaches pushed researchers working on the topic of intelligent vehicles to go for another option: learning systems. Basically, the aim is to avoid coding and calibrating a function to perform a certain task. Instead, we want to teach the system to do what we want him to do by asking him to mimic the desired reaction (output) of the function depending on the situation (input). For example, we could teach the system how to anticipate a lane change maneuver, or how to perform a deceleration at a red traffic light.

There is a huge amount of research on the topic of machine learning. The theoretical approach is not part of this thesis. Instead, we ask ourselves this question: is it possible to apply existing algorithms in order to learn the relation between what the system observes in the scenes, and how a traffic participant acts or will act?

Adaptation to the driver

Each driver has his own way of driving, and this driving style depends on various factors. Some drivers are more skilled than others, or are more confident in their driving ability. Some drivers have a sporty style, while others are more cautious. Some drivers have more experience, and react quickly. An Advanced Driver Assistance System has to adapt to the driving style of each driver. If a system is developed in order to predict the usual behavior of an average driver, it will always consider that sporty or overly cautious drivers are not driving correctly, and trigger warning signals all the time.

One solution would be to develop a system with multiple driving profiles, and let the driver chose which profile he prefers. The system could also detect which profile corresponds to the current driver: given the acceleration profile, is it a sporty driver or a cautious one? The system would then be able to adapt to the driver, within the boundary of the behavior models already integrated into the system.

Another solution is to let the system learn by itself to predict the behavior of the driver depending on the driving scene, thus leading to a finer characterization of driver behavior, reducing quantization errors. This is the reason why we want to let the system learn behavior models by observing driver behavior.

Knowledge acquisition and handling the complexity of the scene

The behavior prediction system has to be trained and has to perform on a huge variety of scenes. This creates issues related to the quantity of knowledge required to operate.

It is not possible to have one learning algorithm that learns all the different possible driving scenes and predicts in each one of them the behavior of the traffic participants. At the same time, it is not possible to have one prediction algorithm per possible driving scene. The quantity of data required to train these types of systems would be enormous.

As an example, the behavior of the ego-vehicle driver triggered by the braking of a following car does not depend on the reason why the following car slowed down. Whether it is for a stop sign, a red traffic light or a pedestrian crossing, the outcome is the same. Should we then train a system for each of these situations? We will propose a solution that decomposes the driving scene into elementary local situations, which are easier to learn, and hold less variability, and predict the evolution of the driving situation in more local, but more frequently occurring situations.

Possibility to apply knowledge to other traffic participants

In order to facilitate the behavior prediction, we want to find out whether the system is able to apply what has been learned in the ego-vehicle perspective to other traffic participants. This implies that the features used for prediction must be based on quantities also observable in other traffic participants.

For example, speed, relative position between vehicles, or the position relative to infrastructure are all quantities which can be extracted using visual processing. In contrast, the status of the gas pedal and the gaze of the ego-vehicle driver, even if they provide a great quantity of information and are strong behavior indicators, can not be observed in other traffic participants. Thus, we chose to concentrate on the former group of quantities. Moreover, studies on behavior prediction using precise monitoring of the driver (e.g. head position or gaze) shows that the system learns to rely mostly on this monitoring [63], rendering the application to other traffic participants impossible.

In this work, the prediction of other traffic participants behavior will be used as an additional source of information for the ego-vehicle behavior prediction system. Ultimately, a complete behavior prediction system would take into account not only the influence of other traffic participants on the ego-vehicle driver behavior, but also the influence of the ego-vehicle on other traffic participants.

Feasibility of behavior prediction in limited sensory setting

It stands to reason that by reducing the field of vision, the possibility for assistant systems and autonomous driving are also reduced. It is however interesting to explore what can be done with what we have. What are the applications that can be developed with limited sensory setting? Is it possible to predict the behavior of the ego-vehicle, even under limited sensing?

We try to answer these question in the last experimental chapter of this thesis. If behavior prediction is performed using a partial observation of the scene, is it possible to derive indicators that tell the system if behavior prediction is possible and reliable?

Chapter 2

How to represent the behavior of the driver

How to correctly define a driver's behavior is a non-trivial problem, because there is no clear definition on what a behavior really is. Presenting a definition of behavior and a method to quantify it seems essential to proceed and study behavior prediction.

2.1 What is a behavior?

2.1.1 Problem

Defining behavior

Intelligent Vehicles is a relatively young scientific field. Only the recent emergence of precise sensors and powerful embedded computational units allowed the field to emerge. There is no consensus yet in the Intelligent Vehicles Area about what constitutes a behavior, how to formulate it and how to quantify it. Researchers often define driving behavior depending on the scientific question they want to answer. In [55], the authors presented a really specific behavior description because the main goal of the paper was to model car-following behavior. Similarly, the prediction of cut-in in [56] influenced the authors in the manner they defined driving behavior.

Multiple definitions of what constitutes a behavior are available from different scientific fields. If we naively look at the first definition of behavior in a dictionary, we might find that a behavior is "the manner of conducting oneself", which is unspecific and not helpful at all in our case. Another more interesting definition, also provided by the dictionary, is "the aggregate of responses to internal and external stimuli" (see [44]). Behaviors are actions and reactions motivated by the aim of a person, and his adaptation to perturbations from the outside world. From a "transportation science" perspective, driver behavior is a set of actions a driver takes in order to reach his destination, and the set of action he takes to cope with environment factors preventing him from reaching this destination.

Capturing behavior

The representation we will implement has to comply with several requirements in order to be suitable for our work on behavior prediction.

First of all, as we want the system to be able to apply the knowledge it acquires from the ego-vehicle perspective to predict the behavior of other vehicles, it would be convenient to use a common behavior representation both for ego-vehicle and other vehicles. From this follows the idea that if we want a common representation, it is necessary that the quantities used to describe the behavior are observable in the ego-vehicle and in other vehicles. For example, if V2V communication is not available, driver monitoring can not be used, because it can neither be observed in other traffic participants, nor acquired by communication.

We aim for a system able to learn and develop online. This approach is not compatible with the use of labeling behavior for training purposes: we can not ask the driver to tell the system what he is doing and why he is doing it, nobody would buy such a car. A valid solution is to let the system learn the cause of a behavior by itself, by observing the driver's behavior and the scene in order to find relations between them.

Representing and quantifying behavior

There are multiple ways of quantifying a driving behavior. The aim is to find a representation that meaningfully captures what a driver is doing, and convert it in quantities which can be used by the system to perform scene understanding and behavior prediction.

The most basic (in the sense of close to sensor data) representation of behaviors is the state of the actuators. Pedal position, steering wheel angle and gearbox states are the most primitive possible representations of a driver's behavior. These quantities are either real numbers or binary values and they measure exactly what the driver is doing. Indeed, they will provide a good idea of what the driver does or intends to do, as they were designed to be the interface between the human driver and the vehicle. These representations, however, can not be observed in other traffic participants without the use of V2V communication, so they are not suitable in our case.

The result of a sequence of actuator states is a sequence of positions, speeds and accelerations. These quantities are measurable using sensors such as, e.g., GPS and inertial sensors. They are from actuator states, because contrary to them, they can be observed in other traffic participants using range sensors.

A succession of position, speed and accelerations can be grouped together in order to form continuous sequences representing elementary actions. As an example, a sequence of positive longitudinal accelerations correspond to an acceleration phase. A sequence of negative lateral speed corresponds to a turn to the left.

Finally, a sequence of these elementary actions can be interpreted as a meaningful action which depends on the driving scene. As an example, passing the frontal predecessor on highway is equivalent to a turn to the left followed later by a turn to the right. This means that we categorize depending on the context of the scene.

All these possible representations are illustrated in Fig. 2.1. As can be seen, we can chose to perform prediction at several level of complexity and abstraction of a behavior.

2.1.2 Related Work

In this section we propose to review approaches that have been adopted in the community of intelligent vehicles. As this scientific field is really close to the one of robotics, we will also present several really relevant papers from the robotic community.

Review of the approaches in the transportation science field

In order to estimate what a suitable behavior representation could be, we studied the literature related to this topic. In this section, we will present five different existing approaches for behavior representation.

In [45], the authors proposed a method for learning motion patterns of vehicles at an intersection, using a fixed top-view camera. They perform a hier-

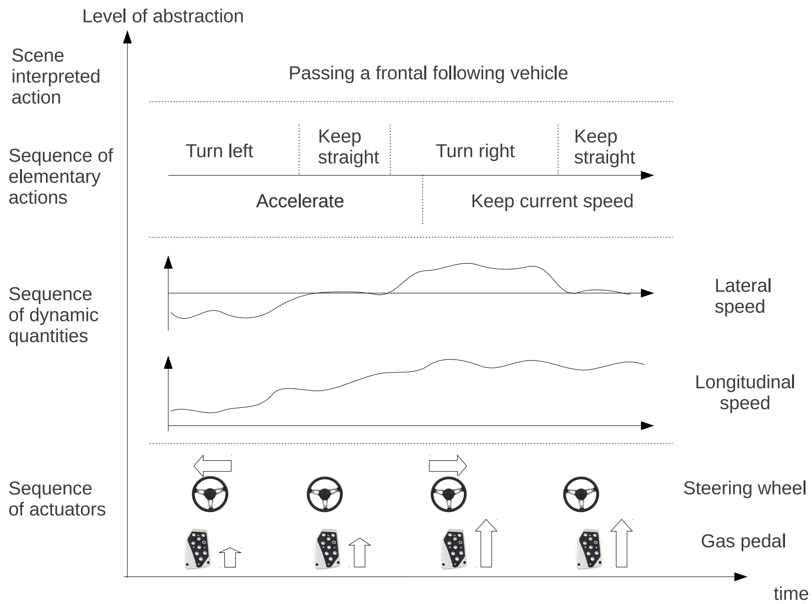

Figure 2.1: Illustration of possible behavior representations for the example of passing a frontal following vehicle.

archical clustering of trajectories based on the sequence of speeds and positions of the observed vehicles. They can extract typical trajectories (which are considered in their work as behaviors) with their system and use them to detect anomalies and predict behavior. The presented results are convincing, however, they are based on a system recording trajectories of multiple vehicles on a fixed scene. Thus, their approach is not applicable to our case, where the scene is constantly changing.

In [74], the authors present an approach for predicting the behavior of the ego-vehicle based on image features extracted using a camera recording the front driving scene. They define the behavior as a set of measurable actuator states: the position of the pedals (brake, acceleration and clutch) and the steering wheel angle. This representation of behavior is not suitable for our work: these quantities are not observable in other traffic participants, so the knowledge acquired from the point of view of the ego-vehicle will not be applicable to other traffic participants.

In [69], the authors use Hidden Markov Models to describe a complete maneuver as the sequence of states and actions. As there is one set of states per driving maneuver, the elementary states are not necessarily comparable from one maneuver to the other. This makes the comparison of behavior and the fusion of information from different traffic participants impossible. Also, it complicates the identification of prediction errors, as the hidden states do not necessarily allow meaningful interpretation.

In [19], the authors try to predict what they term “driving plan”. These complex behaviors, such as passing a frontal successor, are in the authors opinion impossible to predict unless V2V communication is used. Moreover, these behaviors operate at large timescales, and depend on observations that would be available only after the behavior has started. For a passing behavior, the ego-vehicle has no information on the scene in front of his predecessor, so it can not know in advance if the driver will go back on his original lane or stay on the left lane. Therefore long-term prediction is not possible. The authors define the driving plan (a high-level description of driver action) as a composition of driving actions. These action are very similar to the definition of behaviors that we will use in this thesis.

In [36], the authors propose a set of individual driving maneuvers to decompose the driving task, and use a Bayesian network model in order to automatically recognize the driving maneuver. Their definition of behavior is interesting and very similar to the one we will propose. However, their approach necessitates a labeling of the driver behavior for training the Bayesian network. This system reaches good results, however we decided to simplify it by introducing heuristics for segmenting the driving behavior, so that no manual labeling is needed.

Inspirations from robotics

Similar considerations occur in multiple scientific fields. When an abstract quantity can be represented in multiple manners, which one should be chosen for prediction? It is important to have an idea about how the problem of definition and representation of an abstract quantity we want to predict has been solved in other research areas.

In robotics, one of the topics of research is the teaching of actions to a robot using imitation learning, or in a broader sense the interaction between robots and humans. Issues related to the enormous motor and action space the robot evolves in, with multiple degrees of freedom, leading to a lot of parameters to be learned by imitation learning, make any task involving motor control and sequences of actions difficult to parametrize. Moreover, as an action is a succession of these parameters states, the task involves temporal sequences and become even more complicated to formalize. In order to cope with this issue, movement primitives have been developed (see, for an overview, [15, 58, 78]). Their role is to provide an abstraction for motor controls and a building block for full complex movements. Complex actions can be represented as a sequence of movement primitives which are easier to handle.

We can see a clear correspondence between the robotic field and the intelligent vehicle field. In our case, we need an approach to interpret human behavior, and the one we will propose in the following section takes inspiration from movement primitives in robotics.

2.2 Proposal for behavior representation

We represent the behavior of a driver by a succession of elementary maneuvers termed behavior primitives. We separated the different driving scenes into two types: inner-city scenes and highway scenes. We created a set of behavior primitives for each of these two driving scene types: one set of behaviors for longitudinal behaviors and one set of behaviors for lateral behaviors. Each set of behavior primitives is composed of longitudinal and lateral behavior primitives.

For each type of driving scene, we characterize independently the longitudinal behavior and the lateral behavior. We propose an intermediary step between elementary actions and scene-interpreted actions presented in Sec. 2.1.1.

The behavior is described relative to the path the driver is going to take. In the case of the left turn behavior, this can be represented as the driver following a path that takes a left turn, in this representation there is no change to the left lane. Estimating the path the driver will take is not part of the thesis. However it can be achieved if the system knows where the driver wants to go (his aim or destination). For example, this data is accessible if the driver provides this information to the system using a GPS navigating system. Another example is the exiting of the highway, which is not predictable except if the system knows that in order to reach its destination, the driver will have to exit the highway.

Rural road scenes are hybrid situations, between highway and inner-city scenes. The reasons why rural roads scenes are not described in this thesis is because data in rural roads was not available in sufficient quantities to perform behavior prediction.

2.2.1 Inner-city

In inner-city, the driving behavior depends mostly on the reaction to the frontal scene, composed of traffic participants and traffic signs. They will influence the longitudinal behavior, and the driver of the ego-vehicle either keeps speed, decelerates, accelerates or stops. For the longitudinal behavior primitives, we define a segmentation based on the temporal average of the longitudinal acceleration LongAcc. In order to extract the behavior primitives that describe the complex behavior at time t_0, we measure the mean longitudinal acceleration $\widehat{LongAcc}(t_0)$ using a temporal window ΔT around t_0:

$$\widehat{LongAcc}_{\Delta T}(t_0) = \int_{t_0 - \frac{\Delta T}{2}}^{t_0 + \frac{\Delta T}{2}} \mathrm{LongAcc(t)}\, \mathrm{d}t \tag{2.1}$$

where LongAcc(t) is the longitudinal acceleration of the vehicle at time t. Then we segment the longitudinal behavior into four behavior primitives:

- Accelerate, which corresponds to $\widehat{LongAcc}_{\Delta T}(t_0) > \tau_{acc}$
- Decelerate, which corresponds to $\widehat{LongAcc}_{\Delta T}(t_0) < \tau_{dec}$
- Stopped, which corresponds to a speed $LongVel(t_0) < LongVel_{stopped}$
- Keep speed, which corresponds to the default case, when the behavior is none of the three others.

By setting the thresholds τ_{acc} and τ_{dec}, we basically consider small variations of acceleration to be irrelevant, and focus on important variations of accelerations. Averaging the acceleration during a period ΔT allows to have better and cleaner segmentation by smoothing the quantities.

Lateral behaviors in inner city are quite rare, except for changes in direction. For the lateral behavior primitives, we define three behavior primitives:

- change to left lane

- change to right lane
- stay in the ego-lane

In inner-city, lateral behaviors are due to change of direction, a lane change or the avoidance of an obstacle. The prediction of these behaviors will not be considered in this work. Instead, for inner-city scenarios, we will focus on longitudinal behavior.

2.2.2 On Highway

On highway, both longitudinal and lateral behaviors are important. Longitudinal behavior are performed for adaptation to the preceding vehicle behavior, or adaptation to the traffic regulation. Lateral behavior occurs when entering or exiting highway, but these cases are not relevant for our work. However, lateral behavior also occurs during lane changes (for example when passing a vehicle). On highways, we define behavior primitives relative to the ego-lane.

For the longitudinal behavior, we define:

- Accelerate, which corresponds to $\widehat{LongAcc}_{\Delta T}(t_0) > \tau_{acc}$
- Decelerate, which corresponds to $\widehat{LongAcc}_{\Delta T}(t_0) < \tau_{dec}$
- Keep speed, which corresponds to the default case

It is important to note that the thresholds defining behavior primitives and the period of averaging on highway and in inner-city are not necessarily the same, as the scales at which we operate differ greatly.

For the lateral behavior primitives, we define:

- change to left lane
- change to right lane
- stay in the ego-lane

These behaviors can be extracted using a lane marking detection system implemented in the experimental car. This system provides an estimate of the curvature of the highway, and the ego-vehicle position relative to the middle of the lane. This information is given by the system as polynomial curve equation:

$$y = a_3x^3 + a_2x^2 + a_1x^1 + a_0 \tag{2.2}$$

Where y is the lateral position of the center of the lane at a longitudinal distance x in front of the car.

If the car changes lane to the left, it will get closer to the left lane marking, and the coefficient a_0 will increase towards positive values. Then, for a brief period of time, it will be set to 0 (when the system is uncertain about which lane the vehicle is on). Finally, once the lane change is certain, the referential changes and is centered on the left lane, and the car is near to the right lane, so a_0 has now a negative value.

By simply looking at the coefficient a_0, we can derive a simple algorithm to detect if the car changes lane at time t_0:

- if $a_0(t_0 - 1) > \tau_{left}$ and $a_0(t_0) = 0$: Lane change to the left lane
- if $a_0(t_0 - 1) < \tau_{right}$ and $a_0(t_0) = 0$: Lane change to the right lane

The thresholds τ_{left} and τ_{right} are used to make sure that the car is near the left or the right lane, and not in the center of the lane.

2.2.3 Validity of this definition

We have to ensure that this representation of driving behavior as a sequence of behavior primitives meets the requirements presented in Chapter 1.

We made sure that the quantities used for the segmentation into behavior primitives can be observed in other traffic participants through the use of range sensors. We can compute the position, acceleration and speed of surrounding participants using the combination of inertial sensors and GPS measuring the speed and acceleration of the ego-vehicle, and range sensors providing the distance of other traffic participants.

The representation is compatible with online learning, because it does not necessitate manual labeling of the data. The behavior primitive extraction requires heuristics for setting the thresholds segmenting the acceleration profiles. Once a behavior primitive is extracted, the correspondence between past or present representations of the scene and this behavior primitive can be learned. We believe that the heuristics used for extracting behavior primitives can be replaced in the future by unsupervised learning algorithms.

Finally, these behavior primitives, compared to raw acceleration and speed values, facilitate learning by transforming a regression problem into a classification problem.

2.3 Illustration of the segmentation in behavior primitives

2.3.1 Experimental setup

The prototype car used for the experiments is a Honda Legend equipped with cameras and range sensors (radar and laser). The data recordings throughout this thesis were made during test drives in Offenbach am Main and its surroundings, in Germany.

Inertial sensors via CAN-bus provide the lateral and longitudinal speed and acceleration of the ego-vehicle. These quantities are filtered using a Kalman filter in order to reduce the impact of measurement and detection noises. The CAN-bus also provides information about the steering wheel angle, the turning light state, and the different pedal states, but these are not used for behavior segmentation as they are not observable in other traffic participants.

We use the camera images to illustrate the segmentation of behavior primitives. We will show the decomposition for two representative cases. The first case is a traffic-light approach scenario, where the ego-vehicle approaches a traffic light in inner-city. The second case is an example of the ego-vehicle passing a frontal preceding vehicle on highway.

2.3.2 Inner city

Part of the work in this thesis will involve inner-city scenarios. In this section we illustrate the segmentation into behavior primitives of the trajectory of a driver approaching a red traffic light.

The decomposition of the scene in a sequence of behavior primitives is performed following the procedure described in Sec. 2.2, using the following quantities:

- $\Delta T = 1s$
- $\tau_{acc} = 0.03m \cdot s^{-2}$
- $\tau_{dec} = -0.05\ m \cdot s^{-2}$
- $LongVel_{stopped} = 1\ m \cdot s^{-1}$

These quantities were chosen observing several driving maneuvers in inner city. This segmentation is subjective, but as illustrated in Fig. 2.2, the result is coherent. We qualitatively verified that this approach provides a meaningful representation using recorded data in inner-city.

As one can observe, small variations of acceleration are smoothed out during the conversion into behavior primitives.

2.3.3 Highway

In order to illustrate behavior segmentation on highways, we present the scenario of the ego-vehicle passing a frontal preceding vehicle. We detect a lane change as described in Sec. 2.2 and provide an illustration in Fig. 2.3. Given that lanes on the highway are about 3 meters wide, we set the thresholds to:

- $\tau_{left} = -0.8m$
- $\tau_{right} = 0.8m$

So if the car is at more than 1 meter from the center and the lane detection system sets the coefficient a_0 to 0, we detect a lane change.

2.4 Conclusion and Outlook

We propose a simple representation of human driving behavior by segmenting the vehicle trajectory into easily identifiable behavior primitives. They efficiently enable to segment the driving trajectory into a set of elementary building blocks that are an abstraction from dynamic quantities. This abstraction allows us to transform a regression problem (predicting future quantities such as speed and acceleration) into a classification problem.

The advantages of such a representation are clear: it allows us to easily represent a complex quantity in a set of simpler behavior primitives. It allows to smooth out variations in the measurement by integrating quantities that are subject to a lot of noise. However, the main disadvantage comes from the fact that the setting of parameters necessitates heuristics, and might depend on the driver which is driving the ego-vehicle.

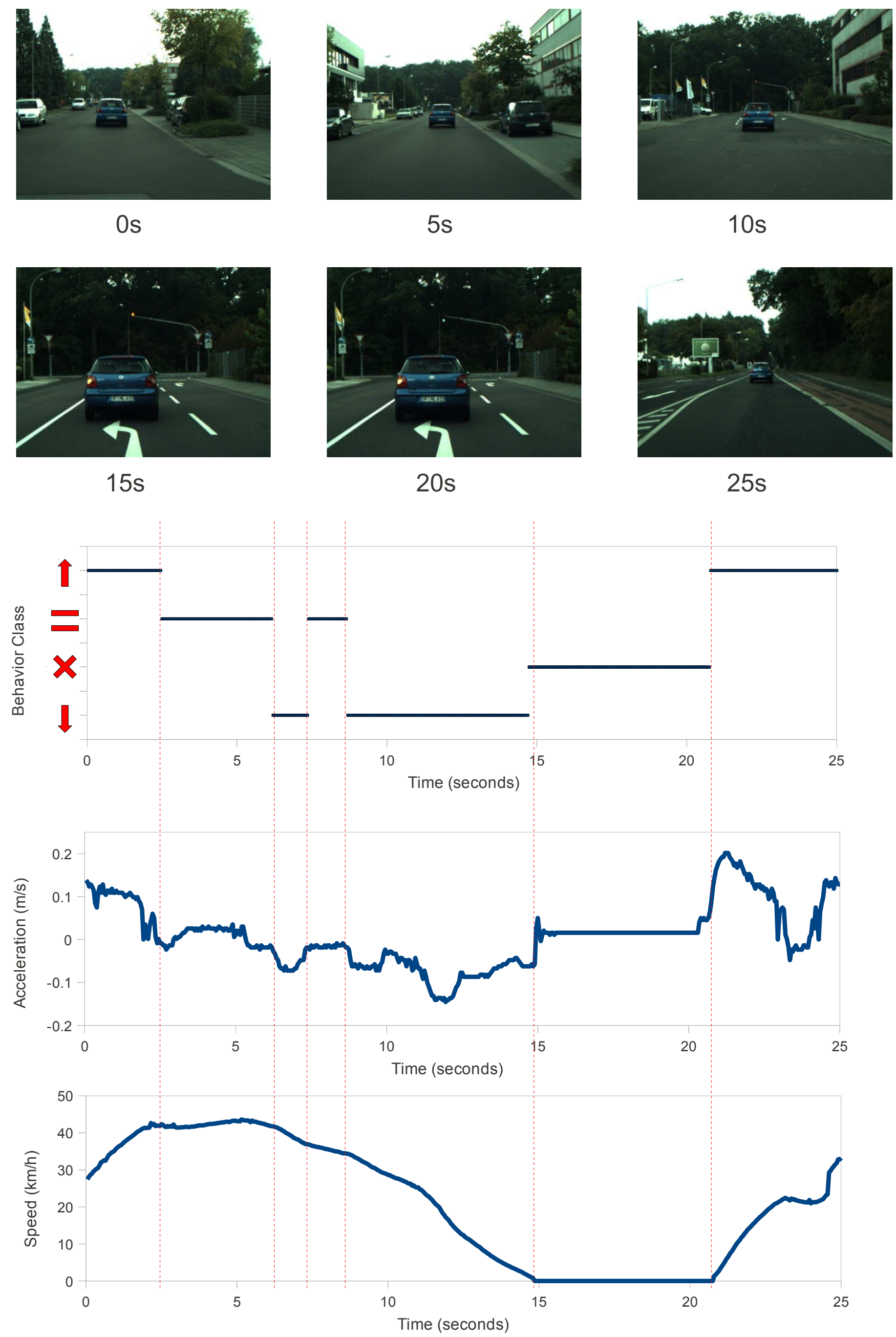

Figure 2.2: Segmentation into behavior primitives for a traffic light approach scenario. First, the car accelerates, then stays at constant speed for some time. It starts decelerating and then stops. Finally, it accelerates again.

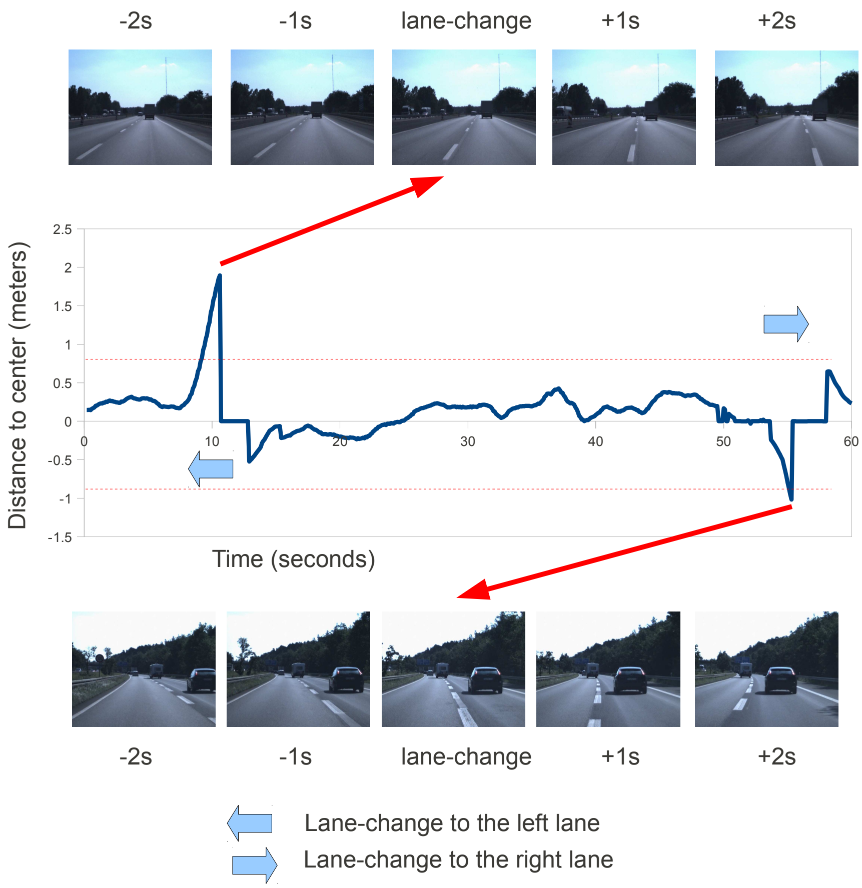

Figure 2.3: Illustration of the sequence of behavior primitives during a passing of a frontal predecessor on highway.

Even if behavior primitives for inner-city and highway are similar, the mathematical quantities to express them are not the same. We decided to separate them in order to emphasis the difference in their extraction. One could argue that the 'stop' behavior could occur on highways, however we decided to exclude traffic-jam situation on highway, because it is not a usual situation. In a system taking into account all situations, it will be necessary to include systems for specific situations like traffic jams, in inner-city and on highway. An indicator for estimating the driving condition will be proposed in this sense in the Chapter 5.

This representation can be further extended if necessary. We can decide, for example, to decompose the acceleration behavior primitive into slight acceleration and strong acceleration, in order to have a finer subdivision. Also, the thresholds used for segmenting acceleration profiles in behavior primitives can be modified depending on the driving style of the driver. The segmentation would then be depending on the result of the behavior prediction. The system would learn the behavior representation along with the behavior prediction. If the behavior segmentation does not allow precise prediction, modifying the thresholds τ_{acc} and τ_{dec} can lead to better results. We decided to not explore this topic, because this interdependency between two learning systems and the mutual influence that affects them over time is not a trivial problem at all.

Chapter 3

Ego-vehicle driver prediction and application to other traffic participants

In the previous chapter, we proposed a definition for behaviors in the context of human drivers and a method to quantify them. We will proceed and present our approach for behavior prediction. We will first present the questions we want to answer in this chapter, and review what has been done in the field. Then we will present our approach and compare it to existing work. Finally, we will demonstrate the capabilities of our approach in the scenario of a vehicle approaching a traffic-light in inner-city.

3.1 Predicting the behavior of the ego-vehicle driver

In this chapter, we start by characterizing behavior prediction. Afterward we propose an approach to perform behavior prediction at multiple timescales.

3.1.1 Questions about behavior prediction

We are looking for a solution in order to perform behavior prediction of the ego-vehicle driver. Knowing the future behavior of the driver as well as other traffic participants is necessary for anticipating danger, or for optimizing the trajectory of the car for eco-driving or comfort driving.

What do we want to predict?

To evaluate if the driver behaves appropriately in a given scene, the system can compare what it observes with what it predicts. By predicting what the behavior of the driver should be given the current driving environment, the system can estimate if a driver acts as expected, and by doing so, detect unusual behaviors. In addition, to anticipate what the driver should do in the near future, the system also has to predict the future behavior of the ego-vehicle driver. Multiple approaches can be implemented in order to do so.

One approach is to predict the behavior "state". We could predict, given a scene, what the current behavior should be or what the future behavior will be in a fixed time-span into the future. As an example, in [74] and [42], the authors predicted the lateral behavior of the ego-vehicle driver depending on the current visual scene. They clustered the behaviors into three classes: "turn left", "go straight" and "turn right", and their system learned the mapping between the current visual scene and the current behavior class. This approach is not applicable to future behavior prediction, because the visual scene alone does not provide enough information about the evolution of the driving scene.

A second option is to predict whether the behavior will change in a fixed amount of time. Given the present scene and the present behavior, will the behavior change to another behavior, and to which one, within the next 2 seconds? This approach is used in [56], [12] and multiple other contributions to predict whether a driver is going to start a certain behavior in the near future, given the current scene representation.

The last approach differs from the previous ones, which were formulated as classification problems. Instead of asking ourselves what the behavior will be in a certain amount of time, or whether the behavior will change in the near future, we could transform the problem in a regression problem. More explicitly, we would try to find out when the behavior will change. In [23], the authors proposed a system based on Bayesian methods that tries to mimic human behavior in a traffic-light approach scenario. This is equivalent to estimating online what behavior to expect given the current scene. They argue that this system can be used to predict a sequence of behaviors of real drivers, but this was only demonstrated on simulated data.

What information can be used for behavior prediction?

What information to use for estimating current or future driver behavior is a topic where many approaches are possible, given the variety of sensors, processing algorithms and also strategies (ego-centered sensing or V2V/V2I). The different types of information used for behavior prediction can be separated in two families: intrinsic and extrinsic data.

Intrinsic data is data concerning the vehicle targeted for behavior prediction. For the ego-vehicle, CAN-bus provides dynamic information as well as some simple driver monitoring by measuring the state of the actuators (gear, pedals and steering wheel) or turning signals. Also, recording the face of the driver, and the position of his foot [88], provides valuable information for behavior prediction. If we want to predict the behavior of other traffic participants, driver monitoring is only available through V2V communication. So if V2V communication is not available, the system can not rely on driver monitoring to predict the behavior of other traffic participants.

Extrinsic data is data about the driving environment measured by the ego-vehicle sensors. It is acquired using range sensors (laser, radar, lidar) and cameras. This data is processed in order to extract descriptors of important information about the driving scene. The level of complexity of the algorithms and the level of abstraction of the descriptors ranges from extracting visual or appearance features [51] to computing abstract and complex properties of a detected object such as object identity, distance, or dynamic properties.

Most research on behavior prediction is based on systems using a combination of extrinsic and intrinsic measurements. The reason is simple: the behavior of a driver depends on the scene as well as on his current state. It is important to note that with V2V and V2I communications, both extrinsic and intrinsic data are available to all vehicles equipped with this communication technology, and can be used for prediction. If V2V and V2I are not implemented, the system has to perform using only what can be observed with range sensors and visual processing from the ego-vehicle point of view. Then, many approaches using driver monitoring are not applicable in our case, because the behavior prediction from the point of view of the ego-vehicle driver would not be transferable to other traffic participants.

As an illustration of this, we can consider the work of [57]. In this work, the authors propose a rear-end collision avoidance system. In the case of the ego-vehicle following a preceding vehicle, they use probabilistic modelling to infer the dangerosity of the situation, the intended action of the ego-vehicle driver, as well as the best course of action for the system. Laser data and CAN-bus data are used to infer the dangerosity of the situation. They use driver monitoring (head and foot recording, as well as pedal status) in order to predict the intended behavior of the driver, then they estimate if the intended behavior is suitable given the situation, and propose that the system warns the driver or takes action by triggering emergency braking. Their system is able to predict driver intent up to 3 seconds in advance, and the success of this prediction is due to driver monitoring features.

How can we perform the prediction of behavior?

Several options are available in order to perform the mapping between a current scene representation and a present or future behavior. We can either hard-code a system, use a system trained offline, or create a system able to learn the relations between scene and behavior online.

The first solution is to create a fixed controlled system which reacts to the environment in pre-determined cases, in a manner that have been implemented by hand. There is no adaptation possible, and it requires significant efforts to develop.

The second solution, offline learning, has the advantage of having a fixed evaluated state. The design issues are reduced, however this approach is not adaptable to changes in the driver's behavior. The vehicle, once equipped with such a system, will always react similarly to situations, independently of the experience the system has with a driver. It can not improve, but it can not become defective either. Its qualities and defects are known and quantified before bringing the vehicle on the market.

The third solution is to introduce learning algorithms in order to adapt the system to the specificity of a driver. For example the system could learn exactly when the driver starts a behavior given a certain scene, which can differ between different types of drivers. This learning approach basically allows to create behavior models for certain scenes based on what the driver usually does. The main disadvantage of this approach is that the system convergence depends on the experience the driver provides. Basically, the system could learn bad driving habits.

A last approach is a compromise where behavior prediction is trained on a set of representative drivers. Then the system could refine the prediction to adapt to a specific driver, within certain boundaries that guarantee the safety of the driver and the respecting of traffic rules.

How to assess the reliability of a prediction?

Our approach will predict a present or future behavior for each data-point provided by the perception system. That means that for each scene representation, we perform one behavior prediction. But how to make sure that the prediction is reliable? How to guarantee that an assistant system can act according to what the behavior prediction provides?

First of all, we must evaluate the whole system, by assessing its prediction capability. By comparing what the system predicts and what the driver does, we can verify whether the system is able to predict the behavior of a driver. If we can evaluate the prediction result for a large amount of examples, we obtain a good measure of the overall confidence of the classifier. This provides the average reliability of the classifier for the set of scenes that were used to train and evaluate the system.

Then, we have to determine the reliability of the individual predictions, because having a system that performs well on average is not sufficient. Evaluating uncertainty can be done following multiple strategies. Most learning algorithms provide a confidence value for each prediction. But we want to develop a reliability measure that is independent of the learning algorithm. This measure should indicate if a prediction should be taken into account by the system or should be discarded.

The combination of those reliability estimates is necessary to determine whether a scene is predictable in general, and whether the result of a prediction is to be discarded.

3.1.2 Presentation of our approach

In this section, we want to present and motivate our behavior prediction approach. We predict the behavior state at multiple time scales in order to evaluate the prediction capability depending on how far into the future we predict. We will also propose an approach to estimate the reliability of predictions and to determine which behavior is selected.

Scene-dependent learning of the future behavior

We chose to develop a system able to learn the mapping between a present scene and a present or future behavior at a fixed time scale. In order to obtain behavior prediction at multiple timescales, we simply train multiple instances of behavior prediction, each of them with a different timescale. The behavior can be a behavior state, or the start of a behavior, depending on the type of application we consider for our system. However, we keep behavior prediction formulated as a classification problem.

The behavior prediction system performs a mapping between situation representation (see Sec. 3.2.2) at time t, and future behavior primitive, predicted on several time scales for times $t + T_1, t + T_2, \dots, t + T_n$. Our goal is to keep the possibility to perform learning in the running system. This implies that we train a learning algorithm, for a given time t and a time scale T_k, to represent the relationship between the situation at $t - T_k$ and the behavior primitive at t. After convergence, the algorithm is used to predict the behavior primitive at time $t + T_k$ using the situation representation at time t. This process is illustrated in Fig. 3.1.

For our current work, we perform learning and prediction in an offline fashion with data from recordings of real driving situations. Therefore our current implementation does not consider real time or storage issues inherent to real on-vehicle processing. However, the training of the system is the same as as in the real-time case, so the performance of the learning is not altered.

Once the mappings between scene representation and behavior have been learned, the output of the prediction system is a vector formed by the set of activations produced by the system. We implement one learning algorithm per predicted behavior primitive, and the outputs for each behavior primitive are compared in a final stage of our system. In order to be compared, these activations are normalized by mean and variance over time. This normalization can also be implemented in an online fashion using self-adaptation mechanisms as can be seen in [72]. In this contribution, we used exponential moving averages to estimate the mean and variance of the output of a neural map and use this

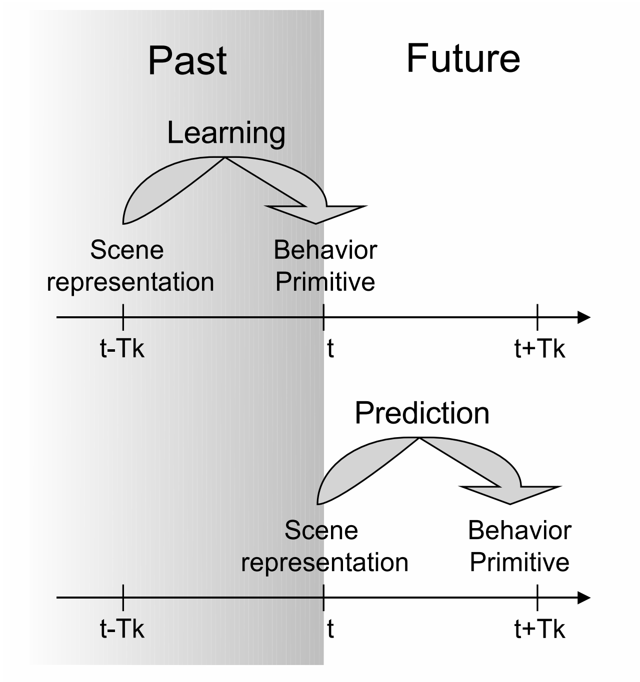

Figure 3.1: Visualization of the learning paradigm: The learning mechanism maps the past situation representation (at time $t - T_k$) and the present behavior primitive (at time t). Then it predicts the future behavior primitive (at time $t + T_k$) using the present situation representation (at time t).

as a feedback to regulate it to a constant mean and variance over time. It is able to regulate online and to adapt to statistical changes in the input data.

Measure of uncertainty

The result of behavior prediction is a set of normalized activations of outputs which we term A_i. In order to prevent the use of unreliable predictions, we derive an estimate of the confidence of this prediction C^{conf} by measuring its variance:

$$C^{\text{conf}} = var(A_i) \tag{3.1}$$

Theoretically, the entropy would be also an attractive measure, because it characterizes the amount of information delivered by a set of quantities. However it is inapplicable here because $\sum_i A_i = 1$ does not hold in general, as the outputs from different classifiers are independent. We can now set a confidence threshold τ^{conf} and determine whether the prediction is reliable or not:

$$\begin{aligned} \text{if } C^{\text{conf}} > \tau^{\text{conf}} &: \text{ the prediction is confident} \\ \text{else} &: \text{ the prediction is not confident} \end{aligned}$$

We chose to use a confidence estimation that does not depend on the type of classifier. The variance of $\{A_i\}$ is highest when there is a single dominant

A_{i^*}, which means that the result of the classification is reliable. In contrast, variance is lowest when all activations are similar; as behavior primitives usually are mutually exclusive, this signals high prediction uncertainty.

This measurement of prediction confidence is a very important step, especially when concurrently predicting a large number of behavior primitives. We assume that recognizing uncertain predictions and taking no decisions is preferable to taking wrong decisions.

3.2 Application on a traffic light approach scenario

The aim of the experimental part of this chapter is to verify in a simple scenario that the learning approach we developed is suitable for behavior prediction.

The test case is the following: the ego-vehicle approaches a traffic light, without other obstacles or preceding vehicles. Depending on the status of the traffic light, the driver will either decelerate or keep his speed.

We decided to use this scenario because traffic light approach is similar to several other situations in inner-city. Approaching a red traffic light is similar to approaching a stop sign, or a fixed obstacle. Also, it is one of the most common traffic elements that can affect the driving in inner-city.

3.2.1 Scenario and data

Scenario

We used a data-set containing approximately 16000 samples of traffic light approach scenes (see Fig. 4.5), extracted from video streams recorded in inner-city environment.

The videos are recorded at 20Hz, so the recorded data corresponds to 13 minutes of driving in inner-city. We split this data-set into 16 subsets, in order to perform N-fold cross validation.

Sensory data

Speed and acceleration of the ego-vehicle are obtained by reading streams of data sent via the CAN-bus. They are used to extract behavior primitives representing the longitudinal behavior of the driver.

Labeling of traffic lights

As we have not yet implemented a robust algorithm for detecting traffic lights, we manually annotated the presence and the status (green, yellow, or red) of traffic lights in image data. In order to estimate the distance to the traffic light, we extract the moment when the ego-vehicle crosses the traffic light. We then calculate previous distances to the traffic light by integrating the speed of the ego-vehicle obtained from the CAN bus. We compute the distance to the traffic light and the status of the traffic light for each sample of the data-set.

For our on-car system, this labeling can be replaced by an ego-vehicle based traffic-light detector. Such algorithms exist (see, e.g., [29] and are performing well. If the infrastructure is equipped for V2I communication, then the traffic

Figure 3.2: Example of inner-city traffic light approach scene.

light equipped with communication devices can send his status and his absolute position to approaching cars [16] and vice versa. The relative position can be obtained by comparing GPS coordinates.

3.2.2 Methods

We predict future behavior primitives ahead in time, on different time scales, depending on the current ego-vehicle status and scene properties.

Encoding of the situation and behavior representations

For the simple scenario considered in this study, input data consist of ego-vehicle speed and acceleration, as well as status and distance of nearby traffic lights.

We compute the behavior primitive for each sample of this data-set according to the procedure described in Sec. 2.3.2. For simplicity, we group the "accelerating" and "keep speed" behavior together. It is encoded as a 3-element binary array, one element for each possible behavior primitive.

The distance is encoded as a single real number, whereas the status of the traffic light is encoded into a 3-dimensional binary array, each element corresponding to one possible status of the traffic light (green, yellow, red).

The distance and status of the traffic light, together with the speed of the ego-vehicle, form a 6-dimensional input vector for each sample in the data-set. The 4-element behavior primitive associated with each sample corresponds to the learning target.

All quantities represented by real numbers are normalized by mean and variance over time.

Multilayer Perceptron for Behavior Prediction

In order to learn the mapping between the current situation representation and future behavior primitives, we use a multi-layer perceptron (MLP). The MLP model [39] is a standard nonparametric regression method using gradient-based learning. It is a rather simple neural model, composed of multiple layers of neurons, with each layer fully connected to the next one (see Fig. 3.3).

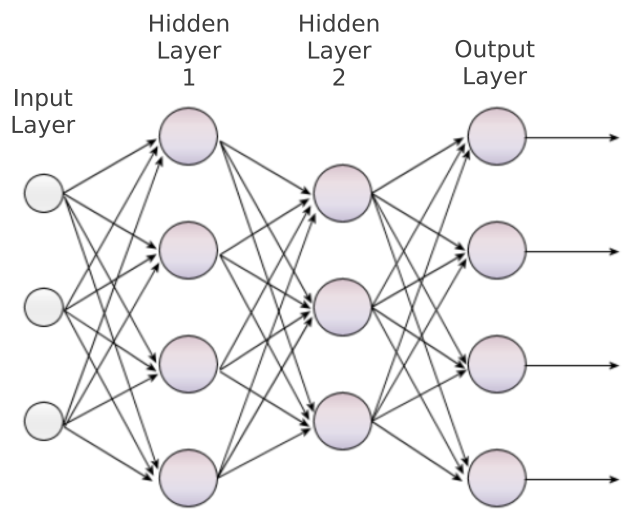

Figure 3.3: Example of a Mutli-Layer Perceptron with two hidden layers.

The MLP we used is composed of an input layer, receiving the indicators from the scene representation, a hidden layer, and an output layer, connected to the learned or predicted behavior primitive. The hidden layer of the MLP may be viewed as an abstract internal representation which captures the behavior model for the scene. For network training, we employ the back-propagation algorithm with weight-decay and a momentum term (see, e.g., [76]). We configure the MLP to produce three real-valued outputs $A_{stopped}$, $A_{decelerating}$ and A_{other} corresponding to the predicted behavior primitives. In order to compensate the different frequencies of occurrence of the three behaviors, we normalize these activations over time to have the same mean and same variance for the evaluation of the quality of the prediction. Learning and evaluation are performed on separate data-sets. In an online learning scenario, normalization would have to be performed using the exponential moving average for normalization as described in [72] .

We used the pyBrain-library [80] for all described MLP experiments. The MLP training algorithm depends on the learning rate parameter ϵ^{MLP} and the momentum parameter ν^{MLP}. The choice of the learning technique is based on a study of different learning approaches in [34, 32]. MLP is a generic and simple method, which can scale to a wide range of problems, and can be easily adapted for online learning.

Evaluation of prediction

The classification value for any output neuron i is obtained by computing $C_i^{\text{class}} = A_i - \sum_{j \neq i} A_j$. This approach is a form of one-against-all classification where the outputs of different classification systems are compared (see [54]). As an example, for decelerating behavior, this becomes:

$$C_{\text{decelerating}}^{\text{class}} = A_{decelerating} - A_{stopped} - A_{other} \tag{3.2}$$

This allows us to compute the likelihood of a behavior compared to the others. We can set a classification threshold τ^{class}, and make a classification decision for each prediction which also depends on the prediction confidence measure C^{conf} described in Sec. 3.1.2:

$$\begin{aligned}
&\text{if } C_i^{\text{class}} > \tau^{class} \text{ and } C^{\text{conf}} > \tau^{\text{conf}}: \\
&\quad \text{behavior primitive is predicted} \\
&\text{if } C_i^{\text{class}} \leq \tau^{class} \text{ and } C^{\text{conf}} > \tau^{\text{conf}}: \\
&\quad \text{absence of behavior primitive is predicted} \\
&\text{if } C^{\text{conf}} \leq \tau^{\text{conf}}: \\
&\quad \text{unreliable prediction is rejected}
\end{aligned} \tag{3.3}$$

For each value pair of the thresholds τ^{class}, τ^{conf} and for each output neuron i, we compute the detection rate ν_i^{correct}, the false positive rate $\nu_i^{\text{incorrect}}$ and the rejection rate ν_i^{reject} which are defined as

$$\begin{aligned}
\nu_i^{\text{correct}} &= \frac{\#(\text{reliable correct classifications})}{\#(\text{reliable positive examples})} \\
\nu_i^{\text{incorrect}} &= \frac{\#(\text{reliable incorrect classifications})}{\#(\text{reliable negative examples})} \\
\nu_i^{\text{reject}} &= \frac{\#(\text{rejected examples})}{\#(\text{all examples})}
\end{aligned}$$

By varying the classification threshold τ^{class}, a receiver-operator-characteristic (ROC) can be generated. This performance measure is a standard tool in machine learning and has been previously used to evaluate behavior prediction systems (see [56]). In the presented ROCs, we plot the detection rate against the false positive rate; as the rejection rate is a function of τ^{conf} which is not varied, we display the rejection rate along with each plotted ROC. Omitting this information could be misleading since a ROC may be of very high quality while accounting only for a small part of test examples, i.e. those who were not rejected.

We employ N-fold cross-validation to assess prediction results, splitting the data set into N subsets, each containing an equal amount of successive samples. We train the system using N-1 subsets and we present the samples from the remaining subset to the trained prediction system. We obtain a sequence of activations for the three output neurons which we normalize according to Sec. 3.2.2.

We then use the activations obtained from the N possible combinations of training and evaluation subsets, in order to evaluate the quality of the prediction over the whole data-set.

3.2.3 Experiments and Results

In this study, we focus on evaluating the possibilities of learning for behavior prediction, and especially on learning in the presence of multiple timescales and behavior classes. We will first analyze the effect of our prediction confidence measure on prediction accuracy by evaluating instantaneous prediction of braking behavior (i.e., we "predict" the present). We then go on to present the results of braking behavior prediction at several time scales ranging from 0s to 6s, again demonstrating the value of prediction confidence estimation as described in Sec. 3.1.2. The activation of the MLP outputs over time, for 4s prediction, can be observed in Fig 3.4, while approaching a red traffic-light.

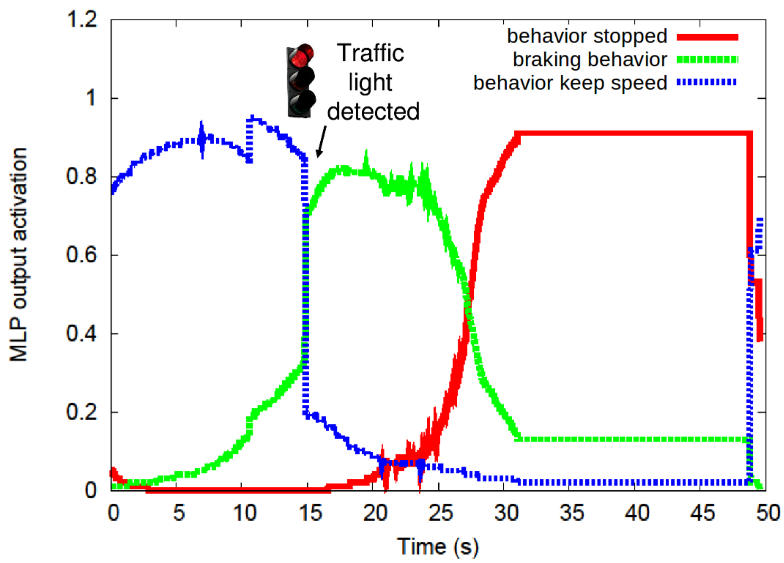

Figure 3.4: Activations of the MLP over time, for a prediction 4s ahead of time, in a red traffic-light approach scene.

Effects of the rejection of non-confident predictions

For this experiment, we train our system to learn the mapping between the current situation representation and the current behavior primitive. This serves to demonstrate the importance of the confidence threshold τ^{conf} which is set to different constant values while ROCs are obtained by varying τ^{class}. In this way, a number of ROCs at different rejection rates is obtained which can be viewed in Fig 3.5. It is apparent that, up to a point, the removal of less confident predictions increases overall prediction quality. Beyond this point, if we set the confidence threshold too high, the overall quality of the prediction will decrease again. Therefore, an optimal value has to be determined in order to reach good performance while avoiding the rejection of too many predictions. Another way of visualizing this effect is presented in Fig. 3.6, where we analyze in more detail the influence of τ^{conf} on the quality of the prediction. Here, we plot the false positive rate against the amount of samples discarded by the

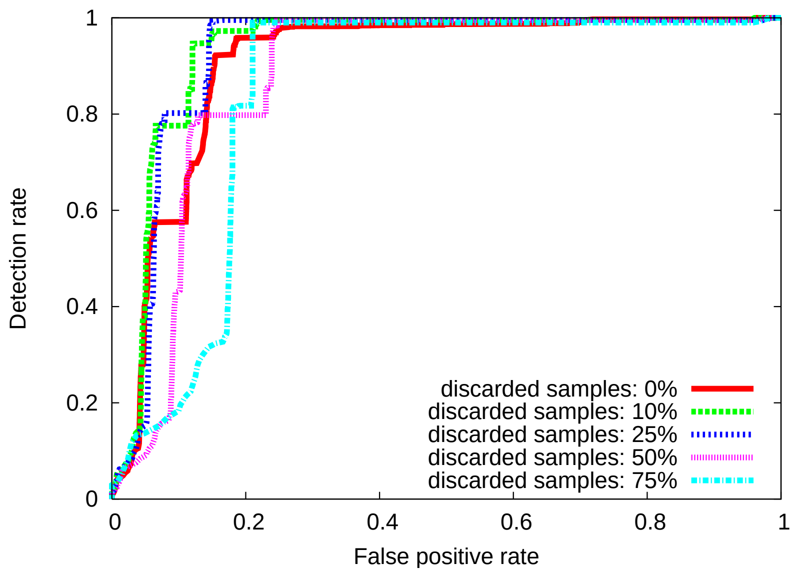

Figure 3.5: Instantaneous prediction of current behavior. Shown are ROCs obtained by variation of the classification threshold τ^{class}. For each ROC, a different confidence threshold τ^{conf} is applied.

confidence evaluation, where the value of τ^{class} has been always chosen such that the detection rate is at 85 %. As one can observe from Fig. 3.6, there is a clear optimal setting for the confidence threshold. We therefore decide for future experiments to set τ^{conf} such as to remove 10% of samples corresponding to the least reliable predictions.

Prediction of future behavior at multiple time scales

We now compare the prediction of the future braking behavior at different time scales. We set the confidence threshold τ^{conf} as determined in the previous experiment, discarding 10% of samples. As can be observed in Fig. 3.7, it is possible to predict the future behavior primitive for the different time scales presented. The quality of the prediction decreases when the timescale of prediction increases. As expected, the quality of braking behavior prediction is best at $T_0 = 0s$ and decreases continuously with increasing prediction horizon. This is an additional cross-check for the validity of the approach, since we expect larger prediction horizons leads to higher uncertainty. We also verified that the quality of the prediction decreases strongly to reach roughly chance level at 10s. One also has to consider that at 50 km/h, a car approaches a traffic light 100 meters away in 7.2 seconds. So 10s prediction would be based on information far away from the traffic light, limiting its quality.

3.3 Conclusion

In this chapter, we showed that it is possible for a system to learn the prediction of present and future behavior primitives at multiple time scales. We also showed

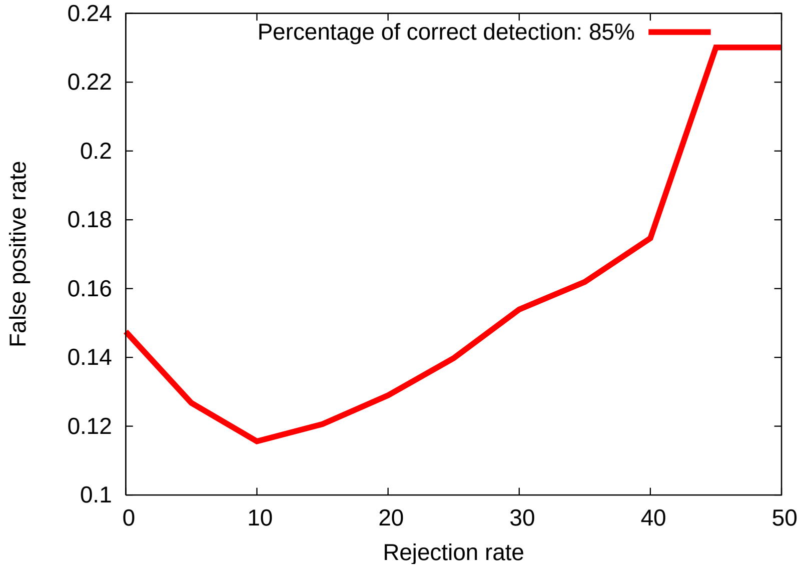

Figure 3.6: Instantaneous prediction of current behavior. Shown is the false positive rate plotted against the amount of rejected samples which varies due to the variation of the confidence threshold τ^{conf}. The value of τ^{class} is fixed to produce (for each distinct value of τ^{conf}) a detection rate of 85%.

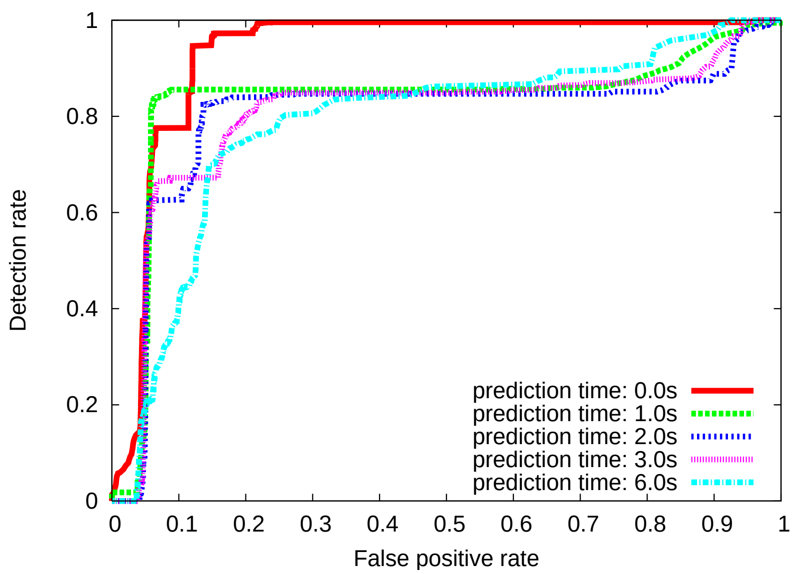

Figure 3.7: Behavior prediction at several timescales using a fixed confidence threshold. As can be seen, the system is able to perform behavior prediction up to 6s into the future.

that a simple learning algorithm operating on low-dimensional representations of situation and behavior is sufficient to predict braking behavior with very good accuracy at time scales up to 3s. As expected, the quality of the prediction decreases with the time scale, and becomes almost meaningless for time scales larger than 6s. We also presented a measure of the prediction confidence in case of multiple behavior classes, allowing to disregard uncertain predictions. We showed that it is possible to increase the quality of the prediction by using a moderate threshold on prediction uncertainty, thus disregarding approximately 10% of predictions. The evaluation of prediction confidence is a key element toward building a large-scale behavior prediction system with a higher number of behavior primitives, as might be expected in complex inner-city traffic.

It could certainly be argued that the learning problem considered here is too simplistic as it has only 6 input and 3 output dimensions. It is certainly true that we are considering a restricted scenario for learning. However, it is precisely our point that it may not be *necessary* to consider very complex or very high-dimensional learning problems for successful behavior prediction: for most ADAS applications (such as warning the driver in case of unusual maneuvers), a really precise and therefore high-dimensional prediction of behavior is not required.

Another objection to our approach is that the neural network might actually just predict based on current and recent ego-velocity values, i.e., perform a very simple physical prediction. This can however not be the case if the scene induces a change in the behavior of the driver. For example, if a traffic light turns red. In these cases, we need information from the scene, and the dynamic state of the vehicle is not sufficient to anticipate future behavior. The distance and state of the traffic light are quantities which make prediction of future behavior primitives a non-trivial and non-physical task.

Finally, our choice for confidence measure can be discussed. It is not guaranteed that we can use classic methods using probabilities or statistical approaches, because the behavior predictions are performed independently and the sum of the prediction scores does not represent a complete probabilistic space. Also, the use of a confidence measure adapted specifically to a certain kind of learning methods forces us to change it if we change the learning approach. That is why we chose a generic, always applicable approach that do not depend on the learning algorithms selected. Other methods can be implemented and evaluated. In [83], the authors propose 4 criterions to evaluate if a prediction is sensible or not, which, applied to our formalism, translate to:

- Max criterion: the maximum activation must be above a threshold
- Dif criterion: the difference between the maximum and second maximum activation must be above a threshold
- Rad criterion: the euclidean distance between the target vector and the prediction must be below a threshold
- Min criterion: the minimum activation must be above a certain threshold

The Max and Dif criterions serve to identify which activation is maximal and to check that it is above other activations. Our classification criterion $C_i^{\text{class}} > \tau^{class}$ is similar to a combination of the Dif criterion and Max criterion. The Rad and Min criterion are used to remove inconsistent predictions (termed "garbage reject" in [83]), which we do using the confidence measure. These criterions do not measure the same kind of errors as the confidence measure we propose, so it would be possible in future works to estimate the prediction quality we would gain by adding the Rad and Min criterion to the confidence measure.

3.4 Possible Future Works

An important topic for future research will be to modify the system to cope with online operation, i.e, implement online normalization of the input, online learning, and online neural output normalization. We showed in [32] that an online version of the MLP is actually performing as well as the offline version.

A further improvement of the current method will be to actively exploit the presence of multiple prediction timescales. After all, there exists a set of predictions from different times in the past for any given point in time, which might be exploited for stabilizing predictions. We could also compare the current system which predicts behavior primitives, i.e., states, to a system predicting *changes* of states to see whether a better prediction quality can be obtained.

In the next chapter, we will investigate behavior prediction in a more complex scenario. For this purpose, we will use information about the position and speed of a possible preceding vehicle. To cope with the increased complexity, situation-specific learning subsystems will be introduced.

Chapter 4

Situation-specific learning

In this chapter, we will try to answer several questions. The first one concerns the issue of scalability of the system: given the possible complexity of the driving scene, and the diversity of combinations of traffic participants and driving infrastructures, can we guarantee that the system will scale? Can we propose an approach that will facilitate the scaling? The second question concerns the transfer of acquired knowledge from one point of view to the other. Is the knowledge obtained from the point of view of the ego-vehicle applicable to predict behavior of other traffic participants ?

A scene can involve multiple traffic participants, as can be seen from Fig. 4.1. In addition to influence the ego-vehicle behavior, the other traffic participants potentially influence each other, which complicates the task of behavior prediction. Prediction of individual traffic participants becomes a difficult task, and subsequently, so is the prediction of the future scene. Also, the amount of traffic participants is always varying. An input vector of varying size is a problem for learning algorithms. How to take these issues into account during the prediction system?

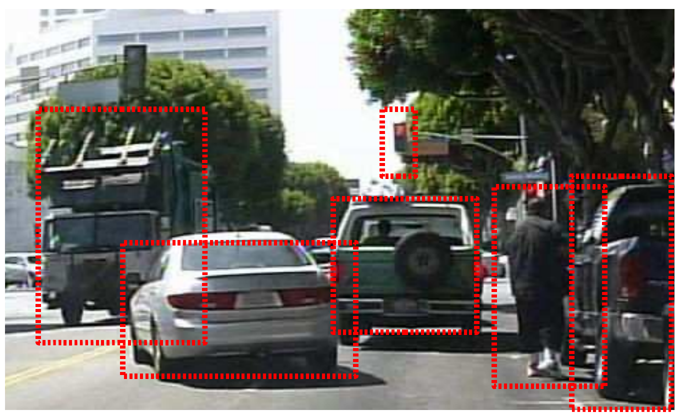

Figure 4.1: Example of a complex scene where multiple traffic participants possibly influence each other.

The future behavior of the ego-vehicle does not depend only on the present scene, but can also be influenced by probable future behaviors of other traffic participants. Is it possible to apply knowledge acquired from the point of view of the ego-vehicle and apply it to other traffic participants?

In this chapter, we propose an approach that addresses these issues by converting the possibly complex driving scene into a set of elementary situations between traffic participants and infrastructure. We believe that decomposing the perceived complex situation into a combination of simpler ones, each of them with a dedicated prediction, will allow the system to reach a performance equivalent to a system training on the whole scene. We believe that this is advantageous for the scalability of the approach to the number of possible situations that the driver will encounter.

The approach is tested on a real world scenario, using streams recorded in inner-city scenes, evaluated for a prediction horizon of 3s into the future. The quality of the prediction is measured using established evaluation methods.

4.1 Handling the complexity of the driving scene

In Chapter 3, we successfully predicted behavior of the ego-vehicle driver on a rather simple scenario, involving only the ego-vehicle and a traffic light. How can we perform behavior prediction if the scene becomes more complex?

4.1.1 Issues with the driving environment

In the previous chapter, the scenario used for evaluating our behavior prediction approach was rather simple. We manually extracted the scenes which were interesting for our demonstration. However, this is not a valid approach for real systems, which drive in environments where the topology of the scene, the infrastructure and the traffic participants are not known in advance. The scene can therefore become much more complex.

Variability in the type of scenes

A traffic scene can involve multiple different traffic participants (e.g. cars, pedestrians, motorbikes or bicycles). Each traffic participant encountered by the ego-vehicle driver has his own behavior. A bike does not act in the same manner as a car, or a pedestrian. Moreover, the driver approaches different intersections, drives on different types of roads (inner-city, highway, or rural roads). The variety of possible scenes, and especially the possible combinations of traffic participants, increases the complexity of the problem to solve, and we have to take this into account if we want our system to scale to different driving environments, as illustrated in Fig. 4.2

Figure 4.2: Example of variability in the driving scene. Driving scenes are different, infrastructures and traffic participants are different. How to cope with this variability?

Is it possible to have one learning algorithm that can predict behavior for all kinds of traffic participants and all kinds of scenes? Do we have to differentiate between different types of participants?

The issue of scene variations for the learning approach

The amount of traffic participants in the scene is not fixed, neither is the amount of infrastructure elements influencing the behavior of the traffic participants (e.g. road signs, number of lanes). Learning the mapping between the scene and the behavior of the driver poses issues because the input to the system (the scene representation) is not fixed in size and does not represent the same traffic participants. For example, for a particular scene, the system will have to predict based on the influence of a preceding vehicle and a stop sign, whereas in another scene it will have to predict based on information about a crossing car and a pedestrian on a zebra crossing.

Therefore, we need to propose a solution in order to take into account a variable amount of traffic participants.

Inter-relationships between traffic participants

Predicting how a traffic scene will evolve is difficult. Each traffic participant has an active role in the scene, so his actions and decisions will influence other traffic participants. At the same time, his actions and decisions are influenced by other traffic participants.

In the previous chapter, we took into account one type of traffic infrastructure (a traffic light) influencing the ego-vehicle behavior. In more complex scenarios, involving more traffic participants, this approach needs to be extended in order to take into account not only the influence multiple traffic participants have on the ego-vehicle, but also the influence they have on each other.

Indeed, the future behavior of the ego-vehicle can depend, as we will show in the experimental section, on the future behavior of other traffic participants. This inter-relationship between traffic participants has to be formalized in order to allow the prediction of the future behavior to be reliable.

4.1.2 Possible approaches

How to handle a varying amount of traffic participants?

The diversity of possible scenes is large, and learning algorithms with a large input vector are very slow to converge to a stable solution. Moreover, the inputs vectors representing the scene will not always characterize the same traffic participants, or the same traffic environment properties. We need to find an approach to handle these issues.

One can consider first a system taking into account all possible traffic participants around the ego-vehicle. For example, in [91], the authors propose a behavior prediction system on highways where all possible surrounding vehicles have a dedicated input slot in the behavior prediction algorithm. If one of the vehicle is not present, the inputs related to this vehicle are set to a default value. Even if this solution is valid for generic highway scenes, it is not guaranteed that it will scale to specific cases, especially in inner city where generic models of the configuration of the road and the position of traffic participants are hard to formalize. It is not realistic to apply this approach to more complex scenes. Building a system having input slots for every possible traffic participant and scene elements would require an enormous amount of training data to be able to discriminate between different situations and perform behavior prediction.

The second possible approach is to develop one learning system per possible situation. Usually, in contributions about behavior prediction, authors propose a system able to perform on a single scene type, so it makes sense to apply it to every scene type. It is a valid approach if the system we are looking to propose has to handle a finite set of specific driving situations. For example, predicting the cut-in behavior of a preceding vehicle on highway, or predicting the crossing of a pedestrian at a zebra in inner-city. However, the scalability of the approach is once again not guaranteed. The developer would have to create one scene-specific learning algorithm for each different scene it encounters. As we want our approach to be applicable to a great number of scenarios, we need a more generic approach.

The last solution is to consider individual contributions of each traffic participants. The system predicts the behavior of the ego-vehicle depending on each traffic participant separately and then combines the contributions of each traffic participant. This has the advantage to prevent combinatorial explosion compared to learning every type of scene separately. This approach can be suitable if we take into account the influence between other traffic participants. To our knowledge, no contribution used this kind of approach for behavior prediction.

How to represent influence between traffic participants?

As stated before, the future behavior of a traffic participant depends on the influence and on the future behavior of other traffic participants. How to represent these inter-dependencies is a problem that needs to be solved in order to reach a behavior prediction of good quality.

In multiple contributions, researchers have been representing "macroscopic" behavior of the traffic participants (such as traffic flow) where the influences between traffic participants have been modeled by a set of interactions. These approaches, inspired by fluid mechanics science, are usually developed in order to evaluate the traffic conditions at some key infrastructure, in order to optimize it. However, the local "microscopic' behavior of individual traffic participants, which is of interest for us, is not studied and not predictable with this approach.

Works related to behavior prediction aim at estimating the behavior of the ego-vehicle depending on the current observation of the surrounding scene. They do not take into account the possible future changes in the driving scene, that can be anticipated by predicting the behavior of other traffic participants. We believe that this anticipation is a valuable information, so we want our system to be able to take into account these cascades of behavior prediction.

How to predict the future behavior of other traffic participants?

Several works show that it is possible to derive generic behavior models, providing reliable behavior prediction. In particular, when behavior prediction is based on observation of a fixed scene, it has been shown that it was possible to estimate average behaviors of traffic participants. As an example, in [52], authors derive a generic behavior prediction model for cars approaching a crossing, in order to predict if they will continue straight, turn left or turn right. However, in our case, the data is provided by ego-vehicle sensors. So the examples available for training the system depend on what has been observed by the ego-vehicle. In order to maximize the use of information acquired, it makes sense to ask ourselves if the knowledge acquired from the ego-vehicle point of view can be applied to other traffic participants.

Few works propose to predict the behavior of traffic participants surrounding the ego-vehicle. One of them, presented in [12], proposes to predict the behavior of a car entering the highway observed from distance by the ego-vehicle already on the highway, approaching the intersection. One of the main issue raised in this contribution is the low quantity of training data. In order to have more training data, we want to use knowledge acquired from the ego-vehicle behavior perspective and apply it to other traffic participants.

Applying knowledge acquired from the ego-vehicle point of view to other traffic participants is not a trivial problem. It is not certain that the behavior prediction learned to fit to the ego-vehicle driver habits will be suitable as a generic behavior prediction system, applicable to any driver. After all, we have no information about the driving style, or the driving habits of other traffic participants. The transfer of knowledge from the ego-vehicle point of view to other traffic participants has to be investigated. Does it make sense to learn one generic model for other traffic participants and one specialized model for the ego-vehicle driver? Can the knowledge obtained from the ego-vehicle perspective be transferred even if it is tuned to the ego-vehicle driver's specific behavior?

4.2 Proposal for acquisition of behavior models and prediction

4.2.1 Scene understanding mechanism

Scene understanding plays a major role in the conception of intelligent vehicles. In our view, it is the step where a virtual representation of the environment is created, allowing behavior prediction to operate at an object level. This is an approach different from approaches where behavior prediction operates at sensor level. Even if developing or even researching scene understanding approaches is not part of this thesis, it is important to mention how to possibly build it, and how we think it should interact with a behavior prediction system.

Scene understanding takes detected objects, infrastructures, and their relative and absolute positions in the scene as an input. It has to consolidate knowledge about the driving scene and provide the behavior prediction system with an abstract representation allowing it to operate on an object level. As we want to take into account the behavioral relations and the influences between traffic participants, it could represent the scene as a network of interconnected objects. Moreover, it has to differentiate between different driving scenes.

All in all, scene understanding provides the behavior prediction system with:

- the context
- the traffic participants and their embedding in the context
- the relationships between traffic participants and the possible influence they have on each other.

4.2.2 Situation-specific learning and prediction

In order to tackle the issues discussed previously, we propose to decompose the driving scene into a set of elementary local situations involving a limited number of traffic participants. These elementary situations would have to be provided by the scene understanding mechanism.

As an example, we have shown the decomposition of a scene in elementary situations in Fig. 4.3.

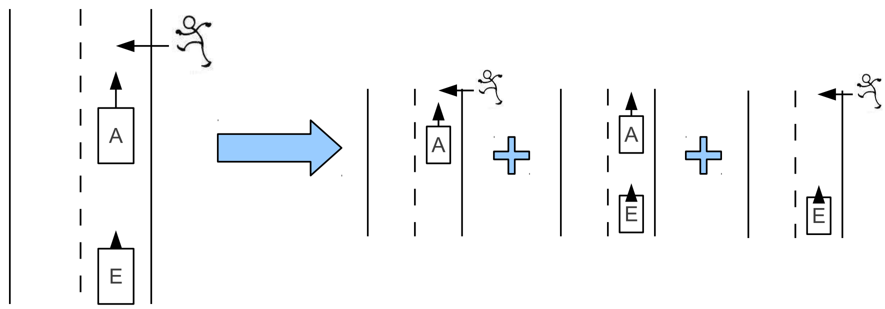

Figure 4.3: Example of decomposition of a complex scene in a set of simple elementary situations. E is the ego-vehicle, A a preceding vehicle, and a pedestrian is crossing in front of A.

If every situation can be decomposed in this manner, the transition from the complete scene to the collection of local scenes does not pose any problem. We believe that the simple situations will be easier to learn because they are less complex to model. For example, we can consider the following different driving scenes:

- the ego-vehicle follows a car which approaches a red traffic light.
- the car in front of the ego-vehicle approaches a zebra with a pedestrian crossing.
- the ego-vehicle approaches a zebra with a pedestrian crossing.
- the ego-vehicle approaches a red traffic light

Classic approaches would require to learn all four kinds of situations separately. However, if we decompose the scenes in elementary situations, we realize that we actually have:

- a vehicle follows a car.
- a vehicle approaches a red traffic light.
- a vehicle approaches a zebra with a pedestrian crossing..

We now have a smaller number of different scenes to learn. Also, the scenes are more simple, because they involve less participants. Behavior prediction is applied to such local situations.

In order to represent the influence between traffic participants, the behavior of a traffic participant is predicted and used when necessary as an additional input in order to predict the behavior of other traffic participants. For example, if an infrastructure element or a traffic participant A influences a traffic participant B which influences a traffic participant C, then A is used to predict the behavior of B, and the predicted behavior of B is used to improve the prediction of C. As an example, we have shown this process in Fig. 4.4, where the ego-vehicle behavior depends on two traffic participants and an infrastructure element.

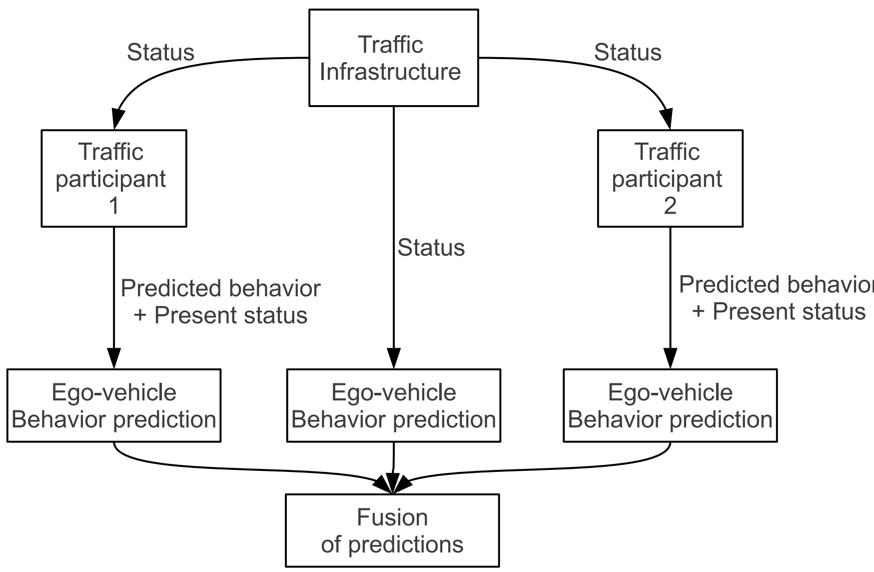

Figure 4.4: Example of the influence of multiple local situations.

If a traffic participant is involved in multiple local situations, we propose to fuse the predictions provided by situation specific modules that link him with infrastructure elements or other traffic participants. This combination can take into account a prioritization of influences of traffic participants depending on their distance, or a measurement or the relative importance of the information they provide.

We believe that this approach can lead to benefits for scalability, because the amount of elementary situations is relatively small compared to the overall amount of possible scenes. Once these elementary situations are learned, complex scenes can be predicted by composing elementary situations.

Embedding of the work: relations with the approach from Sarah Bonnin

A second PhD work has been started in Honda Research Institute Europe GmbH with a focus on behavior prediction. The approach developed in this work also aims for solutions in order to predict behavior of traffic participants in diverse scenes and scenarios.

In [12], the authors propose to represent possible scenes by categorizing them as a hierarchy, where the top nodes represent generic scenarios and the bottom nodes represent specific ones. Every node of this hierarchy would contain a classifier used to predict the behavior.

This approach is motivated by the fact that, by creating a system able to predict generic situations as well as specific ones, this system can cope with the large variety of scenes. Indeed, we agree that this hierarchy-based system allows the incremental complexification of the behavior prediction, by adding more nodes to represent new situations.

We believe that this hierarchy representation can be combined with our situation-specific approach, where new prediction nodes can be added to the hierarchy easily by composing already trained and predictable situations. So, in this sense, situation-specific learning and scene hierarchy based approaches are complementary.

4.3 Experiments

We want to show that it is possible to decompose a driving scene into a set of elementary situations. In order to prove that this decomposition does not affect performance, we will compare the prediction from a decomposed situation compared to the prediction from a system without decomposition.

Also, we want to show that in order to reach similar results with the decomposed system, it is necessary to perform future behavior prediction of other traffic participants in order to improve the performance of future behavior prediction of the ego-vehicle.

4.3.1 Scenario

We put our focus on inner-city traffic, and the scenario is the following: the ego-vehicle approaches a traffic light and is sometimes following a vehicle also approaching the traffic light. So in this study we encounter three kinds of situations:

- "traffic light situation" : the ego-vehicle is approaching a traffic light
- "preceding vehicle situation": the ego-vehicle is following a vehicle driving in the same direction
- "combined situation": the ego-vehicle is following a vehicle driving toward a traffic light

An example of a combined situation can be observed in Fig. 4.5.

If we decompose these scenes, we obtain 2 elementary situations:

- "traffic light approach" : a vehicle is approaching a traffic light
- "preceding vehicle situation": a vehicle is following a vehicle driving in the same direction

We will study this decompositions and research the conditions under which it is valid.

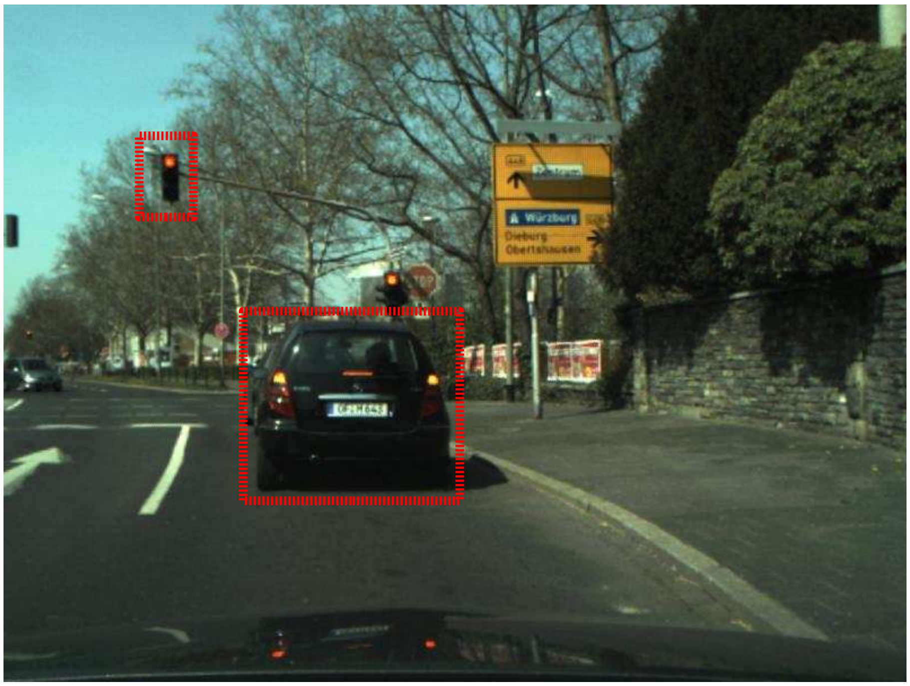

Figure 4.5: Example of inner-city traffic light approach scene: the ego-vehicle behavior can be a reaction to the traffic light or to the preceding vehicle.

4.3.2 Methods

In order to evaluate the quality and advantages of situation-specific learning, we created three different behavior prediction systems that are trained in the same manner as in the previous chapter. The aim is to predict the future longitudinal behavior of the ego-vehicle driver. This behavior is expressed in term of behavior primitives as described in the Chapter 2. We focus on the "decelerating" behavior.

Full learning

The full learning system is a naive system which uses, as an input, all the information available about the scene: the speed and distance of the possible preceding vehicle, the status and distance of the possible traffic light, and the speed and acceleration of the ego-vehicle. It learns the mapping between the complete scene representation at time t and the future behavior primitive at time $t + T_{pred}$.

We know that this basic system can not scale to complex scenes, because then the input vector would be enormous. But it will serve as a baseline to assess the capability of situation-specific learning.

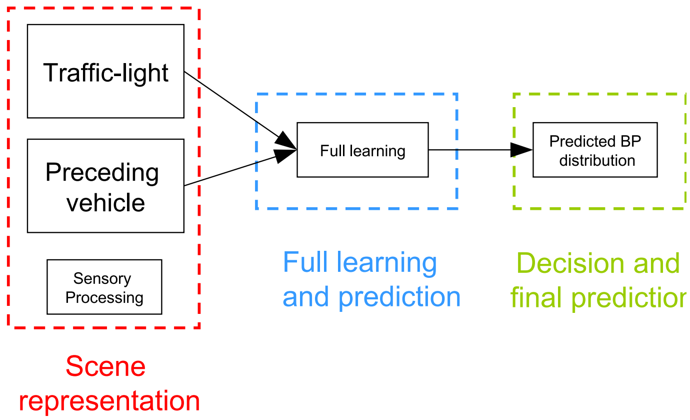

Figure 4.6: Overview of the full learning.

Situation-specific learning

The situation-specific learning system presented in this contribution and illustrated in Fig. 4.7 is first composed of a *situation prioritization* module, which analyzes the scene and triggers *situation-specific learning modules*. The strategy of the situation prioritization has been kept simple in this contribution:

- if there is a preceding vehicle and no traffic light, then the situation-specific learning module "preceding vehicle" is activated.
- if there is a traffic light and no preceding vehicle, then the situation-specific learning module "traffic light approach" is activated.
- if there is a traffic light and a preceding vehicle, then the situation-specific learning module which corresponds to the nearest traffic participant is activated.
- if there is no traffic light and no preceding vehicle, then no situation-specific learning module is activated, and no prediction is performed.

The situation-specific learning module "preceding vehicle" uses the speed and distance of the preceding vehicle as well as the speed and acceleration of the ego-vehicle as an input. In the same way, the situation-specific learning module "traffic light approach" uses the status and distance of the traffic light and the speed and acceleration of the ego-vehicle as an input. The triggered situation-specific module learns the mapping between the specific scene representation at time t and the future behavior primitive at time $t + T_{pred}$. Once the mapping is learned, these situation-specific modules predict the future behavior of the ego-vehicle.

Finally, one *fusion of predictions* module selects which prediction is relevant depending on the scene. In this contribution, the fusion of prediction is also very simple, since two situation-specific learning modules can not be activated together. We take the output of the activated situation-specific module as the output of the predictions.

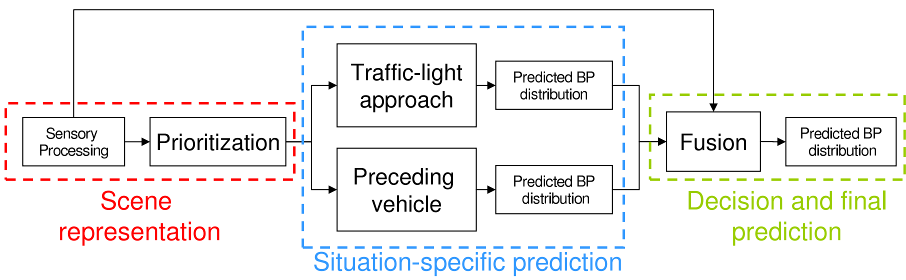

Figure 4.7: Overview of the situation-specific learning.

This approach would correspond to the case when all traffic participants and traffic infrastructure elements are used independently for behavior prediction.

Advanced situation-specific learning

The advanced situation-specific learning system (see Fig. 4.8) follows the same principle as the situation-specific learning presented previously. Additionally, when the preceding vehicle is approaching a traffic light, we apply the already trained "traffic light approach" module to predict the future behavior of the preceding vehicle. We want to assess if using the prediction trained on the ego-vehicle can be beneficial for predicting the future behavior of the preceding vehicle, and be used to improve the prediction of the ego-vehicle behavior.

We use the distance between the preceding vehicle and the traffic light, the status of the traffic light, and the speed and acceleration of the preceding vehicle at time t to obtain an estimate of the preceding vehicle future behavior at time $t + T_{pred}$. The result of this prediction is used as an additional input for the "preceding vehicle" module, which then uses the predicted behavior of the preceding vehicle, the distance and speed of the preceding vehicle, and the acceleration and speed of the ego-vehicle at time t, in order to predict the future behavior of the ego-vehicle driver at time $t + T_{pred}$.

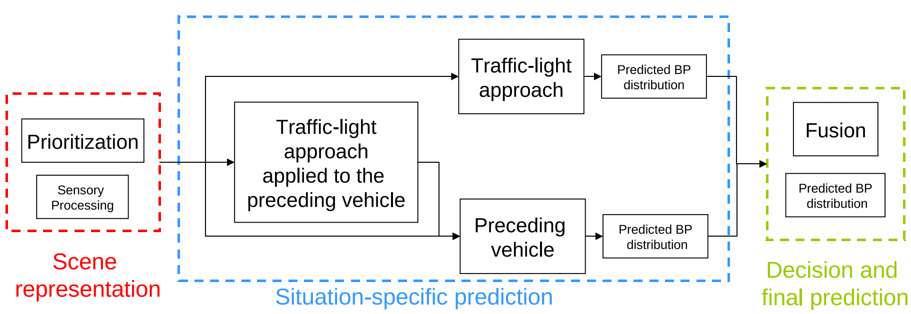

Figure 4.8: Overview of the advanced situation-specific learning. The preceding vehicle elementary situation is different compared to the one in Fig. 4.7, as it receives, as an additional input, the predicted behavior of the preceding vehicle.

This approach corresponds to the case when the influence between traffic participants is taken into account.

4.3.3 Experimental setup

We created a data-set containing 80 traffic light approach scenes, for a total of 50000 samples (image and data) of inner-city driving (see Fig. 4.5). Approximately 14000 samples are a “preceding vehicle situation”, 25000 are a “traffic light situation”, and 11000 are a situation including both traffic light and preceding vehicle. As the videos are recorded at 20Hz, this corresponds to 40 minutes of driving. We split this data-set into 6 subsets of roughly 8000 samples, in order to evaluate our systems using N-fold cross-validation.

For our behavior learning, we train MLPs according to what was presented in the chapter 3. We configure the MLP to produce four real-valued outputs $A_{stopped}$, $A_{decelerating}$, $A_{accelerating}$ and $A_{keepingspeed}$ corresponding to the predicted behavior primitives.

All MLPs have one hidden layer of size 30, and we verified that the results obtained were equivalent from 20 to 50 hidden units. They have 4 output neurons, applying a sigmoid non-linearity for hidden layer and output neurons and a bias neuron for the hidden layer and the output layer. Standard training of the MLP requires 4 *rounds* (gradient steps) before early-stopping [76] occurs (one round is one iteration over the whole data-set). We work with $\epsilon^{\mathrm{MLP}} = 0.01$.

Extraction of the preceding vehicle using laser data

Our experimental vehicle uses two ibeoLUX sensors mounted left and right under the front bumper of the experimental vehicle. Data from the sensors are integrated into a binary, metric “laser image” where filled pixels indicate the presence of a laser target (i.e., an obstacle). We use a simple template-based detection approach for horizontal segments in the metric laser image in order to detect vehicles. Tracking is used to stabilize detections and to determine the relative speed of detected vehicles. Parked vehicles are excluded by computing the absolute speed relative to the road, using the known speed of the ego-vehicle. Vehicles coming from the opposite direction are detected, by extracting objects with a negative speed, but they are not used in this work.

Encoding of the situation and behavior representation

The input data for the prediction are restricted to speed and acceleration of the ego-vehicle, distance and status of the possible traffic light, distance and speed of the possible preceding vehicle, and distance between the preceding vehicle and the traffic light, when both are present in the scene.

We compute behavior primitives for each sample of this data-set in an offline fashion, and they are encoded as a 4-element binary array, one element for each possible behavior primitive.

We compute the distance to the traffic light and the status of the traffic light for each sample of the data-set using the same approach as in chapter 3. The distance is encoded in a single real number, whereas the status of the traffic light is encoded into a 3-dimensional binary array, each element corresponding to one possible status of the traffic light (green, yellow, red).

The speed and distance of the preceding vehicle is obtained processing laser data (see Sec. 4.3.3). They are encoded in two real numbers. When both the traffic light and the preceding vehicle are present in the scene, the distance between the traffic light and the preceding vehicle is computed using the distance between the traffic light and the ego-vehicle, and the distance between the preceding vehicle and the ego-vehicle. An estimate of the acceleration of the preceding vehicle is obtained by differentiating its speed.

4.3.4 Experiments and Results on recorded data

For the following experiments, we set the threshold responsible for discarding unreliable predictions τ^{conf} to 0 (see Chapter 3). We chose not to discard unconfident prediction, in order to better assess the prediction capability and to have a fair comparison between the different systems. We verified that the results presented in Chapter 3 are still valid: discarding 10% of the most unconfident samples increases the probability of correct detection by 5% on average, for a given probability of false detection of 0.05.

We displayed ROCs for a probability of false detection up to 20%, because probability of false detection higher than 20% is not realistic for real inner-city applications.

The results presented in this section, except for the baseline, were obtained for a prediction horizon of 3s. We verified that the conclusions are also valid for 1s and 2s.

Baseline

In order to evaluate the quality of the prediction, we perform a simple prediction from the vehicle state at time t to the behavior primitive at time $t + T_{pred}$. As can be seen in Fig. 4.9, the quality of the prediction is high for an instantaneous prediction (0s), and it decreases depending on the timescale of prediction.

Behavior prediction for the traffic light situation

We first evaluated the situation-specific learning for the traffic light situation. In order to have a fair comparison between the systems, we trained the full learning system regardless of the situation, and evaluated it only on traffic light situations.

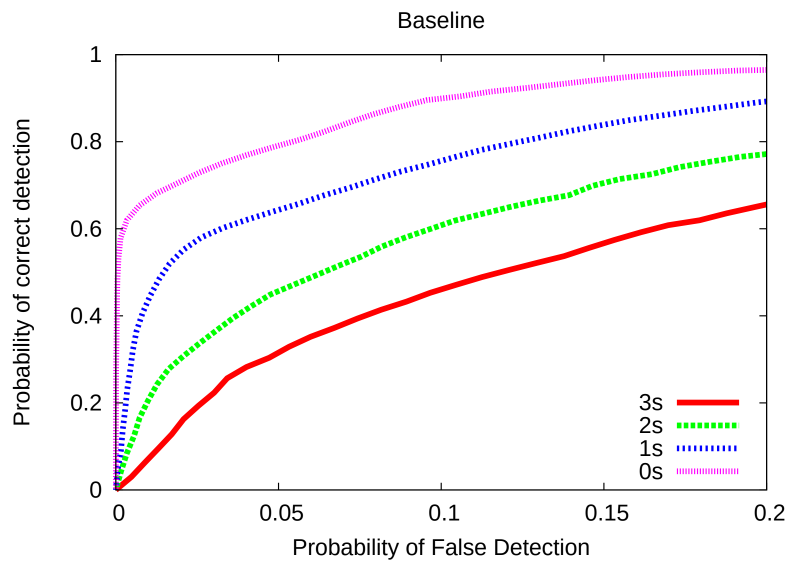

Figure 4.9: Results for the baseline.

The evaluations of the different systems can be observed in Fig. 4.10, where ROCs for a timescale of prediction of 3s are displayed. As can be seen, the prediction using all features (full learning) and the prediction using only features related to the traffic light (situation-specific learning) are equivalent.

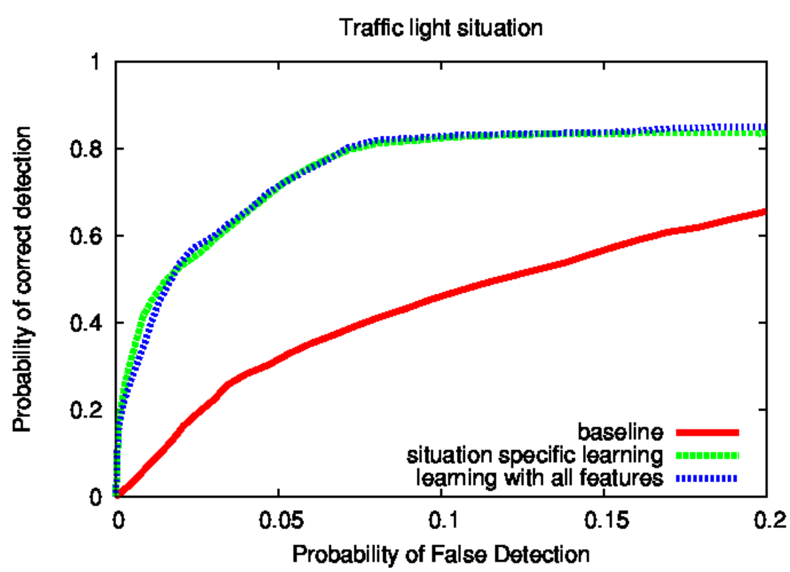

Figure 4.10: Result for the "traffic light approach" situation, T_{pred} =3s.

Behavior prediction for the "preceding vehicle" situation

We evaluated the situation-specific module for the "preceding vehicle situation", as well as the advanced situation-specific module, which uses the predicted behavior of the preceding vehicle as an additional input. The traffic light module used to predict the future behavior of the preceding vehicle was trained beforehand. We trained the full learning system regardless of the situation, and evaluated it only on preceding vehicle situation.

The evaluations can be seen in Fig. 4.11, where ROCs for a timescale of prediction of 3s are displayed. We can observe that the results for the situation-specific module and the baseline are equivalent. We verified that a learning system using only information about the preceding vehicle, without information about the ego-vehicle, reaches the same result. This means that the behavior of the driver in the car-following situation in inner-city is reactive and instantaneous most of the time. We verified this hypothesis by observing the speed and acceleration curves of the preceding and ego-vehicle over time. The comparison between the full learning and the advanced situation-specific learning shows that taking into account the prediction of the preceding vehicle behavior (advanced situation-specific learning) improves the prediction quality, which becomes equivalent to the quality of the prediction using all features.

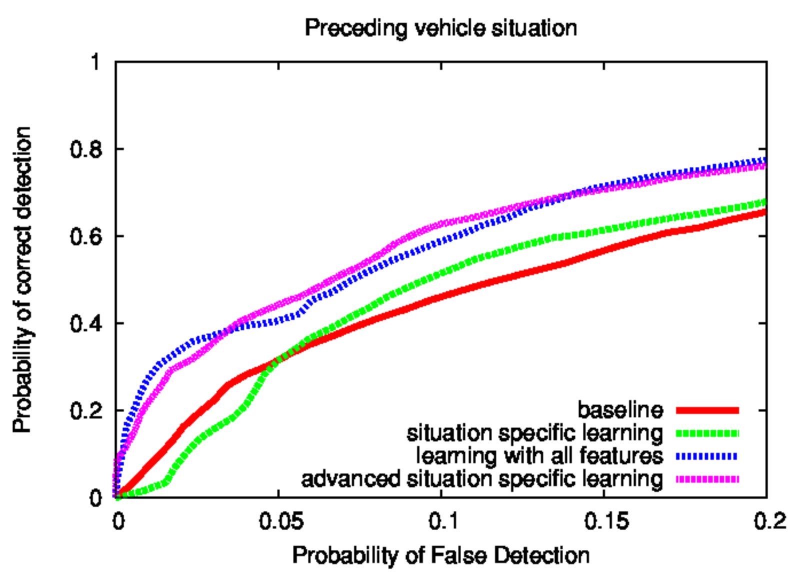

Figure 4.11: Result for the "preceding vehicle" situation, T_{pred} =3s.

4.3.5 Experiments and Results on simulated data

In order to provide additional results to support certain claims of this Chapter, we performed a simulation of a traffic light approach. We used the Simulation of Urban MObility (SUMO) simulator [49] in order to create a scenario for applying the algorithms presented in Chapter 4.

Presentation of the simulation

SUMO is a simulation environment that can be used to represent portions of routes and traffic flows. We simulated a simple single lane inner-city road of 1 km long, with a traffic light positioned at 500 m from the border of the road. The traffic light has 3 states (Green, Yellow, Red) and the dynamics of the traffic light are as follows:

- 30 seconds at Green
- 6 seconds at Yellow
- 30 seconds at Red, then it shifts to Green again.

The vehicles approaching the traffic light behave according to driving behavior already implemented. The behavior of the vehicles are generated using the default model provided in SUMO. These models are implemented following [50].

For each simulation, we initialize the scenario with 2 cars driving toward the traffic light. Their initial position, as well as the initial traffic light state, are random. By doing so, we have variability in the scenario, and we can cover multiple different situations. We record the information about the vehicles and the traffic light at 20 Hz, the vehicle closer to the traffic light being the preceding vehicle. We simulated 1000 traffic light approach trajectories, corresponding to approximately 1.6 million data points (2.2 hours). We use this data to predict the behavior of the ego-vehicle simulated driver.

For the following experiments, we predict the future behavior of the driver using the same MLP as in Chapter 4. The parameters are similar, and the quantities used for prediction are the same.

Comparison between full learning and advanced situation specific learning

In order to estimate the effect of the prediction horizon on the learning quality, we predict the future behavior of the driver in the case of the full learning and the advanced situation specific learning.

The result of the prediction for the full learning is presented in Fig. 4.12, and the result for the advanced specific learning is presented in Fig. 4.13. We present the results of the braking maneuver prediction over multiple timescales in the form of a ROC, similarly to Chapter 4.

As can be seen on these results, the results for the advanced situation specific learning are better than those with full learning. When the individual ROCs evaluating the same prediction horizon are compared, the results for the advanced situation specific learning approach are always better than the one with full learning. Moreover, the quality of the prediction decreases more strongly

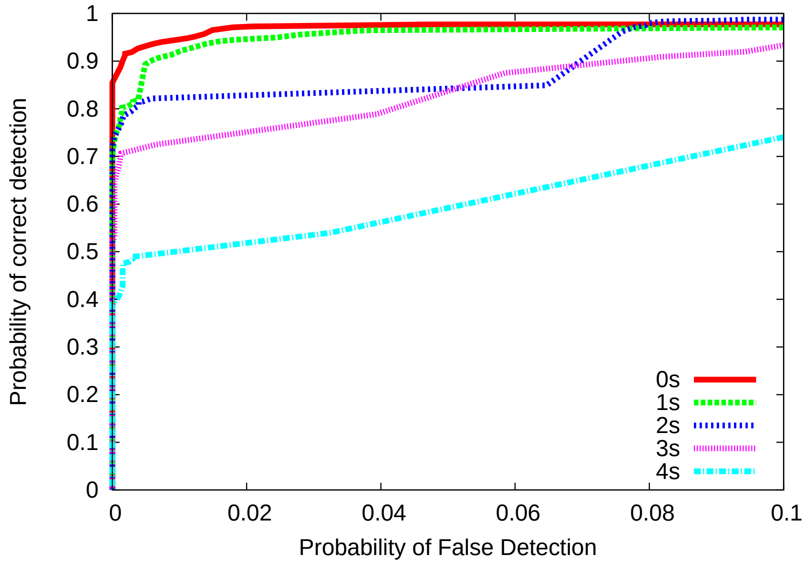

Figure 4.12: Result of the learning on simulated data for a full learning approach at multiple prediction horizons.

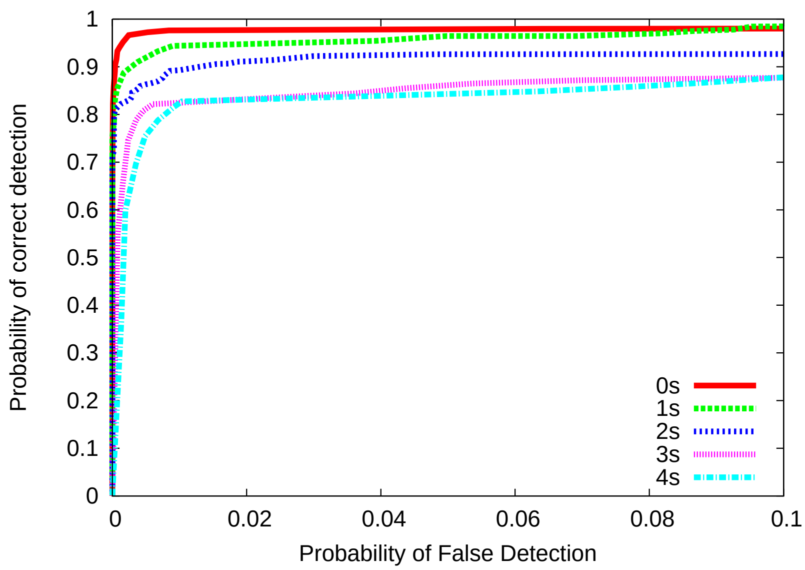

Figure 4.13: Result of the learning on simulated data for an advanced situation specific learning approach at multiple prediction horizons.

with the prediction horizon for full learning compared to advanced specific learning. If the prediction is still of good quality for 3 and 4 seconds of prediction horizon in the case of advanced specific learning, it is not for the case of full learning.

Also, one can note that as expected, the quality of the learning decreases with the prediction horizon. This can be explained by the fact that the change in traffic light status is not predictable. When the prediction horizon increases, the uncertainty concerning the future traffic of the traffic light increases also.

Finally, the prediction on simulated data seems to be of better quality in general than the prediction on simulated data. This can be explained by the fact that one behavior model was used for the simulation whereas multiple driver participated in the experiments that provide the data for experiments on real traffic light approach scenes.

Comparison of fusion approaches

In Chapter 4, we proposed to fuse the behavior predictions coming from different situation specific predictors. In order to compare possible fusion mechanisms, we predict the behavior of the ego-vehicle driver depending on the preceding vehicle and depending on the traffic light separately, and then we fuse these prediction using different strategies.

Each of the two situation specific learning system provides a set of predictions for the 4 behavior primitives, similarly to Chapter 4. If we considere that we have M behaviors predicted by N situation specific learning algorithms, we can write the result of this prediction as a set of predictions C_i^k where i is the index of the predicted behavior and k is the index of the situation specific learning predictor. A set of confidences $Conf^k$ is also available.

We propose 4 methods to fuse these predictions:

- Mean: The prediction of corresponding behaviors are averaged accross the different situation specific predictors.

$$C_i^{fusion} = \frac{1}{N} \sum_{k=0}^{N} C_i^k \tag{4.1}$$

- Pairwise Maximum: We keep the maximum of each behavior

$$C_i^{fusion} = max\{C_i^k, \text{k in 1 .. N}\} \tag{4.2}$$

- Confidence Maximum: We keep only the prediction coming from the situation specific prediction with the maximum confidence.

- Confidence Sum: we weight the prediction using the confidence, so that predictors with better confidence participates more strongly in the decision:

$$C_i^{fusion} = \frac{1}{N} \sum_{k=0}^{N} C_i^k * Conf^k \tag{4.3}$$

Once the behavior predictions provided by different situation specific predictors are aquired and fused, they provide a single vector composed of one classification value per behavior.

We apply these rules to the preceding vehicle and traffic light situation specific learning. The prediction horizon is 2 seconds. The results of the fusion of behavior predictions are shown in Fig.4.14.

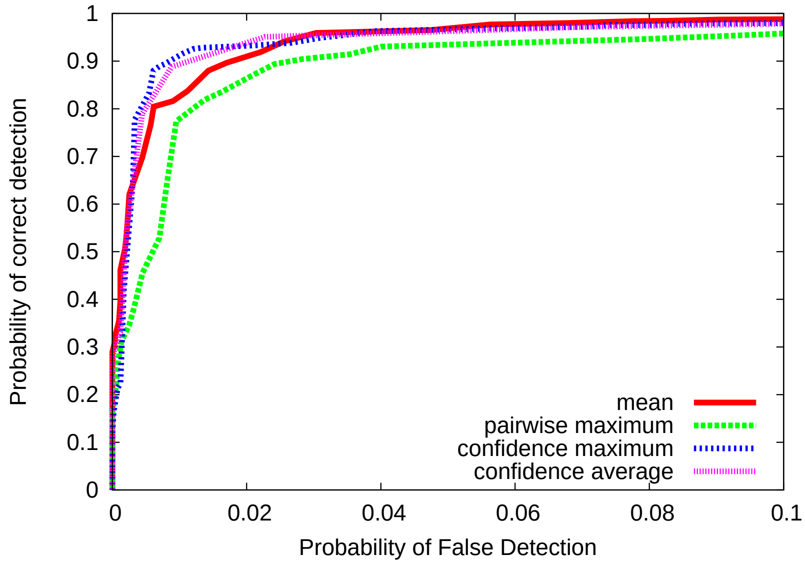

Figure 4.14: Comparison of different fusion strategies.

As one can observe, taking into account the confidence when fusing multiple predictions lead to superior results compared to Pairwise maximum and simple average of prediction. The result are approaching the quality of advanced situation specific learning when the confidence is taken into account.

4.4 Conclusion and Future Works

4.4.1 Conclusion

In this contribution, we compared several architectures for behavior prediction in an inner-city environment. We showed that it is possible to predict the future behavior of the driver using current scene information. We presented a system that uses the decomposition of a complex situation into simpler situations: situation-specific learning.

The complexity of the scene in inner-city traffic can grow very large, because of the number of traffic participants possibly interacting with each other, and thus influencing the ego-vehicle driver behavior. A system which encounters new situations is not trained to interpret them and thus to predict the future behavior of the driver. However, if we can represent this complex situation using a composition into several simpler situations which have already been encountered, we believe it becomes possible to predict the future behavior of the driver. We showed on a simple scenario that a complex situation can be decomposed into a set of simpler situations without loss of prediction quality. However, it is not always that easy to decompose a situation. On highways, for example, when the ego-vehicle approaches a frontal preceding vehicle, the behavior prediction needs the knowledge about the presence or absence of a left preceding vehicle to correctly predict the behavior. So the decomposition needs to be performed with caution.

More importantly, we showed that we could use what has been learned from the point of view of the ego-vehicle, and apply it to other traffic participants. This can be applied for anticipating the behavior of other traffic participants, in order to obtain more advanced predictions. If we can predict what other traffic participants will do, it stands to reason that this will improve ego-vehicle behavior prediction.

4.4.2 Future works

The benefits of situation-specific learning regarding the scalability will have to be demonstrated by applying this concept to more complex situations. Even if this proof of concept shows the feasibility of the approach, applying it to more complex scenarios will prove that the approach can scale.

Finding an approach to decompose a scene and to find relations between traffic participants is not a trivial task, and is a research topic on its own. It has to be part of a scene understanding module, and it has to be investigated. It is needed to explore the ways of generalizing the approach presented in this Chapter, so that it can adapt to all possible scenes.

We can estimate the number of elementary situations that could be necessary for a correct representation of complex scenes in usual driving scenarios. In inner-city, we need elementary situations to represent the influence of dynamic actors of the scene: possible vehicles surrounding the ego-vehicle, pedestrians, bicycle and motor bikes. If we count one elementary situation per class of dynamic actor, we need at least 4 elementary situations, that can be refined if needed (multiple types of vehicles for example). We also need behavior prediction depending on static parts of the scene, which englobe traffic signals in general (road signs, traffic lights, speed bumps, zebras). It is not sensible to consider that each kind of traffic sign is a separate elementary situation, so we would group them as one elementary situation. On highway, we would need one prediction system for vehicles, one for motor cycles. Then, one would be used for traffic signs. Similarly to the inner-city case, elementary situations could be refined in order to express the difference in usual behavior of different classes of vehicles (e.g. sports car, truck or touring car). In this case, at least 3 behavior primitives would be necessary.

The fusion of predictions in this scenario was rather simple. Even if we believe that it is most of the time possible to prioritize behavior prediction coming from different situation-specific modules, it would be interesting to develop a generic approach for taking decisions based on multiple predictions.

Chapter 5

Behavior prediction on a limited sensory setting

In the previous chapter, we studied behavior prediction in inner-city scenarios. In this chapter, we will study behavior prediction on highways, in the scenario of a lane change by the ego-vehicle in order to pass a frontal preceding vehicle. Compared to the previous inner-city scenario, where the predicted behavior was longitudinal, we will predict the future lateral behavior of the ego-vehicle. We will also present the issue of coping with a limited sensory setting. We will try to predict the left lane change behavior using only front sensing. Also, we will propose an indicator for estimating whether the behavior prediction should be used depending on the scene.

5.1 Introduction

In this chapter, we will predict the behavior of the ego-vehicle driver based only on front sensing on highways. Predicting lane-change is important to prevent accidents. First of all, if a driver does not perform a lane change when he should, it is a signal for dangerous situations. Then, when applied to other traffic participants, it can be used to detect lane change maneuvers of the right preceding vehicle before it occurs, and then act accordingly.

5.1.1 Limiting the sensory setting to front sensing

Most of the important information for predicting the behavior of the driver on highways is concentrated on the front scene. The longitudinal behavior mostly depends on whether the driver is free to reach his desired speed or if he has to slow down because of a front vehicle. If there is a vehicle preventing him from reaching his desired speed, the ego-vehicle driver will try to change lane, and this change of lane depends on whether the lane target lane is occupied or not, and whether the speed of the vehicles on the other lanes is higher than the one on the current lane. The lane change behavior is naturally conditioned by the presence or absence of an incoming follower on the destination lane. If we want to reduce the sensory setting, we have to answer the following question: Is it possible to predict whether the driver is going to perform a lane change even if the system has no information about the rear scene? Is rear sensing mandatory? Basically, we want to find out if the way a driver approaches a preceding is also an indicator whether or not he will try to change lane.

5.1.2 Using the system only when possible

The behavior of the driver highly depends on the congestion of the highway. If the driver is in congested traffic, the likelihood of performing a lane change will be reduced. Indeed, there will be more incoming followers delaying the lane change. Also the difference of speed between the current lane and the destination lane will be lower, which will then decrease the advantage the ego-vehicle driver will gain by performing a lane change. In the case of congested traffic, the usual behavior prediction system is not valid anymore. Is it possible to derive an indicator that tells the system when the conditions for behavior prediction are not met?

5.1.3 Avoiding the use of driving monitoring

As we will see in the next section, most contributions on lane-change detections use driver monitoring. Quantities derived from the observation of the driver (e.g. head motion or gaze) give good indicators about the intent of the driver: a driver who wants to change lane will turn his head first. But these systems learn to rely solely on driver monitoring in order to predict the behavior of the driver. This is not acceptable: what if the driver forgets to turn his head, or forgets the turning lights?

Moreover, we want the prediction from the ego-vehicle perspective to be applicable to other traffic participants, which is not possible in the case of driver monitoring except if we use V2V and V2I communication. Once again, if a car is not equipped with such technologies, predicting the behavior of his driver becomes impossible.

5.2 Related Works

Numerous contributions these past few years presented systems to infer the driver future behavior ([37], [77], [36], [19], [42] or [56]). We present three contributions that we consider related to the approach we propose in this study, and working on scenarios similar to ours.

In [40] the authors present a study of indicators for lane-change intent extracted by observing the surroundings as well as by monitoring the driver. Even if they do not perform prediction, they conclude their study by affirming that the driver head monitoring and the left turn signal are strong indicators for lane change intent. This explains why so many contributions presented in [22], a review about driver intent inference, use driver monitoring.

In [68], the authors equipped a car with limited sensory device: one front camera and one rear camera. They also use driver monitoring in order to extract the position, orientation and viewpoint of the driver. They can predict different types of driving behaviors on highway on average 1 second before they occur, and in particular the lane change on average 0.1 seconds before it occurs.

Finally, recent works by [63] present a lane change intent detection system on highway. They perform detection of rear, side and front vehicles using a combination of different sensory inputs and obtain a reliable scene representation. Driver monitoring (head orientation and blinker state) is used in addition to the surrounding monitoring. Using these sensory inputs, they can predict the lane change up to 3 seconds before it occurs. As stated in their contribution, the performance of the system is mainly due to the head monitoring.

These works are of great value to explore the possible capabilities of behavior intent prediction systems. However, they all include monitoring of the driver and rear sensing. We want to evaluate the possibility of driver behavior prediction using only front sensing, limiting as much as we can the amount of sensors. Another reason for avoiding driver monitoring is that using only transferable quantities to build the system (e.g. speed or acceleration) allows the knowledge learned from the point of view of the ego-vehicle driver to be applied to other traffic participants behavior prediction (see Chapter 4), which is not possible if part of the system is based on driver monitoring.

5.3 Methods

The prediction of the future behavior of the driver is based on the current front scene. Inputs from sensors are filtered and used to compute indicators, which represent relevant information about the current observed scene. We did not implement Situation-specific learning as described in Chapter 4, because the amount of available training data did not allow us to train 3 different systems (one for each following vehicles).

The behavior of the driver is a binary target: Is the driver going to perform an action in the future prediction horizon? In this sense, we have to solve a classification problem.

We used Locally-Weighted Projection Regression (LWPR) as the learning algorithm [92]. It is fast and suitable for online learning. This algorithm is designed to build incrementally a knowledge representation in the form of receptive fields (internal representation). We believe that the algorithm used is not important. It was shown in [32] and [34] that LWPR and MLP reach similar results given that the input vectors are presented in a suitable form. What matters most is the discriminative power of the indicators for the learning task. We used the publicly available implementation of [48] by the authors for all described experiments.

In the training phase, we train the LWPR using the indicators as an input and the labeled future behavior as a learning target. A representation of the system can be observed in Fig. 5.1.

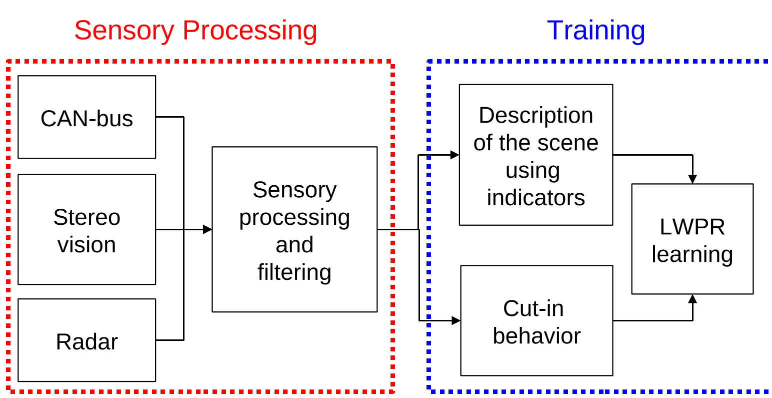

Figure 5.1: The mapping between present descriptors of the scene and the future behavior is learned by the LWPR.

In the prediction phase, we use the output of LWPR to estimate if the driver is going to perform a lane change in the next 2 seconds. We then can compare the prediction to the ground truth and evaluate our prediction system (see Fig. 5.2).

5.3.1 Sensory Inputs

The CAN-bus provides information from which we derive filtered quantities about the ego-vehicle. At each timestep t_0, we have:

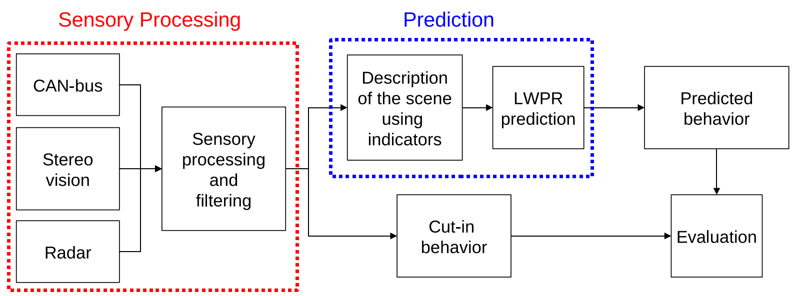

Figure 5.2: The mapping learned by the LWPR is used to predict the future behavior for scenes that were not previously encountered.

- the longitudinal and lateral absolute velocity of the ego-vehicle: $AbsLongVel_{ego}(t_0)$ and $AbsLatVel_{ego}(t_0)$
- the longitudinal and lateral absolute acceleration of the ego-vehicle: $AbsLongAcc_{ego}(t_0)$ and $AbsLatAcc_{ego}(t_0)$

The sensory system provides information about the surrounding vehicles. We consider only the frontal preceding vehicle $V_{frontal}$, the left preceding vehicle V_{left} and the right preceding vehicle V_{right}, as described in Fig. 5.3.

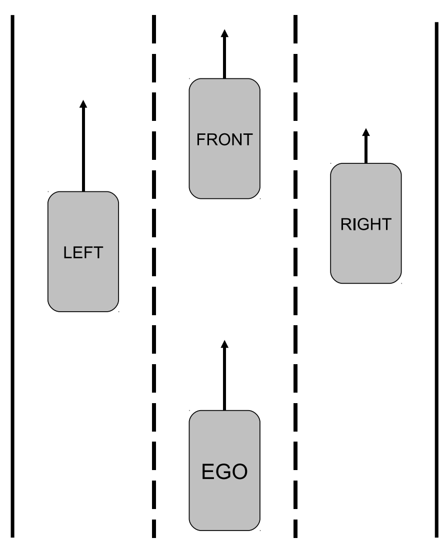

Figure 5.3: The sensors detect the frontal preceding vehicle, the left preceding vehicle and the right preceding vehicle as well as the lanes. In most cases, only one or two of these preceding vehicles are present in the scene.

The sensory system provides for each vehicle i at time t_0:

- the relative lateral and longitudinal positions: $LatPos_i(t_0)$ and $LongPos_i(t_0)$
- the lateral absolute and lateral relative velocity: $AbsLatVel_i(t_0)$ and $RelLatVel_i(t_0)$

- the longitudinal absolute and longitudinal relative velocity: $AbsLongVel_i(t_0)$ and $RelLongVel_i(t_0)$
- the longitudinal absolute and longitudinal relative acceleration: $AbsLongAcc_i(t_0)$ and $RelLongAcc_i(t_0)$

The sensory system also detects vehicles in front of these preceding vehicles. The presence of those vehicles will be used to determine if the highway is congested or not.

5.3.2 Indicators

We selected indicators in order to include knowledge in the system: these quantities are descriptors of the current scene. Their level of abstraction is higher than basic sensory inputs. For each timestamp t_0 we compute:

- The time to contact to the frontal preceding vehicle which quantifies the way the driver is approaching the frontal preceding vehicle. It represents the time, at constant speed, before the ego-vehicle reaches the frontal preceding vehicle:

$$TTC_{frontal}(t_0) = \frac{LongPos_{frontal}(t_0)}{RelLongVel_{frontal}(t_0)} \tag{5.1}$$

- The difference to the desired speed, measured by calculating the average past speed in a time window of δT (5 minutes in this contribution) and comparing it to the speed of the frontal preceding vehicle. It indicates if the frontal preceding vehicle goes slower (and how much slower) than the cruise velocity of the ego-vehicle, and provides a good measure of the incentive to change lane:

$$\begin{aligned}\Delta Vel(t_0) = \frac{1}{\delta T}\int_{t_0-\delta T}^{t_0} & AbsLongVel_{ego}(t)\,\mathrm{d}t \\ & -AbsLongVel_{frontal}(t_0)\end{aligned} \tag{5.2}$$

- The distance to the left preceding vehicle gives information on the availability of the left lane. If a car is too close to the ego-vehicle, it prevents the cut-in :

$$LongPos_{left}(t_0) \tag{5.3}$$

- The difference between the left and frontal preceding vehicle speed. If this difference is high, the gain the driver will have by changing lane increases:

$$\begin{aligned}\Delta Vel_{left-frontal}(t_0) = & AbsLongVel_{left}(t_0) \\ & -AbsLongVel_{frontal}(t_0)\end{aligned} \tag{5.4}$$

All the indicators are scaled between 0 and 1. If needed, the indicator distribution is inverted so that its value increases together with the probability of lane change. Extremas related to division by zero or errors in detections are set to 0. An example of the evolution of the modified $TTC_{frontal}$ can be observed in Fig. 5.4.

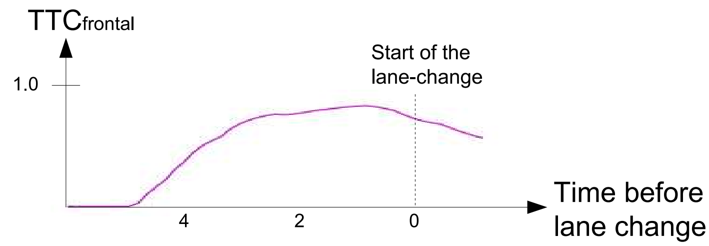

Figure 5.4: Evolution of the normalized $TTC_{frontal}$ before a lane change.

5.3.3 Target of the prediction system

We want to predict the left lane change, and we define the start of the maneuver as the moment when the car starts turning. Instead of using the behavior segmentation proposed in Chapter 3, we decided to label by hand the moment when the car starts turning by observing the video streams. The proposed method for behavior segmentation on highways was proposed after that the experiments concerning behavior prediction were performed. We labeled the frames before the lane change as follows:

- up to 2 seconds before the lane change starts: label 1
- 2 to 4 seconds before the lane change: not used for training
- more than 4 seconds before the lane change: label 0

We do not perform learning from 4 to 2 seconds before the lane change in order to facilitate the task of the learning algorithm, which will identify more easily the difference of features by comparing their distribution before 4 seconds and between 0 to 2 seconds. The temporal window when learning does not occur has been chosen empirically. The labeling of a lane change scene can be observed on Fig. 5.5.

5.3.4 Measure of the congestion of the highway

In order to determine if the system can function properly given the current driving condition, we derive an estimate of the congestion of the highway $\tau_{congestion}$. We measure the exponential moving average of the amount of vehicles counted in the front scene at time t_0:

$$\tau_{congestion}(t_0) = (1-\alpha)\#vehicles(t_0) + \alpha\tau_{congestion}(t_0 - 1) \qquad (5.5)$$

Where $\#vehicles(t_0)$ is the amount of vehicles detected at time t_0, and α is a coefficient determining the importance of past measures compared to new ones. It is similar to an integration period.

This can be used as an indicator or as a switch to turn off the system. Basically, if the driver is in a traffic jam and wants to change lane, he will do it as soon as he can, but knowing that there is a traffic jam will not help to determine the precise moment. So we believe it is better to use this indicator as a validation step for the prediction, answering the question: Are we in good conditions to perform prediction?

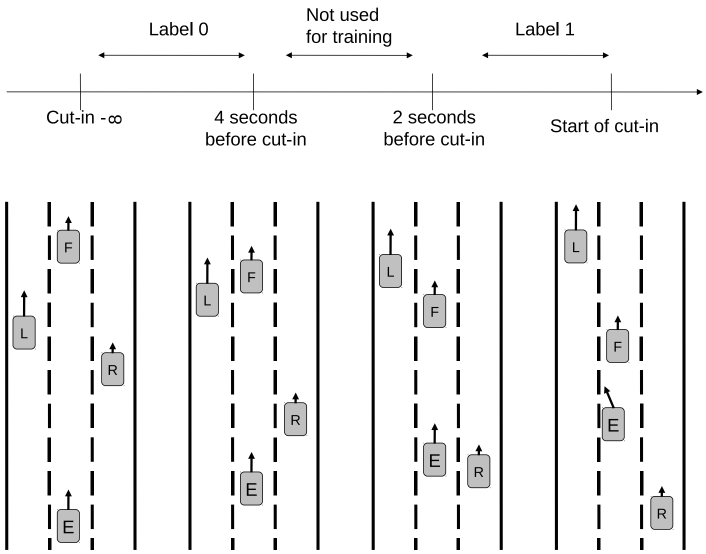

Figure 5.5: Labeling of the scene before a lane change event.

5.3.5 Building of the data-sets

As stated before, one of the aims of this contribution is to evaluate the importance of rear sensing by comparing prediction of scenes with and without contribution from rear vehicles. For this we created two data-sets:

- the first one includes all the previous indicators
- the second one includes all the previous indicators in addition to a label representing the presence of an incoming left follower, simulating rear sensing.

With these two data-sets, we can evaluate the importance of rear sensing, and we can evaluate whether a prediction system based only on front sensing can be reliable enough.

5.4 Experiments: predicting lane change on highway

In this contribution, we present a lane-change prediction system on highway based only on front sensing. One of these lane changes is presented in Fig. 5.6

The presented system should solve a classification task answering the following question: Is the driver going to perform a lane change in the next 2 seconds? We avoid using any form of driver monitoring (like turning lights, head orientation, or gaze) for reasons explained in Sec. 5.1.3.

We present several indicators representing behaviorally relevant knowledge. We show that front sensing is sufficient to give satisfactory results, and that the front scene holds most of the information necessary to predict lane change behavior. Of course, we also show that, as expected, additional information about the rear scene helps the system to achieve better predictions. We test our system using event-based as well as frame-based evaluation, and we use N-Fold cross validation in order to ensure a reliable and fair evaluation.

Figure 5.6: Example of an approach of a preceding vehicle followed by a lane change. One image every 2 seconds are displayed, and organized from left to right and top to bottom.

5.4.1 Evaluation Measures

The evaluation in several contributions about future behavior prediction is performed frame-based: the system evaluates for each timestamp if the prediction is correct or not. This evaluation has a real scientific value, because it is a direct measurement of the learning capability of the system. However, it is not suitable if we want to evaluate the real-world capability of the system. We add to this standard method the event-based evaluation: we group consecutive predictions together, forming temporally consistent events. N successive predictions of a future behavior form an event, and we check if it intersects with a target event. This type of evaluation is very useful to identify clusters or incorrect predictions at the scene level. Also, it enables us to measure the amount of false positive events per hour, a measurement commonly used in the automotive industry [22].

N-Fold cross-validation

We use N-fold cross-validation in order to evaluate our system. We split the data-set into N subsets, each containing a scene of the ego-vehicle approaching a frontal predecessor. We train the system using N-1 subsets and we present the samples from the remaining subset to the trained prediction system.

We then use the activations from the N evaluation subsets, obtained from the N possible combinations of training and evaluation subsets, in order to evaluate the quality of the prediction over the whole data-set.

The output of LWPR algorithm gives us a sequence of activations which is normalized by its long-term mean and variance. Then, for each timestep t_0 we obtain a classification value $C^{\text{class}}(t_0)$.

Frame-based evaluation

We used frame-based evaluation as one of the means of assessing the quality of the prediction. We use the same evaluation approach as in the previous chapters. This performance measure is a standard tool in machine learning and has been used to evaluate behavior prediction systems (see, e.g., [56, 63]). In the presented ROC, we plot the detection rate against the false positive rate.

Event-based evaluation

For each threshold τ^{class}, we obtain a sequence of decisions for the prediction. For each timestep the future lane change is either predicted or not. We group together consecutive positive predictions of lane change behavior. One interval of these predictions is called an event.

If this event intersects with the 2 seconds interval before the lane change, we consider that the prediction is correct, and we count one true positive event. If this event does not intersect with the 2 seconds interval before the lane change, the prediction is incorrect, and we count one false positive event. We count the amount of of false positive events per hour for each value of the threshold. In particular, we can decide to set the threshold τ^{class} in order to obtain a certain percentage of correctly detected events, and evaluate the amount of false positive per hour for this percentage of correctly detected events. In addition, we calculate the average length of the prediction event.

5.4.2 Experimental setup

Our experimental vehicle is equipped with a front camera and a front radar recording data at a frequency of 10 Hz. A combination of radar and video is used to detect traffic participants and their position relative to the ego-vehicle. From this information, we compute their position, speed and acceleration, processed using a Kalman Filter. The cameras are used to detect lane markings, in order to assign a lane to each traffic participant. The sensory setting is described in [67], and illustrated in Fig. 5.7. The lateral position extracted with radar is refined using video processing.

We use the CAN-bus data in order to obtain information about driver behavior: the speed and acceleration of the ego-vehicle are also processed using Kalman filtering. We choose not to use turning lights status or driver head monitoring, as motivated in 5.1.3.

The start of the lane change behavior has been annotated. In the future, if we want the system to learn online and adapt to the driver behavior, it can be replaced by a detection of change in lane orientation, as shown in Chapter 2.

Our data for this study has been acquired by recording the driving of the ego-vehicle on German highways. 50 minutes of driving were recorded under classic driving conditions. We labeled 20 examples of lane change situations:

- 10 scenes are normal lane change scenes: the ego-vehicle approaches the frontal predecessor and cuts in
- 5 scenes are lane changes caused by the right preceding vehicle cutting in in front of the ego-vehicle
- 5 scenes are delayed lane change situations: the ego-vehicle approaches the front preceding vehicle but slows down and waits before cutting in, because of an incoming left follower.

These lane change scenes represent usual situations encountered during driving. As an example, we show one of the normal lane change scenes in Fig. 5.6

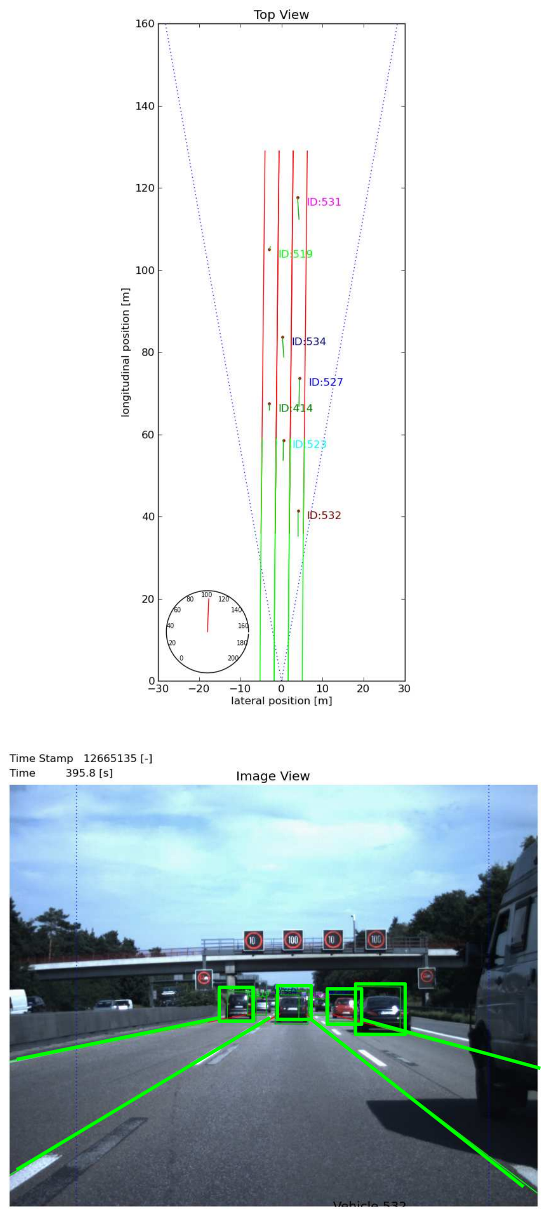

Figure 5.7: Result of sensory processing: vehicles and lanes detection illustrated with ego-vehicle perspective and top-view perspective.

5.4.3 Experiments and Results

Results for future behavior prediction

The correct detection rate against the false detection rate is presented in the ROC curve in Fig. 5.8. This frame-based evaluation shows that the prediction capability of the system improves slightly if the data-set includes information about rear vehicles. The evaluation also shows that the prediction capability even without relying on driver monitoring is of good quality.

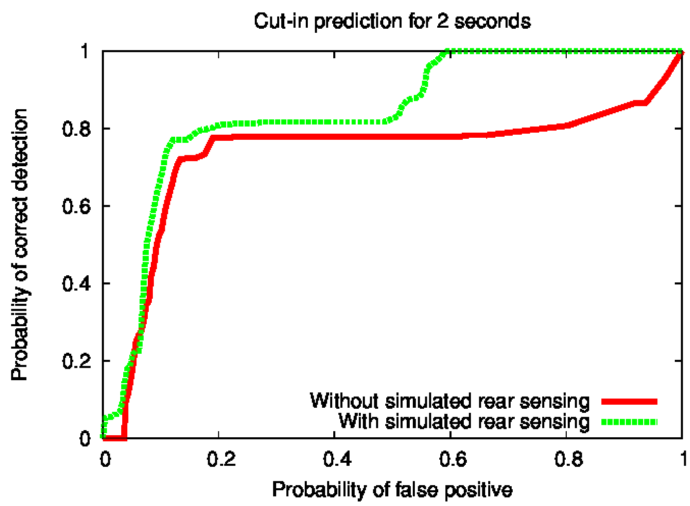

Figure 5.8: ROC presenting the results for the frame-based evaluation. As expected, the results are better using a simulated rear sensing.

We evaluated the amount of false positive events per hour in the same condition. We set the classification threshold in order to obtain 80 % of correct detected events.

If we use the data-set for which a label representing an incoming left follower was included, we measure 32 false positive events per hour. The average length of the events is 33 timesteps (3.3 seconds) for positive events, and 37 timesteps (3.7 seconds) for negative events.

Without this labeling of an incoming left successor we measure 45 false positive events per hour. The average length of the events is 40 timestamps (4 seconds) for the positive events, and 48 timestamps (4.8 seconds) for the negative events. It is important to note that a similar system using only the time to contact $TTC_{frontal}(t_0)$ reaches a performance of 52 false positive events per hour. That means that a huge part of the necessary information is contained in this indicator.

The results demonstrate that the system can learn the future behavior of the driver even without driver monitoring. The prediction without rear sensing is possible, but the results are not as good as with simulated rear sensing.

Characterization of errors

Using event-based evaluation, we could identify several factors responsible for false positive events that can be taken into account in future experiments. First of all, multiple detection errors occurred during the processing of the sensory inputs, and were responsible for 5 false positive events. We will have to explore the possible improvements for the sensory setting. In particular, the detection can be improved by having a better fusion of information. A second possible reason for false positive is the case when the driver is in a traffic jam, or in a congested traffic. 5 false positive events occurred during traffic jams: the driver does not change lane, even if he can, because there is no gain for him. A third reason is the case when the driver wants to exit the highway in the near future (3 false positive events). He stays on the right lane, and again, even if he has the occasion to change lane, and even if in normal conditions he would have changed lane, he does not.

Using the estimation of congestion

We measure the congestion of the highway using the method described in 5.3.4. For each timestamp, we have an estimate of the congestion of the highway $\tau_{\text{congestion}}$. If it is superior to $\tau_{\text{congestion}}^{max}$, the system can not perform prediction, and is shut down.

We study the effect of this strategy depending on several values of $\tau_{\text{congestion}}^{max}$. For these experiments, we chose a coefficient $\alpha = 0.999$. The results of this approach can be seen in Tab. 5.1. We displayed the amount of correct detection and false positive events given that we obtain 80 % of detected events.

$\tau_{\text{congestion}}^{max}$	False Positive events	Correct detection
3	8	8
3.5	16	12
4	32	12
4.5	40	15
5	45	19

Table 5.1: Results of lane change prediction depending on the value of $\tau_{\text{congestion}}^{max}$, for 80 % of detected events.

As we can see, if we limit the use of the system to cases when the traffic is fluent, the ratio between correct detection and false positive events increases.

5.5 Discussion and Future works

5.5.1 Discussion

We presented a set of indicators which are used to perform behavior prediction on highway. The system predicts whether the ego-vehicle driver is going to start a change to the left lane in the next 2 seconds. This prediction does not require any driver monitoring. We evaluated the importance of rear sensing, and showed that although the prediction was not completely dependent on rear sensing, knowledge about the rear scene improves the results of the prediction. We believe that even a simple rear sensing detection will improve the results greatly.

The use of an indicator to determine whether the conditions of utilization of the system are correct improved the ratio between correct detections and errors. We believe it is an important finding: the system has to know when the conditions for a reliable predictions are not met. This measure is different from the reliability measure proposed in Chapter 3. Instead of relying on the uncertainty of prediction, it directly derives an indicator from the observed scene. Of course it does not suppress all false positives, but it shows that context indicators can be used to modulate the result of a prediction system.

5.5.2 Future Works

First of all, it is necessary to record more driving scenes on highway, in order to show that our approach generalizes well. The data used for this contribution were recorded with two different drivers. However, the quantity of available data did not allow us to compare the influence of driving styles properly. So it would be interesting to record driving sessions with different drivers, in order to evaluate the adaptation capability of our system, as well as the influence of the driving style of the driver.

When a driver approaches a preceding vehicle, he can either change lane (left or right) or stay in the lane but reduce his speed. A necessary improvement would be to predict both the lane change behavior and the braking behavior in order to introduce competition between the classification outputs. This competition mechanism is observable in the human brain, and has been research in the community of Neural Network (see [4, 5]). We believe that it could lead to better results.

Finally, an important extension would be to perform the prediction at multiple time-scales and identify the time of prediction leading to the best results. Also, this prediction at multiple timescales can be used as a temporal integration, or a verification for temporal consistency, in order to remove false positive events.

Chapter 6

Conclusion and Outlook

In my opinion, there is a clean and safe way to transport people from one point to the other with reasonable speed if they have time to spare. It is a really cheap technology, it is based on the improvement of built-in capabilities of human beings. It allows us to take our time, and it is even good for health. It is called a good pair of shoes. I went to work nearly every day during the last six months of my PhD using this means of transport in order to thoroughly examine it, and submit it to extensive testing. The conclusion is that it works. Of course I only had six kilometers to walk. It is not too fast, but provided that you have time, you are never late. There are no traffic jams for pedestrians, nothing to block the road. It uses food as energy, and food, as everybody knows, is not a rare resource in our world, contrary to gas or (soon-to-be) electricity. It does not pollute at all, and if you need to go faster, you can still take a bike.

However, in our society, people need to take their vehicle for multiple reasons, and they often can not do without it. It is part of their daily life, and good sentiments about walking, taking a bike or taking the bus will not change their view. We live in a world where everything needs to be done fast, where taking one hour to go to work on foot is not acceptable. People need their car to go to the shopping mall, bring their kids to school, or go see their relatives every once in a while. Honestly, I have no arguments against that. The technology exists, and is quite useful, so why not use it? The problem is that it comes with a cost: pollution, traffic accidents, and the discomfort of having to drive at 10 $km.h^{-1}$ when your car can go 140, while being stuck in a traffic jam at 7 in the morning. Even if the best way to avoid traffic accidents, reduce energy consumption, and improve the comfort (and health) of people is to simply reduce the use of vehicles, it is not a suitable way, because people are dependent on their cars and often do not have a choice. So we have to find other ways to improve transportation means so that they are more energy efficient, more safe, and more comfortable.

This thesis is based on this conclusion. And in order to do so, we need to be able to predict what traffic participants are doing or will do, as explained in Chapter 1.

6.1 Conclusion and Discussion about the work presented in this thesis

In this contribution we highlighted scientific directions that needed to be explored for developing intelligent vehicles in order to provide a safer, less polluting, and more enjoyable way of transporting people. This requires that the system perceives its environment, reasons and decides what to do, and acts by warning the driver or by taking partial or total control of the vehicle. We focused our work on the reasoning part, with the aim to research the possibilities of driver behavior prediction.

In Chapter 2, we proposed an approach for describing human driver behavior, and quantify it in a way that is suitable for behavior prediction. As explained, multiple approaches can be formalize. The one we propose, based on the abstraction from quantities such as speed, acceleration, or vehicle status, derives behavior primitives that are also observable and quantifiable for other traffic participants, allowing to apply what has been learned in the point of view of the ego-vehicle to other traffic participants. It is possible to adopt a finer way for describing human driver behavior, and it is to be determine which representation is more suitable for the task at hand.

In Chapter 3, we showed on a simple example that the present and future behavior of the driver could be predicted using the perception of the scene. The prediction was performed at an object level, which means that it uses already processed quantities giving abstract information about the driving scene. We presented indicators that are used to evaluate the quality and reliability of the prediction.

In Chapter 4, we proposed a solution to handle multiple traffic participants possibly influencing each other. We showed on an inner-city example that the approach could be used to combine knowledge already acquired on different scenes and use them on new scenes. We also showed that it is possible to transpose what was learned from the point of view of the ego-vehicle to other traffic participants, given that the quantities used for prediction are observable in other traffic participants. It is clear that our approach has to be evaluated on more complex scenarios. Also, it requires a system which autonomously translates the visual scene in a list of inter-connected objects. This scene understanding not yet exists, and the quality and reliability of behavior prediction for more complex scenes will greatly depend on it.

In Chapter 5, we studied the possibility of performing behavior prediction on a limited sensory setting. We showed that it is possible to predict the change lane behavior of the ego-vehicle even without knowledge about the vehicles behind the ego-vehicle. We also proposed a measurement for traffic congestion which enables to characterize the reliability of the behavior prediction depending on the amount of traffic participants. Limiting the sensory setting reduces the information we have about the scene, and it is reasonable to think that the more knowledge we have about the driving environment, the better the prediction will become. To summarize Chapter 5, we could say that behavior prediction on a limited sensory setting is possible under certain conditions and up to a certain quality, but that no safety systems could work properly without a strong perception of the surroundings of the car, because these systems need to be perfect.

6.2 Outlook: what is still to be done

At the time when I finished my PhD at Honda Research Institute GmbH, several projects involving driver behavior prediction were on their way. In most of them, it was required to use annotations of the positions and characteristics of traffic participants because a reliable perception system in order to acquire knowledge about the scene was missing. It is a really difficult scientific problem, and a lot of research is being done in multiple teams around the world to improve the perception of the vehicle. I think that providing a reliable perception system is a mandatory step, because all works involving reasoning are depending on this perception. Also, as most intelligent systems are based on learning, we are very dependent on the quality and quantity of training data. I think it would be necessary to obtain more data-sets about driving scenes in order to be able to train the systems on more examples.

Concerning behavior prediction, this contribution is based on exploratory research, and there is still a great amount of work to be done. As we explained in this thesis, driver behavior prediction is a stepping stone towards assisted driving or autonomous driving. It is necessary to further ensure safety and comfort of traffic participants, and once engineered, it could provide solutions for eco-driving and the reduction of pollution. In this thesis, we showed the feasibility of behavior prediction on certain scenes, but it is necessary to cover more scenarios. Also, we predicted the usual behavior of traffic participants given a certain configuration of the driving environment, which was occurring often enough so that it could be learned using machine learning techniques. But a system which handles unusual situations is also required. It is not yet sure that it could be trained using techniques similar to the ones we presented. Indeed, the occurrence of such unusual events are by definition scarce, so statistical learning is not compatible with this kind of behaviors. This is closely related to the way humans drive, and how they perceive immediate danger.

Concerning my future works, I plan to continue working on behavior prediction on my free time. It is such an interesting topic! I plan to focus on highway scenes in the first place, as data are much easier to obtain. I plan to build a benchmark data-set with scenes recorded on highway with multiple cameras, and to work on the vehicle detection and characterization. I will start by implementing existing techniques for visual detection of vehicles, and improve them if necessary. I will continue the work on behavior prediction once a reliable detection system is available. I would like to build a system for recording data-sets easily, so that researchers of the field could easily implement it, contribute by adding new data-sets, and share and compare the results of their research.

There is still a lot of work to be done.

Bibliography

[1] E. Ackerman. Stanford's autonomous racer hits 120 mph on the track. `http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/stanfords-autonomous-racer-hits-120-mph-on-the-track`, 2012.

[2] A. Alessandrini, C. Holguin, and M. Parent. Advanced transport systems showcased in la rochelle. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 896 –900, oct. 2011.

[3] A. Alin, M.V. Butz, and J. Fritsch. Incorporating environmental knowledge into bayesian filtering using attractor functions. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 476 –481, june 2012.

[4] S-I Amari. Neural theory of association and concept-formation. *Biological Cybernetics*, 26(3):175–185, 1977.

[5] Shun-Ichi Amari and Kenjiro Maginu. Statistical neurodynamics of associative memory. *Neural Networks*, 1(1):63–73, 1988.

[6] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[7] M. Barth and K. Boriboonsomsin. Energy and emissions impacts of a freeway-based dynamic eco-driving system. *Transportation Research Part D: Transport and Environment*, 14(6):400 – 410, 2009. ¡ce:title¿The interaction of environmental and traffic safety policies¡/ce:title¿.

[8] A. Bazzi and B.M. Masini. Taking advantage of v2v communications for traffic management. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 504 –509, june 2011.

[9] M. Bertozzi, A. Broggi, A. Fascioli, A. Tibaldi, R. Chapuis, and F. Chausse. Pedestrian localization and tracking system with kalman filtering. In *Intelligent Vehicles Symposium, 2004 IEEE*, pages 584–589, 2004.

[10] F. Beruscha, K. Augsburg, and D. Manstetten. Haptic warning signals at the steering wheel: A literature survey regarding lane departure warning systems. *Haptics-e: The Electronic Journal of Haptics Research*, 2011.

[11] B. Beusen, S. Broekx, T. Denys, C. Beckx, B. Degraeuwe, M. Gijsbers, K. Scheepers, L. Govaerts, R. Torfs, and L. Panis. Using on-board logging

devices to study the longer-term impact of an eco-driving course. *Transportation Research Part D: Transport and Environment*, 14:514–520, 2009.

[12] S. Bonnin, F. Kummert, and J. Schmudderich. A generic concept of a system for predicting driving behaviors. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, pages 1803–1808, 2012.

[13] M. Bou Farah, D. Mercier, E. Lefevre, and F. Delmotte. Towards a robust exchange of imperfect information in inter-vehicle ad-hoc networks using belief functions. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 436 –441, june 2011.

[14] J. Brembeck and P. Ritzer. Energy optimal control of an over actuated robotic electric vehicle using enhanced control allocation approaches. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 322 –327, june 2012.

[15] J. Buchli, F. Stulp, E. Theodorou, and S. Schaal. learning variable impedance control. 2011.

[16] C. Cai, Y. Wang, and G. Geers. Adaptive traffic signal control using vehicle-to-infrastructure communication: a technical note. In *Proceedings of the Second International Workshop on Computational Transportation Science*, IWCTS '10, pages 43–47, New York, NY, USA, 2010. ACM.

[17] V. Charissis and M. Naef. Evaluation of prototype automotive head-up display interface: Testing driver's focusing ability through a vr simulation. In *Intelligent Vehicles Symposium, 2007 IEEE*, pages 560 –565, june 2007.

[18] Yi ching Lee and John D. Lee. Visual attention in driving: The effects of cognitive load and visual disruption, 2007.

[19] I. Dagli, M. Brost, and G. Breuel. Action recognition and prediction for driver assistance systems using dynamic belief networks. In *Agent Technologies, Infrastructures, Tools, and Applications for E-Services*, Lecture Notes in Computer Science, pages 179–194. Springer Berlin / Heidelberg, 2003.

[20] I.M. de Diego, O.S. Siordia, C. Conde, and E. Cabello. Optimal experts' knowledge selection for intelligent driving risk detection systems. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 896 –901, june 2012.

[21] B. Dittes, M. Heracles, T. Michalke, R. Kastner, A. Gepperth, J. Fritsch, and C. Goerick. A hierarchical system integration approach with application to visual scene exploration for driver assistance. In *International Conference on Computer Vision Systems (ICVS)*, 2009.

[22] A. Doshi and M.M. Trivedi. Tactical driver behavior prediction and intent inference: A review. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 1892–1897, 2011.

[23] M. Eilers and C. Mobus. Learning the human longitudinal control behavior with a modular hierarchical bayesian mixture-of-behaviors model. In *Proc. of the IEEE Intelligent Vehicles Symposium*, 2011.

[24] M. Enzweiler and D.M. Gavrila. Integrated pedestrian classification and orientation estimation. pages 982–989. IEEE, 2010.

[25] A. Ess, B. Leibe, K. Schindler, and L. Van Gool. A mobile vision system for robust multi-person tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.

[26] L. Spinello et al. Multiclass multimodal detection and tracking in urban environments. *Journal of Robotics Research*, 2010.

[27] Institute for Energy European Commission, Joint Research Center. Mobile measurement of pollutant emissions and fuel consumption of road vehicles in real-world driving situations using portable emission measurement systems (pems), 2007.

[28] Institute for Energy European Commission, Joint Research Center. Analyzing on-road emissions of light-duty vehicles with portable emission measurement systems (pems), 2011.

[29] N. Fairfield and C. Urmson. Traffic light mapping and detection. In *ICRA*, pages 5421–5426, 2011.

[30] J.R. Flanagan, P. Vetter, R. S. Johansson, and D. M. Wolpert. Prediction precedes control in motor learning. *Current Biology*, 13(2):146 – 150, 2003.

[31] C. Frese and J. Beyerer. A comparison of motion planning algorithms for cooperative collision avoidance of multiple cognitive automobiles. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 1156 –1162, june 2011.

[32] M. Garcia Ortiz, B. Dittes, J. Fritsch, and A. Gepperth. Autonomous generation of internal representations for associative learning. In K. Diamantaras, W. Duch, and L.. Iliadis, editors, *Artificial Neural Networks ICANN 2010*, volume 6354 of *Lecture Notes in Computer Science*, pages 247–256. Springer Berlin Heidelberg, 2010.

[33] A. Gepperth. Co-training of context models for real-time vehicle detection. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 814 –820, june 2012.

[34] A. Gepperth, B. Dittes, and M. Garcia Ortiz Ortiz. The contribution of context information: A case study of object recognition in an intelligent car. *Neurocomput.*, 94:77–86, October 2012.

[35] A. Gepperth, M. Garcia Ortiz, and B. Heisele. Dynamic attention priors for object detection in intelligent vehicles. In *submitted to International Conference on Intelligent Robots and Systems (IROS) 2013*, 2013.

[36] A. Gerdes. Automatic maneuver recognition in the automobile: the fusion of uncertain sensor values using bayesian models. In *Proceedings of the 3rd International Workshop on Intelligent Transportation*, pages 129–133, 2006.

[37] T. Gindele, S. Brechtel, and R. Dillmann. A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments. In *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2010.

[38] E. Guizzo. Toyota's semi-autonomous car will keep you safe. `http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/toyota-semi-autonomous-lexus-car-will-keep-you-safe`, 2013.

[39] S Haykin. *Neural networks: a comprehensive foundation.* Prentice Hall, 1999.

[40] M. J. Hennings, O. Georgeon, T. Wynn, and J. F. Krems. Modelling driver behavior in order to infer the intention to change lanes. In *Proc of European Conference on Human Centered Design for Intelligent Transport Systems*, 2008.

[41] M. Heracles, U. Körner, T. Michalke, G. Sagerer, J. Fritsch, and C. Goerick. A dynamic attention system that reorients to unexpected motion in real-world traffic environments. In *International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, 2009.

[42] M. Heracles, F. Martinelli, and J. Fritsch. Vision-based behavior prediction in urban traffic environments by scene categorization. In *British Machine Vision Conference (BMVC)*, 2010.

[43] M. Himmelsbach and H. J Wuensche. Tracking and classification of arbitrary objects with bottom-up/top-down detection. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 577–582, 2012.

[44] Random House. *Random House Webster's College Dictionary.*

[45] T. Hülnhagen, I. Dengler, A. Tamke, Thao Dang, and G. Breuel. Maneuver recognition using probabilistic finite-state machines and fuzzy logic. In *Intelligent Vehicles Symposium*, pages 65 – 70, 2010.

[46] R. Kastner, F. Schneider, T. Michalke, J. Fritsch, and C. Goerick. Image-based classification of driving scenes by hierarchical principal component classification (hpcc). In *IEEE Intelligent Vehicles Symposium (IV)*, Xi'an, China, 2009. IEEE.

[47] K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura. Pedestrian recognition using high-definition lidar. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 405–410, 2011.

[48] S. Klanke, S. Vijayakumar, and S. Schaal. A library for locally weighted projection regression. *J. Mach. Learn. Res.*, 9:623–626, June 2008.

[49] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. Recent development and applications of SUMO - Simulation of Urban MObility. *International Journal On Advances in Systems and Measurements*, 5(3&4):128–138, December 2012.

[50] Stefan Krauß. *Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics.* PhD thesis, 1998.

[51] T. Kuhnl, F. Kummert, and J. Fritsch. Monocular road segmentation using slow feature analysis. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 800 –806, june 2011.

[52] S. Lefevre, C. Laugier, and J. Ibanez-Guzman. Risk assessment at road intersections: Comparing intention and expectation. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 165–171, 2012.

[53] T.R. Lim and A.T. Guntoro. Car recognition using gabor filter feature extraction. In *Circuits and Systems, 2002. APCCAS '02. 2002 Asia-Pacific Conference on*, volume 2, pages 451 – 455 vol.2, 2002.

[54] Yi Liu and Y.F. Zheng. One-against-all multi-class svm classification using reliability measures. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 2, pages 849–854 vol. 2, 2005.

[55] Xiaoliang M. A neural-fuzzy framework for modeling car-following behavior. In *Systems, Man and Cybernetics*, pages 1178 – 1183, 2006.

[56] J. C. Mccall, D. Wipf, M. M. Trivedi, and Bhaskar Rao. Lane change intent analysis using robust operators and sparse bayesian learning. In *IEEE Trans. Intell. Transp. Syst*, pages 431–440, 2005.

[57] J.C. McCall and M.M. Trivedi. Driver behavior and situation aware brake assistance for intelligent vehicles. *Proceedings of the IEEE*, 95(2):374–387, Feb 2007.

[58] F. Meier, E. Theodorou, F. Stulp, and S. Schaal. movement segmentation using a primitive library. In *ieee/rsj international conference on intelligent robots and systems (iros 2011)*, 2011.

[59] T. Michalke, J. Fritsch, and C. Goerick. Enhancing robustness of a saliency-based attention system for driver assistance. In *The 6th International Conference on Computer Vision Systems (ICVS)*, Lecture Notes in Computer Science (LNCS), Santorini, Greece, 2008. Springer Verlag Berlin Heidelberg New York.

[60] T. Michalke, R. Kastner, J. Adamy, S. Bone, F. Waibel, M. Kleinehagenbrock, J. Gayko, A. Gepperth, J. Fritsch, and C. Goerick. An attention-based system approach for scene analysis in driver assistance. *at - Automatisierungstechnik*, 56(11):575–584, 2008.

[61] T. Michalke, R. Kastner, M. Herbert, J. Fritsch, and C. Goerick. Adaptive multi-cue fusion for robust detection of unmarked inner-city streets. In *IEEE Intelligent Vehicles Symposium (IV)*, Xi'an, China, 2009. IEEE.

[62] F. Moosmann and C. Stiller. Velodyne slam. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 393 –398, june 2011.

[63] B. Morris, A. Doshi, and M. Trivedi. Lane change intent prediction for driver assistance: On-road design and evaluation. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 895 –901, june 2011.

[64] D. Muller, J. Pauli, M. Meuter, L. Ghosh, and S. Muller-Schneiders. A generic video and radar data fusion system for improved target selection. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 679 –684, june 2011.

[65] United Nations. Kyoto protocol to the united nations framework convention on climate change, 1998.

[66] United Nations. Regular budget 2012-2013, 2012.

[67] M. Nishigaki, S. Rebhan, and N. Einecke. Vision-based lateral position improvement of radar detections. In *Intelligent Transportation Systems Conference*, 2012.

[68] N. Oliver and A. Pentland. Driver behavior recognition and prediction in a smartcar. In *Proc. of SPIE Enhanced and Synthetic Vision*, 2000.

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

[70] World Health Organization. World report on road traffic injury prevention, 2008.

[71] World Health Organization. Global status report on road safety, 2009.

[72] M. Garcia Ortiz and A. Gepperth. Neural self-adaptation for large-scale system building. In *Proceedings of the First International Conference on Cognitive Neurodynamics*, 2009.

[73] N. Ovcharova, M. Fausten, and F. Gauterin. Effectiveness of forward collision warnings for different driver attention states. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 944 –949, june 2012.

[74] N. Pugeault and R. Bowden. Learning pre-attentive driving behaviour from holistic visual features. In *Proceedings of the 11th European conference on Computer vision: Part VI*, ECCV'10, pages 154–167, Berlin, Heidelberg, 2010. Springer-Verlag.

[75] A. Rauch, F. Klanner, and K. Dietmayer. Analysis of v2x communication parameters for the development of a fusion architecture for cooperative perception systems. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 685 –690, june 2011.

[76] R.J. Reed and R.J. Marks II. *Neural smithing: Supervised Learning in Feedforward Artificial Neural Networks.* MIT Press, 1999.

[77] M. Reichel, M. Botsch, R. Rauschecker, K. Siedersberger, and M. Maurer. Situation aspect modelling and classification using the scenario based random forest algorithm for convoy merging situations. In *Proc. of the IEEE Intelligent Transportation Systems Conference*, pages 360 – 366, 2010.

[78] J. Nakanishi S. Schaal, J. Peters and A. Ijspeert. Learning movement primitives. In *International Symposium on Robotics Research (ISRR2003)*, 2004.

[79] F. Saust, J.M. Wille, B. Lichte, and M. Maurer. Autonomous vehicle guidance on braunschweig's inner ring road within the stadtpilot project. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 169 –174, june 2011.

[80] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, and J. Schmidhuber. PyBrain. *Journal of Machine Learning Research*, 2010.

[81] K. Schindler, A. Ess, B. Leibe, and L. Van Gool. Automatic detection and tracking of pedestrians from a moving stereo rig. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2010.

[82] J. Schmuedderich, N. Einecke, S. Hasler, A. Gepperth, B. Bolder, R. Kastner, M. Franzius, S. Rebhan, B. Dittes, H. Wersing, J. Eggert, J. Fritsch, and C. Goerick. System approach for multi-purpose representations of traffic scene elements. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 1677 –1684, sept. 2010.

[83] Jürhen Schürmann. *Pattern Classification: A Unified View of Statistical and Neural Approaches*. Wiley-Interscience, 1996.

[84] H. Singh, J. S. Bhatia, and J. Kaur. Eye tracking based driver fatigue monitoring and warning system. In *Power Electronics (IICPE), 2010 India International Conference on*, pages 1–6, 2011.

[85] P. Steinemann, J. Klappstein, J. Dickmann, H.-J. Wunsche, and F.V. Hundelshausen. Determining the outline contour of vehicles in 3d-lidar-measurements. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 479 –484, june 2011.

[86] P. Themann and L. Eckstein. Modular approach to energy efficient driver assistance incorporating driver acceptance. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 1023 –1028, june 2012.

[87] S. Thrun. What were driving at. `http://googleblog.blogspot.fr/2010/10/what-were-driving-at.html`, 2010.

[88] C. Tran, A. Doshi, and M. Trivedi. Modeling and prediction of driver behavior by foot gesture analysis. *Computer Vision and Image Understanding*, 116(3):435–445, 2012.

[89] G. Underwood, P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers, 2003.

[90] European Union. Questions and answers on interinstitutional agreement on budgetary discipline and sound financial management 2007-2013, 2006.

[91] B. Vanholme, D. Gruyer, S. Glaser, and S. Mammar. A legal safety concept for highly automated driving on highways. In *Intelligent Vehicles Symposium*, pages 563–570, 2011.

[92] S. Vijayakumar, A. D'souza, and S. Schaal. Incremental online learning in high dimensions. *Neural Comput.*, 17(12):2602–2634, December 2005.

[93] S. Worrall, D. Orchansky, F. Masson, and E. Nebot. Improving vehicle safety using context based detection of risk. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 379 –385, sept. 2010.

[94] G. Xu, W. Li, K. Xu, and Z. Song. An intelligent regenerative braking strategy forelectric vehicles. *Energies*, 4(9):1461–1477, 2011.