Institutionen för systemteknik
Department of Electrical Engineering

Examensarbete

Algorithm Design for Driver Attention Monitoring

Examensarbete utfört i Fordonssystem
vid Tekniska högskolan vid Linköpings universitet
av

Olle Sjöblom

LiTH-ISY-EX--15/4883--SE

Linköping 2015

Linköpings universitet
TEKNISKA HÖGSKOLAN

Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

Linköpings tekniska högskola
Linköpings universitet
581 83 Linköping

Algorithm Design for Driver Attention Monitoring

Examensarbete utfört i Fordonssystem
vid Tekniska högskolan vid Linköpings universitet
av

Olle Sjöblom

LiTH-ISY-EX--15/4883--SE

Handledare: **Peter Nyberg**
ISY, Linköpings Universitet
Christer Ahlström
VTI

Examinator: **Erik Frisk**
ISY, Linköpings Universitet

Linköping, 24 augusti 2015

Avdelning, Institution
Division, Department

Division of Vehicular Systems
Department of Electrical Engineering
SE-581 83 Linköping

Datum
Date

2015-08-24

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English
☐ ________

Rapporttyp
Report category

☐ Licentiatavhandling
☒ Examensarbete
☐ C-uppsats
☐ D-uppsats
☐ Övrig rapport
☐ ________

ISBN
—

ISRN
LiTH-ISY-EX--15/4883--SE

Serietitel och serienummer
Title of series, numbering

ISSN
—

URL för elektronisk version

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-120762

Titel
Title

Algoritmutveckling för detektion av förardistraktion
Algorithm Design for Driver Attention Monitoring

Författare
Author

Olle Sjöblom

Sammanfattning
Abstract

The concept *driver distraction* is diffuse and no clear definition exists, which causes troubles when it comes to driver attention monitoring. This thesis takes an approach where eye-tracking data from experienced drivers along with radar data has been used and analysed in an attempt to set up adaptive rules of how and how often the driver needs to attend to different objects in its surroundings, which circumvents the issue of not having a clear definition of *driver distraction*. In order to do this, a target tracking algorithm has been implemented that refines the output from the radar, subsequently used together with the eye-tracking data to in a statistical manner, in the long term, try to answer the question *for how long is the driver allowed to look away in different driving scenarios?* The thesis presents a proof of concept of this approach, and the results look promising.

Nyckelord
Keywords

target tracking, radar, Kalman filter, eye-tracking, driver attention, algorithm

Abstract

The concept *driver distraction* is diffuse and no clear definition exists, which causes troubles when it comes to driver attention monitoring. This thesis takes an approach where eye-tracking data from experienced drivers along with radar data has been used and analysed in an attempt to set up adaptive rules of how and how often the driver needs to attend to different objects in its surroundings, which circumvents the issue of not having a clear definition of *driver distraction*. In order to do this, a target tracking algorithm has been implemented that refines the output from the radar, subsequently used together with the eye-tracking data to in a statistical manner, in the long term, try to answer the question *for how long is the driver allowed to look away in different driving scenarios?* The thesis presents a proof of concept of this approach, and the results look promising.

Acknowledgments

I would like to thank Christer Ahlström and Katja Kircher for the possibility to perform this thesis and for valuable discussions. Thanks to Erik Frisk for examining the work and for great input to questions and problems of mine, and to Peter Nyberg for fruitful comments on the report and for very quickly responding to e-mails.

Finally, thank you Karin for your love and support and for proofreading this report (unintentional rhyme).

Linköping, August 2015
Olle Sjöblom

Contents

1 Introduction

"Can more road signs be put up without impairing the driving ability among the road users?" was the coarse question that funded a project at the Swedish National Road and Transport Research Institute, VTI, on how to make quantitative, unbiased assessments of driver distraction. Much effort has been put into trying to define the concept distraction, but yet today there exists no general applicable definition, and there are almost as many definitions as there are researchers. Many distraction detection algorithms have been proposed, but due to the fact that the notion itself is not well-defined, inconsistencies and validity issues are commonplace among them.

Most of all, the so called hindsight problem needs to be solved, meaning that the outcome of a certain situation should not need to be known in order to determine whether or not the driver was distracted. An accident might occur despite driver attentiveness, and driver distraction does not necessarily imply an accident.

Researchers at VTI have proposed a new approach to this issue - instead of focusing directly on driver distraction, one should start with a clear definition of attention. In order to eliminate the hindsight problem, there is a need for setting up rules on when and how often a driver is required to sample certain information in different traffic situations, and as soon as these rules are broken, the driver should considered distracted. However, if the requirements are fulfilled, the driver should be allowed to use spare capacity to attend to secondary tasks such as looking at navigational device or tuning the radio, without being classified as distracted. They call this generic idea *Minimum Required Attention*, MRA, presented in [Kircher and Ahlström, 2014].

Currently there is an algorithm developed at VTI called AttenD, which uses eye-tracking data for driver attention monitoring. A *Field Relevant for Driving*, FRD, is defined, and when the driver is looking outside FRD, a buffer counts down. When the driver is looking inside, the buffer counts up. When the buffer runs empty, the driver is considered distracted. The problem is that AttenD is completely unaware of the current traffic situation and the exact same attention requirements apply when, for instance, driving 5 km/h on an empty parking lot, as when driving 80 km/h on a busy highway, which is of course unreasonable.

1.1 Objectives

The overall problem that is being studied and (partly) solved in this thesis is the issue of how to make use of acquired data in order to incorporate situational awareness into AttenD. After a start-up phase containing a brief investigation of the available data, the thesis work was divided into the following tasks:

- Tracking improvement. The data received from the radar are not raw measurements, but output from an internal filter. This means that the available data from the radar consists of lateral and longitudinal position of each object within the field of view of the radar, as well as relative longitudinal velocity, accessed through making use of the Doppler effect. As it turned out, the performance of the internal tracking is debatable. It uses up to 32 channels to store information on up to 32 objects - that is, one channel per object visible to the radar at any given time. The rule determining which object to be stored in which channel however seems obscure, and each track is not stored in the same channel throughout its life time. Considering the overall goal with this project - to set up adaptive rules of how often the driver needs to attend to certain objects in the environment - this structure of the data is insufficient. The objects need to be tracked individually without the channel switching phenomenon. Only then is it possible to keep track of whether or not the driver has attended to an object according to the rules (the rules that this project aims to set up, that is). Figure 1.1 illustrates the channel switching issue, and the goal with this task is to develop an algorithm that solves the channel switching phenomenon.

- Conduct eye-tracking analysis. The eye-tracking data is delivered as an angle at each time step, giving the gaze direction of the driver, and an interesting question is how well that direction can be matched with the situational awareness data provided by the radar.

- Propose a novel method for setting up attentional demands in certain traffic situations. The outcome of the previous tasks will obviously set the foundation for the possible complexity of the model.

1.2 VTI and MFT

VTI is an independent and internationally prominent research institute within the transport sector, with offices located in several cities in Sweden. Their main

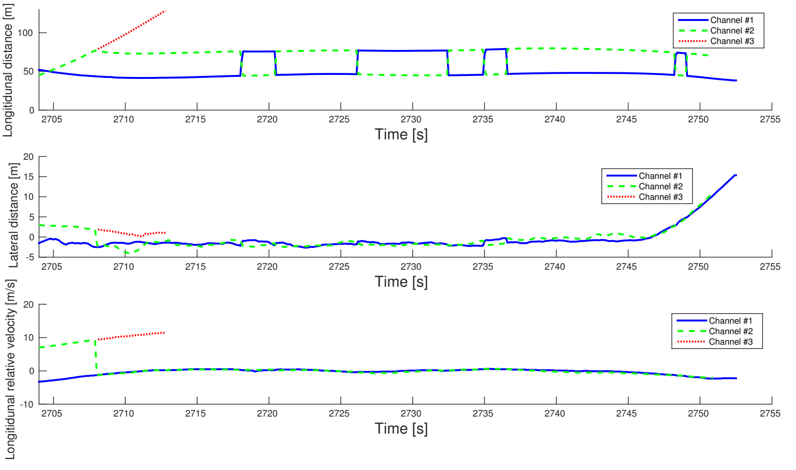


Figure 1.1: *Illustration of the target association issue. The plot of the longitudinal distance (uppermost) most clearly demonstrates the problem of how the signals make numerous jumps between the the targets.*

objective is to conduct research and development of infrastructure, traffic and transport. Areas included among others are road and rail engineering, maintenance, vehicle technology, road safety, traffic analysis, environment and transport economy.

This thesis was carried out at the *Human, Vehicle, Transport System Interaction* unit, MFT, which aims to conduct research concerning road users and their interaction with the transport system with main focus on questions concerning road user's limitations and possibilities in their interaction with other road users.

1.3 Thesis Outline

Chapter 2 briefly describes the system that was used for data acquisition, and presents the available data. It also provides background to different methods of assessment of driver capacity. **Chapter 3** presents the theoretical framework for target tracking including filter theory and a representation of the Kalman filter, while **Chapter 4** describes how the tracking concept was applied to this specific problem and presents the results. The possibilities to improve the performance of the target tracking was investigated in the final phase of the thesis work - **Chapter 5** presents the conclusion on this. In **Chapter 6** it is being theorised around how the data and the results so far can be used to define situational demands in different traffic situations and derive, in some sense, a proof of concept. Finally, **Chapter 7** states concluding remarks and proposes future work.

2

Data Acquisition

The vast majority of the data used in this thesis was collected prior to its start (towards the end of the project more data was collected using a camera sensor, which was not used during the first effort of collecting data), but the course of action will still be described. This chapter will start off with a quick review of the project background in order to provide context to the description of different methods for assessment of driver capacity, which is provided afterwards. A brief description of the system setup and finally a detailed description of the acquired data follows.

2.1 Background

As already been mentioned, the project that this thesis is a part of centralises around a new approach for monitoring driver attention: the outcome of a situation should not have to be used in order to determine whether the driver was distracted or not. Instead, *attentional demands* for any given situation should be defined. If the driver's *spare capacity* also could be determined at any given time, then these two could be compared and a decision on attentiveness could then be stated. The idea is that as long as the attentional demands are fulfilled, the driver can use any spare capacity for other tasks, such as looking at navigational device or interacting with passengers, and can not be blamed for being distracted, even if an accident should occur. On the other hand, as soon as the attentional demands are not fulfilled the driver should be considered distracted.

The difficulties of course lie in identifying the attentional demands of any given situation. They can be either model based or data driven. A few comments on that:

- A model based approach would mean utilising physics and mathematics in order to build a model representing how the driver gains and loses information depending on if he or she attends to the road or not. Senders et al. [1967] give this a try by considering a straight road carrying a uniform distribution of information while the host vehicle is pervaded at constant velocity. They manage to obtain a parametrised model valid at that specific situation. The parameters are then determined via test runs, and it turned out that the parameters vary individually, suggesting that this approach is not very suitable for the objective of setting up situational dependent attentional demands. Not much more research has been conducted on this, which reinforces that suggestion. Additionally, traffic rules are not formulated in a way that would make them easily translated into mathematical equations.

- The other approach would mean acquiring different data and utilise that in some way to define the attentional demands. It can be assumed that experienced drivers after years and years of driving have established internal versions of the traffic rules (with "traffic rules" we here mean partly the actual written rules that state for instance what car that has priority at a highway entrance, but also the more vague rules that stem from common sense and mutual respect), and hence can be used as benchmarks when it comes to obtaining successful attention allocation patterns in different situations. This means that an experienced driver should "feel" that he or she should pay more frequent attention the less predictable the situation is. Driving just behind a heavy truck being overtaken by another heavy truck why self being overtaken by another vehicle requires a completely different attention allocation pattern than when driving solo on the road, and it is these "rules" that experienced drivers are assumed to have internalised.

2.1.1 Driver Capacity Assessment

Well before this master thesis was announced it was decided to use a data driven approach. The data acquisition will be described later in this chapter but first, a few well-known methods for assessment of driver capacity will be described below.

Eye-Tracking

Several cameras mounted on the dashboard record a video of the movement of the driver's head and eyes, from which the gaze direction can be estimated in real-time. The concept is well-established and commercial alternatives are available, but they are relatively expensive, and it requires frame-by-frame evaluation of the surrounding. In order to catch saccades the sample frequency must be high, which makes the evaluation possibly tedious. Furthermore, only the foveal vision is considered, and the peripheral impressions are left out. On the opposite, all targets that are glanced at are not necessarily attended to - so called *vacant staring* or *looking without noticing* might occur. Also, it is hard to in real-time determine whether or not the glanced targets is relevant for safety. This is a problem since the project aims to set up rules that *do not* require hindsight knowledge.

Self-Paced Visual Occlusion

Whenever the driver feels that the current traffic situation is predictable, he or she activates occlusion goggles, preventing him or her from seeing the road, which in real-time gives access to how the driver assesses its spare capacity. Driving fast with occluded sight may strike the driver as uncomfortable or scary, consequently leading to conservative handling.

Verbal Protocol

Whilst driving, the driver reports what he or she pays attention to, giving access to not only gaze allocation but also to attention allocation. Due to the issue of vacant staring described previously, this is a good complement to eye-tracking. This does however require a certain effort from the driver, and people may or may not be proficient enough at simultaneously driving and reporting as the reporting might interfere with other cognitive tasks. Also, fast sequences of events will not be consummately retold.

Expert Judgement

An experienced driver identifies necessary gaze targets based on a film or a situation description, that may provide information that motivates certain behaviour patterns. It is however obviously not a real-time method, and there is no guarantee that the expert judgement agrees with the behaviour intended by the driver.

The methods are summarised in Table 2.1.

2.2 System Setup

The test vehicle was, in addition to conventional sensors measuring e.g. wheel speed, yaw rate etc. (see Table 2.2 for a complete list of available data), equipped with a radar sensor mounted in front of the vehicle, and a number of cameras shooting out through the front and rear windshield. For eye-tracking, the commercial system SmartEye Pro 6.1 was used, with hardware consisting of 5 cameras mounted on the dashboard along with a PC, IR lights, external processors, and a chessboard for camera calibration.

2.2.1 CAN-Bus

CAN, Controller Area Network, is a bus standard devised to allow for serial and asynchronous communication between micro controllers without the use of a host computer. Its main application is the automotive industry where it is used

Table 2.1: Methods for driver distraction assessment.

Method	Advantages	Disadvantages
eye-tracking	• real-time	• no explicit consideration of peripheral vision
	• well established	• requires hindsight judgement whether glance target is traffic relevant or not
	• unintrusive to moderately intrusive	• expensive apparatus and cumbersome evaluation
		• not all targets that are fixated are necessarily attended to
self-paced visual occlusion	• real-time	• somewhat cumbersome apparatus
	• direct assessment of spare capacity	• only allows assessment of spare capacity, not of glance targets
		• may lead to conservative handling
verbal protocol	• access to attention allocation, not only gaze allocation	• might interfere with other cognitive tasks
	• simple and cheap	• fast sequences of events will not be reported completely
		• people might be more or less proficient at verbal reporting
expert judgement	• may provide background information to motivate certain behaviours	• not real-time
	• simple and cheap	• rather remote from actual behaviour

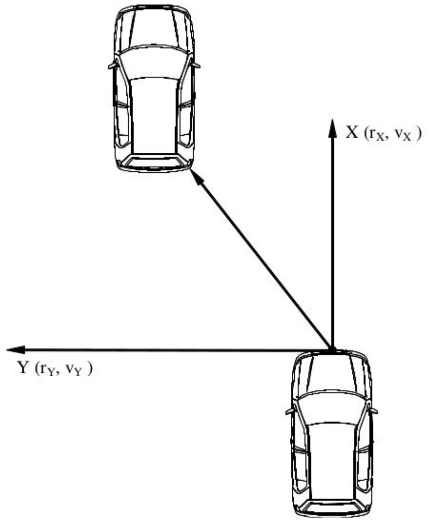


Figure 2.1: *Local coordinate system used by the radar, c.f. [Smartmicro, 2012].*

by the control units in vehicles that need to exchange information and it is well fitted for e.g. logging sensor data.

2.2.2 Radar

The radar sensor was a Smartmicro UMMR Automotive Type 29, which was via a CAN-bus connected to a power supply and a visualisation module. The CAN-bus gives the radar module access to the vehicle dynamic data in Table 2.2. The radar uses a Cartesian coordinate system as in Figure 2.1.

2.3 Experimental Design

A total of 12 test drivers, whereof the majority were driving instructors and thus can be considered very experienced drivers, were involved during the data acquisition. They each drove a path along the highway E4 outside Linköping, more precisely the path between the east and west exit. The whole driven path recorded by the GPS, starting and ending at the parking lot at VTI, can be seen in Figure 2.2, where the highway part, which is where focus is in this project, has been highlighted.

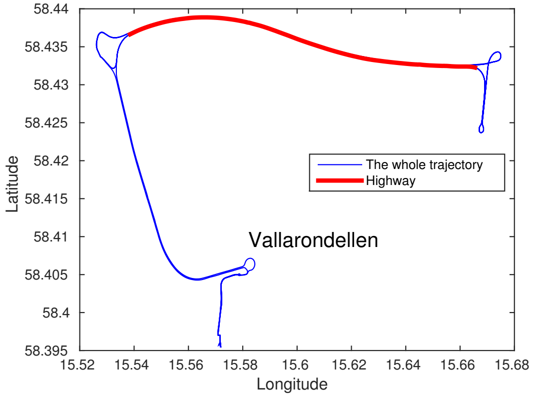


Figure 2.2: *The trajectory driven, with the highway part highlighted.*

Most drivers did four runs, where each run consisted of three laps:

- One run during daylight.
- One run during night or dusk.
- One run during daylight where occlusion goggles were used.
- One run during daylight with verbal protocol.

Some drivers only did the daylight run and the occlusion goggles run. During each run a large amount of data was collected, which will be described in more detail in the following section.

2.4 Available Data

The nature of the collected data is diverse, although the vast majority is available as standard numerical signals. Vehicle data is delivered from the CAN bus. Table 2.2 states the majority of the data variables along with a description where considered necessary.

Table 2.2: *The available data signals.*

Variable	Description	Unit
Vehicle data		
Brake	Are brakes applied?	1/0
Steering wheel angle	-	degrees
Engine speed	-	rpm
Heading	-	degrees
Lateral acceleration	-	m/s^2
Longitudinal acceleration	-	m/s^2
Acceleration pedal position	-	degrees
Clutch pedal position	-	degrees
Velocity	Host vehicle velocity	m/s
Yaw rate	-	degrees/s
Gear	Current gear	$\{0 \cdots 6\}$
Cruise	Is cruise control active?	1/0
AC	Is air conditioner active?	1/0
Outdoor temperature	-	°C
Indoor temperature	-	°C
Radar data		
Range X	Longitudinal distance to object	m
Range Y	Lateral distance to object	m
Velocity X	Longitudinal velocity of object	m/s
Eye-tracking data		
Gaze X	Horizontal gaze angle	rad
Gaze Y	Vertical gaze angle	rad
Other		
Occlusion goggles	Occlusion goggles active?	1/0
Latitude	From GPS	°N
Longitude	From GPS	°E

Figure 2.3: *Front view with overlay gaze allocation (blue dot).*

Additional data that was acquired during the test runs:

- Verbal protocols and expert judgements (see Table 2.1).
- GoPro movies forward and backward.
- A front view movie where the gaze allocation delivered by the eye-tracker has been overloaded. In other words, it shows the point where the driver looks compared to ground truth. Figure 2.3 illustrates this. The image is in black and white and of a lower quality than the GoPro movies that do not have overloaded gaze allocation.

3

Target Tracking Theory

The target tracking objective is to make use of sensors to collect data of an environment containing one or several possible targets, and partition the data into sets of observations, one set for each possible target. The channel switching phenomenon inherent with the radar data hence is a suitable problem to apply target tracking theory to. This chapter will first reproduce the problem and then provide a theoretical framework of target tracking based on different approaches in literature.

3.1 Problem Formulation Revisited

Although a tracking module is integrated with the radar, that refines the raw measurements and turns them into interpretable signals, the nature of the data is not of desired character. As of now, a maximum of three objects are tracked. At any given time, information of each object that is visible to the radar is stored in one of three channels according to the nominal rule that channel #1 shall contain information on the object within the shortest range to the host vehicle, although ocular investigation has proved that this is not always the case. That is, there is no strict functioning rule of how the targets are partitioned, and this causes the signals to appear as in Figure 1.1: the tracked objects are not persistent upon which radar channel they are stored in. As explained in Chapter 1, the overall goal with these project requires that each object can be tracked individually, so records can be kept regarding whether or not an object has been attended to according to the rules that this project aims to set up. Hence, there is a need of improving the tracking with a formal target tracking algorithm to partition the data into one signal per target present, and as mentioned in the introduction of this chapter target tracking is a method that approaches this kind of problem. It is a non-trivial task, due to e.g.

- multiple targets present
- false alarms
- multiple measurements close to target
- detection uncertainties
- hardware failure.

Techniques to tackle these problems will be presented in the following sections.

3.2 Tracking Approaches

Target tracking is a wide concept, and plenty of different approaches are described in literature; [Blackman and Popoli, 1999] is a rigorous book on the topic, and another one is [Bar-Shalom and Fortmann, 1988]. The general patterns however persist:

- The basic idea of tracking is to predict future states of the present targets and associate them with available observations. It therefore requires a filter for state prediction. Section 3.3 will cover this topic.
- Given a state prediction, it must be determined which measurements that can be considered possible candidates to be associated with that particular state. The theory behind this is called *gating* and is described in Section 3.4.
- When a gate has been set up, it must be decided which of the candidates within the gate that actually should be associated to the state prediction. Section 3.5 contains the theory behind two popular approaches for this.
- If there are observations that could not be associated to any current target, it is possible that it means that a new target has appeared. As false alarm also is a possible explanation, it requires a rule that separates false alarms from new tracks. On opposite, there is also a need for a rule to determine when a track should be deleted. Methods for track maintenance are described in Section 3.6.

3.3 Filter Theory

A filter for state prediction is required in a target tracker, and the type being used in this particular problem is described below.

3.3.1 State-Space Model

A generic linear state-space model in continuous time, with random noise inputs, is

$$
\begin{aligned}
\dot{x}(t) &= Ax(t) + w(t) \\
y(t) &= Cx(t) + e(t),
\end{aligned}
\tag{3.1}
$$

where A is the dynamic model determining how the states in the state vector x propagate over time, and C is the measurement function describing how the measurements y relate to the states. The process noise $w(t)$ and measurement noise $e(t)$ are assumed to be uncorrelated zero mean Gaussian noise with covariance matrices $\mathrm{Cov}(w(t)) = Q$ and $\mathrm{Cov}(e(t)) = R$.

3.3.2 Discretisation

A discretised form of the state-space model,

$$
\begin{aligned}
x_{k+1} &= Fx_k + Gw_k \\
y_k &= Hx_k + e_k,
\end{aligned}
\tag{3.2}
$$

can be acquired accordingly using zero-order hold, see [Glad and Ljung, 2004]:

$$F = e^{AT} \tag{3.3}$$

$$G = \int_0^T e^{A\tau} d\tau \tag{3.4}$$

$$H = C. \tag{3.5}$$

3.3.3 Kalman Filter

Given a linear system as in (3.2), the best linear unbiased filter is given by the Kalman filter, presented in e.g. [Gustafsson, 2012]:

1. **Measurement update:**

$$
\begin{aligned}
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}(y_k - H\hat{x}_{k|k-1}) \\
P_{k|k} &= P_{k|k-1} - P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}HP_{k|k-1}.
\end{aligned}
$$

2. **Time update:**

$$
\begin{aligned}
\hat{x}_{k+1|k} &= F\hat{x}_{k|k} \\
P_{k+1|k} &= FP_{k|k}F^T + GQG^T,
\end{aligned}
$$

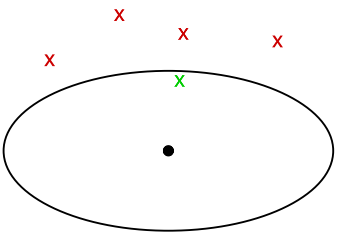


Figure 3.1: A target along with an ellipsoidal gate and five available measurements at a given scan. The measurement within the gate is accepted and the ones outside the gate are neglected.

where P is the state covariance. Time index k for time-invariant matrices are omitted for convenience.

By introducing the innovation ε_k, the innovation covariance S_k and the Kalman gain K_k according to

$$\varepsilon_k = y_k - H\hat{x}_{k|k-1} = y_k - \hat{y}_k \tag{3.6}$$

$$S_k = HP_{k|k-1}H^T + R, \tag{3.7}$$

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}, \tag{3.8}$$

the measurement update can be written more concisely:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k\varepsilon_k, \\ P_{k|k} &= P_{k|k-1} - K_k S_k K_k^T. \end{aligned} \tag{3.9}$$

The filter is assumed to have been initiated with

$$\begin{aligned} \hat{x}_{1|0} &= \mathbf{E}(x_0) \\ P_{1|0} &= \mathrm{Cov}(x_0). \end{aligned} \tag{3.10}$$

3.4 Gating

Gating is the issue of determining whether or not a measurement should be considered valid for a certain target. Each tracked target is given a region that defines a subspace of the measurement space where measurements should be accepted, referred to as the target's gate. Figure 3.1 illustrates the concept where five measurements are available for a target, among which only one is inside its gate and therefore is the only one considered valid.

The gate is formed from the predicted state of the target, and it is simply only a matter of forming a distance measure between the measurement and the predicted state.

The most frequently used gate in literature is an ellipsoidal one. Let the normalised statistical distance of an observation-to-track assignment be

$$d^2 = \varepsilon_k^T S_k^{-1} \varepsilon_k. \tag{3.11}$$

Constructing the gate then becomes an issue of finding the threshold γ_G such that

$$P(d^2 < \gamma_G) = \alpha_G, \tag{3.12}$$

where α_G is the desired confidence level of the test. Under assumptions of correctly chosen motion and measurement models, it holds that

$$\varepsilon_k \sim \mathcal{N}(0_{n_y}, S_k). \tag{3.13}$$

If S_k can be Cholesky decomposed, that is, if $S_k = U_k U_k^T$ (see [Bar-Shalom and Li, 1993] for details), then

$$d^2 = \varepsilon_k^T S_k^{-1} \varepsilon_k = \varepsilon_k^T U_k^{-T} U_k^{-1} \varepsilon_k = \|U_k^{-1} \varepsilon_k\|_2^2. \tag{3.14}$$

Let $U_k^{-1} \varepsilon_k \triangleq u_k$. It holds that

$$\varepsilon_k \sim \mathcal{N}(0_{n_y}, S_k) \Rightarrow U_k^{-1} \varepsilon_k \sim \mathcal{N}(0_{n_y}, \overbrace{(U_k U_k^T)^{-1}}^{=S_k^{-1}} S_k), \tag{3.15}$$

and thus,

$$u_k \sim \mathcal{N}(0_{n_y}, I_{n_y}). \tag{3.16}$$

Since a sum of squares of n independent identically distributed $\mathcal{N}(0, 1)$-variables is χ^2-distributed with n degrees of freedom, it can be concluded that

$$d^2 = \|u_k\|_2^2 \sim \chi^2_{n_y}, \tag{3.17}$$

whereupon the gate threshold γ_G can be calculated using the cumulative distribution function of the χ^2 distribution. Measurements that should be accepted is then the ones that lie within the elliptic disc $d^2 \leq \gamma_G$.

3.5 Data Association

A question that is naturally raised when implementing a target tracking algorithm is that of how to associate measurements with tracks. Figure 3.2 illustrates a scenario where a conflict has occurred. Each target has two measurements within its gate, which of course means that there are two detections for each object that both with high probability could be associated with it. Furthermore, each measurement lies within more than one gate. Different approaches to address this problem are presented below.

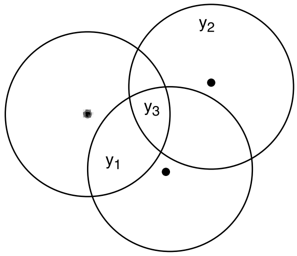


Figure 3.2: *Illustration of an association conflict. The dots represent the predicted state of three different objects whereas the circles are their corresponding gates. y_1, y_2 and y_3 are three available measurements at the given scan.*

3.5.1 Nearest Neighbour Association

One approach, and possibly the simplest one, is the nearest-neighbour filter (see [Aziz, 2013]). The nearest-neighbour measurement $\tilde{y}_k^i$ for target i at scan k is simply the measurement that lies closest to the predicted state, or mathematically,

$$\tilde{y}_k^i = \arg\min_j (d^2)^{ij}, \qquad j = 1, 2, \cdots, n_v, \tag{3.18}$$

where n_v is the number of measurements within the gate corresponding to target i, and $(d^2)^{ij}$ is the distance between track i and measurement j. Using the measure defined in (3.11), it becomes

$$(d^2)^{ij} = (\varepsilon_{k+1|k}^{ij})^T (S_{k+1|k}^i)^{-1} \varepsilon_{k+1|k}^{ij}, \tag{3.19}$$

where ε^{ij} is *the innovation due to measurement j and target i*.

An obvious drawback is that there is a probability that the nearest neighbour is not the correct measurement, meaning that this filter might use incorrect measurements believing they are true.

3.5.2 Probabilistic Data Association

In contrast to the hard-decision based nearest-neighbour filter, a soft-decision approach is the probabilistic data association filter. Using N measurements within the gate of track i, there are $N + 1$ hypotheses that can be formulated for that track: $\mathcal{H}_1, \cdots, \mathcal{H}_j, \cdots \mathcal{H}_N$ stating that measurement j is the valid one, and $\mathcal{H}_0$ stating that no observation within the gate is valid. The probability p_{ij} that observation j belongs to track i is calculated as

$$p_{ij} = \begin{cases} \frac{b}{b+\sum_{l=1}^{N} a_{il}}, & j = 0 \text{ (no valid observation)} \\ \\ \frac{a_{ij}}{b+\sum_{l=1}^{N} a_{il}}, & 1 \leq j \leq N, \end{cases} \tag{3.20}$$

where

$$b = (1 - P_D \alpha_G)(\beta_{NT} + \beta_{FA}) 2\pi \sqrt{|S^i|} \tag{3.21}$$

$$a_{ij} = P_D e^{-(d^2)^{ij}/2}. \tag{3.22}$$

As previously, α_G is the confidence level of the gate test and S^i the innovation covariance of track i. For a full description of the probabilistic data association filter, see [Bar-Shalom et al., 2009].

3.6 Track Maintenance

The big initialisation issue is the decision on whether a new detection should be accepted or not, as it could be a false alarm. Therefore, there is a demand of having some sort of rule for making this decision. There are several well-known approaches described in the literature, and a few are presented here.

3.6.1 M/N-Logic

The M/N-logic approach (see [Bar-Shalom et al., 1989]) is intuitive and straightforward and goes as following:

1. Following a first detection, that is, a detection that has not been associated to any existing track (more on data association in Section 3.5), a gate is set up. If a detection in the next sample is made within the gate, go to step 2. If not, delete the initiator.

2. For each of the N following sample times, set up a new gate based on the assumed motion model of the tracked object. A detection in the corresponding gate has to be made in M samples in order for the sequence of detections to be accepted.

An illustration of the method is shown in Figure 3.3, and an explanation to the figure is provided in Table 3.1. Concerning track deletion, there are several possibilities. The simplest one is deleting a track when no measurement has been available for a certain amount of consecutive samples, but one could also implement an M/N-logic in the exact same way as for track initiation.

Table 3.1: *How to interpret Figure 3.3.*

Box parameter	Explanation
Square	Initiator still tentative
Circle	Confirmed initiator
Rhombus	Deleted initiator
+	Detection within gate
-	No detection within gate

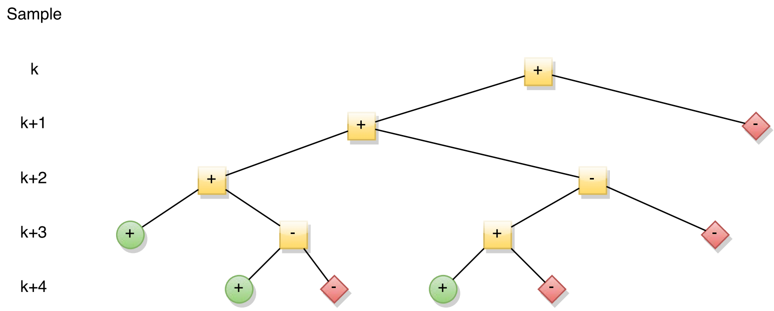


Figure 3.3: *Illustration of the M/N-logic method for track initialisation with $M = 2$ and $N = 3$. Two consecutive observations within the gate must first be made, whereat two out of the three (or generally M out of N) following observations must also lie within the gate at each time.*

3.6.2 Score Based Approach

Another method for track initiation is a score-based approach, for instance the sequential probability ratio test. Consider the hypotheses

$$\begin{aligned} &\mathcal{H}_0 : \text{Measurements originate from false alarms} \\ &\mathcal{H}_1 : \text{Measurements originate from a true target.} \end{aligned} \tag{3.23}$$

Using Bayes' rule, the *likelihood ratio* LR is defined as

$$LR = \frac{p(\mathcal{H}_1|y_{0:k})}{p(\mathcal{H}_0|y_{0:k})} = \frac{p(y_{0:k}|\mathcal{H}_1)P_0(\mathcal{H}_1)}{p(y_{0:k}|\mathcal{H}_0)P_0(\mathcal{H}_0)}, \tag{3.24}$$

where

- $y_{0:k}$ is a sequence of measurements up to time k
- $p(y_{0:k}|\mathcal{H}_i)$ is the probability density function evaluated with the received data under the assumptions of $\mathcal{H}_i$ being correct
- $P_0(\mathcal{H}_i)$ is the a priori probability of $\mathcal{H}_i$.

The *score* $\mathcal{L}$ is obtained from taking the natural logarithm of the likelihood ratio. A new score is calculated recursively at each sample, and by forming the thresholds

$$T_1 = \log \frac{P(\text{rejecting a true track})}{P(\text{rejecting a false track})}$$

and

$$T_2 = \log \frac{P(\text{accepting a true track})}{P(\text{accepting a false track})},$$

the track maintenance test is as following:

$$\begin{aligned} \mathcal{L} \leq T_1 &: \text{Accept } \mathcal{H}_0. \\ \mathcal{L} \geq T_2 &: \text{Accept } \mathcal{H}_1. \\ T_1 < \mathcal{L} < T_2 &: \text{continue monitoring.} \end{aligned} \tag{3.25}$$

The details regarding the derivations can be found in [Blackman and Popoli, 1999].

A drawback with this soft type of approach is that a target that has been tracked for a long time will also take long time to be deleted, since the score value becomes biased towards good measurements. Blackman and Popoli [1999] suggest a solution to this issue that includes using the deviation from the maximum value of the score in the thresholding.

4

Radar Data Processing

The idea behind Chapter 3 was to provide a general theoretical framework for the concept of target tracking, and different established approaches were presented. This chapter will start off with a quick data analysis, as it early turned out that the lateral values provided by the radar were inherent with some unrealistic properties. After that, motivations to chosen methods among those described in Section 3 will be provided, followed by results illustrating the performance of the tracker in a few scenarios. The chapter will finish off with a discussion about limitations.

4.1 Data Analysis

A first survey of the radar data quickly reveals some potential issues. There are recurring segments where the lateral distance appears to be larger than physically possible. By plotting the azimuth angle to the objects for all available data points, Figure 4.1 is obtained. The manufacturer (Smartmicro [2012]) claims that the radar has a field of view reaching from -18° to 18° (with the angle being measure anti-clockwise against the longitudinal axis), which is also the interval with the highest density of objects according to Figure 4.1. Without the knowledge of the nominal value, the figure would suggest a field of view reaching from approximately -40° to 40°.

Correlation Analysis

In order to attempt to acquire a better perception of the phenomenon, a correlation analysis was performed. The correlation ρ between two signals X and Y is defined as

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbf{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \tag{4.1}$$

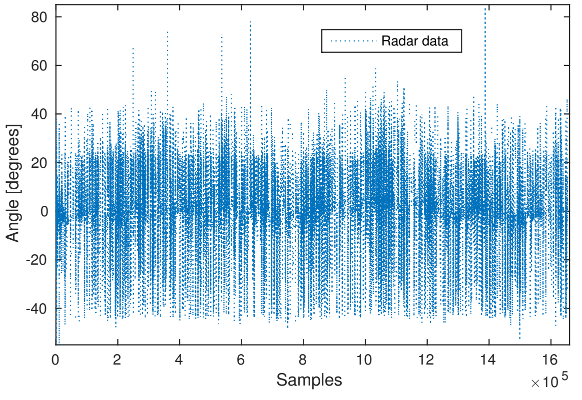


Figure 4.1: *Calculated azimuth angle to objects based on the measurements provided by the radar. The nominal field of view for the radar is [-18,18] degrees. It is clear that the highest density of measurements is within that interval, but there are allegedly suspiciously many objects outside the nominal interval.*

where σ denotes standard deviation, μ expected value and **E** is the expected value operator. It holds that $\rho \in$ [-1,1], where $\rho = 1$ would mean a perfect increasing linear relationship, and $\rho = -1$ a perfect decreasing linear relationship. Also, $\rho = 0$ would mean that there is no direct relationship between the signals.
Samples holding remarkable lateral distances, that is, an absolute value of the lateral distance larger than some threshold T, were extracted, and correlated with corresponding values in numerous other signals. The results are shown in Table 4.1. For assessment assistance, the number of data points that the correlation was based on is also presented.

Discussion

A few comments on the results presented in Table 4.1:

- Firstly, lateral distances less than 5 metres should be considered normal due to road curvature and knowledge of the lane width; a typical width of a Swedish highway lane is 3.5 metres (Trafikverket [2004]).
- There appears to be practically no connection between the odd lateral measurements and the variables brake, lateral acceleration and longitudinal acceleration.
- A correlation with absolute longitudinal velocity of the host vehicle seems to exist. However, it changes from being a negative correlation for $T = 5$, to being practically non-existing at $T = 10$ to be a positive correlation at $T = 15$ and $T = 20$, so no conclusions can probably be drawn from there.

***Table 4.1:** Parts of the signal for lateral distance to objects that are claimed to have an absolute value greater than T were correlated with other selected signals to see if there was any relationship. The results for different values of T are shown here. P denotes the number of data points the correlation was based on. For reference, the total number of available data points is 500072.*

	$T = 5$ $P = 133641$	$T = 10$ $P = 23353$	$T = 15$ $P = 6303$	$T = 20$ $P = 1378$
Longitudinal distance to object	0.13	-0.22	-0.36	-0.41
Longitudinal velocity relative to object	-0.23	-0.47	-0.58	-0.73
Longitudinal velocity	-0.18	0.05	0.20	0.18
Longitudinal acceleration	-0.01	-0.01	-0.07	-0.10
Acceleration pedal position	-0.25	-0.32	-0.32	-0.24
Lateral acceleration	0.03	0.07	0.12	0.20
Yaw rate	0.15	0.30	0.38	0.41
Steering wheel angle	-0.13	-0.26	-0.34	-0.37
Outdoor temperature	0.25	0.35	0.34	0.37
Brake	0.05	0.04	0.07	0.15
Time	0.16	0.22	0.13	0.01

- For yaw rate and steering wheel angle as well as longitudinal distance and relative velocity to object, a higher correlation for a larger T value can be noticed. This suggests that these signals may have something to do with "ghost" lateral distances. On the other hand, large lateral distances rise naturally in situations when exiting the highway while the front vehicle remains on it so it is not very surprising that there is a correlation here.

- Outdoor temperature does indeed hold a significant correlation with the odd lateral distances, but since it seems unaffected by increasing threshold it is probably nothing there. Also, it is not really reasonable that the radar should be affected by temperature, especially since the temperature is constant during one run as it takes only about 40 minutes to complete.

- Looking at the acceleration pedal position, the correlation coefficient is high, although it does not really increase with higher T value. What is also odd is that the longitudinal acceleration does not hold a large correlation while there is one for the acceleration pedal. What this stems from is not clear.

To summarise, there is no doubt that there is some connection between the odd lateral values and some of the other variables. However, Table 4.1 does not immediately provide a clear explanation and hence a deeper survey is required before any conclusions can be stated. As reverse engineering of the internal filter was outside the scope of this project, it was left out for future work and the remaining of the work in this thesis is done with the knowledge that the lateral distances are not trustworthy.

4.2 Choices of Approaches

Chapter 3 presented the theory behind the use of a target tracker and described a few basic ideas. The performance of the tracker depends on the choice of track initiation and data association methods, as well as the filter type and motion model. Justifications to the techniques selected to solve the problem in this thesis are presented throughout this section.

4.2.1 Track Initiation

Using an M/N-logic initialisation method, the design parameters needed to be tuned are M, N and N_{dead}, which is how many consecutive samples without observations are allowed to pass before deletion of the track is decided. In the case of score based initiator on the other hand, the design parameters are the probabilities which we can allow to confirm a false track and the probability which we can allow to delete a true track. Besides, prior knowledge of the rates of new targets and false alarms as well as the detection probability need to be known for good performance. The M/N-logic approach is more intuitive and easier to implement, requires lower computational power, the tuning parameters are fewer and have an easier physical interpretation, and no prior knowledge of any rates or probabilities are required. Furthermore, as stated when presenting the method, the longer the track has lived, the longer it takes for the score to decrease to the deletion threshold, which is a property that is not always desired. Based on these facts, the M/N-logic initiator was chosen. As it turned out, satisfying results were accomplished with this method which is why there was never a need for trying out the score based approach, or any other track initiation method.

4.2.2 Data Association

As in the case of deciding on a track initiation method, the choice of data association method came down to the decision of starting with the one easiest to implement. Based on the theoretical description in Section 3.5, this is beyond all reasonable doubt the nearest-neighbour filter as the association then only is a matter of using the measurement being closest to each track. The nature of this particular tracking problem, that is highway driving, is such that the plausibility that this method would yield false association patterns can be assumed fairly low as cars on the highway tend to keep a fair distance to one another. Again, as in the choice of initiation approach, a simple choice gave results that were satisfying enough. However, in situations were the data is available as raw measurements (as oppose to this particular application were the measurements are output from an internal tracking filter), with presence of clutter, a more sophisticated approach would possibly be required. In that case, the probabilistic association method would have been the first-hand choice.

4.2.3 Motion Model

When focusing on highway driving, the vast majority of the motion is in the longitudinal direction. Generally, people convey their cars with constant speed on the highway and therefore, a (nearly) constant velocity model was first tried. As it turned out, the assumption of a pervading constant velocity was not enough as vehicles with a non-neglectable relative acceleration from time to time were present. As a matter of fact, situations where even a third order model appeared to be useful were found during the progress of this thesis work. Figure 4.2 illustrates this type of scenario. The longitudinal relative velocity of a vehicle takes what seems to be a quadratic shape, which indicates a linear acceleration $a(t)$ and hence a constant jerk $j(t) = \frac{\mathrm{d}a(t)}{\mathrm{d}t}$. Consequently, in order to catch the full behaviour of this mode, and other holding the same properties, the longitudinal model should be augmented with the jerk as a state. However, in literature this seems to be extremely rare, not to say non-existent. Consensus appears to be that a constant velocity model often is sufficient and is, to name one, utilised by Naesseth [2013] with what appears to be satisfactory results. An example of where a constant acceleration is used to model the longitudinal dynamics is Möbus et al. [2003]. In fact, all papers, articles, books and theses on target tracking that were consumed during the progress of this thesis used either a constant velocity or constant acceleration model and hence, it was decided that a more advanced model should not be used here neither. As it turned out, after a bit of tuning this mode actually could be caught with an acceleration model, not entirely or perfectly but sufficiently. Figure 4.8 in Section 4.3 shows this.

Moreover, the lateral motion of a highway driving vehicle is practically neglectable which is why an acceleration model in that direction was assumed to be redundant. In fact, if the coordinate system would have been road-aligned (as in e.g. Eidehall [1996]), the lateral dynamics would be simplified to $v_{\mathrm{y}} = 0$. For this particular situation with car-aligned coordinates, there are however some lateral dynamics present that need to be modelled since the road is not entirely straight (see Figure 2.2) and so a constant velocity model was assumed in the lateral direction.

This means that to each confirmed track i, the state vector

$$x^i = \begin{pmatrix} r_{\mathrm{x}}^i \\ r_{\mathrm{y}}^i \\ v_{\mathrm{x}}^i \\ v_{\mathrm{y}}^i \\ a_{\mathrm{x}}^i \end{pmatrix}$$

is introduced, where

- r_{x}^i is longitudinal distance
- r_{y}^i is lateral distance
- v_{x}^i is longitudinal relative velocity

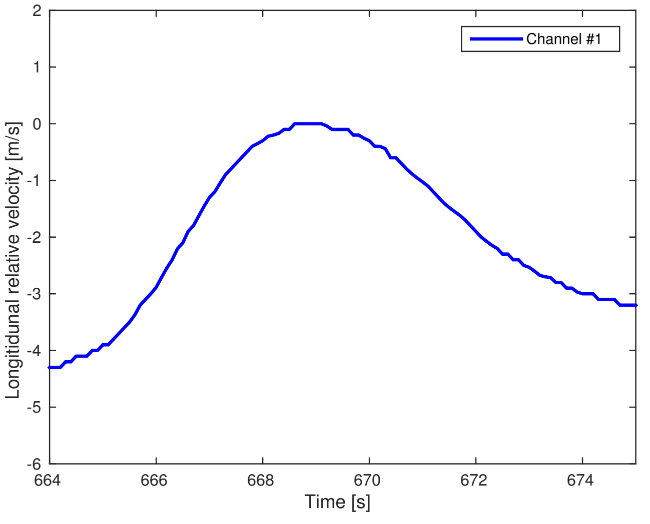


Figure 4.2: *A situation where constant acceleration model is not ideal as the velocity takes a quadratic shape. A jerk model would be required to catch the complete behaviour of this mode.*

- v_{y}^i is the lateral relative velocity
- a_{x}^i is the longitudinal relative acceleration

with the coordinate system being aligned as in Figure 2.1. Then the following linear state-space is obtained:

$$\begin{aligned}
\dot{x}^i(t) &= \overbrace{\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}}^{=A} x^i(t) + \overbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}}^{=B} w^i(t), \\
y^i(t) &= \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}}_{=C} x^i(t) + e^i(t).
\end{aligned} \tag{4.2}$$

Discretisation subsequently yields the following discrete-time state-space model, assuming the noise being constant between samples:

$$
\begin{aligned}
x^i_{k+1} &= \begin{pmatrix} 1 & 0 & T & 0 & T^2/2 \\ 0 & 1 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & T \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} x^i_k + \begin{pmatrix} T^3/6 & 0 \\ 0 & T^2/2 \\ T^2/2 & 0 \\ 0 & T \\ T & 0 \end{pmatrix} w^i_k, \\
y^i_{k+1} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} x^i_k + e^i_k.
\end{aligned}
\tag{4.3}
$$

T is the sample time of the sensor. The radar has a sample frequency of 10 Hz, which means that $T = 0.1$ s.

4.2.4 Filter

The standard Kalman filter is by far the most frequently used filter in automotive applications for tracking, and therefore it was a natural choice to apply to this problem too. Another alternative would have been for instance the particle filter, but no effort was put into investigating other options.

Filter Initialisation

As described in Chapter 3, the common way to initiate the Kalman filter is to use the expected value and covariance of the state, if its distribution is known. To this particular problem however, this is not the case since an object might appear very close to the host vehicle during e.g. an overtaking situation, as well as very far ahead. Mallick and La Scala [2008] propose that for a such problem, a so-called *single-point track initialisation algorithm* can be used. It suggests that the first state estimate shall consist of the first valid measurements, that is,

$$
\hat{x}_{1|1} = \begin{pmatrix} y_{1,\mathrm{x}} & y_{1,\mathrm{y}} & y_{1,v_\mathrm{x}} & 0 & 0 \end{pmatrix}^T, \tag{4.4}
$$

whereas the corresponding covariance is given by

$$
P_{1|1} = \begin{pmatrix} \sigma^2_\mathrm{x} & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_\mathrm{y} & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{v_\mathrm{x}} & 0 & 0 \\ 0 & 0 & 0 & \frac{v^2_{\mathrm{y,max}}}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{4.5}
$$

where $v^2_{\mathrm{y,max}}$ is the maximum possible lateral speed of a target, which is provided as an estimate by the user.

Filter Tuning

Tuning the filter to obtain good performance involves adjusting Q and R, that is, the process and measurement noise covariance. Since the noise is assumed to be decoupled so that the matrices become diagonal, five parameters need to be tuned: lateral and longitudinal process noise, and measurement noise for each of the three measurable signals. Essentially, the tuning procedure is a trade-off between noise sensitivity and speed of the filter. A slightly more detailed exposition on the impact of Q and R is presented below:

- The process noise $w(t)$ determines how the confidence interval grows with time, so its covariance Q determines how much randomness is incorporated in the states. As Q grows, the larger the covariance ellipse becomes and the less can predicted about the future. In other words, the larger Q the less the dynamic model is trusted in the filter.
- The measurement noise $e(t)$ determines how the measurements should affect the confidence intervals. The smaller R, the more trustworthy are the measurements, and $R = 0$ means that the measurements are perfect, whereas $R \to \infty$ means that the measurements do not add any information at all. In other words, the larger R the less the measurements are trusted.

An initial guess was made based on the physical insight of the dynamics. On highway driving, the motion is mainly in the longitudinal direction and the lateral motion arises mostly from minor wheel corrections and during lane switches. Therefore, the longitudinal motion was assumed to be fairly consistent with the model, i.e. low value of Q, whilst the lateral motion was assumed to have more inherent randomness, i.e. higher value of Q. There is a bit of a dilemma here though. As mentioned before, the lateral measurements provided by the radar appear strange and from time to time unreasonably large. This would indicate that the measurements should not be trusted, and hence R should be large. However, this approach would not be consistent with the aim of the task which was to reorganise the data to eliminate the phenomenon of objects switching channels randomly (see Figure 1.1 for a reminder). The task was *not* to do reverse engineering of the internal filter in order to understand and improve the original data. Hence, for the objective of this task, the measurements were considered trustworthy in the filter tuning. Its measurement covariance was first set to the identity matrix, whereupon semi-optimisations were made on both matrices until satisfying performance was achieved. The tuning results are presented below:

$$Q = \begin{pmatrix} 0.05 & 0 \\ 0 & 1 \end{pmatrix}, \tag{4.6}$$

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}. \tag{4.7}$$

4.2.5 Pre-Processing

Within the concept of MRA, a driver cannot be accused of being inattentive in non-predictable situations, such as if an oncoming vehicle on the highway sud-

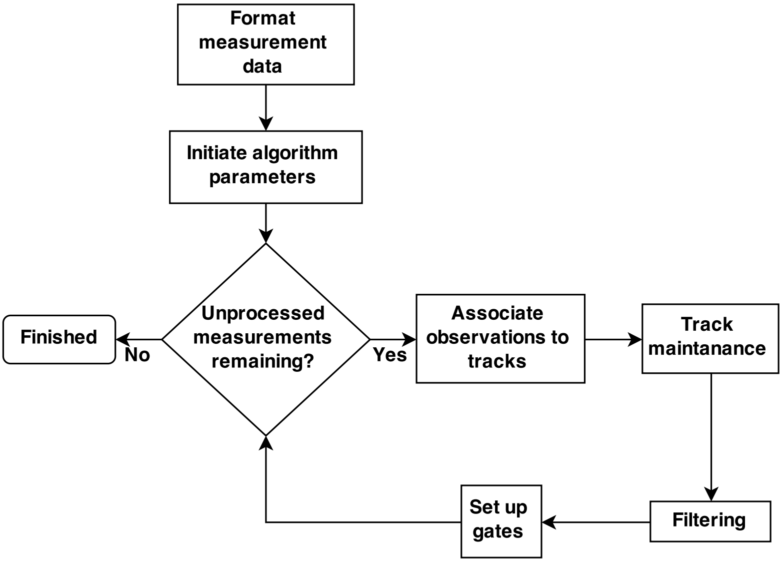


Figure 4.3: *Work flow of the tracking algorithm.*

denly appears in the wrong lane due to loss of driver control. Therefore, detections of vehicles driving in the opposite direction can be interpreted as false alarms, and by removing these prior to applying the tracking algorithm, its performance can be improved. In practise, this was done by ignoring all data points holding a relative velocity greater than some threshold. Sweden has a maximum allowed speed of 120 km/h on highways, whereas for a car with a trailer it is limited to 80 km/h. Assuming people generally convey their vehicles at the maximum allowed speed, this would mean that relative speed should not exceed 40 km/h at any time. Accounting for e.g. speeding cars, the threshold was set to 55 km/h.

4.3 Results

Below, a few results produced by the target tracker illustrating the performance are presented. The general work flow is as in Figure 4.3. Using the M/N-logic $M = 2$ and $N = 3$, and a dead-reckoning time of N_{dead} 30 samples (that is, 3 seconds), examples of the tracker performance is presented in Figure 4.5 and 4.7. Figure 4.4 and 4.6 show the original data. The improvements gained from the designed filter is clearly proven.

What can be seen in the figures is that the results are satisfying. The tracking algorithm has used the available samples and partitioned the data into one signal per present vehicle. This now allows to keep track of how often each individual vehicle is being attended to, which was the aim of this task.

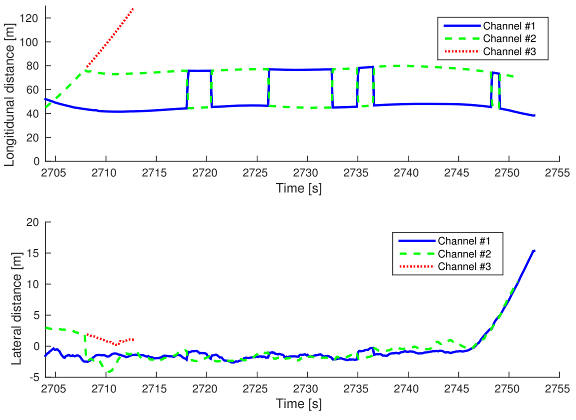


Figure 4.4: *Original data. This is the same figure as Figure 1.1, but the velocity plot has been omitted.*

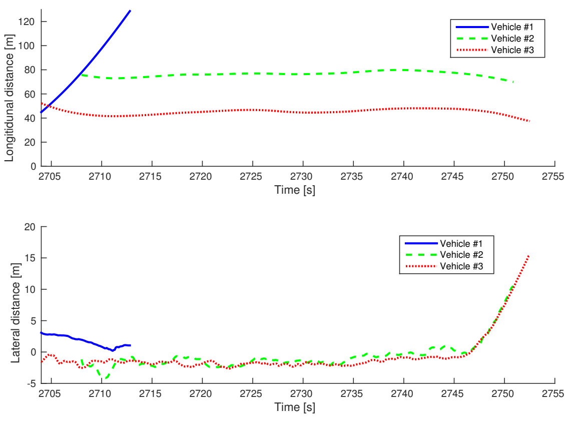


Figure 4.5: *Output from the target tracker on the segment in Figure 4.4. Especially in the longitudinal distance signal the results are clear; the present vehicles are now separated, which was the aim of this task. Note that the numbering of the vehicles in the legend is irrelevant outside the context of this plot.*

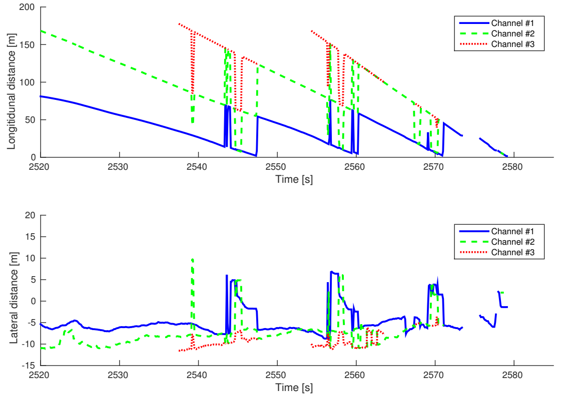


Figure 4.6: *Another set of the original data. Note the unreasonably large lateral distances - the objects in the channels are in a steady-state mode at five to ten metres to the right of the host vehicle, which is physically impossible on a highway.*

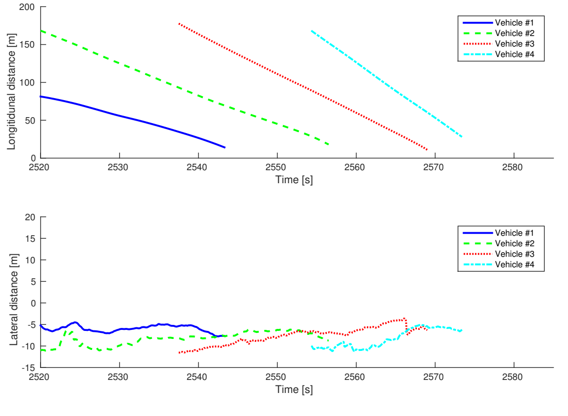


Figure 4.7: *Output from the target tracker from the data segment in Figure 4.6 again proving the performance.*

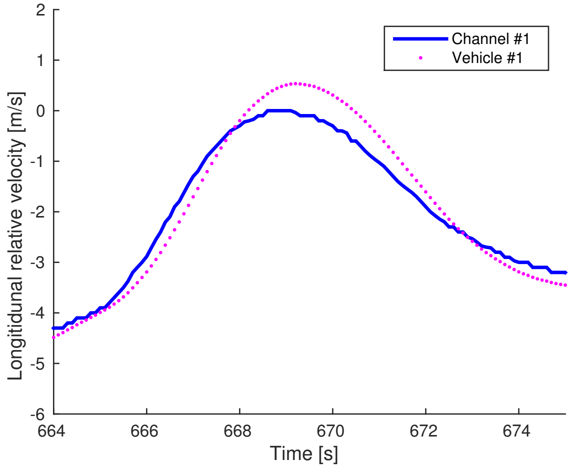


Figure 4.8: *Here it can be seen how a constant acceleration model can catch modes where a jerk is present (not perfectly but sufficiently).*

4.4 Limitations

The motion model is chosen based in the prior information that the algorithm will be applied to data collected on highway driving, on which the algorithm also renders satisfying results. There is no guarantee for the algorithm performance on other road types where the traffic flow and road shape differ.

Further, and perhaps more importantly, a major limitation of the general accuracy of the tracked vehicles is due to the fact that only a radar sensor has been used as a single-sensor system generally often lacks reliability and robustness. A radar provides accurate measurements of range and range rate at far distances while it is robust against bad weather conditions, but fails in providing good angle measurements as they are subject to noise [Amditis et al., 2004] . This means that the lateral position using radar only is not very trustworthy. The lack of robustness only applies when data measurements appear as raw data, that is, the lateral uncertainties should be accounted for in the internal filter. The data provided by it, which is the data used as measurements here, are as said before the results of a tracking performed by the radar module. To make a guess, the odd lateral distances returned by the radar could (partially) stem from the filter being badly tuned.

Another common sensor in automotive tracking applications is a camera. A camera provides highly accurate lateral position of targets, but is in opposite to the radar sensitive to weather such as rain and fog, and cannot provide accurate longitudinal position estimates at distances exceeding 50-60 meters [Gern et al., 2001].

The advantages and drawbacks described above suggest that a combination of these two would be a possibly good idea. Amditis et al. [2004] for one, use sensor fusion of camera and radar information to create artificial angle measurements (in that article they use polar coordinates for tracking) as $\theta = \arctan(y_{\mathrm{camera}}/x_{\mathrm{radar}})$. This obviously gives rise to issues regarding association between radar and camera measurements, but the drawbacks from each individual sensor are circumvented.

The uncertainties in lateral position from radar only could play a role in explaining the large lateral distances described earlier in Section 4.1. It is however doubtful that it is the only explanation. The angular errors that radar naturally give rise to stem from the fact that cars lack well-defined reflection points such as corner reflectors, and therefore the radar beam may slide from one side of the vehicle to another (again, see [Gern et al., 2001]). The error therefore increases with increasing angle to the tracked vehicle, and uncertainties of a few meters due to the sliding behaviour of the radar is to be expected. As presented in Table 4.1, lateral distances greater than 10 meters occur at over 20000 data points, and distances greater than five meters at over 130000 data points. There is also a noteworthy amount of samples that hold a lateral distance greater than 20 meters. They constitute only approximately 0.03% of total number of collected samples, but still: 1378 data points sampled at 10 Hz means roughly two minutes of accumulated time. That is too much to be considered neglectable.

It should be emphasised that these seemingly bad lateral measurements do not affect the tracking itself; the tracking filter performs well based on the available data and fulfils its purpose of refining the output from the internal filter to create individual signals of each and every vehicle. The issue is that the output from that the internal filter, that the tracking algorithm developed here considers as input, apparently does not entirely resemble the truth.

5 Vision Data

The discussion carried out in Section 4.4 concerning the odd lateral distances produced by the radar, drawbacks and benefits of different sensors and the improvements that theoretically could be gained from performing sensor fusion of radar and vision data induced the idea that it should be tried out if this would be the case here. Four months into this project a camera was installed in the test vehicle and a test run was made where both radar and vision data was collected. It was decided that if the data turned out to be good enough - that is, if the camera and radar produced results that suggested that a sensor fusion of them both would improve performance of the target tracker - an attempt would be made to incorporate the vision data into the target tracker. The outcome of the test runs will be described in this chapter.

5.1 Vision System

The vision system used was from the a Mobileye AWS-2000. As opposed to the radar which delivers lateral distance, longitudinal distance and longitudinal relative velocity, the vision system delivers range and angle to each tracked object. The range rate, that is the radial relative velocity, is not calculated. Note that no image processing was required in this project as it was made by an internal filter in the camera module.

5.2 Data Analysis

The radar and vision data were plotted together so that a naive survey could be made by comparing the signals visually. Plots showing some typical behaviour modes of the signals are presented below, which illustrates the overall appearance.

- Figure 5.1 depicts a typical situational in the longitudinal distance plot and it is clear that the signals are inherent with mutual inconsistencies. The camera is either late to capture an object that the radar already has detected, oscillates vigorously or has a large offset.
- Figure 5.2 shows how extremely odd lateral values - well over 200 metres - are occasionally provided by the camera.
- Figure 5.3 depicts a scenario where the radar does not at all match the camera in the lateral distances. Furthermore, the camera tracks an object that appears to slide up to 50 metres to the left while the radar instead tracks two objects almost 10 metres to the right. The true scenario, from the front view camera, is that the host vehicle is driving behind another vehicle in the left lane, overtaking a vehicle in the right lane. This means that apparently only camera channel #1 delivers reasonable data here. An object sliding 50 metres to the right is nowhere to be seen in the movie.
- Figure 5.4 depicts a rare scenario where the camera and the radar follow the same trend with a clear offset in between them. The true scenario is that the host vehicle is driving right begind another vehicle, that eventually takes an exit to the right. That is, the camera appears to deliver the truth here while there is an offset in the radar.

5.3 Conclusion

The figures in this chapter show typical behaviour modes, and these kind of inconsistencies and abnormalities are commonplace throughout the entire data set, and hence it was decided not to proceed to include the vision data into the tracking algorithm. The test with the vision system was done with the aim of hoping to see whether it could provide a simple answer to the issues with the odd lateral distances delivered by the radar; for example, if the data would be as in Figure 5.4 while longitudinal distances being consistent, it could have been concluded that the radar is just inherent with an offset. That information could then be fused into the target tracking algorithm enhancing its performance. Now, this was not the case as the camera as well performed badly and so it was decided that spending more time on the camera data would not be worthwhile.

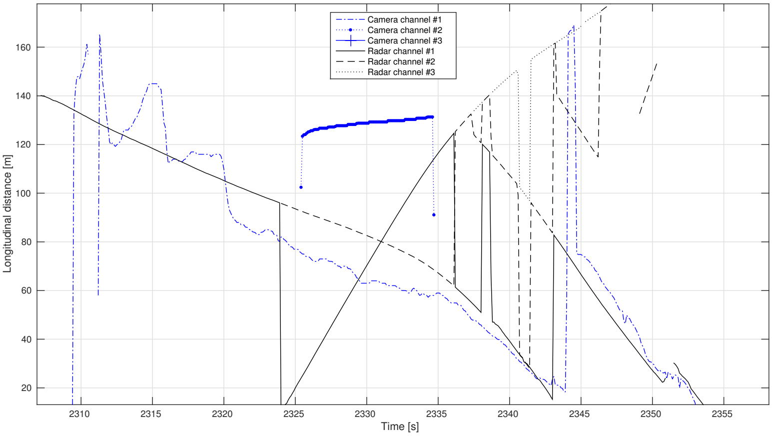


Figure 5.1: *Typical scenario. Between 2310 and 2345 seconds it shows how the camera data in principal follows the radar data, but it oscillates and there is a clear offset. For instance, between 2320 and 2330 seconds, the offset is around 20 metres which is extremely large in the context. Further, the radar sees an object at 120 metres away at 2335 seconds, and starts tracking it (although channel shifts occur). At first when the object is 80 metres away - at 2345 seconds - the camera catches it. Thereafter, the radar-camera match appears reasonably good for that object. However the object caught by the radar at ∼2325 seconds and continuously tracks for over 20 seconds is ignored by the camera. Note that the figure may strike the reader as intricate and hard to interpret. This itself is an indication that the data is not very good - if the radar and the camera had delivered consistent data, it would have been easier to interpret the figures.*

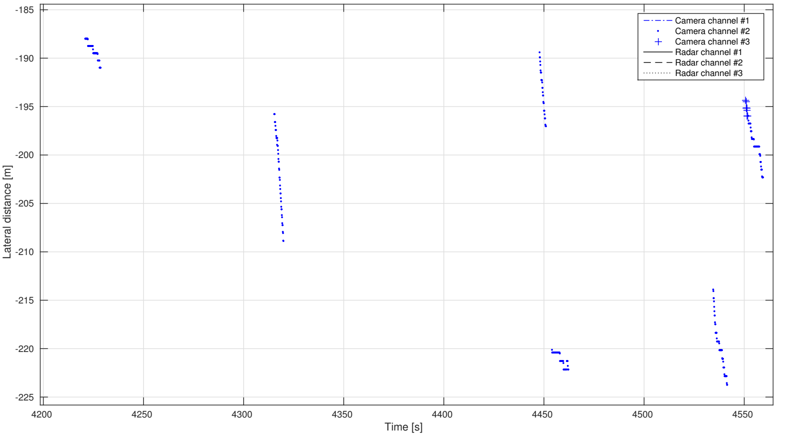


Figure 5.2: *Multiple extremely odd lateral values are occasionally delivered by the camera. What they stem from is unclear.*

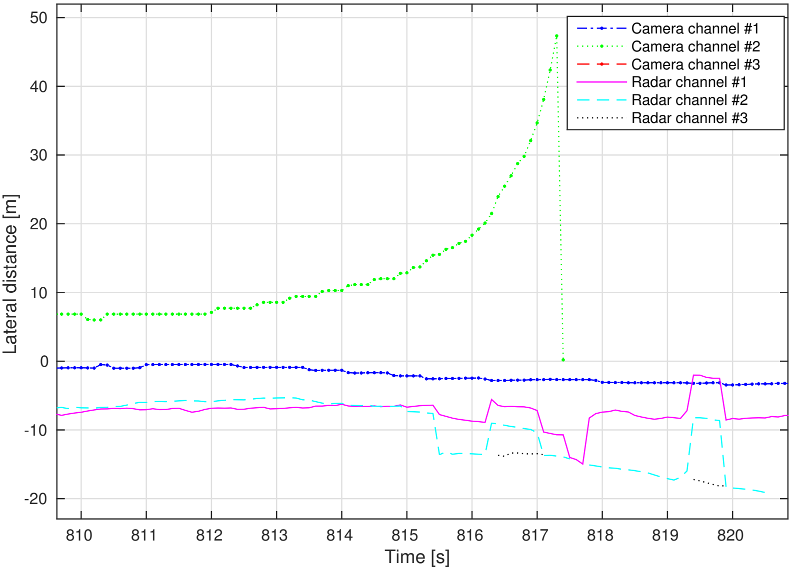


Figure 5.3: *Scenario where the vision and radar data differ considerably. The camera tracks an object ~8 meters to the left with an exponentially growing distance. The radar does not catch this object at all. The camera also tracks an object that seems to be driving steady right in front of the host vehicle. The radar on the other hand tracks two, and at times even three, objects, that appear to be ~8 meters or more to the right. True scenario: driving right behind a vehicle in the left lane with another vehicle being in the right lane.*

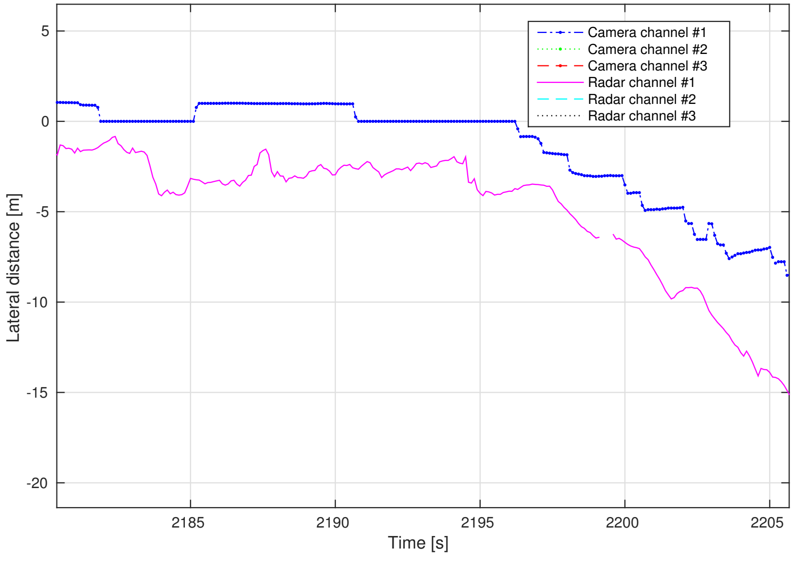


Figure 5.4: *Scenario where radar and camera seem to track the same object. The lateral distances adopt the same shape, but there is an offset between them. The camera claims that the object is practically in front of the host vehicle during the first 15 seconds - which is what to be expected on a highway - while the radar delivers a distance what is a couple of metres to the right. True scenario: driving right behind a vehicle, that eventually exits the highway to the right.*

6

Modelling of Driver Behaviour

The initial main purpose of this thesis was to propose a novel attention monitoring algorithm based on the acquired data. VTI wanted a data driven method rather than a result of physical modelling. The model should reflect how experienced drivers behave while driving on highways, people tend to often drive close to other vehicles, even though it is nominally dangerous. In case something happens, a close distance can cause severe damage as the driver will not have time to react and brake or steer away. However, people do this anyway as they from experience know that highway accidents are extremely rare according to Taieb-Maimon and Shinar [2001]. This means that a physical model that would require a theoretically safe distance in real life probably would warn all the time. This is why a model based on how very experienced drivers behave in traffic is desired. As it turned out, the reliability of the radar was debatable, which complicated the task. In order to be able to proceed, limitations needed to be set. This chapter will walk through the incitement of developing a context adaptive attention monitoring algorithm. The limitations that were set will be presented along with benefits and drawbacks induced by them. Next, it will be discussed how it might be possible to utilise the data in a novel and efficient way. The possibilities are restricted by the limited amount of good data (and time), so the chapter will mostly revolve around theories and observations that can be concluded.

6.1 Background

Driving is a task where loads of different modes are inherent. Although highway driving is the most simple case, there are still plenty of scenarios where the gaze pattern nominally should differ:

- Depending on how many vehicles are present, the attentional demand differs.

- During entrances and exits the driver needs to be extra attentive.
- If the maximum allowed speed limit is lowered, e.g. from 120 km/h to 100 km/h as the highways traverses a city, the driver needs to be pay extra attention as the behaviour of other drivers will plausibly change within the next seconds as they will lower their speed.
- Overtaking situations are inherent with other attentional demands than regular, monotonous driving.

More situations could be mentioned. All the attentional demands that differ between situations are purely cognitive. The driver - at least an experienced one - identifies each unique situation and tailors his or her gaze pattern thereafter. Catching all the different behaviour modes based on only physical modelling would be a hard task, surely way outside the scope of a master thesis within electrical engineering as it probably would require deep knowledge of psychology and cognitive science.

Consequently, it was decided to analyse the gaze pattern of the test persons. One possible drawback, apart from the fact that no person is obviously fully-learned when it comes to driving, is that all glances may not stem from the fact that they feel that they *need* to attend to a specific target at the given time. For instance, assume a situation where a heavy truck is performing an overtaking of another heavy truck far away from the host vehicle. As the distance is large, the driver does not feel insecure about the situation, and hence does not feel the urge to pay close attention to what is going on. However, the driver has been driving monotonously for a long time, and hence joyfully observes the situation. That is, the driver does not attend to the trucks because he or she feels that he or she needs to, but because there is nothing else that demands the attention. Conclusively, the natural behaviour of the driver implies that not all gazes stem from attentional demands and it is hard to distinguish what the purpose of a gaze is - a consequence of an attentional demand, or just natural behaviour. Verbal protocols may help with this task, but it is impossible for a driver to explain each and every gaze and saccade.

6.2 Eye-Tracking

A Matlab script for visualising the tracking results in the xy-plane was written; a snapshot of the simulation view is presented in Figure 6.1, and the exact same frame of the front view is provided in Figure 6.2. The simulation view claims that the driver is looking at an object almost straight ahead, which agrees what is seen in the front view film. However, the simulation view also claims that objects are being present about seven and nine metres to the left of the host vehicle, which is not agreed upon by the film. It has been suggested before that the radar delivers non-robust data, and by doing this type of simultaneous survey of the simulation and real view, there is no longer any doubt at all the lateral measurements are not trustworthy enough to be used.

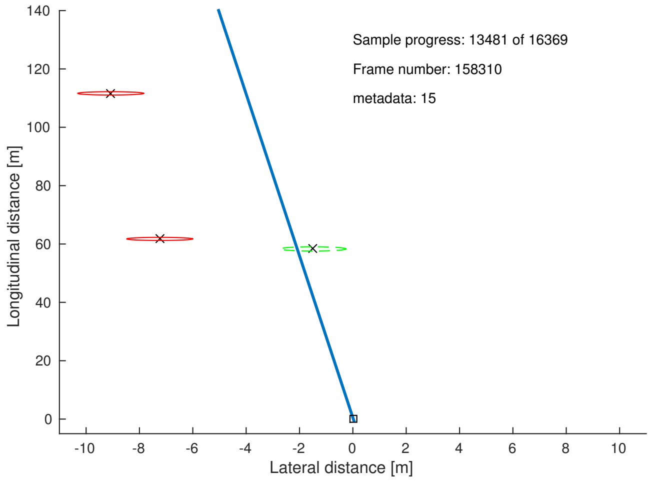


Figure 6.1: *The scenario in Figure 6.2 in a simulation view. The black square at the origin represents the host vehicle and the blue straight line is the eye-tracking vector i.e. the blue dot in Figure 6.2, but seen from an above perspective. The black crosses are the tracked targets along with a 90% confidence interval (ellipses). A green dashed ellipse means that the eye-tracking vector intersects with the covariance ellipse at the current sample, which should be interpreted as that the target in question is being attended to. A red, solid ellipse in oppose means that the target is not being attended to at the current sample. The text strings are auxiliary information for synchronisation purposes. As mentioned before, the large lateral distances are doubtful when comparing to ground truth (that is Figure 6.2).*

An initial part of this thesis was to analyse how well the eye-tracking data can be matched with the radar data. The answer turned out to be that it is basically impossible due to the lack of robustness in the radar.

Under the assumptions of trustworthy measurements, an intuitive approach to determining if an object is being looked at is to use the covariance ellipse as in Figure 6.1. Since there are uncertainties both in the state predictions and in the eye-tracking data, this seemed like a natural way to address the problem; if the eye-tracking vector intersects with the covariance ellipse, then the object in question is being attended to. Obviously, as the radar lacks trustworthiness this is not applicable.

6.3 Suggested Approach

In order to evade the problem with bad data, a major limitation had to be added; instead of incorporating all available data to the method on modelling driver behaviour, it was decided to aim our focus towards situations with only one vehicle

Figure 6.2: *A screen shot of the film where the gaze location (blue dot) has been overlayed to a front view film. The number in the upper left corner is the frame number, which has been used in order to synchronise these films with the simulation script.*

present. Pros and cons with this restriction is presented below:

- If only one vehicle is present, its lateral position can be ignored. As opposed to the lateral measurements, the longitudinal ones seem to be perfectly fine. Additionally, the problem of determining whether or not an object was payed attention to is far less sensitive to longitudinal than to lateral errors; for example, if an object 60 metres straight ahead is being attended to, then that object would also have been attended to if it would have been 50 or 70 metres straight ahead (or with even larger margins than so), while these numbers of course do not apply to lateral errors. With that said, if only one vehicle is present, its longitudinal position can be used to assume that it is located somewhere on the road at that longitudinal position. This illustrated in Figure 6.3 (note that two vehicles are depicted in the figure, but that is only to illustrate that a vehicle will be covered by the triangle no matter what lane it is in). If the driver looks within that triangle he or she also looks at the present vehicle. It is a simplification, but it evades the issue with bad lateral measurements. On the other hand, this approach implicitly accounts for peripheral vision, so the discussion regarding fixated and peripheral objects is circumvented. Additionally it covers possible minor uncertainties in the eye-tracking data.

- An argument in favour of using situations where only one vehicle is present is the fact that this field is rather uncharted territory; physical models are

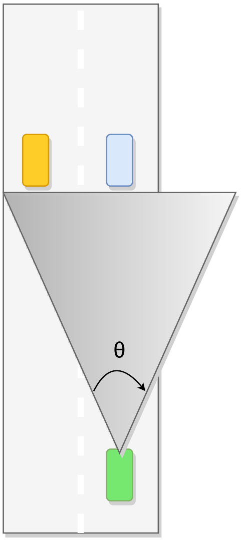


Figure 6.3: *When a vehicle is detected, a triangle as in this figure is set up. Its angle θ such that the entire width of the road is precisely covered at the longitudinal distance at which the vehicle is. If the eye-tracking vector is within that cone, it is considered that the vehicle is looked at. Note that this method works since only the scenario where one vehicle is present is studied. The figure depicts a situation where two vehicles appear to be present, but that is just for illustrational purposes, in order to show that it will not matter what lane the detected or host vehicle is in, as it will always be within the triangle.*

widely used in e.g. threat assessment for advanced driver assistance systems where the safety margins need to be large enough to avoid presumptive collisions at any time, but the approach of consulting experienced drivers and trying to concretise rules that based on their behaviour patterns in different situations is new. Therefore, one should start out simple, and by initially focusing on situations where only one vehicle is present, conclusions regarding whether or not it is worth investing time in continued work with more complex scenarios could be drawn. That is, the work about to be presented could be seen as an early-stage proof of concept of the approach.

- Delimiting the work to include one vehicle scenarios only will decrease the number of available data points vigorously, which is always a drawback in statistical context (as long as all data stem from the approximately same distribution, that is). It turned out that excluding all scenarios with multiple vehicles present reduced the data points by five sixths. More on this in Section 6.4.

6.4 Course of Action

In Section 2.3 it was stated that test runs were made during four different conditions: regular driving during day time, regular driving during night time, day driving with occlusion goggles and day driving with verbal protocols. The night and verbal protocol runs were only carried out by six test persons out of the total twelve. Therefore it was decided to focus towards regular day driving and driving with occlusion goggles, from now on referred to simply as "day" and "occlusion". The radar data, acquired at a sampling frequency of 10 Hz was up-sampled to 60 Hz to match the eye-tracking data and then data points from day- and occlusion scenarios with one vehicle visible to the radar were cropped out. The work flow was accordingly:

- The longitudinal distance to the vehicle was used to calculate a triangle covering the entire road at that distance as in Figure 6.3
- It was investigated whether the eye-tracking vector fell inside the calculated triangle; if so, the vehicle was noted as being attended to at that data point. Concerning the scenarios with occlusion goggles, it was also required that the goggles were not active in order for the vehicle to be considered attended to.
- Each time the driver looked away from the vehicle, the longitudinal distance and relative velocity at that time was noted. For the following consecutive samples that the driver did not attend to the vehicle, the relative velocity was stored, and a counter was increased.
- When the driver eventually re-attended to the vehicle, the accumulated look-away time was stored. Look-away times shorter than 150 ms were ignored as those more plausibly were blinks than self-aware gaze drifts. Similarly, if the driver after re-attending to the vehicle did not keep attending for more than 100 ms, those gazes were also ignored.

By following this work flow for all one-vehicle scenarios for all twelve test persons, the results were lots of data saying things like *when a vehicle was 40 metres ahead with a relative velocity of 5 km/h, the driver looked away for 2 seconds*. They idea was then to use this data to see if any patterns regarding how different parameters affect the total time that a driver can look away (recall that the behaviour patterns of these experienced drivers is to be considered "correct") would emerge. The number of one-vehicle scenarios for the twelve test persons is shown in Table 6.1. The reason that the number of one-vehicle scenarios differ is simply that the runs were not performed at the same time of the day, and hence at different traffic densities.

It is now assumed that the longitudinal distance and relative velocity are the variables causing the most influence on driver's gaze patterns during highway driving, and hence only these will be considered from now on. Other variables probably have a larger impact in more complex environments.

Table 6.1: *Number of one-vehicle scenarios for the test persons that the analysis is based on.*

Test person	No. one-vehicle scenarios (day, occlusion)
1	24, 0
2	15, 20
3	14, 21
4	0, 15
5	21, 13
6	12, 22
7	12, 15
8	18, 15
9	4, 19
10	4, 17
11	10, 18
12	4, 20

Regarding how distance and relative velocity to a vehicle affects the drivers gaze behaviour, it is not obvious, and it is most certainly not some linear relationship. A first thought is that it is obvious that the faster the host vehicle drives compared to the other vehicle and the shorter longitudinal distance, the more seldom we can allow ourselves to look away. By making more thorough inquiry it can be realised that the word *predictability* plays a key role, and the following discussion arises (note that this discussion applies to highway driving. Other environments could raise a different discussion):

- The lowest possible relative velocity - that is, the maximum negative relative velocity - occurs in the event of another vehicle standing still on the highway (if we assume that the risk of encountering a reversing vehicle on a highway is non-existent). Depending on what has caused the event and what the situation looks like, the action the driver performs will probably either be a lane change and keep driving at the approximately same speed, or start slowing down. In both scenarios, the future is fairly predictable - the other vehicle is standing still, so it cannot do anything unpredictable like a poorly signaled lane change.

- The smallest longitudinal distance occurs during overtakning situations. If the host vehicle is the one being overtaken, it can be assumed that the overtaking vehicle is fully aware of the situation, and hence will not to anything unpredictable so it should be fine to look away even while the overtaking vehicle is fairly close.

- On the other hand, in the event of the host vehicle approaching another vehicle, it can not be guaranteed that the driver in the front vehicle is aware of that is happening behind him or her. Hence, the host vehicle driver probably feels the urge to pay a little extra attention to the front vehicle as the situation now is unpredictable.

To summarise, the situational demands theoretically clearly differ depending on longitudinal distance and relative velocity to other vehicles, but, again, the relationship appears to be fairly complex. The discussion just previously suggests that the sign of the relative velocity should have a large impact on the situational demands - approaching another vehicle is a less predictable situation than one where the front vehicle moves away. Therefore, a first step was to categorise the one-vehicle scenarios into

- host vehicle approaching front vehicle
- front vehicle moving away from host vehicle
- host vehicle and front vehicle driving with the approximately same speed.

The motivation behind this partition was to see whether the actual gaze patterns would agree with the theoretical attentional demands - that is, that shorter times of gaze allocation outside the road are allowed when approaching another vehicle than when the front vehicle moves away. This means that for each type of test run - day and occlusion, there are data points saying things like *when a vehicle was 40 metres away, moving farther away, the driver looked away for 2 seconds* and *when a vehicle was 30 meters away, while being approached by the host vehicle, the driver looked away for 0.9 seconds* and so on.

It should now be emphasised that it is not theoretically motivated to merge the data stemming from the two different types of test runs:

- During day, the driver is told to behave exactly as normally. This means that the driver may, and probably will, attend to the road even though the situational demands do not force him or her to do it, since looking straight ahead simply is the most convenient thing to do while driving.
- During occlusion on the other hand, the driver is instructed to activate the occlusion goggles as often as possible. Therefore it can be assumed that the driver will only look straight ahead then the situational demands ask for this.

Nominally, these two runs would provide different information: the day runs should reflect normal behaviour, whereas the occlusion runs should (given enough data) render information on what should be theoretically allowed regarding the time allowed to look away in different situations. Merging the data from the two runs would result in more data points, which under the assumptions of the data stemming from the same distribution would render a more nuanced picture, but since this assumptions cannot be made based on the premises, they were kept distinguished.

Figure 6.4 shows an example of what the scatter plots look like. With a little determination, some patterns can be recognised that support the discussion held previously regarding how the situational demands theoretically should vary with distance and relative speed; when front vehicle moves away from host vehicle, the situation is predictable, and hence long gaze drops (that is, the time that the road is not being attended to) are "allowed" even when the other vehicle still is close, while the opposite hold when approaching the front vehicle. The plots are obviously not extremely convincing, but the patterns are there, and with plenty of more data points the anticipation is that the patterns will evolve.

The midmost figure - when driving at the approximately same speed as the front vehicle - is not very interesting as the individual gaze preferences of the drivers override the "statistical patterns". Possibly, more data points could help revealing some pattern. From now on, we omit this scenario and focus only towards the other two.

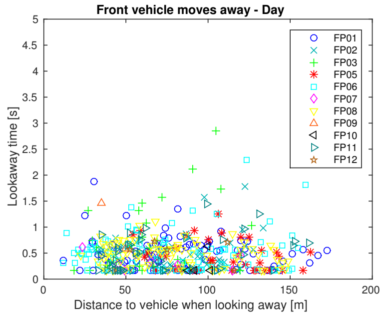
6.5 Trend Estimation

As already been stated, the scatter plots in Figure 6.4 suggest that some pattern exists of how the situational demands vary with distance and relative velocity. In this section, different techniques for estimating these variations are proposed.

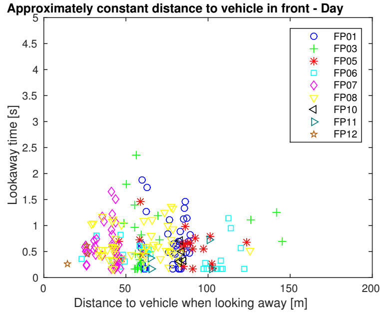
6.5.1 Regression

A first step to estimate how the variations of the situational demands would be to estimate a regression curve that follows the longest gaze drops for the different speed profiles That is, we try to fit a curve to the maximum gaze drop durations at different speeds. The distances were categorised into bins of 5 metres, i.e. 0-4, 5-9, 10-14. 15-20 and so on. In each bin, the 95-percentile (to avoid extremely long gaze drops having too much influence of the outcome of the regression) was found and a regression curve was estimated on these data points. One might raise the question on why only the long gaze drops are considered, and why a regression of *all* data points is not computed. The reason is, as already been mentioned, that we are interested in the extremes - that is, we are interested in investigating how the *longest* gaze drops vary with the distance and relative velocity to the vehicle, and more specifically, we are interested in whether or not they agree with the intuition. A regression of all data points would show how the *general* gaze behaviour varies with distance and relative speed, but that would probably only be some line approximately parallel to the horizontal axis. We are here interested in the *maximum* gaze drop durations in order to determine the *maximum* allowed time to remove attention from the road, and hence include only these in the regression computation. It was assumed that at some point the distance to the vehicle no longer has an impact on the situational demands; for instance, a vehicle 130 metres ahead and 150 metres ahead probably will not give rise to different gaze behaviour. A brief look at Figure 6.4a and 6.4c indicates that this distance occurs at somewhere between 100 and 150 metres. Therefore, the data points with distances larger than 125 metres were not used when forming the regressions curves. More effort could be put into this survey, but that is left for future work.

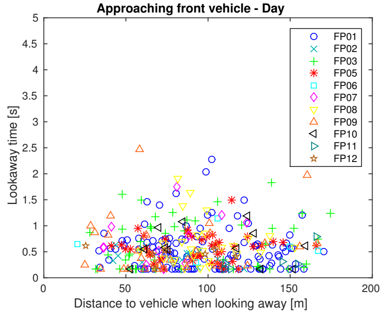
Both a linear regression $y = kx + m$ and a power regression $y = ax^b + c$, y being and gaze drop duration and x being the distance when looking away, were



(a) Front vehicle moving away from host vehicle.



(b) Host vehicle and front vehicle driving with the approximately same speed.



(c) Host vehicle approaching front vehicle during day.

Figure 6.4: *Scatter plots of distance to vehicle when looking away versus the time looked away, for the three different relative speed profiles during day run. FP means test person. With a slight determination, patterns resembling the theoretical ones could be recognised: longer gaze drops at short distances are more likely to be found when front vehicle moves away than when approaching the front vehicle.*

estimated using the data up to 125 metres. The results are shown in Figure 6.5 and 6.6. A few remarks:

- The curves has been estimated on the maximum gaze drop durations in each bin. That is, the curves aim to estimate the longest allowed time to look away at different distances to front vehicle, for the different relative speed profiles approaching host vehicle and host vehicle moving away. As this is just an effort to derive a proof of concept, all data is considered estimation data, so that all data was used when estimating the curves. The suggestion is *not* to just take this curves straight away and say that they resemble the context adaptive situational demands. It is rather so, that if these investigations agree with the intuition, it can be seen upon as a "proof" that this approach may actually work, and then put more effort in exploiting the concept and find more robust ways to compute these curves, and also to incorporate more variables.

- Generally, the gaze drops are much longer in the occlusion runs. This stem from the fact that during day run, the drivers should behave "normally", and the gaze drops stem only from looking beside the road. During occlusion run they were told to make the effort of using the occlusion goggles as much as possible, and the gaze drops stem both from regular looks beside the road and from occlusion data. This shows why the data from the two runs should not be merged together in order to achieve more data points - the data is simply collected during different conditions.

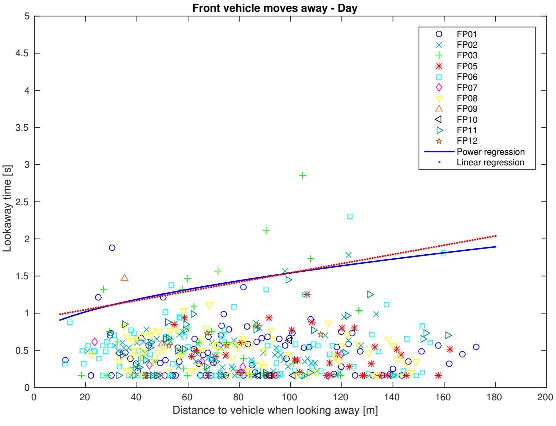
- The day plots (Figure 6.5) seem to agree with the intuition - the regression curves are much steeper when approaching front vehicle, which would mean the attentional demands are higher there.

- The power regression in Figure 6.6a adopts a little different shape than the other power regression curves since it is more eager to catch the early gaze drop durations than the linear regression. After ~20 metres it levels out and remains constant.

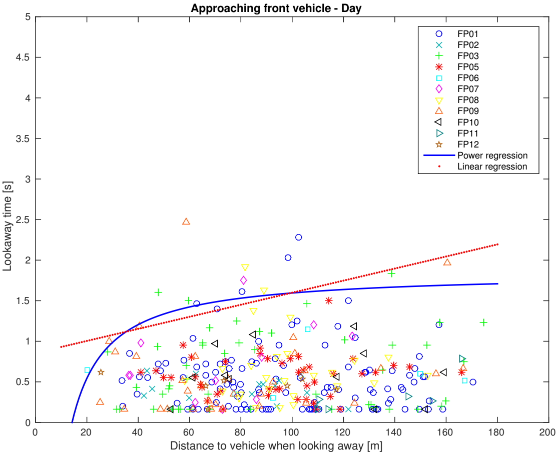
To compile the results presented above, all four linear regression curves were plotted in the same figure, and the same thing was done for the power regression curves. They can be seen in Figure 6.7 and 6.8. It is clear that the same pattern can be seen during occlusion as during day, but longer gaze drop durations occur during occlusion. This is, as already been mentioned, since the test persons are explicitly told to use the goggles as often and as long as they find safe, while the day plots only resemble normal behaviour.

Including Velocity Measurements

The next natural step would be to exploit the velocity measurements as numerical values instead of grouping them together in three different speed profiles, and then visualise the gaze drop durations as a function of both distance and relative velocity. The most convenient way to display the results would be a level

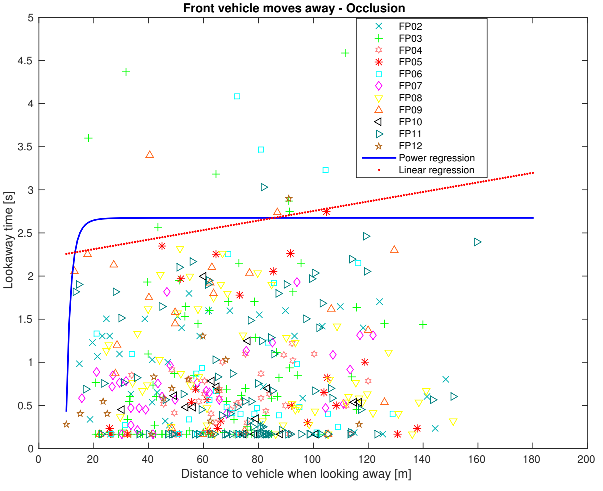


(a) Front vehicle moving away from host vehicle during day run, with regression curves estimated from the 95-percentile of the gaze drop durations for data up to 125 metres.

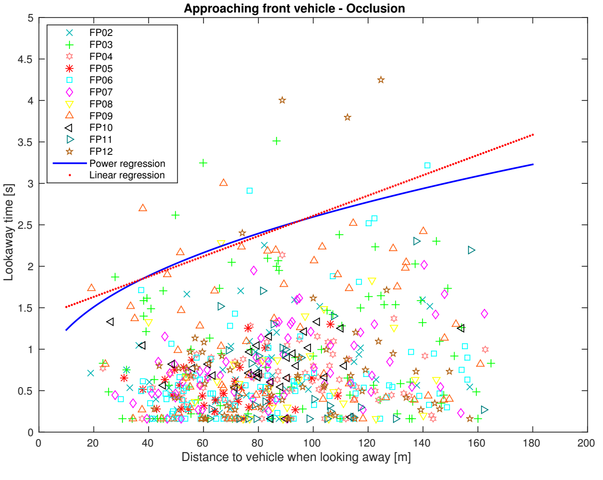


(b) Host vehicle approaching front vehicle during day run, with regression curves estimated from the 95-percentile of the gaze drop durations for data up to 125 metres.

Figure 6.5: *Linear and power regression curves were estimated based on the 95-percentile of the gaze drop durations. They show that longer gaze drops are allowed when the front vehicle moves away, than when approaching the front vehicle. These figures are from the day runs.*



(a) Front vehicle moving away from host vehicle during day run, with regression curves estimated from the 95-percentile of the gaze drop durations for data up to 125 metres.



(b) Host vehicle approaching front vehicle during day run, with regression curves estimated from the 95-percentile of the gaze drop durations for data up to 125 metres.

Figure 6.6: *Linear and power regression curves were estimated based on the 95-percentile of the gaze drop durations. They show that longer gaze drops are allowed then the front vehicle moves away, than when approaching the front vehicle. These figures are from the occlusion runs.*

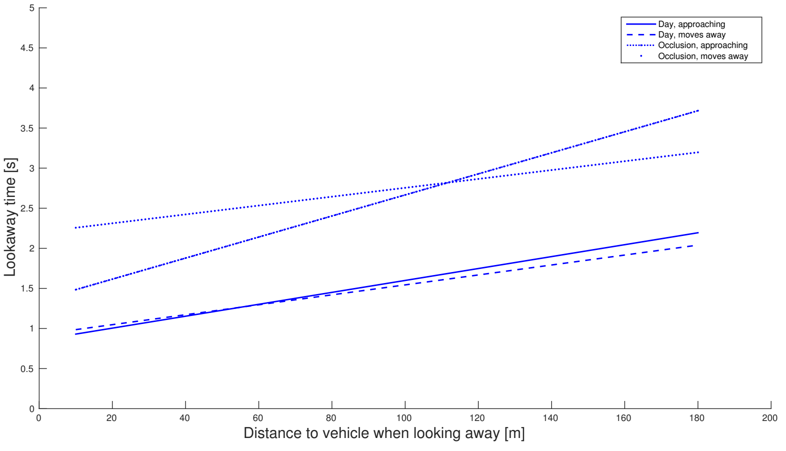


Figure 6.7: *The linear regressions for each of the four scenarios in the figure legends.*

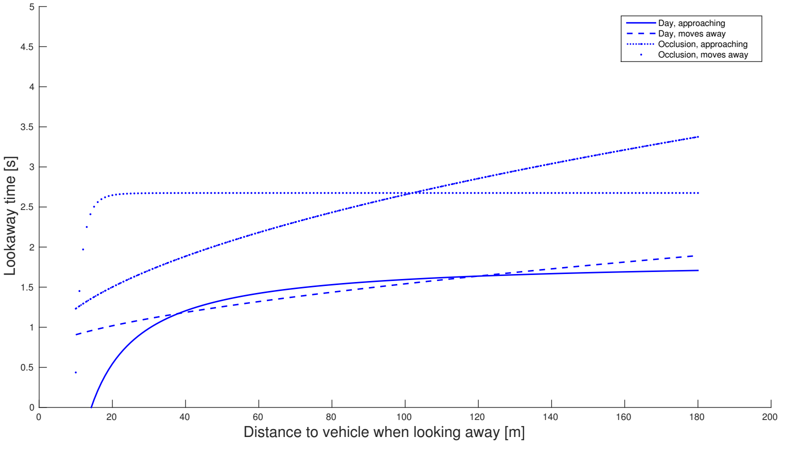


Figure 6.8: *The power regressions for each of the four scenarios in the figure legends.*

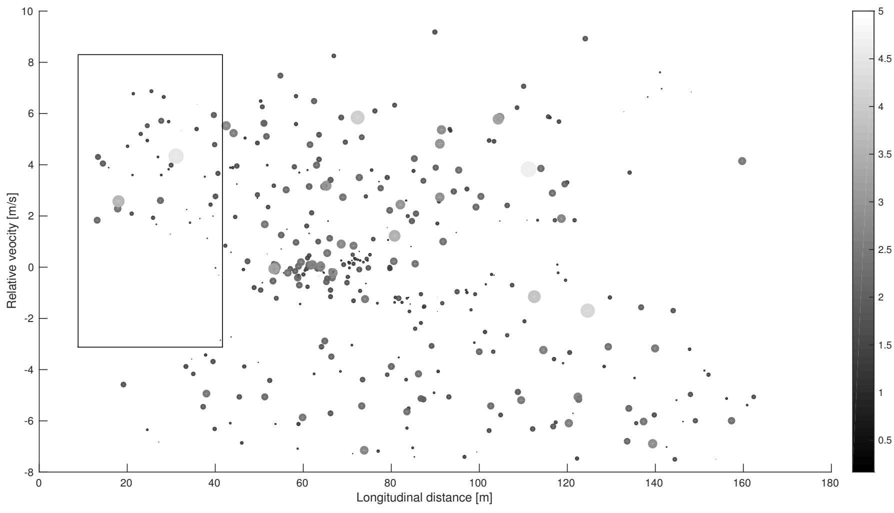


Figure 6.9: *Time of use of occlusion goggles (in seconds) plotted against distance and relative velocity and illustrated by the colour scale. Also, the bigger circles, the longer durations (the circle radii are proportional to the duration squared). The fact that the density of large circles is higher for short distances (up to ∼60 metres) for positive relative velocities, that is, front vehicle moves away, agrees with what has been discussed concerning the theroretical situational demands.*

plot where the gaze drop durations are colour coded and plotted against the two independent variables. Such a plot is provided in Figure 6.9. In this case only occlusion data has been used, that is, the plot illustrates the use of occlusion goggles while normal gaze drops (i.e. looking beside the road) have been ignored. This is interesting since it depicts the longest durations that the test persons are prepared to look away during driving. It is clear that for short distances (within the black rectangle), the density of occlusion goggles being active is significantly higher for positive relative velocities, that is, when the front vehicle moves away. When approaching a front vehicle, use of occlusion is extremely rare. This agrees with the intuition.

A regression plane could of course be estimated in the same way as the curves for the two dimensional case, but if more dimensions were to be introduced that would not be a good idea as visualisation would be impossible. Therefore, a more sophisticated approach to distinguish "normal" gaze behaviour from occasional longer gaze drops is desired. In the next subsection one such method will be suggested and discussed, although it was never tested out so it should be seen upon only as a suggestion of a method that could be investigated in the future.

6.5.2 Support Vector Machines

Using a regression line as a "look-up table" for the maximum allowed gaze drop durations during different conditions as described previously, would only be an available option if the number of dependent variables does not exceed two since higher dimensions cannot be visualised. A less heuristic approach would be to consult the machine learning community and feed the data into an algorithm that trains to recognise the patterns that may or may not be visible to the human eye. The support vector machine is a popular technique for doing this. In its standard use, it is used to separate data into one of two classes. More specifically, given a training data set

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \ : \mathbf{x}_i \in \mathbb{R}^p, \ y_i \in \{-1, 1\}\}_{i=1}^{N},$$

with N being the number if samples and p the number of features (i.e. the dimension of $\mathbf{x}_i$), the aim is to find the hyperplane that most effectively separates the points in $\mathcal{D}$ that belong to class +1 from the ones that belong to class -1. Test data with unknown class belonging can then be categorised into either class -1 or 1 depending on where it is located in relation to the hyperplane generated by the algorithm from the training set.

As seen in the figures, the distribution generally seems to be such that the vast majority of the data points lie within a certain range, with occasional points standing out. Put into physical context, most gaze drops are short, but with occasional longer ones. If we consider a long gaze drop to be an "outlier", then some method for outlier detection could perhaps be utilised in order to separate "normal" data from outliers. Such a method is suggested in the following section.

One-Class Support Vector Machine

According to Manevits and Yousef [2001], a first introduction to the one-class support vector machine, from now on referred to as OCSVM, was made in Schölkopf et al. [1999]. The abstract of this report provides a background for the adoption of this classifier as following:

> *Suppose you are given some dataset drawn from an underlying probability distribution P and you want to estimate a "simple" subset S of input space such that the probability that a test point drawn from P lies outside of S equals a priori specified value between 0 and 1.*
>
> *We propose a method to approach this problem by trying to estimate a function f which is positive on S and negative on its complement. [...].*
>
> *The algorithm is a natural extension of the support vector machine to the case of unlabelled data.*

Consider a situation where we monitor some industrial machine that during the vast majority of the time works as intended, but where abnormalities from time to time occur. The abnormalities manifest themselves differently each time and their origins are not necessarily known. Putting the system into the different fault modes in order to collect training data of class both -1 and +1 (that is, class "normal" and class "abnormal") to train a standard two-class support vector machine is hence difficult not to mention potentially dangerous, depending on the result of a faulty behaviour. What is desired is a classifier that can be trained only on data collected during normal conditions, and this is exactly what is being introduced in [Schölkopf et al., 1999].

Available Software

The Statistics Toolbox for Matlab includes functionality for two-class classification using support vector machines but as of now, it does not support OCSVM. LIBSVM (see [Chang and Lin, 2011]) is a free library for support vector machines available for a large number of programming languages including Java, Python, C, C++, Lisp, Ruby, R and Matlab(1). It provides automated functions for training different types of support vector machine classifiers, including OCSVM. The user provides the training data set along with choice of kernel function and design parameters. LIBSVM then computes a model that can be used to classify data in a test set.

6.6 Discussion

The one-vehicle scenario data was first partitioned into i) host vehicle approaching front vehicle, ii) front vehicle moving away from host vehicle and iii) front vehicle and host vehicle driving with the approximately same speed. It was discussed how the situational demands theoretically should differ within these scenarios, and by estimating a regression curve through the maximum gaze drop durations it was found that the results agreed with the intuition. Then, the partitioning was skipped and the relative velocity was used explicitly so that the gaze drop durations could be plotted against both distance and relative velocity. The density of the gaze drop durations continued to agree with the theoretical insights. No "regression plane" was attempted to be fitted to the data, but instead it was suggested that a one-class support vector machine could perhaps be used to identify "outliers" - that is, the longest gaze drop duration allowed at different relative velocity and distance to object. Trying OCSVM out went outside the time scope of this project, so it was left as a proposal for future work, with a motivation to why it was found to be a well fitted method to solve the problem.

The work done in this chapter has aimed to derive a proof of concept regarding a data driven attention monitoring algorithm. Further investigations and development of this suggested method would exceed the time scope of this thesis, but the results look quite promising as what has been produced so far agrees with the intuition of how the driver should behave during different conditions. With more data, the patterns found would most plausibly evolve even more, and then other variables could be introduced as well. With more accurate tracking data, more complex scenarios than one-vehicle situations could be investigated.

(1)A complete list can be found on the website http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

7

Concluding Remarks and Future Work

As this thesis report now approaches its end, it is time to present some conclusions that have been made, and to give some ideas on future work based on this thesis:

- The tracking algorithm works well for highway driving, which is the simplest type of driving in the sense that the roads are fairly straight and the speed profiles among the cars are mostly easily predictable. It remains to be seen how well it works when being applied to data collected on rural roads as they tend to have more frequent curvature. On the other hand, they are less trafficked and there is only one lane in each direction, which simplifies the work for the tracking algorithm.
- As of now, the driver behaviour model only uses longitudinal distances due to bad lateral data. Investing in new measurement instruments, i.e. a radar and a camera that delivers consistent, robust data, would imply great possibilities for improvements and further development.
- In this work, eye-tracking and radar data have been analysed for driving scenarios with one front vehicle being present, with promising results. The next step would be to given proper measurement instruments (as mentioned earlier) perform further tests in order to acquire more data. Then, the algorithm could be extended to handle more complex driving scenarios.
- Finally, another suggested proceeding of this work is to dig into the OCSVM and investigate how well the proposal of that being a sophisticated method for identifying abnormalities.

Conclusively, the results from this work provide a good basis for further development of formulating the situational demands in different driving scenarios based on eye-tracking data from experienced drivers.

Appendix

A

Appendix A

Mathematical Formulation of the OCSVM

A comprehensible mathematical problem formulation can be found in e.g. Zhang et al. [2012], which is briefly reproduced here.

Given a set of samples $\mathcal{T} = \{\mathbf{x}_i\}_{i=1}^{N}$, $\mathbf{x}_i \in \mathbb{R}^p$ where p is the dimension of $\mathbf{x}$ or the number of features, the OCSVM maps $\mathcal{T}$ into a possibly infinite-dimensional feature space $\mathcal{F}$ by a nonlinear function $\phi(\mathbf{x})$, and then constructs a hyperplane

$$f(\mathbf{x}) = \boldsymbol{\omega} \cdot \phi(\mathbf{x}) - \rho, \tag{A.1}$$

with $\boldsymbol{\omega}$ being its normal vector and ρ its offset from the origin. The hyperplane f aims to separate "normal" data from abnormalities and the parameters are found by solving the optimisation problem

$$\begin{aligned} \min_{\boldsymbol{\omega}\in\mathcal{F},\ \xi\in\mathbb{R}^N,\ \rho\in\mathbb{R}} \quad & \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho \\ \text{subject to} \quad & \boldsymbol{\omega}\cdot\phi(\mathbf{x}_i) \geq \rho - \xi_i,\ \xi_i \geq 0, \end{aligned} \tag{A.2}$$

where the design parameter $\nu \in [0, 1]$ can be interpreted as an estimate of the fraction of outliers in the training set and ζ_i are slack variables that penalise error rejection of $\phi(\mathbf{x})$.

Omitting the details, the dual problem is obtained as

$$\begin{aligned} \min_{\alpha_i,\ \alpha_j} \quad & \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu N},\ \sum_{i=1}^{N}\alpha_i = 1, \end{aligned} \tag{A.3}$$

with α_i and α_j being Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_j) \cdot \phi(\mathbf{x}_j)$ the *kernel function* that can e.g. linear, polynomial or Gaussian, whichever is found most convenient for the specific application. Subsequently, the hyperplane f is transformed into a decision function

$$f(\mathbf{x}) = -\log \frac{\sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x})}{\rho}. \tag{A.4}$$

If $f(\mathbf{x}) > 0$, then the sample $\mathbf{x}$ is outside the region defined by the hyperplane and is hence to be considered an outlier.

Bibliography

A. Amditis, N. Floudas, and A. Polychronopoulos. Lateral Motion Tracking of Automobiles. 2004. Cited on page 34.

A.M. Aziz. A New Nearest-Neighbor Association Approach Based on Fuzzy Clustering. *Aerospace Science and Technology*, 26:87–97, 2013. Cited on page 18.

Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, Inc., 1988. Cited on page 14.

Y. Bar-Shalom and X.R. Li. *Estimation and Tracking: Principles, Techniques and Software*. Artech House, 1993. Cited on page 17.

Y. Bar-Shalom, K.C. Chang, and H.M. Shertukde. Performance Evaluation of a Cascaded Logic for Track Information in Clutter. In *IEEE Transactions on Aerospace and Electronic Systems*, volume 25, pages 873–876, November 1989. Cited on page 19.

Y. Bar-Shalom, F. Daum, and J. Huang. The Probabilistic Data Association Filter. In *IEEE Control Systems Magazine*, pages 82–100, December 2009. Cited on page 19.

S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, 1999. Cited on pages 14 and 21.

C-C. Chang and C-J. Lin. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`. Cited on page 59.

A. Eidehall. *Tracking and Threat Assessment for Automotive Collision Avoidance*. PhD thesis, Linköping University, 1996. Cited on page 27.

A. Gern, U. Franke, and P. Levi. Robust Vehicle Tracking Fusing Radar and Vision. In *Multisensor Fusion and Integration of Intelligent Systems*, pages 323–328, 2001. Cited on pages 34 and 35.

T. Glad and L. Ljung. *Modellbygge och Simulering*. Studentlitteratur, 2004. Cited on page 15.

F. Gustafsson. *Statistical Sensor Fusion*. Studentlitteratur, 2012. Cited on page 15.

K. Kircher and C. Ahlström. Minimum Required Attention - an Operationalisation of Driver Distraction, 2014. Not yet published. Cited on page 1.

M. Mallick and B. La Scala. Comparison of Single-point and Two-point Difference Track Initiation Algorithms Using Position Measurements. *Acta Automatica Sinca*, 34(3), 2008. Cited on page 29.

L.M. Manevits and M. Yousef. One-Class SVMs for Document Classification. In *Journal of Machine Learning Research*, volume 2, pages 139–154, 2001. Cited on page 58.

R. Möbus, A. Joos, and U. Kolbe. Multi-Target Multi-Object Radartracking. In *2003 IEEE Intelligent Vehicles Symposium*, 2003. Cited on page 27.

C. Andersson Naesseth. Vision and Radar Sensor Fusion for Advanced Driver Assistance Systems. Master's thesis, Linköping University, 2013. Cited on page 27.

B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and A.C. Williamson. Estimating the Support of a High-Dimensional Distribution. Technical report, 1999. Cited on pages 58 and 59.

J.W. Senders, A.B. Kristofferson, W.H. Levison, C.W. Dietrich, and Ward J.L. The Attentional Demand of Automobile Driving. pages 15–21, 1967. Cited on page 6.

Smartmicro. *UMRR Automotive Type 29 Data Sheet*, 2012. Cited on pages 9 and 23.

M. Taieb-Maimon and D. Shinar. Minimum and Comfortable Driving Headways: Reality versus Perception. *The Journal of the Human Factors and Ergonomics Society*, 43(1):159–172, 2001. Cited on page 43.

Trafikverket, May 2004. URL `http://www.trafikverket.se/TrvSeFiler/Foretag/Bygga_och_underhalla/Vag/Vagutformning/Dokument_vag_och_gatuutformning/Vagar_och_gators_utformning/Sektion_landsbygd-vagrum/05_vagtyper.pdf`. Cited on page 24.

Q. Zhang, J. Wang, J. Zeng, and G. Xu. A Dynamic Memory Model for Mechanical Fault Diagnosis using One-Class Support Vector Machine. In *8th IEEE International Conference on Automation Science and Engineering*, August 2012. Cited on page 65.