Sensor data fusion for automated driving

Toward robust perception in adverse weather conditions

Domhof, J.F.M.

DOI
10.4233/uuid:ee7b5513-3917-44d0-9307-7689be201153

Publication date
2022

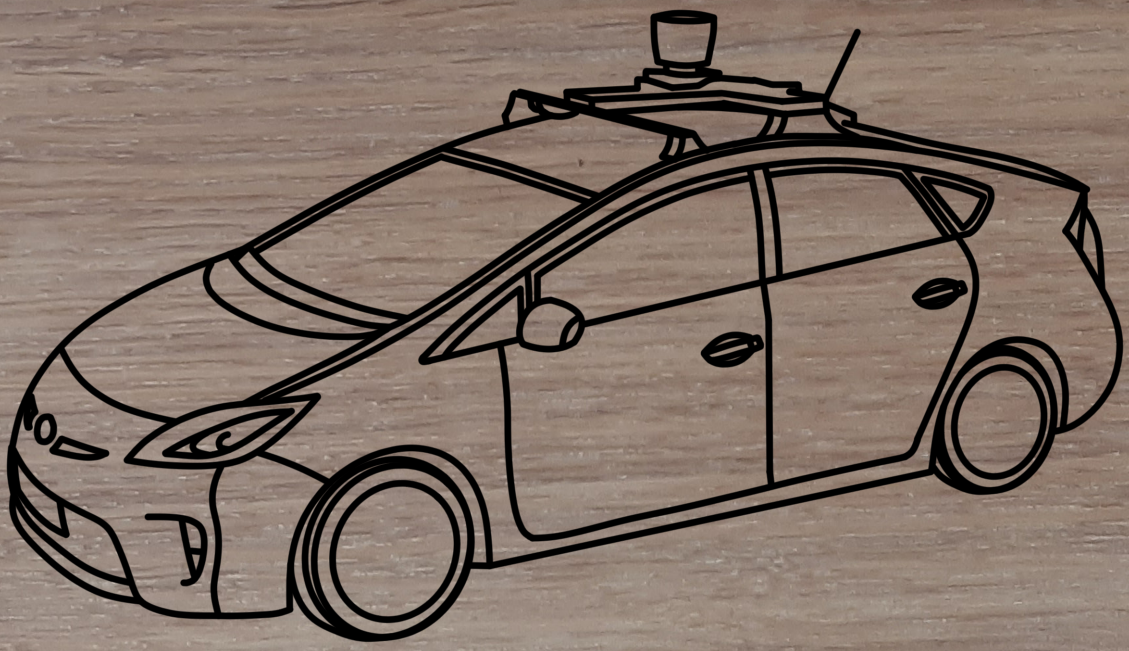
Document Version
Final published version

Citation (APA)
Domhof, J. F. M. (2022). *Sensor data fusion for automated driving: Toward robust perception in adverse weather conditions*. [Dissertation (TU Delft), Delft University of Technology].
https://doi.org/10.4233/uuid:ee7b5513-3917-44d0-9307-7689be201153

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

SENSOR DATA FUSION FOR AUTOMATED DRIVING

TOWARD ROBUST PERCEPTION IN ADVERSE WEATHER CONDITIONS



Joris Ferdinandus Maria Domhof

SENSOR DATA FUSION FOR AUTOMATED DRIVING

TOWARD ROBUST PERCEPTION IN ADVERSE WEATHER CONDITIONS

SENSOR DATA FUSION FOR AUTOMATED DRIVING

TOWARD ROBUST PERCEPTION IN ADVERSE WEATHER CONDITIONS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op woensdag 21 december 2022 om 15:00 uur

door

Joris Ferdinandus Maria DOMHOF

Master of Science in Mechanical Engineering,
Technische Universiteit Delft, Nederland,
geboren te Winterswijk, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. D.M. Gavrila
copromotor: Dr. J.F.P. Kooij

Samenstelling promotiecommissie bestaat uit:

Rector Magnificus,	voorzitter
Prof. dr. D.M. Gavrila	Technische Universiteit Delft, promotor
Dr. J.F.P. Kooij	Technische Universiteit Delft, copromotor
onafhankelijke leden:	
Prof. dr. ir. J. Hellendoorn	Technische Universiteit Delft
Prof. dr. ir. B. van Arem	Technische Universiteit Delft
Prof. dr. G.C.H.E. de Croon	Technische Universiteit Delft
Dr. F. Garcia Fernandez	University Carlos III of Madrid, Spain
Dr. G. Dubbelman	Technische Universiteit Eindhoven

Keywords: sensor data fusion, adverse weather, object tracking

Printed by: Gildeprint

Front & Back: Image by L. A. Heinsius

ISBN 978-94-6419-651-1

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

Summary vii

Samenvatting ix

1 Introduction 1
- 1.1 Sensor Data Fusion . . . 2
- 1.2 Research Questions . . . 3
- 1.3 Contributions . . . 5

References . . . 5

2 Sensor Selection 7
- 2.1 Introduction . . . 7
- 2.2 Related Work . . . 8
  - 2.2.1 Contributions . . . 8
- 2.3 Proposed Approach . . . 9
  - 2.3.1 Multi-Sensor Data Fusion . . . 9
  - 2.3.2 Non-Linear Filtering . . . 9
  - 2.3.3 Motion Model . . . 9
  - 2.3.4 Sensor Measurement Models . . . 10
  - 2.3.5 Cramér-Rao Lower Bound (CRLB) . . . 11
  - 2.3.6 Initialization . . . 11
  - 2.3.7 Performance Measures . . . 11
  - 2.3.8 Implementation Details . . . 11
- 2.4 Simulations . . . 12
  - 2.4.1 Process Noise Variance . . . 12
  - 2.4.2 Measurement Noise Variance . . . 12
  - 2.4.3 Urban Environment . . . 13
  - 2.4.4 Highway Environment . . . 17
  - 2.4.5 Tracking Filter Convergence . . . 17
  - 2.4.6 Imperfect Detection and Clutter . . . 17
- 2.5 Discussion . . . 22
- 2.6 Conclusion . . . 22

References . . . 23

3 Extrinsic Sensor Calibration 25
- 3.1 Introduction . . . 25
- 3.2 Related Work . . . 28
  - 3.2.1 Pairwise Calibration . . . 28
  - 3.2.2 Joint Calibration . . . 30
  - 3.2.3 Contributions . . . 30

3.3 Proposed Approach . . . 30
3.3.1 Calibration target design . . . 31
3.3.2 Detection of calibration target . . . 32
3.3.3 Pairwise calibration . . . 32
3.3.4 Joint calibration with more than two sensors . . . 34
3.3.5 Pose estimation of body reference frame . . . 35
3.4 Experiments . . . 36
3.4.1 *Relative calibration* . . . 37
3.4.2 Absolute calibration . . . 39
3.4.3 Outdoor experiments . . . 42
3.5 Discussion . . . 45
3.6 Conclusion . . . 48
References . . . 48

4 Object Tracking **53**
4.1 Introduction . . . 53
4.2 Related Work . . . 54
4.2.1 Contribution . . . 56
4.3 Methodology . . . 56
4.3.1 Original p_D-CPHD Filter . . . 57
4.3.2 Multi-Sensor p_D-CPHD Filter . . . 60
4.3.3 Bootstrap filters . . . 62
4.4 Experiments . . . 63
4.4.1 Experiments in Fog Simulator . . . 64
4.4.2 Recorded data with Prius . . . 66
4.5 Discussion . . . 69
4.6 Conclusion . . . 70
References . . . 71

5 Conclusion **75**
5.1 Sensor Selection . . . 75
5.2 Extrinsic Sensor Calibration . . . 75
5.3 Object Tracking . . . 76
5.4 Final remarks . . . 76
References . . . 78

Acknowledgements **81**

A CRLB **83**

B Hellinger distance **87**
References . . . 87

C Experimental results **89**

Curriculum Vitæ **91**

List of Publications **93**

SUMMARY

The aim of the thesis is to develop methods and algorithms for the development of a robust perception system that is capable of dealing with adverse weather conditions. Robust environmental perception is important in order to guarantee safety for the automated vehicle and the road users in the neighborhood. To create a robust perception system, a sensor setup should be selected with multiple sensing modalities. Commonly used sensing modalities in the field of intelligent vehicles are lidar, camera and radar sensors. This thesis addresses three subjects that are important for robust perception, namely *sensor selection*, *extrinsic calibration* and *object tracking*.

The first subject that this thesis addresses is *sensor selection*. The sensor setup should be able to fulfill the requirements for automated driving applications. The position and velocity of the objects should be estimated accurate enough. Therefore, the predicted tracking performance limits of the sensor setup should be estimated in an early design phase. A systematic approach is presented that is able to predict tracking performance limits for a setup. This approach can also be used to predict the performance in case of sensor failure, by computing the tracking performance limits for the sensor setup without the failing sensor.

The second addressed subject is *joint extrinsic calibration* of sensor setups consisting of one or multiple lidar, camera and radar sensors. Extrinsic calibration estimates the orientation and the position of the sensors to express the sensor data in a common coordinate frame. This is relevant for robots and intelligent vehicles. The requirement for joint extrinsic calibration is the availability of a calibration target that facilitates accurate detections for all sensing modalities. A calibration target has been designed for that purpose. For joint extrinsic sensor calibration, three optimization configurations are identified. The first configuration optimizes all sensor-to-sensor errors with respect to a reference sensor. The second configuration optimizes all sensor-to-sensor errors in combination with a loop closure constraint. The last configuration jointly estimates the sensor poses and calibration board poses in a probabilistic model. Experiments with a sensor setup consisting of a lidar, a radar and a stereo camera suggest that the configuration that includes all sensor-to-sensor errors in combination with a loop closure constraint performs best. Apart from estimating the relative pose of the sensors with respect to each other, it is also important to know where exactly in the vehicle the sensors are located. Therefore the sensors need to be calibrated with respect to the body reference frame of the robot. For that, two requirements are identified, namely the need for an external sensor that is able to 'see' the car and the calibration board and a set of 3D reference points. The set of 3D reference points are needed to determine the pose of the body reference frame and these points can be obtained by the method *Markers* that uses visual markers placed in the environment or by the method *Human labeling* using geometrical shape fitting. Experiments showed that *Human labeling* using geometrical shape fitting provides more accurate results, since the median rotation error around the

vertical axis 0.33° for the method *Markers* and 0.02° for the method *Human labeling*.

The third and last subject of this thesis is object tracking in adverse weather and illumination conditions. Object tracking is the process of estimating the position, velocity and other properties of all objects using detections of consecutive moments in time. Detecting road users in camera images is more challenging in adverse weather and illumination conditions. This means that the number of false negatives increases. The tracking performance is affected if this is not taken into account. The thesis proposes an adaptive filter comprising of two filters; a filter that estimates the detection probabilities, and a tracking filter that uses the estimated detection probabilities. By using the estimated probability of detection of both sensors in the tracking filter, the effect of a failing or affected sensor can be mitigated. This is shown both in a controlled environment (fog simulator) and in real world experiments with the Toyota Prius vehicle. The experiments in a controlled environment, for which the visibility range in fog decreases from 300 m to 50 m, showed improved tracking performance compared to tracking filters using fixed detection probabilities. Furthermore, experiments with the vehicle prototype showed improved tracking performance for scenarios recorded after sunset (twilight/night) and in scenarios where there was precipitation.

Finally, this thesis ends with a chapter that addresses the conclusions of all chapters. In addition, this chapter also addresses the final remarks and recommendations.

SAMENVATTING

Het doel van dit proefschrift is om methodes en algoritmes te maken voor de ontwikkeling van een robuust perceptiesysteem voor intelligente voertuigen dat met slechte weersomstandigheden om kan gaan. Om de veiligheid van de passagiers en andere weggebruikers te waarborgen is het belangrijk dat het systeem voor de perceptie van de omgeving robuust is. Om dit te bewerkstelligen moet er een set van sensoren geselecteerd worden die meerdere modaliteiten van waarneming bevat. Veelgebruikte modaliteiten op het gebied van intelligente voertuigen zijn lidar, radar en camera sensoren. Dit proefschrift behandelt drie belangrijke onderwerpen die nodig zijn voor robuuste perceptie; *selectie van sensoren, extrinsieke kalibratie* en *object tracking*.

Het eerste onderwerp van dit proefschrift is de *selectie van sensoren*. De sensor set van het intelligente voertuig moet kunnen voldoen aan de vereisten voor geautomatiseerde rijtoepassingen. De locatie en de snelheid van andere weggebruikers moet met een zekere nauwkeurigheid geschat kunnen worden. Hiervoor moet in een vroeg stadium van de ontwerpfase een schatting worden gemaakt van hoe nauwkeurig de locatie en de snelheid van andere weggebruikers kan worden bepaald. Er wordt een systematische benadering gepresenteerd die in staat is om de prestatie limieten van object tracking te voorspellen voor een sensor set. Deze benadering kan ook worden gebruikt om de prestaties te voorspellen in het geval van een sensor storing, namelijk door de object tracking prestatie limieten te berekenen voor de sensor set zonder de falende sensor.

Het tweede onderwerp is *gezamenlijke extrinsieke kalibratie* van sensor opstellingen die bestaan uit één of meerdere lidar, camera en radar sensoren. Tijdens extrinsieke kalibratie worden de posities en oriëntaties van alle sensoren bepaald zodat alle sensorgegevens in een gemeenschappelijk coördinatenstelsel kunnen worden gerepresenteerd. Dit is relevant voor zowel robots als intelligente voertuigen. De vereiste voor gezamenlijke extrinsieke kalibratie is de beschikbaarheid van een kalibratie object dat nauwkeurige detecties geeft voor alle modaliteiten (lidar, camera, radar). Daarom is hiervoor een specifiek kalibratie object ontworpen. Er zijn drie optimalisatie configuraties geïdentificeerd voor gezamenlijke extrinsieke kalibratie. De eerste configuratie optimaliseert alle sensor-naar-sensor fouten ten opzichte van een referentie sensor. De tweede configuratie optimaliseert alle sensor-naar-sensor fouten door middel van een lus sluitings beperking (loop closure constraint). De laatste configuratie schat gezamenlijk de positie en de oriëntatie van de sensoren en de kalibratie objecten in een probabilistisch model. De sensoren moeten niet alleen ten opzichte van elkaar worden gekalibreerd, maar het is ook belangrijk om te weten waar de sensoren zich bevinden in het voertuig. Daarom wordt er in dit proefschrift onderzocht hoe de sensoren kunnen worden gekalibreerd ten opzichte van het referentie coördinatenstelsel van het voertuig. Om dit mogelijk te maken zijn er twee vereisten geïdentificeerd; de behoefte van een externe sensor die in staat is het voertuig en het kalibratie object te 'zien', en een set van 3D-referentiepunten om de positie en de oriëntatie van het referentie coördinatenstelsel te

bepalen. De set met 3D-referentiepunten kan worden verkregen door twee methodes, namelijk de method *Markers* en de methode *Human labeling*. In het geval van *Markers* worden er visuele markeringen geplaatst in de omgeving. De methode *Human labeling* maakt gebruik van handmatig gelabelde punten in de sensor data waardoor er een geometrische vorm geschat wordt (*geometrical shape fitting*). Experimenten laten zien dat de methode *Human labeling* de beste resultaten geeft, omdat de mediaan van de rotatie fout om de verticale as van het voertuig 0.33° is voor de methode *Markers* en 0.02° is voor de methode *Human labeling*.

Het derde en laatste onderwerp van dit proefschrift is *object tracking* in ongunstige weers- en lichtomstandigheden zoals mist en schemering. Object tracking is het proces waarbij de positie, snelheid en andere eigenschappen van alle objecten geschat worden met behulp van detecties gedurende opeenvolgende momenten in de tijd. In ongunstige weers- en lichtomstandigheden is het detecteren van weggebruikers in camerabeelden uitdagender dan in gunstige omstandigheden. Dit betekent dat het aantal ontbrekende detecties (false negatives) toeneemt. Een ontbrekende detectie is een object dat niet gedetecteerd wordt terwijl het in werkelijkheid wel bestaat. Als de toename van ontbrekende detecties niet bekend is dan beïnvloedt dat de object tracking prestaties. Een adaptief filter wordt voorgesteld dat bestaat uit twee componenten. Het eerste component is een filter dat de detectie kansen schat voor iedere sensor. Het tweede component is een object tracking filter dat gebruik maakt van de geschatte detectie kansen. Door in het object tracking filter gebruik te maken van de geschatte detectie kansen, kan het effect van een defecte sensor of verminderde detectie condities worden beperkt. Dit wordt aangetoond zowel in een gecontroleerde omgeving als in praktijkexperimenten met een Toyota Prius voertuig. In de gecontroleerde omgeving zijn er experimenten uitgevoerd in een mist simulator waarbij de zichtafstand in mist afneemt van 300 m tot 50 m. Het adaptieve filter laat betere resultaten zien dan filters die gebruik maken van vaste detectie kansen. Tevens laten de praktijkexperimenten met het voertuig zien dat er verbeterde resultaten worden gevonden voor de opnames die zijn gemaakt na zonsondergang (schemering/nacht) en waarbij er neerslag viel.

Ten slotte eindigt dit proefschrift met een hoofdstuk waarin de conclusies van alle hoofdstukken worden behandeld. Daarnaast komen in dit hoofdstuk ook de slotopmerkingen en aanbevelingen aan de orde.

1

INTRODUCTION

Intelligent vehicles are equipped with multiple sensors to perceive the surroundings of the vehicle. The main three sensing modalities are lidar, camera (monocular and stereo) and radar. The data from these sensors are used for environmental perception [1].

In environmental perception, road users and other important objects (e.g. traffic signs, road boundaries) are observed and interpreted. The locations of these road users and other objects are required for the intelligent vehicle to be able to act in the world. For instance, the vehicle should react on crossing pedestrians or plan a safe route through other traffic participants. From now on, all road users and other relevant objects on the road are simply referred to as *objects*. To detect all objects in the environment, the vehicle utilizes machine learning techniques to classify and localize all objects in raw sensor data (point clouds, images, etc.). This process is called *object detection*.

Using the detections from the *object detector* in consecutive frames, the state of all (moving) objects can be estimated using estimation and tracking techniques. This is the responsibility of the *object tracker*, since it estimates the state of these objects and the number of objects in the environment. The object's state consists of the spatial location and kinematic state of the object. The kinematic state describes the motion parameters of the object, for instance linear and angular velocity and accelerations. Furthermore, another important state parameter is the shape extent, which describes the object's dimensions.

In the *motion planning* module, the motion for every object is predicted to anticipate on potential collisions. For that, the motion planner is using the location and velocity estimates from a state estimator, i.e. object tracker. The motion planner plans a sequence of actions to move from A to B on the road, while avoiding collisions and complying with traffic rules.

Now that the main elements of the perception pipeline have been mentioned, the next section will deal with the importance of sensor data fusion for intelligent vehicles. Furthermore, the prerequisites that are needed to perform sensor data fusion are discussed as well as the main challenges. The last two sections in this chapter are devoted to the outline of the thesis and to the contributions of this work.

1

1.1. SENSOR DATA FUSION

As mentioned in the previous section, intelligent vehicles are equipped with a sensor setup consisting of one or more lidars, cameras and radars. Intelligent vehicles use multiple sensing modalities for environmental perception. One important reason is to improve the estimates of perception system, as there is more data available to estimate the state. Each sensing modality has its strengths and weaknesses. For instance, for a radar the Doppler velocity can be used as a measurements, however the data is not as dense as a lidar sensor. Furthermore, a perception system consisting of a single sensor has a single point of failure. Hardware failures or other environmental conditions like adverse weather conditions could deteriorate the sensor data, resulting in that the perception system fails. A failing perception system means that the intelligent vehicle cannot guarantee safety of the passengers. Adding a second sensing modality, the perception system is more robust and it is better equipped to deal with these circumstances.

To fuse data from multiple sensors, a central place should be assigned where the fusion takes place. Two prerequisites should be met in order to be able to fuse the data from multiple sources. The first prerequisite is that the sensors should be intrinsically and extrinsically calibrated. This way the sensor data can be expressed in a common coordinate frame. The second prerequisite is that the measurements or processed sensor data should be communicated along with time stamps for temporal integration.

When the prerequisites are met, the sensor data can be fused. The system should deal with several challenges, for instance:

- *Data association*: Identifying which observations originated from the same object.
- *Asynchronous sensors*: Each sensor has a certain update rate, the frame rate. Furthermore, each sensor sends its data to the fusion center with a delay. Because the raw sensor data usually needs to be processed and transmitted to the processing unit, a delay between observing something and receiving the data in the fusion center arise. Due to the fact hat each sensor has a different frame rate and delay, measurements arrive asynchronous in the fusion center.
- *Imperfect measurements*: The ideal sensor does not exist. Sensor data contains noisy measurements, false positives and false negatives and it depends on many factors like environmental conditions and the environment (urban, highway, rural, etc.). This means that the observation noise, probability of detection and false alarms ratio are not constant, but varying with time and state dependent, e.g. are depending on distance of the object.

To ensure safety in intelligent vehicles, a robust perception system is required that can deal with adverse weather conditions. Each of the three sensing modalities (lidar, radar and vision) use a different part of electromagnetic spectrum, therefore these sensing modalities are not affected by adverse weather conditions (rain, fog, snow, etc.) in the same way. Some sensing modalities are more sensitive to different weather and illumination conditions than others. Detection of road users using vision is sensitive to illumination and weather conditions [2]. For instance, adverse weather conditions like haze and fog degrade images due to atmospheric absorption and scattering [3, 4]. A methodology to benchmark imaging sensors was proposed in [5], and the authors show

that cameras lose contrast in fog due to air-light and attenuation. Also, laser scanners perform poorly in dense fog conditions and the maximum viewing distance is reduced to a fraction of the clear-weather viewing distance[6]. Benchmarking of lidar sensors in snow and icy conditions was performed in [7]. In turbulent snow tests, the viewing distances of the lidars were shortened [8].

To deal with adverse weather conditions, one option is to incorporate all adverse weather conditions in annotated datasets, such that the *object detectors* could be trained to recognize object appearances in adverse weather conditions. State of the art datasets (e.g. [9–11]) cover a variety dynamic objects in various weather and illumination conditions at different geographical locations with sensor data from various sensing modalities used in automotive; lidar, camera and radar. Furthermore, the dataset of [12] contains data from a variety of harsh weather and illuminations conditions such as heavy fog, heavy snow, and severe rain. As the training data in harsh conditions such as dense fog and snow is rare, the authors only train their object detector in clear conditions and test on adverse weather data. Training a state of the art detector requires a large annotated dataset, therefore it is challenging to obtain robust detections by incorporating adverse weather conditions in a training dataset. Apart from recording a diverse dataset, manually labeling ground truth is a time consuming task.

Alternatively, a classifier could be trained to recognize the adverse environmental conditions and act upon these recognized conditions. For instance, in case of bad sensing conditions for the camera sensor (e.g. direct sunlight or foggy conditions), the camera observations could be excluded in the sensor data fusion. Another option is to use techniques to improve the raw sensor data in software. For instance, image enhancement techniques in camera images for bad weather conditions [3, 4, 13].

Furthermore, an alternative approach would be to design a perception system that is able to estimate the quality of the data arriving in the sensor data fusion module. The estimated quality of the inputs to fusion module can be used to monitor in real time the quality of perception system (e.g. sensors). For instance, if one sensor always detects an object and the other sensor does not, it could mean that one sensor is failing or is affected. Furthermore by estimating the sensing quality, the algorithm could benefit from this information by means of a different fusion strategy, e.g. relying fully on the sensor with a high detection probability. Monitoring the health of the sensor data is of importance, therefore identifying the quality metrics is of interest. The components that use the output of the perception system should be able to know how reliable the output is. Therefore the reliability of all the perception sensors needs to be assessed in real time to inform the users about the reliability. Some of the metrics could be estimates of the observation noise, clutter, detection probability and frame rate of the individual sensors.

1.2. RESEARCH QUESTIONS

The aim of the thesis is to develop methods and algorithms for a robust perception system that is able to deal with adverse weather conditions.

In order to create a robust perception system, the sensing modalities and the sensors need to be selected that is be able to fulfill the requirements of the automated driving application. To evaluate if a sensor setup is able to achieve these requirements, a systematic evaluation is required to determine the tracking performance limits in the de-

sign phase. How to compute the tracking performance limit for a certain sensor setup? In addition, to develop a redundant system, the performance of the perception system should be known in case of sensor failures. What are the predicted tracking performance limits in case of a single sensor failing? A systematic approach is required to estimate the tracking performance limits for a given sensor setup, which is the subject of chapter 2.

After selecting the sensing modalities and sensors in the early design phase, the sensors can be mounted on the mobile platform. The sensors need to be intrinsically and extrinsically calibrated to be able to use the sensor data and to express the sensor data in a common reference frame. Therefore, extrinsic sensor calibration of the sensor setup is the main focus of chapter 3. This chapter addresses the following research questions. What are the different optimization configuration to jointly extrinsically calibrate a multi-sensor setup? Which of the optimization configuration performs best? The sensor poses should not only be estimated with respect to other sensors, the sensor poses should also be known with respect to the (robot) vehicle's geometrical model, which is a geometrical model that describes the dimensions, joints and parts. To avoid collisions with other road users, the automated vehicle requires knowledge where other road users are located with respect to its own geometrical shape. To express the estimated locations of road users in a body reference frame, the sensor poses should be known with respect to a body reference frame of the vehicle. This way the automated vehicle can plan a route to overtake a cyclist without hitting the cyclist with, for instance, the side mirror. Chapter 3 also answers the questions: What are the requirements to calibrate the ego sensors with respect to the robot's body reference frame? What is the best method to determine the pose of the body reference frame?

In adverse weather conditions, detection and tracking of objects in more challenging. Chapter 4 will address the subject of object tracking in adverse weather conditions. For instance, the camera detector has more difficulties in detecting objects far away due to a decreased visibility in fog, which leads to more false negatives. The increase of false negatives impacts the tracking performance if the detection probabilities are considered fixed, therefore the detection probabilities need to be continuously estimated. How can the detection probabilities be estimated in adverse weather and illumination conditions? How can the estimated detection probabilities be used in the tracking filter to mitigate the effect of degraded detection performance in adverse weather and illumination conditions? Chapter 4 presents a tracking filter that is able to monitor the probability of detection of the sensors, and the estimated detection probability is used within the tracking filter to realize improved tracking performance.

Finally, chapter 5 will summarize the findings of this thesis. Furthermore, it will also discuss open issues and directions for future research.

1.3. CONTRIBUTIONS

The main contributions of the thesis are:

- A systematic approach to evaluate tracking performance limits for different combinations of lidar, camera and radar sensors. In an early design phase, this approach can be used to select sensing modalities and sensors for automated driving applications.
- An open-source extrinsic calibration tool to jointly calibrate sensor setups consisting of lidars, radars and cameras. This tool consists of three optimization configurations to jointly calibrate a sensor setup of lidars, radars and cameras. Furthermore, this thesis proposes two methods to estimate the pose of the body reference frame of the robot, which is needed to calibrate the sensors with respect to the vehicle.
- A tracking filter is proposed that is able to deal with affected detection probabilities in adverse weather conditions. The multi-modal bootstrap filter that online estimates the probability of detection of each sensor and the filter uses these detection probabilities to take into account a failing of affected sensor.

REFERENCES

[1] E. Marti, M. A. de Miguel, F. Garcia, and J. Perez, *A review of sensor technologies for perception in automated driving,* IEEE Intelligent Transportation Systems Magazine **11**, 94 (2019).

[2] S. Sivaraman and M. M. Trivedi, *Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis,* IEEE Transactions on Intelligent Transportation Systems **14**, 1773 (2013).

[3] K. He, J. Sun, and X. Tang, *Single image haze removal using dark channel prior,* IEEE Transactions on Pattern Analysis and Machine Intelligence **33**, 2341 (2010).

[4] R. T. Tan, *Visibility in bad weather from a single image,* in *2008 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2008) pp. 1–8.

[5] M. Bijelic, T. Gruber, and W. Ritter, *Benchmarking image sensors under adverse weather conditions for autonomous driving,* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2018) pp. 1773–1779.

[6] M. Bijelic, T. Gruber, and W. Ritter, *A benchmark for lidar sensors in fog: Is detection breaking down?* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2018) pp. 760–767.

[7] M. Kutila, P. Pyykönen, M. Jokela, T. Gruber, M. Bijelic, and W. Ritter, *Benchmarking automotive lidar performance in arctic conditions,* in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2020) pp. 1–8.

[8] M. Jokela, M. Kutila, and P. Pyykönen, *Testing and validation of automotive point-cloud sensors in adverse weather conditions,* Applied Sciences **9**, 2341 (2019).

[9] A. Geiger, P. Lenz, and R. Urtasun, *Are we ready for autonomous driving? The KITTI vision benchmark suite,* in *2012 IEEE conference on computer vision and pattern recognition* (IEEE, 2012) pp. 3354–3361.

[10] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, *nuscenes: A multimodal dataset for autonomous driving,* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020) pp. 11621–11631.

[11] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, *Eurocity persons: A novel benchmark for person detection in traffic scenes,* IEEE transactions on pattern analysis and machine intelligence **41**, 1844 (2019).

[12] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, *Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather,* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020) pp. 11682–11692.

[13] S. G. Narasimhan and S. K. Nayar, *Contrast restoration of weather degraded images,* IEEE Transactions on Pattern Analysis and Machine Intelligence **25**, 713 (2003).

2

SENSOR SELECTION

This chapter presents an systematic approach to evaluate the tracking performance limits for different sensing modalities (lidar, radar and vision) and for combination of these sensing modalities. For that, the Cramér-Rao lower bound (CRLB) is used to predict the tracking performance limits for state of the art sensors such as the Continental ARS408, Velodyne HDL-64E lidar and monocular/stereo camera. The performance is evaluated by computing the theoretical CRLB in urban and highway environments. Based on numerical study in an urban environment, a CRLB σ_x and a CRLB σ_y of less than 0.1 *m was found for all sensor configuration consisting out of two sensors, within an observation time of* 0.5 *seconds. In highway environments, the best performance can be achieved by sensor data fusion of radar and lidar.*

This chapter is based on the paper [1].

2.1. INTRODUCTION

Intelligent vehicles are equipped with different sensing modalities such as lidar, radar and vision to detect and track road users such as cars, cyclists and pedestrians. Each of the sensing modalities has its advantages and disadvantages. A systematic evaluation is required to select the best sensor(s) for an automated driving application. Therefore, a design tool is needed in an early stage in the design process to predict the tracking performance for multi-sensor systems. Since state of the art sensors are costly, the cheapest set of sensors could be selected that meets the system requirements. In this way the sensing modality, sensor type (camera A vs camera B) and sensor placement can be selected to achieve the required tracking performance. This is achieved by systematic evaluation of the output of the multi-sensor perception system.

There are three challenges to objectively compare the tracking performance of different sensors. The first challenge is data heterogeneity; each sensor has a specific observation model with a number of observable parameters in a coordinate frame (pixels, polar, planar, etc.). This means that the measurement uncertainty is defined in different coordinate frames and units. Furthermore, intelligent vehicle systems do not rely on a

single sensor observation, but instead the systems will react on tracked objects. Therefore, the second challenge is to compare the tracking performance for sensors with different sampling rates that results in that two sensors might have an unequal number of measurement updates for a track within a period of time (tracking/observation time). The third challenge is that the measurement uncertainty often depends on the state of the object. These three challenges result in that there is no clear and objective answer which sensor or sensor set is best for a particular automated driving application (e.g. automated people mover, urban scenarios, highway scenarios, etc.).

Based on these challenges, the following research questions are defined and addressed in this chapter. What are the predicted tracking performance limits for each of the different sensing modalities (radar, vision and lidar)? What are the predicted tracking performance limits of multi-sensor data fusion systems?

2.2. RELATED WORK

The Cramér-Rao lower bound (CRLB) provides a lower bound on the variance of an estimator. It can be used to estimate the tracking performance limits of estimators for different levels of probability of detections, false alarm rates, measurement noise uncertainty and process noise [2, 3]. Furthermore, the CRLB can be used for optimal sensor placement in (mobile) sensor networks [4, 5]. It can also be used to select a subset of sensors for optimizing tracking performance in sensor networks [6]. In addition, the CRLB can be used to quantify the performance of an estimator, because if an estimator extracts all information from the data the state covariance matrix equals the inverse of the Fisher Information Matrix (FIM) [7]. For instance, Blanc et al. [8] use the CRLB to estimate the tracking performance of estimators that fuse the data of a radar and lidar by assuming that the sampling rates are identical.

Previous work either present object tracking approaches using a single sensing modality or present approaches using sensor data fusion of different modalities. For instance, stereo vision is used in [9] for pedestrian tracking and prediction. Pedestrian tracking using radar and lidar is proposed in [10] and [11], respectively. Cho et al. [12] proposed a multi-sensor system using lidar, radar and vision. However, there is not a clear answer which sensing modality or sensor set fulfills best the perception system's requirements. Therefore, this thesis proposes to use the CRLB to select sensing modalities for automated driving by evaluating the complete perception system instead of selecting based on the (individual) sensor specification sheets.

2.2.1. CONTRIBUTIONS

The main contribution of this chapter is twofold. First, this chapter proposes a design tool to systematically evaluate tracking performance limits for single sensor and multi-sensor perception systems in order to select sensing modalities, sensor type and sensor placement for automated driving. Second, the tracking performance limits are analysed for a state of the art lidar, radar and vision sensors as well as combinations of these sensors.

2.3. PROPOSED APPROACH

In this work, the tracking performance limits are analyzed for an intelligent vehicle. The assumption is made that there are optimal atmospheric and light conditions. Furthermore, the measurement noise covariance matrices are mainly derived from documented sensor specification sheets.

It is considered that a static point object, being a passenger car or pedestrian, suddenly appears in front of the vehicle. This object is modeled as a point object. The thesis assumes that the object is tracked for $0.5s$. Subsequently, the effect of shorter or longer observation time is evaluated in section 2.4.5. The uncertainty of the ego motion of the car is ignored, because the aim is to compare of different sensor configurations. Furthermore, it is assumed that there is no measurement origin uncertainty, which means that there is no doubt which measurement belongs to which object. In addition, it is considered that there are no false alarms and that the probability of detection, $p_D = 1$. However, in section 2.4.6 the influence of a probability of detection $p_D < 1$ and clutter is quantified.

2.3.1. MULTI-SENSOR DATA FUSION

In the centralized tracking architecture (type IV configuration [7]), all measurements from N sensors are processed centrally. In this case, the measurements are fused centrally resulting in optimal tracking performance [7, 13]. In reality, most sensors are asynchronous therefore sequential measurement updates take place. This means that the tracks are predicted using different intervals. The origin of the coordinate system is located at the center of the rear axle of the car at ground level. The x-axis points in the forward driving direction and the y-axis to the left side of the car.

2.3.2. NON-LINEAR FILTERING

For a non-linear filtering problem with additive Gaussian noise, the state transition model and observation model are equal to:

$$x_{k+1} = f(x_k) + v_k \tag{2.1}$$

$$z_{k+1} = h(x_{k+1}) + w_{k+1} \tag{2.2}$$

where v_k and w_{k+1} are the process noise $v_k \sim \mathcal{N}(0, Q_k)$ and observation noise $w_{k+1} \sim \mathcal{N}(0, R_{k+1})$, respectively. Furthermore, $\hat{x}_k$ denotes the state estimate that is obtained using a estimator (e.g. a non-linear Kalman filter), and $P_{k|k}$ its state covariance matrix. The subscript of $P_{n|m}$ denotes that the estimate of P is at time n, given the measurements integrated up to and including time m. In the next sections, the used motion model and observation models are addressed.

2.3.3. MOTION MODEL

The state vector $x_k = [x, y, v_x, v_y]^T$ consists of the positions (x and y) and the velocities in x and y (v_x and v_y). Instead of an non-linear motion model as defined in equation 2.1, an linear motion model is considered, namely the constant velocity model is used [14]:

$$x_{k+1} = F_k(\Delta t)x_k + w_k, w_k \sim \mathcal{N}(0, Q(\Delta t)) \tag{2.3}$$

where

$$F_k(\Delta t) = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4}$$

2

2.3.4. SENSOR MEASUREMENT MODELS

In this section, the sensor measurement models are elaborated on. It is assumed that the observation noise w is zero mean Gaussian noise. Furthermore, it is assumed that the measurement covariance matrix is a diagonal matrix, so all sensor errors are independent.

To deal with the non-linear observation models, the models are are linearized using the Jacobian matrix, which contains the first order partial derivatives of the non-linear model. The Jacobian matrix of the observation model is denoted as H_{k+1}.

LIDAR

The lidar observes the range (r), azimuth (α) and elevation angle (ϵ):

$$\begin{bmatrix} r \\ \alpha \\ \epsilon \end{bmatrix} = h_l(x_k) + w_l \tag{2.5}$$

RADAR

The observation vector for a radar consists of the range (r), azimuth (α) and Doppler velocity ($\dot{r}$):

$$\begin{bmatrix} r \\ \alpha \\ \dot{r} \end{bmatrix} = h_r(x_k) + w_r \tag{2.6}$$

The polar radar observations, range and azimuth, can be converted to Cartesian coordinates when the following is valid [7]:

$$\frac{r \cdot \sigma_\alpha^2}{\sigma_r} < 0.4 \tag{2.7}$$

This is called Converted Measurement Kalman Filter (CMKF).

VISION

In case of vision, the observation vector consists of the pixel location (u, v) for a monocular camera. For a stereo camera, the disparity (d) is the third observation:

$$\begin{bmatrix} u \\ v \\ d \end{bmatrix} = h_v(x_k) + w_v \tag{2.8}$$

2.3.5. CRAMÉR-RAO LOWER BOUND (CRLB)

The Cramér-Rao lower bound (CRLB) can be used as an analysis tool [15] in order to predict the performance of estimation problems. The expected value of the mean squared error is greater than or equal to the inverse of the Fisher Information Matrix (FIM). This means that the state covariance matrix $P_{k|k}$ has a lower bound (CRLB) [7]:

$$P_{k|k} \triangleq E\{[\hat{x}_k - x_k][\hat{x}_k - x_k]^T\} \geq J_k^{-1} \tag{2.9}$$

where $\hat{x}_k$ is the state estimate and x_k is the true state at time k. Furthermore, J_k represents the FIM, and its inverse is the CRLB [15]. For the linear state transition model and a non-linear observation model, the FIM is computed by [15]:

$$J_{k+1} = \left(Q_k + F_k P_{k|k} F_k^T\right)^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1} \tag{2.10}$$

where F_k is the state transition matrix and Q_k is the covariance matrix of the process noise, as defined in equation 2.3. Furthermore, H_{k+1} is the Jacobian of the observation models as defined by equations 2.5, 2.6 and 2.8.

When the state covariance matrix equals the inverse of the FIM, the estimator (tracking filter) has extracted all information from the data [7]. In this case, the estimator is called a (statistically) efficient estimator, however the CRLB might not be attainable.

2.3.6. INITIALIZATION

The state covariance matrix is initialised using the one point initialization [7]. A maximum velocity is defined for initialization of velocity component of the state covariance matrix and for the radar the Doppler velocity is utilized.

In the multi-sensor case, the initial state covariance matrix is estimated using the covariance intersection (CI) algorithm that fuses the state covariance matrices of all available sensors. This estimate is optimal if the cross-covariance between two estimates is unknown [16, 17]. Note that in the multi-sensor case, it is considered that the sensors start synchronously, i.e all sensors start at time $t = 0$. After that, the sensors run at their specific sampling rates. The covariance intersection method for N sensors provides a state covariance matrix (P_{CI}) that equals [18]:

$$P_{CI}^{-1} = \sum_{i=1}^{N} \omega_i P_i^{-1} \tag{2.11}$$

$$\omega = argmin(det(P_{CI})) \tag{2.12}$$

subject to $\sum_{i=1}^{N} \omega_i = 1$ and scalar parameter $\omega_i \in [0,1]$ for $i = 1, \ldots N$ sensors.

2.3.7. PERFORMANCE MEASURES

The tracking performance limit is quantified using CRLB. The first measure is the L_2 norm of the Cramér-Rao lower bound. The second performance measure is the σ for x, y, v_x and v_y in order to determine main differences between sensors. Finally, the filter convergence for a sensor or a combination of sensors is plotted.

2.3.8. IMPLEMENTATION DETAILS

In algorithm 1, the pseudo-code can be found for computing the CRLB for an object located at x, y for a observation time of t_{end}.

Algorithm 1 Compute Cramér-Rao lower bound

```
function COMPUTECRLB(x, y, t_end)
    P_0: initialize using available sensors
    t_i: Schedule all available sensors   ▷ To determine all prediction intervals Δt_i for equation 2.3.
    for t = 0 : Δt_i : t_end do
        Compute R_{k+1}                   ▷ State and sensor dependent
        Compute H_{k+1}                   ▷ State and sensor dependent
        Update J_{k+1}                    ▷ see equation 2.10
    end for
end function
```

2.4. SIMULATIONS

A virtual car is equipped with a lidar, radar and (stereo) camera. The lidar is mounted on top of the vehicle, the stereo camera in front of the windscreen and the radar in the bumper of the vehicle. The (stereo) camera has a pitch angle of 2° (downward).

2.4.1. PROCESS NOISE VARIANCE

In urban environments, the vehicle has to deal with vulnerable road users such as pedestrians and cyclists.

In order to compute the performance limits in an urban environment, the object of interest is considered to be an adult pedestrian. In the work of [9], the process noise parameter q is optimized for the constant velocity (white noise acceleration) model, where q is defined as $q = \sigma_v^2$ with the optimal $\sigma_v = 0.77$ was found. In highway scenarios, the vehicle only has to deal with cars, therefore a normal passenger car is considered and the process model of [19] is used with $S_x = (0.629 m/s^2)^2\ s^{-1}$ and $S_y = (0.472 m/s^2)^2\ s^{-1}$

$$Q = \begin{bmatrix} \frac{\Delta t^3}{3} S_x & 0 & \frac{\Delta t^2}{2} S_x & 0 \\ 0 & \frac{\Delta t^3}{3} S_y & 0 & \frac{\Delta t^2}{2} S_y \\ \frac{\Delta t^2}{2} S_x & 0 & \Delta t S_x & 0 \\ 0 & \frac{\Delta t^2}{2} S_y & 0 & \Delta t S_y \end{bmatrix}. \tag{2.13}$$

2.4.2. MEASUREMENT NOISE VARIANCE

LIDAR

According to the user manual [20] of the Velodyne HDL-64E LiDAR, the standard deviation in range equals $\sigma_r = 2$ cm. It is assumed that it is independent of the range of the object. For σ_α and σ_ϵ, it is assumed that these are equal to the maximum expected error (resolution/2). Moreover, σ_α depends on the frame rate and σ_ϵ depends on the elevation angle. The frame rate of the lidar is set to 10Hz. Furthermore, it is assumed that there is no minimum number of lidar hits required to detect objects.

RADAR

The measurement noise variance of the Continental ARS408 radar is based on the Continental specification sheet [21]. This radar consists of a far range radar sensor and a near range radar sensor. The accuracy in azimuth is interpolated using the provided azimuth accuracy values. Furthermore, it is assumed that these values equal 1σ.

VISION

The camera sensor has a 8mm lens, resolution of 1936×1216 pixels with a pixel size of $5.86\ \mu m$. The baseline for the stereo camera is equal to $30cm$. The standard deviation of the measurement noise for stereo vision equals $\sigma_u = 6.15$ pixels and $\sigma_d = 0.32$ pixels according to [9]. Furthermore, it is assumed for simplicity that $\sigma_v = \sigma_u$. In all cases, limits on minimum and maximum distance for object detection are ignored.

2.4.3. URBAN ENVIRONMENT

In urban environment scenarios, a maximum range of $50m$ is considered to compute CRLB. For initialization (see section 2.3.6), a walking speed of 5km/h is used. Figure 2.1 shows the single sensor tracking performance limits for stereo, radar and lidar.

Stereo vision provides the best the σ_x in short range ($\approx$ 15 m). Since σ_x scales quadratically with distance, at larger distances the performance decreases. Due to very accurate range observations ($\sigma_R = 2$ cm), the lidar sensor performs best in all other ranges. In urban environment, the lidar most accurately estimates the lateral position (y) of an object. Within field of view (FOV) of the radar, the radar is the best sensor for the velocity in x-direction (v_x). For the lateral velocity (v_y), the lidar is the most precise sensor as shown in Figure 2.1. Moreover, the radar is not performing well in both the lateral position as the lateral velocity.

Figure 2.2 shows the CRLB performance for multi-sensor multi-modal systems. The top two rows show the configuration with radar and monocular camera and radar and stereo camera. The multi-modal performance is significantly better than using a single sensor (modality). This is because radar and camera are complementary to each other. Furthermore, the configuration with stereo vision is slightly better in the short range as indicated in Figure 2.2. Adding a stereo camera instead of a monocular camera to the lidar, only (slightly) improves the performance limits in the short range (see row three and four). The fifth row visualises the performance of radar and lidar, whereas the last row shows the performance of lidar, radar and stereo vision. It can be seen that the stereo vision hardly improves the tracking performance compared to the lidar + radar when the σ_x, σ_y, σ_{vx} and σ_{vy} of the fifth row and sixth row are compared. However, any configuration with lidar provides accurate tracking performance according to the norm of the CRLB. Moreover, using monocular or stereo vision the sensing region is reduced. This relates to the resolution of the used camera(s) and the selected camera lens(es) (focal length). The latter is selected based on a compromise between accuracy and FOV.

However, the main finding of Figure 2.2 is that with any configuration of multi-sensor data fusion the σ_x and σ_y is within 0.1 m in the tested FOV. Compared to the single sensor performance (Figure 2.1), all multi-sensor systems provide a significant improvement. In the case that an accurate velocity estimate is required, the best sensor is the radar as it can directly observe the Doppler velocity.

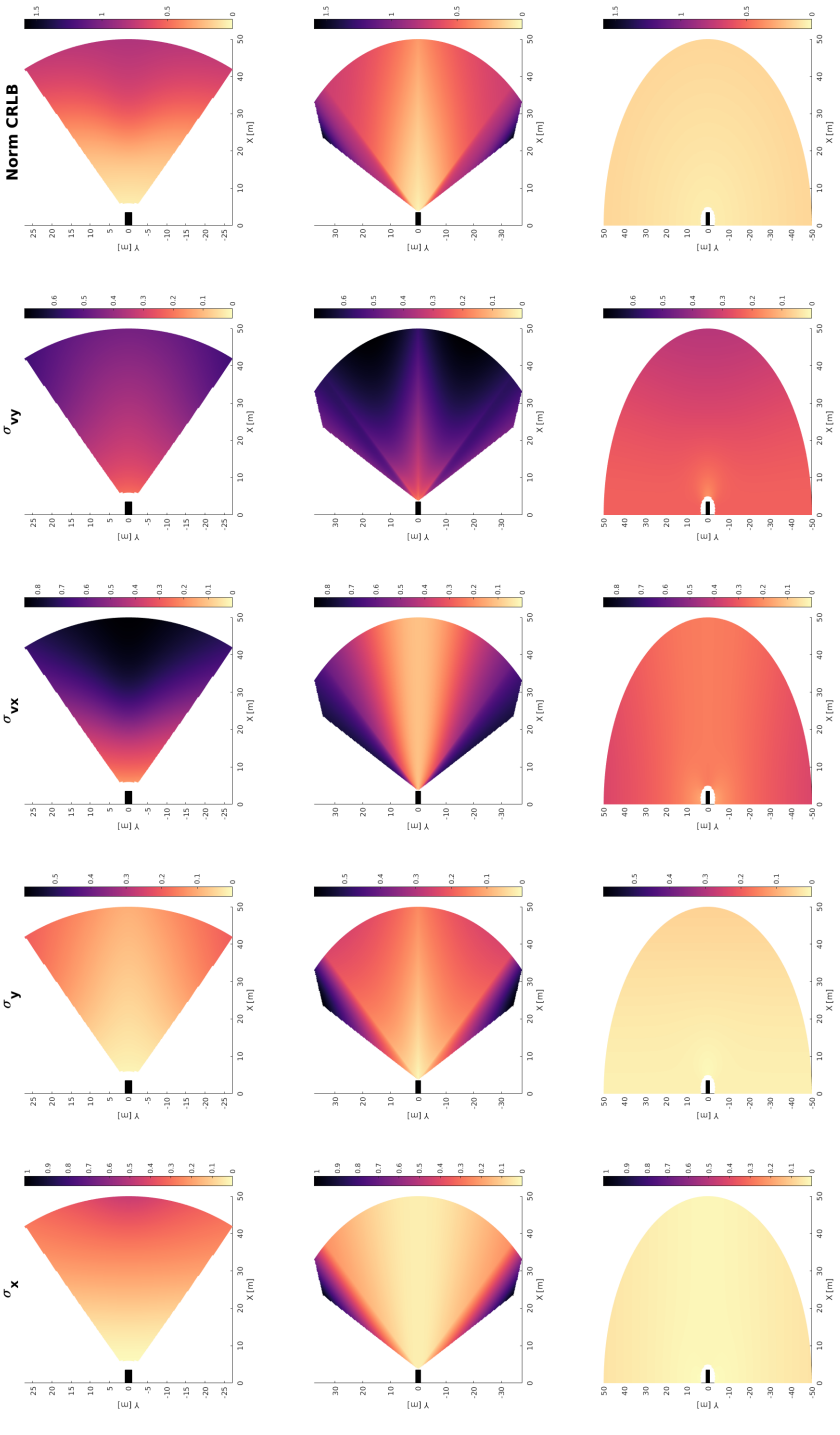


Figure 2.1: Cramér-Rao lower bound results for single sensor object tracking with, from top to bottom, stereo camera (S), radar (R) (near range sensor) and lidar (L), respectively. From left to right the standard deviation in x, y, v_x and v_y is plotted. The last column respresents the L_2 norm of the CRLB. The units in the colourbar are [m] for σ_x and σ_y, and [m/s] for σ_{vx} and σ_{vy}. These results are obtained after a tracking (observation) time of 0.5 seconds. The lower the value in the plot the better the predicted tracking performance limit. In Appendix A, a larger version of this figure can be found.

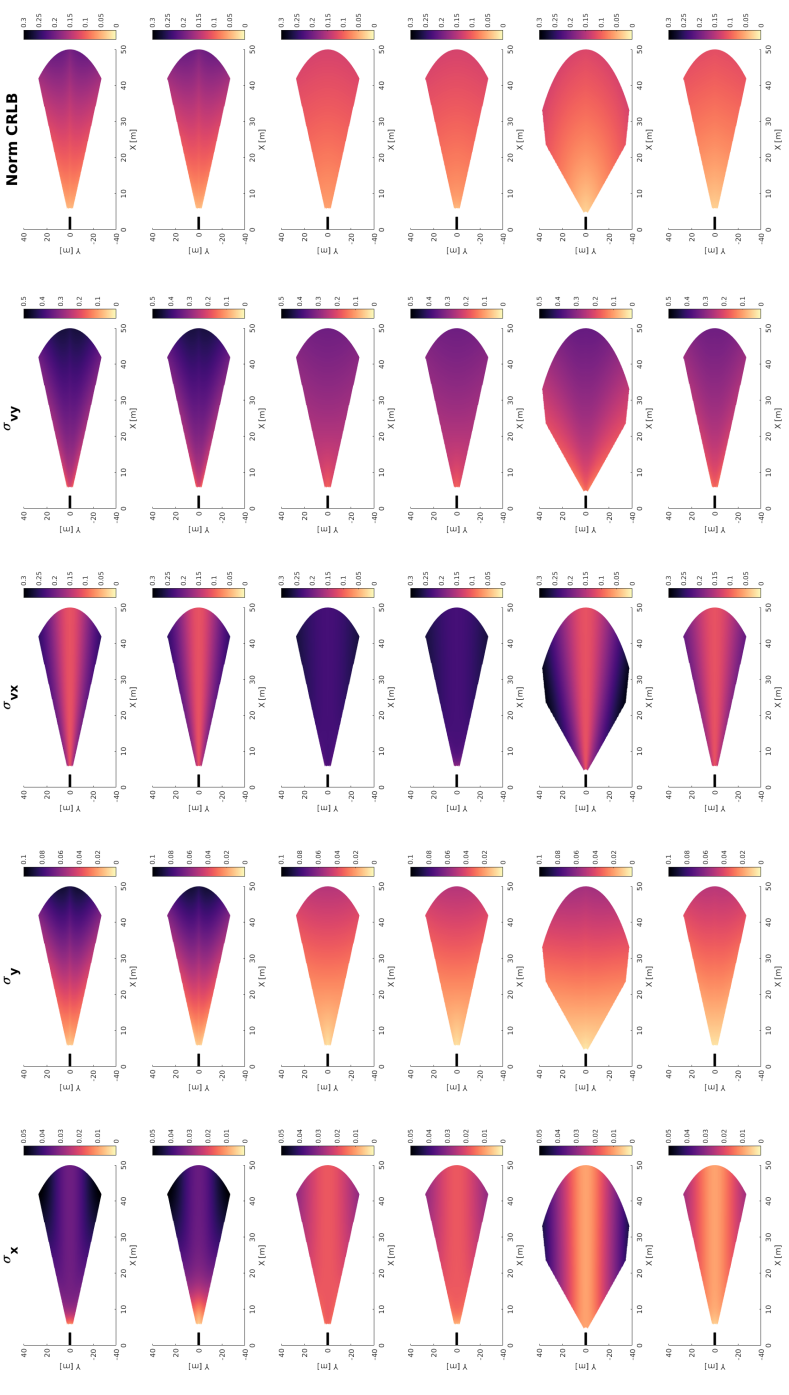


Figure 2.2: Cramér-Rao lower bound results for multi-sensor object tracking with from top to bottom radar + monocular camera (R+C), radar + stereo camera (R+S), lidar + monocular camera (L+C), lidar + stereo camera (L+S), lidar + radar (L+R) and lidar + radar + stereo camera (L+R+S). From left to right the standard deviation in x, y, v_x and v_y is plotted. The last column respresents the L_2 norm of the CRLB. These results are obtained after a tracking (observation) time of 0.5 seconds. In Appendix A, a larger version of this figure can be found.

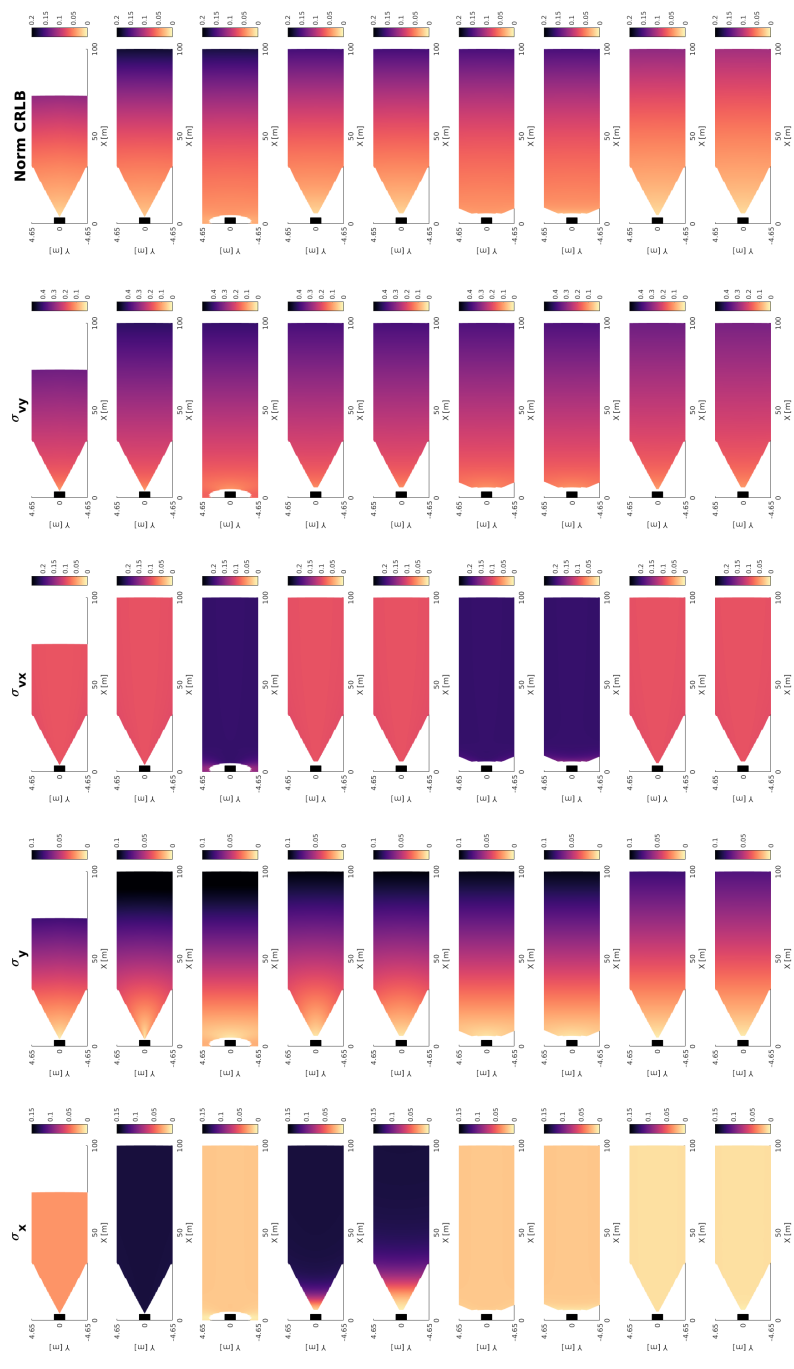


Figure 2.3: The tracking performance limits for different configurations of lidar, radar and vision are visualised. From top to bottom: radar near range + radar far range ($R_{NR} + R_{FR}$), radar far range (R_{FR}), lidar (L), radar far range + monocular camera ($R_{FR} + C$), radar far range + stereo vision ($R_{FR} + S$), lidar + camera ($L + C$), lidar + stereo ($L + S$), lidar + radar far range ($L + R_{FR}$), lidar + radar far range + stereo vision ($L + R_{FR} + S$). These results are obtained after a tracking (observation) time of 0.5 seconds. In Appendix A, a larger version of this figure can be found.

2.4.4. HIGHWAY ENVIRONMENT

In the highway scenario, the vehicle is considered driving on the central lane of a three lane highway and therefore the figures only visualize this sensing region.

For the observed vehicle, a maximum velocity of 100 km/h is considered for initialization for σ_{vx} and σ_{vy} (see section 2.3.6).

Figure 2.3 shows the performance limits for configurations of lidar, radar and vision. Since the distance accuracy scales quadratically with distance, object tracking using only stereo vision is not suitable in the far range. Therefore, the configuration with only stereo vision is not plotted in Figure 2.3.

According to Figure 2.3, the best sensor for σ_x in a highway environment is the lidar. Also the radar is a good option if both far range and near range are combined. Also, for lateral accuracy the radar is a good option. Only the combination of lidar + radar and lidar + radar + vision are performing significantly better. Moreover, the radar sensor is achieving the best results in estimating the σ_{vx}, because it can directly observe the Doppler velocity. The best performance for σ_{vy} can be reached by a combination of radar and lidar. Based on the norm of the CRLB, the best performance for highways can be achieved by sensor data fusion of lidar and radar. Although a stereo camera would improve the results, the performance increase is minimal according the norm of the CRLB.

2.4.5. TRACKING FILTER CONVERGENCE

Figure 2.4 shows the tracking filter convergence for different sensing modalities. In this case, a highway scenario is considered. For two object's distances (20 m and 50 m), the convergence of the tracking filter is plotted as a function of tracking (observation) time.

In all case the norm of the CRLB quickly converges in the first 0.5 second, which was therefore also considered in the previous sections. In most cases, the σ_x, σ_y, σ_{vx} and σ_{vy} have converged in the first 0.5 seconds, whereas in some cases it already converges within 0.2 seconds. For example, in the longitudinal direction (x and v_x) for radar and lidar. In close range, stereo vision quickly converges to the steady state CRLB, however at larger distances the filter needs more time to converge. Especially, the velocities need more time to converge for an object located at 50 m distance.

Since the radar is able to directly observe the object's velocity, the tracking filter is initialized and updated with the Doppler velocity. This results in that longitudinal velocity (v_x) quickly converges if a radar sensor is used. This means that the intelligent vehicle is able to quickly respond to unexpected events, which is important for smooth and responsive control.

2.4.6. IMPERFECT DETECTION AND CLUTTER

In reality, the probability of detection is smaller than one. In addition, clutter measurements are present. The effect of these two factors will be addressed in this section. For predicting the tracking performance of a probability data association (PDA) filter, the information reduction factor (IRF) can be used [7]. The IRF factor quantifies how much the innovation S_k is decreased due to imperfect detection and clutter.

Updating equation 2.10 to take into account the IRF [7] results in

$$J_{k+1} = \left(Q_k + F_k P_{k|k} F_k^T\right)^{-1} + q_2 H_{k+1}^T R_{k+1}^{-1} H_{k+1}. \tag{2.14}$$

where q_2 is the IRF, and it depends on clutter density λ, the probability of detection p_D and the volume V of the innovation matrix S_k. For the exact graph for q_2, the reader is referred to [7]. In this numerical study, the thesis will look into a sensor data fusion of radar (range and azimuth) and camera (pixel location). Furthermore, the curves from [7, 22] are interpolated to acquire the IRF which depends on the p_D and the λ.

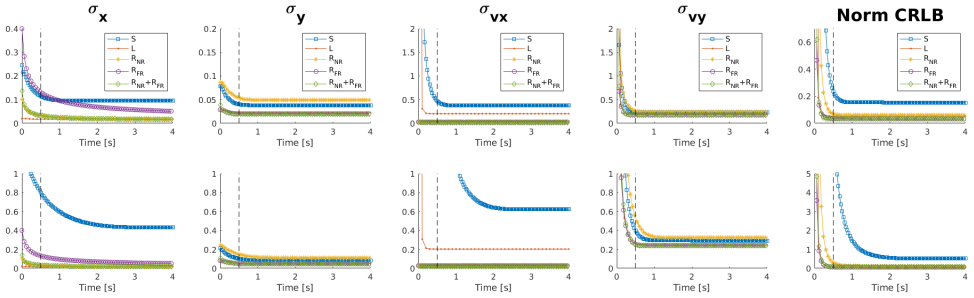


Figure 2.4: Tracking filter convergence for lidar, radar and vision with on the horizontal axis the tracking (observation) time in seconds. From left to right: σ_x, σ_y, σ_{vx}, σ_{vy}, and the L_2 norm of the CRLB. From top to bottom, an object located at 20 m and 50 m. The following sensors are considered: stereo camera (S), lidar (L), radar near range (R_{NR}), radar far range (R_{FR}) and radar near range + radar far range ($R_{NR} + R_{FR}$).

According to Zhang et al. [23], the state of the art pedestrian detectors have approximately 1 false positive per image at miss rate of 0.1. This means that at a probability of detection of 0.9, the clutter density is 1 per image for a 640×480 image. The clutter density per unit volume λ can be computed. Furthermore, for the radar a probability of detection of $p_D = 0.9$ can be achieved at a false alarm probability of 10^{-6} ([24, 25]).

Figure 2.5 shows the norm of the CRLB in an urban environment for the perfect case and for the imperfect detection and clutter case. It can be seen that the predicted tracking performance degrades. The mean difference of the norm of the CRLB equals 4.6% with respect to the perfect case.

In automated driving, it is essential to robustly track objects in all conditions. Missing sensor detections can be caused by sensor failures or by environmental conditions such as rain, snow, fog, illumination, etc. Redundancy is crucial in order to safely drive in all conditions, so it is important in sensor selection for automated driving. The tracking performance reduction due to these type of circumstances can be estimated in advance using the proposed method. The influence of sensor failure can be investigated as well as the effect of environmental conditions (e.g. reduced visibility in fog). Therefore, a numerical study is performed to quantify the tracking performance in these conditions. Figure 2.6 visualizes the σ_x and σ_y in four different conditions: imperfect detection and clutter, adverse environmental conditions (camera $p_D = 0.3$), failing camera sensor ($p_D = 0$ at $t > 0$) and failing radar sensor ($p_D = 0$ at $t > 0$). Figure 2.6 shows that this tool can be used to decide if the tracking performance is satisfactory in cases of sensor failure or changing environmental conditions if the requirements of the perception system are given.

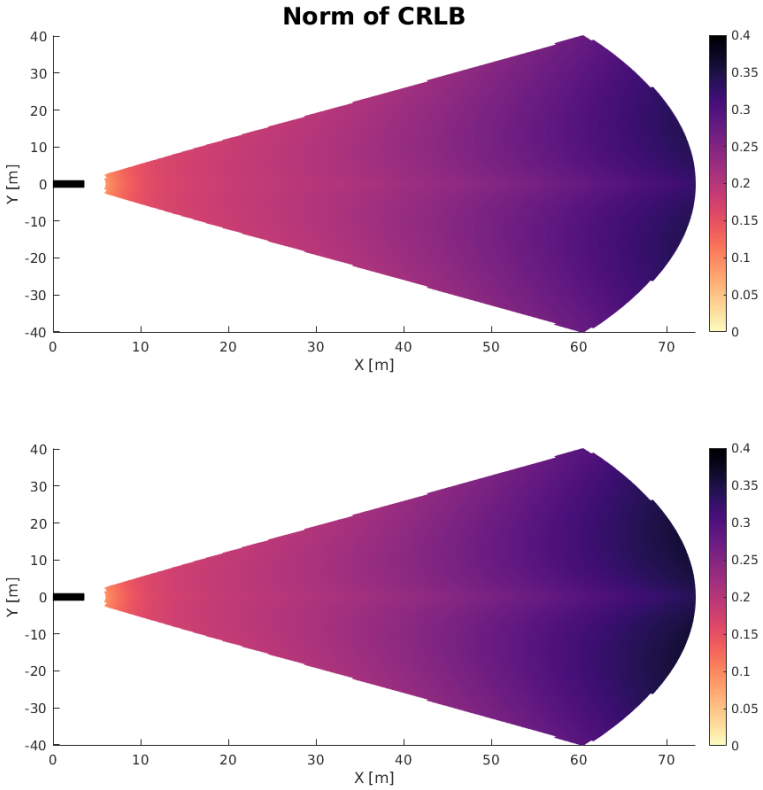


Figure 2.5: Effect of imperfect detection ($p_D < 1$) and clutter ($\lambda > 0$) for a sensor set that consists of a radar and a camera. The top figure shows the norm of the CRLB in optimal conditions, whereas the bottom figure shows the tracking performance in case of imperfect conditions.

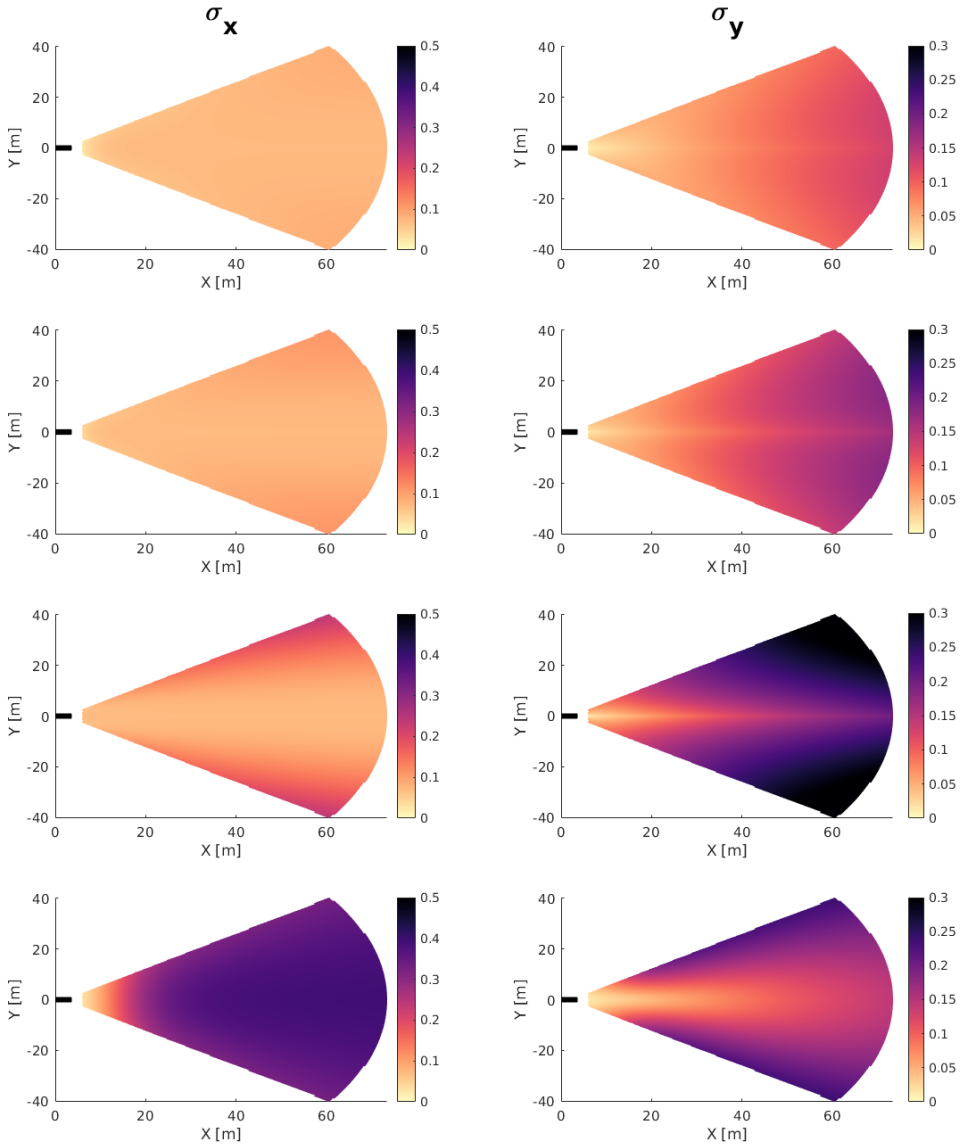


Figure 2.6: The effect of changing environmental conditions and sensor failure on a multi-sensor tracking filter for a sensor set that consists of a radar and a camera. From top to bottom: imperfect detection and clutter, adverse environmental conditions $p_D = 0.3$, failing camera sensor and failing radar sensor.

SENSOR PLACEMENT

This methodology can also be used for sensor placement. Consider that the aim is to position two radars in the bumper of the car. Figure 2.7 shows the norm of the CRLB for two radars which are separated 0.40 m apart and two radars which are separated 1.6 m apart (in y-direction).

Finally, the findings of the different numerical studies are summarized. In highway conditions (see sections 2.4.4 and 2.4.5), the radar is an essential sensor for smooth and responsive control. In order to develop a cheap and redundant sensing system that works as well in an urban environment, a good choice according to section 2.4.3 would be to include a stereo camera. The tracking performance can be found in section 2.4.3 and section 2.4.4. In urban conditions, the CRLB tracking performance limit of σ_x and σ_y is < 0.1 m. Section 2.4.6 gives an example of the influence of imperfect detection and clutter on the tracking performance for a sensor set of radar and camera. It was shown that the performance is affected on average with $\approx 5\%$.

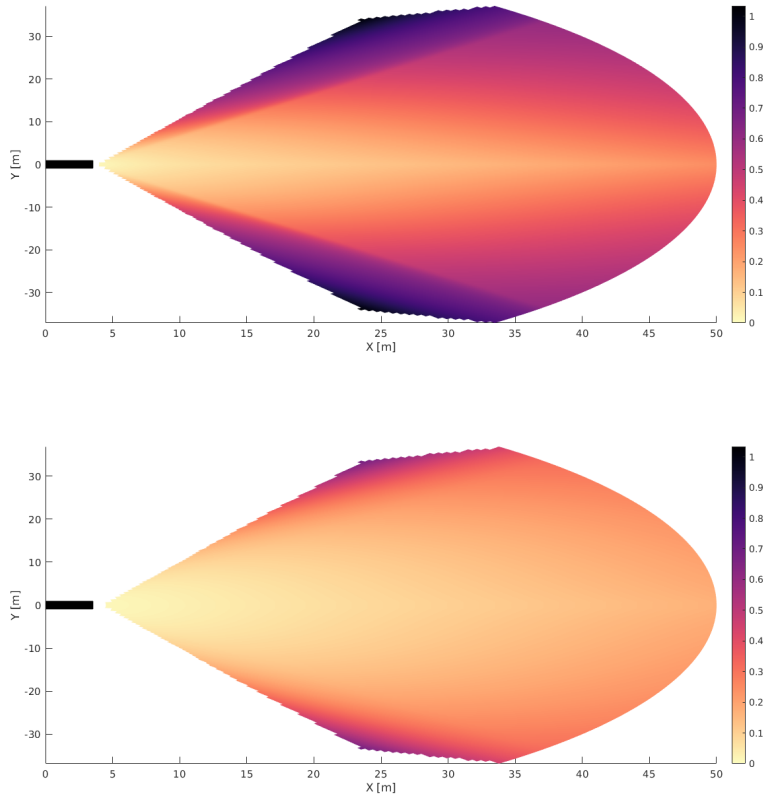


Figure 2.7: Difference between the norm of the CRLB for 2 different sensor positions. The figure shows the norm of the CRLB for 0.4 m and 1.6 m seperation between radars.

2.5. DISCUSSION

The proposed tool can be used to select the sensing modalities, sensor types and sensor placements in multi-sensor perception systems. In order to focus on comparing various sensing modalities and configurations, ego motion uncertainty is ignored. The CRLB provides a lower bound on the performance of an object tracker that might not be attainable. In reality, there are also errors due to measurement origin uncertainty, sensor calibration and sensor placement.

Furthermore, the specification sheets of radar and lidar have been used to estimate sensor noise characteristics. However in an early phase in the design process, experiments cannot be conducted to retrieve the sensor uncertainties because that is too expensive (purchasing sensors and invested time).

Atmospheric conditions and occlusions affect the performance of localizing objects in the surroundings of the car, however these can also be captured by simulating a time or state dependent p_D.

2.6. CONCLUSION

This thesis presents a tool to predict tracking performance limits for multi-sensor configurations in an early phase in the design process. Object tracking performance limits are computed for multi-sensor configuration of state of the art sensors (lidar, radar and vision). The assumption is made that the probability of detection $p_D = 1$ and that there are no false alarms. Subsequently, the influence of imperfect detection ($p_D < 1$) and clutter have been quantified. Furthermore, it was shown that the tracking performance can be estimated in case of sensor failure or changing environmental conditions.

This approach can be used to study the tracking performance limits for various sensor configurations and sensor types. In addition, this approach can be used to analyze what the limits are in case of a sensor failure.

This chapter contains different numerical studies on state-of-the art automotive sensors to analyze the performance for tracking applications. In the close range stereo vision performs well, and lidar provides very accurate positional x and y estimates. In highway environments, the best performance can be achieved by sensor data fusion of radar and lidar. In urban environments, any two sensor combination provides a CRLB σ_x and a CRLB σ_y of less than 0.1 m within an observation time of 0.5 seconds.

REFERENCES

[1] J. Domhof, R. Happee, and P. Jonker, *Multi-sensor object tracking performance limits by the Cramér-Rao lower bound,* in *20th International Conference on Information Fusion (Fusion)* (IEEE, 2017) pp. 128–135.

[2] P. Tichavsky, C. H. Muravchik, and A. Nehorai, *Posterior Cramér-Rao bounds for discrete-time nonlinear filtering,* IEEE Transactions on Signal Processing **46**, 1386 (1998).

[3] X. Zhang, P. Willett, and Y. Bar-Shalom, *Dynamic Cramér-Rao bound for target tracking in clutter,* IEEE Transactions on Aerospace and Electronic Systems **41**, 1154 (2005).

[4] D. Moreno-Salinas, A. Pascoal, and J. Aranda, *Optimal sensor placement for acoustic underwater target positioning with range-only measurements,* IEEE Journal of Oceanic Engineering **41**, 620 (2016).

[5] S. Martínez and F. Bullo, *Optimal sensor placement and motion coordination for target tracking,* Automatica **42**, 661 (2006).

[6] L. Zuo, R. Niu, and P. K. Varshney, *A sensor selection approach for target tracking in sensor networks with quantized measurements,* in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008.* (IEEE, 2008) pp. 2521–2524.

[7] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and data fusion* (YBS publishing, 2011).

[8] C. Blanc, P. Checchin, S. Gidel, and L. Trassoudaine, *Data fusion performance evaluation for range measurements combined with cartesian ones for road obstacle tracking,* in *IEEE International Conference on Vehicular Electronics and Safety, 2007. ICVES.* (IEEE, 2007) pp. 1–6.

[9] N. Schneider and D. M. Gavrila, *Pedestrian path prediction with recursive Bayesian filters: A comparative study,* in *German Conference on Pattern Recognition* (Springer, 2013) pp. 174–183.

[10] M. Heuer, A. Al-Hamadi, A. Rain, and M. M. Meinecke, *Detection and tracking approach using an automotive radar to increase active pedestrian safety,* in *IEEE Intelligent Vehicles Symposium Proceedings* (2014) pp. 890–893.

[11] H. Wang, B. Wang, B. Liu, X. Meng, and G. Yang, *Pedestrian recognition and tracking using 3D lidar for autonomous vehicle,* Robotics and Autonomous Systems **88**, 71 (2017).

[12] H. Cho, Y.-W. Seo, B. V. Kumar, and R. R. Rajkumar, *A multi-sensor fusion system for moving object detection and tracking in urban driving environments,* in *IEEE International Conference on Robotics and Automation (ICRA), 2014* (IEEE, 2014) pp. 1836–1843.

[13] P. Lytrivis, A. Amditis, and G. Thomaidis, *Sensor data fusion in automotive applications* (INTECH Open Access Publisher, 2009).

[14] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software* (John Wiley & Sons, 2004).

[15] B. Ristic, S. Arulampalam, and N. J. Gordon, *Beyond the Kalman filter: Particle filters for tracking applications* (Artech house, 2004).

[16] S. Matzka and R. Altendorfer, *A comparison of track-to-track fusion algorithms for automotive sensor fusion,* in *Multisensor Fusion and Integration for Intelligent Systems* (Springer, 2009) pp. 69–81.

[17] J. Uhlmann, *Dynamic map building and localization for autonomous vehicles,* Ph. D (1995).

[18] M. Liggins II, D. Hall, and J. Llinas, *Handbook of multisensor data fusion: Theory and Practice, Second Edition* (CRC press, 2009).

[19] J. E. Stellet, F. Straub, J. Schumacher, W. Branz, and J. M. Zöllner, *Estimating the process noise variance for vehicle motion models,* in *2015 IEEE 18th International Conference on Intelligent Transportation Systems* (2015) pp. 1512–1519.

[20] *HDL-64E S2 and S2.1 User's Manual*, Velodyne (2017).

[21] *ARS 408-21 Premium, Long Range Radar Sensor 77 GHz*, Continental (2016).

[22] T. Fortmann, Y. Bar-Shalom, M. Scheffe, and S. Gelfand, *Detection thresholds for tracking in clutter–a connection between estimation and signal processing,* IEEE Transactions on Automatic Control **30**, 221 (1985).

[23] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, *How far are we from solving pedestrian detection?* in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).

[24] B. R. Mahafza, *Radar systems analysis and design using MATLAB* (CRC press, 2002).

[25] M. I. Skolnik, *Introduction to radar,* Radar Handbook **2** (1962).

3

EXTRINSIC SENSOR CALIBRATION

This chapter addresses joint extrinsic calibration of lidar, camera and radar sensors. To simplify calibration, the thesis proposes a single calibration target design for all three sensing modalities, and implements the approach in an open-source tool with bindings to Robot Operating System (ROS). The tool features three optimization configurations, namely using error terms for a minimal number of sensor pairs, or using terms for all sensor pairs in combination with loop closure constraints, or by adding terms for structure estimation in a probabilistic model. Apart from relative calibration where relative transformations between sensors are computed, this chapter also addresses absolute calibration that includes calibration with respect to the mobile robot's body. Two methods are compared to estimate the body reference frame using an external laser scanner, one based on markers and the other based on manual annotation of the laser scan. In the experiments, the three configurations for relative calibration are evaluated. The results show that using terms for all sensor pairs is most robust, especially for lidar to radar, when minimum five board locations are used. For absolute calibration, the median rotation error around the vertical axis reduces from 1° before calibration, to 0.33° using the markers and 0.02° with manual annotations.

This chapter is based on the paper [1].

3.1. INTRODUCTION

Nowadays, mobile robots have sensor setups consisting of multiple sensors for environmental perception. To increase robustness, these sensor setups consist of various sensing modalities such as lidars, cameras and radars [2, 3]. For effective sensor data fusion, a geometrical description is needed that describes the location and orientation of all the robot's sensors with respect to each other, and to its body. For that, all sensors need to be calibrated.

One can distinguish two types of calibration tasks, namely intrinsic calibration and extrinsic calibration. Intrinsic calibration involves estimating the internal parameters of the sensor. For a camera, this calibration procedure consists of estimating all entries

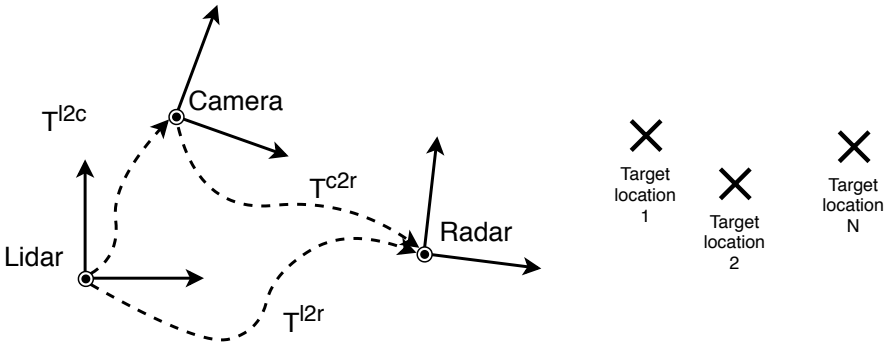
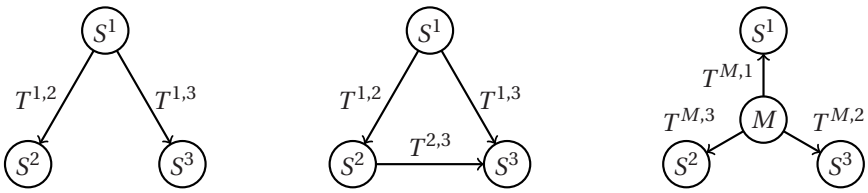


Figure 3.1: Schematic overview of an example sensor setup with three coordinate frames (lidar, camera and radar) with transformation matrices from one reference frame to another, e.g. *l2c* for lidar to camera. Joint multi-sensor calibration requires detections from multiple target locations which can be detected by all sensors simultaneously.



(a) MCPE (b) FCPE (c) PSE

Figure 3.2: Optimization configurations for joint calibration. The symbols S^i stand for sensor reference frames, and $T^{i,j}$ for coordinate frame transformations from sensor reference frame i to j. a Minimally connected pose estimation (MCPE) relies on a reference sensor S^1; b Fully connected pose estimation (FCPE) adds the loop constraint $T^{2,3} \cdot T^{1,2} = T^{1,3}$. c Pose and structure estimation (PSE) also estimates latent variables M that represent the true board locations (i.e. the structure).

of the camera projection matrix (focal length, skew parameter and principal point[4]) and the distortion coefficients of the lens. For a lidar, the intrinsic parameters are range offset, scale factor, vertical offset, elevation angle and azimuth angle [5]. Extrinsic calibration instead estimates the orientation and the position of the sensor (i.e. sensor pose) with respect to a frame of reference, which is also called pose estimation [6] and sensor registration [7].

Extrinsic calibration methods can further be split into two groups: target-less and target-based methods. Target-less methods (e.g. [8–10]) are potentially able to perform online calibration as these methods use natural features in the environment to calibrate the sensors. However, target-less calibration methods are challenging since these methods need to deal with asynchronous and heterogeneous sensors. Target-based methods instead use specifically designed physical calibration objects (i.e. targets) to obtain robust features. A typical example of a calibration target is the checkerboard pattern for intrinsic and extrinsic (stereo) calibration of cameras [4, 6, 11]. Since each sensing modality (lidar, camera and radar) works on a different wavelength and operating principles, it is challenging to find corresponding features across sensing modalities. Therefore, this

chapter focuses on target-based procedures to obtain accurate key points for all involved sensors at once. Multiple correspondences can be found by repositioning the calibration target at various locations in the overlapping Field of View (FOV) of the sensors.

While reasonable initial estimates of all sensor poses can be obtained from technical drawings of the robot (e.g. computer-aided design (CAD) models), an extrinsic calibration considers the sensor measurements to determine their actual poses. In this work, a rigid robot body is considered, which means that the transformations between the sensors and the body coordinate frame are constant (i.e. no relative movement). Extrinsic sensor calibration can be split into two procedures: First, a *relative calibration* procedure estimates the sensor poses relative to all other sensors, see Figure 3.1. Second, an *absolute calibration* procedure estimates sensor poses with respect to a body coordinate frame of the robot. If a *relative calibration* is done first, the *Absolute calibration* only needs to estimate the transformation of one sensor to the robot body to complete the geometric model.

Existing multi-modal calibration methods usually only address combinations of two sensor sensing modalities. Accordingly, each approach uses a calibration target design that only works for their sensor pair, e.g. lidar and monocular camera. For more complex sensors setups involving radar, camera and lidar calibration, such as intelligent vehicles, multiple calibration boards and calibration tools would be needed to calibrate all sensors. However, as this chapter will show, optimization of all sensor pairs jointly should be preferred over separate pairwise calibration. Furthermore, a joint extrinsic calibration procedure reduces the calibration effort and calibration time, since the sensors poses are estimated at once using a single calibration target design. Related work also typically only addresses *relative calibration*, while in practice *absolute calibration* is often needed.

In this thesis, the focus is on a joint extrinsic sensor calibration procedure for sensor setups containing lidars, radars and/or cameras, using a single target design for all these sensing modalities. This thesis considers three configurations to jointly calibrate such multi-modal setups, as shown in Figure 3.2: *Minimally Connected Pose Estimation* (MCPE) estimates sensor-to-sensor transformations with respect to a single reference sensor. *Fully Connected Pose Estimation* (FCPE) provides transformations between all sensor pairs by adding a constraint that forces loop closure. The configuration *Pose and Structure Estimation* (PSE) jointly estimates sensor poses as well as the structure (i.e. calibration board poses). Additionally, the thesis addresses the problem of target-based *absolute calibration* to relate the sensors to a robot's body coordinate frame. The tool is implemented in an open-source tool with bindings to Robot Operating System (ROS).

The next section addresses the related work in detail. After that, the proposed approach is presented that elaborates on the three joint calibration configurations and the procedure to calibrate the sensors with respect to the robot coordinate frame. Finally, the experimental section provides comparisons of these three configurations on real sensor data from a sensor setup with a lidar, a stereo camera and a radar. Furthermore, the two methods are evaluated to determine the body reference frame for *absolute calibration*.

3.2. Related Work

An overview of related work on multi-modal extrinsic calibration is provided in Table 3.1, which is elaborated on in the following subsections. Note that a sensor pair with a stereo camera could be calibrated as two separate monocular cameras, however this is suboptimal if a full point cloud of the stereo camera is available (i.e. in case of a calibrated stereo camera).

3.2.1. Pairwise Calibration

The method of Peršić et al. [12] focuses on lidar to radar calibration. Rectangular shaped objects are inaccurate to detect in a lidar sensor, because nearly vertical or horizontal edges might fall between lidar scan planes (finite resolution issues). Therefore the authors use a triangular shaped Styrofoam calibration target with an attached metal trihedral corner reflector. Corner reflectors are a common target for radar because of their distinct reflectivity, the Radar Cross Section (RCS) value. The reprojection error between point cloud data and radar detections is minimized in their optimization procedure. In addition, the RCS values of multiple target locations are used to refine a subset of the transformation parameters.

Lidar to stereo calibration can be performed using the method of Guindel et al. [13]. This method uses a calibration target with four circles to calibrate a lidar and a stereo camera. Iterative Closest Point (ICP) [14] minimizes the error between the detected circle centers in both sensors.

For lidar to monocular camera calibration there are more methods available, namely [5, 9, 15–22]. Mirzaei et al. [5] perform intrinsic calibration of the lidar as well extrinsic calibration with respect to a monocular camera. The authors refine an analytical solution for intrinsic and extrinsic parameters by an optimization procedure based on iterative least squares. Geiger et al. [15] use data from multiple checkerboard patterns that are positioned in the environment to calibrate a lidar and a monocular camera. A set of initial transformation hypothesis are generated by a global registration procedure that minimizes the distance between the normal vectors and the centroids of the checkerboard patterns. After that, the set of transformation hypothesis is refined using ICP that minimizes the sum of point-to-point distances.

Extrinsic calibration of radar and monocular camera is performed by several methods [12, 23–26]. El Natour et al. [26] solve a system of equations with additional spherical and geometrical constraints to obtain the transformation matrix. Both [23] and [25] estimate a homography projection between the two sensors, which means that the full 3D transformation is not available.

Table 3.1: Related work on multi-modal extrinsic sensor calibration. The abbreviations in column *Optimization (Optim.)* denotes the optimization procedure where *pairwise* refers to optimization of the transformation between a pair of sensors and where *joint* refers to joint optimization of the whole pose graph. In addition, the column *Abs./Rel.* indicates if the work considers *absolute calibration* or *relative calibration*. Furthermore, the letters *L, R, S* and *M* stand for lidar, radar, stereo camera and monocular camera, respectively. For instance, the column L & M stands for calibration of the sensor pair of lidar and monocular camera. Symbols ✓ and ✗ indicate whether the experiments, documentation or software show that the work can calibrate a particular sensor pair. Symbol ~ indicates that a sensor pair with a stereo camera could be calibrated as two separate monocular cameras, in principle. The column *SW* indicates if the software is open-source and available to the community.

	Configuration	**Optim.**	**Abs./Rel.**	**L & R**	**L & S**	**L & M**	**R & S**	**R & M**	**SW**	**Toolbox name**
Peršić et al. [12]	MCPE	Pairwise	Rel.	✓	✗	✗	~	✓	✗	
Guindel et al. [13]	MCPE	Pairwise	Rel.	✗	✓	✗	✗	✗	✓	velo2cam_calibration
Chen and Chien [21]	MCPE	Pairwise	Rel.	✗	✓	✓	✗	✗	✗	
Geiger et al. [15]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	~	Online web toolbox
Velas et al. [16]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	✓	but_calibration_camera_velodyne
Alismail et al. [19]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	✓	calidar (MATLAB)
Zhang & Pless [18]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	✓	RADLOCC (MATLAB)
Dhall et al. [17]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	✓	lidar_camera_calibration
Mirzaei et al. [5]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	✗	
Gong et al. [22]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	✗	
Vasconcelos et al. [20]	MCPE	Pairwise	Rel.	✗	~	✓	✗	✗	✗	
El Natour et al. [26]	MCPE	Pairwise	Rel.	✗	✗	✗	~	✓	✗	
Sugimoto et al. [23]	MCPE	Pairwise	Rel.	✗	✗	✗	✗	✓	✗	
Wang et al. [25]	MCPE	Pairwise	Rel.	✗	✗	✗	✗	✓	✗	
Apollo [27]	MCPE	Pairwise	Rel.	✗	~	✓	~	✓	~	Only executables
Sim et al. [28]	MCPE/FCPE	Joint	Rel.	✗	~	✓	✗	✗	✗	
Pusztai et al. [29]	PSE	Joint	Rel.	✗	~	✓	✗	✗	✗	
Owens et al. [30]	PSE	Joint	Rel.	✗	~	✓	✗	✗	✗	
Proposed	MCPE/FCPE/PSE	Joint	Rel.&Abs.	✓	✓	✓[1]	✓	✓[1]	✓	multi_sensor_calibration

3.2.2. JOINT CALIBRATION

In order to calibrate a multi-modal sensor setup, one could simply pairwise calibrate all sensors with respect to one reference sensor, i.e. *minimally connected pose estimation*.

Alternatively, one could get inspiration from Simultaneous Localization and Mapping (SLAM), where loop closure is applied to readjust a trajectory of poses when the robot revisits the same location[31]. *Fully connected pose estimation*, a loop closure can be added as a constraint in the optimization procedure in extrinsic sensor calibration. In case of loop closure, moving over the edges in the loop (see Figure 3.2b) should result in the original pose, i.e. the multiplication of the transformation matrices of sensors in a loop results in the identity matrix. Sim et al. [28] use this 'loop closure' constraint for calibration of a lidar with multiple cameras.

Visual Odometry estimates the ego-motion based on matched features in consecutive images, and it could include bundle adjustment that refines all poses in a (sub) trajectory [32]. Bundle adjustment simultaneously refines sensor poses and 3D coordinates of landmarks [32]. A similar approach can be applied to extrinsic calibration. Pusztai et al. [29] uses a 'bundle adjustment-like' approach that consists of two steps, where in the first step the lidar errors are minimized and in the second step the camera re-projection errors are minimized. Owens et al. [30] use a graph optimization approach to calibrate a setup consisting of multiple lidars and cameras.

3.2.3. CONTRIBUTIONS

The overview in Table 3.1 reveals several open issues: Existing work only addresses *relative calibration*, is not able to calibrate all combinations of radar, lidar, and (stereo) camera jointly, and the community lacks an open-source tool to jointly calibrate such a multi-modal sensor setup.

The chapter of this thesis addresses these issues with four contributions. First, three extrinsic calibration configurations to jointly calibrate a sensor setup consisting of lidars, cameras and radars are examined. Important factors like configuration choice, required number of calibration board locations and choice for the reference sensor are investigated using a real multi-modal sensor setup. Second, this thesis proposes and compares two methods to estimate the pose of the body reference frame of the robot in order to perform *absolute calibration*. Third, a calibration target design that is detectable by lidar, camera and radar is presented. Fourth, the software is released as an open-source extrinsic calibration tool with bindings to Robot Operating System (ROS)[2]. For ROS users, a tool is provided that updates the Unified Robot Description Format (URDF) file that describes the robot model, to facilitate user-friendly usage of the tool on real robotic platforms.

3.3. PROPOSED APPROACH

In this section, the joint extrinsic calibration tool to calibrate lidar, camera and radar jointly with respect to the body reference frame of the robot is presented. Figure 3.3

[1] The repository contains a calibration board detector for monocular cameras, therefore sensor pairs with a monocular camera can also be calibrated.

[2] github.com/tudelft-iv/multi_sensor_calibration

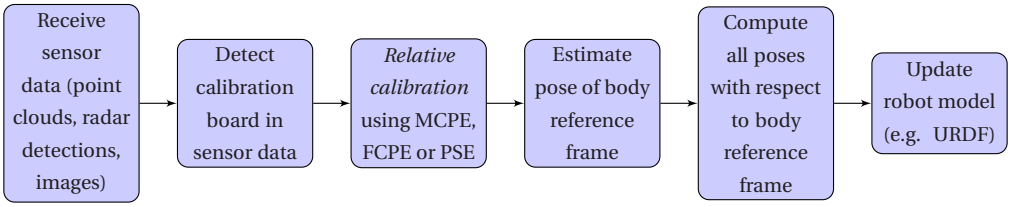


Figure 3.3: Extrinsic multi-sensor calibration pipeline. The first three steps perform *relative calibration* estimating the transformation matrices between all sensors using one of three optimization configurations (MCPE, FCPE, PSE). For *absolute calibration*, the next two steps relate the sensor frames to the robot body frame, by scanning the calibration targets and the robot body with an external laser scanner. The final step updates the URDF file with the new calibration results.

shows the pipeline with all steps to calibrate the sensors with respect to the body reference frame of the robot.

The next section discusses the calibration board design. Then, the detectors are described that extract the key points from this calibration board design. Using the detections, the thesis presents the details on pairwise calibration and then it is extended to joint calibration of a multi-modal sensor setup that consists of more than two sensors. The last part contains the proposed approach for *absolute calibration*.

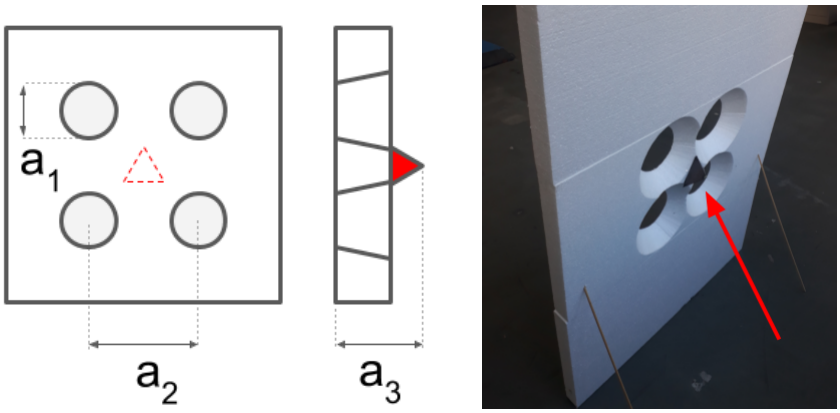


Figure 3.4: From left to right, front view drawing, side view drawing, and an image of the back of the target. The trihedral corner reflector is indicated in red (triangle and arrow).

3.3.1. CALIBRATION TARGET DESIGN

The design of the calibration target should facilitate accurate detections for all sensing modalities. For accurate radar detections, the thesis uses a trihedral corner reflector that facilitates radar reflections with specific RCS values. To limit the effect on detectability of the corner reflector, Styrofoam is chosen as material for the calibration target[33]. As target for lidar and camera, this thesis pursues the approach of [13, 16] and use circular holes. These holes have edges, which are perfect features to detect in both sensors. The layout of the target, with a size of 1.0 m by 1.5 m, with circle diameter $a_1 = 15$ cm, and distance between the centers $a_2 = 24$ cm is shown in Figure 3.4. The reflector is posi-

tioned in the middle of the four circles at the back of the Styrofoam plate (at $a_3 = 10.5$ cm from the front).[3]

3.3.2. DETECTION OF CALIBRATION TARGET

The lidar detector and the stereo detector of [13] have been adapted. For lidar and camera, the 3D location of the circle centers are returned as features. Incorrect detections can be discarded since the geometry of the board is known and there are four feature points. If the ratio between the diagonal and the side of the square is not equal to $\sqrt{2}$, detections can be discarded.

The radar measurements consist of 2D locations in polar coordinates and a RCS value. First, all detections are kept that are within the expected RCS range. From all those detections, the closest measurement to the robot is taken as radar detection as the assumption is made that the calibration board is the closest target in the vicinity of the robot.

For the monocular camera detector, the four circles are detected based on edges in the 2D image plane. Using the known geometry of the calibration board, perspective-n-point algorithm (PnP) [34] can be used to extract the 3D locations of the circle centers.

3.3.3. PAIRWISE CALIBRATION

First, pairwise calibration is explained, which is then extended to joint calibration of a setup with N sensors in section 3.3.4.

The calibration target is positioned at K different locations in FOV of two sensors, referred to as sensor 1 and sensor 2. Each detector returns K detections $\boldsymbol{y}^1 = \{\boldsymbol{y}_1^1, \cdots, \boldsymbol{y}_K^1\}$ for sensor 1 and $\boldsymbol{y}^2 = \{\boldsymbol{y}_1^2, \cdots, \boldsymbol{y}_K^2\}$ for sensor 2. Each calibration board location provides four detections in 3D for lidar and camera: $\boldsymbol{y}_k = (y_{k(1)}, \cdots, y_{k(4)})$. Furthermore, the radar detector only returns a single detection as the target has one trihedral corner reflector. This detection $\boldsymbol{y}_k = (y_{k(1)})$ is defined in 2D Euclidean coordinates. Since a detector might not always detect the target, for instance if the target is not in the sensor's FOV, indicator variables μ_k^i are used to represent if the detector of sensor i was able to successfully detect calibration board location k. This means that $\mu_k^i = 1$ if the target was detected and $\mu_k^i = 0$ otherwise.

Extrinsic calibration between the two sensors aims to estimate the relative rigid transformation $T^{1,2}$. This transformation can be used to project a point from the coordinate frame of sensor 1 to the coordinate frame of sensor 2. The rigid transformation is expressed as a 4×4 matrix for homogeneous coordinates that consists of a 3×3 rotation matrix R and 3D translation t vector,

$$T^{1,2} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}. \tag{3.1}$$

To use this homogeneous representation, each 3D point (x, y, z) is represented as an augmented 4D vector $(x, y, z, 1)$. To parametrize the 6 degrees of freedom of transformation $T^{1,2}$, vector $\theta^{1,2} = (t_x, t_y, t_z, v_x \cdot \alpha, v_y \cdot \alpha, v_z \cdot \alpha)$ is used. The rotation part is expressed by

[3] See README file in the repository for details on the calibration board.

an axis-angle representation (using Rodrigues' rotation formula), namely as a unit vector (v_x, v_y, v_z) for the axis of rotation, and an angle α.

For the k-th target location, the total squared Euclidean distance of the four detected circle centers is used to define the transformation error between lidar and camera detections,

$$\epsilon_k\left(\theta^{1,2}\right) = \sum_{p=1}^{4} \left\| y^2_{k(p)} - T^{1,2} \cdot y^1_{k(p)} \right\|^2. \tag{3.2}$$

If the sensor pair contains a radar, a different error term is used. Let $\boldsymbol{y}^R_k$ represents the radar measurement of target k, then the squared Euclidean error equals

$$\epsilon_k\left(\theta^{1,R}\right) = \left\| y^R_{k(1)} - p(T^{1,R} \cdot g(\boldsymbol{y}^1_k)) \right\|^2. \tag{3.3}$$

Here, function $g(\boldsymbol{y}_k)$ calculates the expected 3D position of the trihedral corner reflector in the reference frame of sensor 1 by using the four circle center locations in detection $\boldsymbol{y}_k$ and the geometry of the calibration board. Then, function $p(q_k)$ first converts 3D Euclidean point q_k to spherical coordinates (r_k, ϕ_k, ψ_k), disregards the elevation angle ψ_k, and converts (r_k, ϕ_k) back to 2D Euclidean coordinates.

In addition, the thesis adds constraints that enforce that the projected 3D points lie within radar Field of View (FOV). To achieve that, the elevation angles ψ_k for all calibration board locations k should be within the maximum view angle ψ_{max} of the radar,

$$|\psi_k| - \psi_{max} \leq 0, \quad \forall k. \tag{3.4}$$

Pairwise calibration is now formulated as an optimization problem that finds the optimal transformation between both sensors by minimizing the total error $f(\theta^{1,2})$ between all K calibration targets,

$$f(\theta^{1,2}) = \sum_{k=1}^{K} \mu^2_k \cdot \mu^1_k \cdot \epsilon_k\left(\theta^{1,2}\right). \tag{3.5}$$

The indicator variables $\mu^2_k \cdot \mu^1_k$ ensure calibration board locations K that are detected by both sensors are included. By minimizing the error criterion $f(\theta)$ subject to zero or more (in)equality constraints (e.g. equation (3.4)), the optimal relative transformation are obtained.

Sequential Least SQuares Programming (SLSQP) from the *SciPy* library [35] is used to solve the optimization problem, which is potentially subject to constraints. To initialize the optimization procedure, an initial solution is required. For that, the optimal rotation between the point cloud containing centroids of the four circle detections for all calibration board locations K is computed by Kabsch algorithm [36]. Using this rotation matrix, the initial translation vector can be determined. To find an initial transformation for a sensor pair containing a radar, it is assumed that detections lie on the radar plane (zero elevation angle).

3.3.4. JOINT CALIBRATION WITH MORE THAN TWO SENSORS

To generalize extrinsic calibration from pairwise calibration to N sensors, three configurations are considered to jointly calibrate a multi-sensor sensor setup, namely MCPE, FCPE, PSE. Instead of estimating a single edge (i.e. sensor-to-sensor transformation), now multiple edges are present. The three configurations for *relative calibration* are visualized in Figure 3.1 and will be discussed in this section.

Minimally connected pose estimation (MCPE) In the first configuration, all sensors are calibrated in a pairwise manner with respect to a selected 'reference' sensor. This results in a *minimally connected* graph, which is visualized in Figure 3.2a. The edges describe the transformation from the 'reference sensor' to the other sensors. Without loss of generality, let's assume that the first sensor is selected as the reference sensor. In this case, the optimization criterion is formulated as

$$f(\boldsymbol{\theta}) = \sum_{i=2}^{N} \left[\sum_{k=1}^{K} \mu_k^i \cdot \mu_k^1 \cdot \epsilon_k \left(\theta^{1,i} \right) \right]. \tag{3.6}$$

Note that transformations between any non-reference sensors i, j can be computed from the known transformations in this graph, i.e. $T^{i,j} = T^{1,j} \cdot (T^{1,i})^{-1}$.

Fully connected pose estimation (FCPE) In the second configuration, the thesis considers optimizing transformations between all sensors at once, without assigning a specific reference sensor. This results in optimizing edges in a fully connected graph (see Figure 3.2b), akin to a loop closure optimization in SLAM. Instead of estimating $N-1$ transformation matrices with respect to a reference sensor, all transformation matrices between all $\binom{N}{2}$ combinations of two sensors are computed. In this case, the error functions equals

$$f(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\sum_{k=1}^{K} \mu_k^i \cdot \mu_k^j \cdot \epsilon_k \left(\theta^{i,j} \right) \right]. \tag{3.7}$$

To ensure that all loops l equal the identity matrix, the loop closure constraint is included in the optimization problem,

$$(T^{s_l,1} \cdot T^{s_l-1,s_l} \cdot \ldots \cdot T^{1,2}) - I = 0, \qquad \forall l \tag{3.8}$$

where s_l equals the number of sensors in this loop l. The thesis only considers all $\binom{N}{3}$ combinations of $s_l = 3$ sensors. The advantage of this optimization is that it is potentially more robust against noisy observations from one reference sensor. The disadvantages are that the number of error terms increases with the number of sensors N and that by adding extra sensors, additional loop constraints must be included as well.

Pose and structure estimation (PSE) The third configuration is called pose and structure estimation and it is visualized in Figure 3.2c. This configuration has similarities to bundle adjustment since it simultaneously estimates all sensor poses and calibration board poses. This means that both the unknown structure $M = (m_1, \cdots, m_K)$ of the true

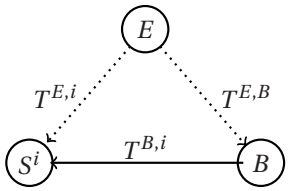


Figure 3.5: For *absolute calibration*, the transformation between body reference frame B and a sensor S^i (solid arrow) is found indirectly by first determining the transformations to both frames from an external reference sensor E (dashed arrows).

target poses in a fixed coordinate frame, and the transformation $T^{M,i}$ from the fixed frame to each sensor i are estimated. Observations are considered samples from a probabilistic measurement model, which uses $\hat{y}^M_{k(p)} = h(m_k, p)$, with zero-mean Gaussian noise,

$$y^i_{k(p)} = T^{M,i} \cdot \hat{y}^M_{k(p)} + \eta^i, \quad \eta^i \sim \mathcal{N}(0, \Sigma^i). \tag{3.9}$$

Therefore, instead of the squared Euclidean distance, the squared Mahalanobis distance is used, which equals

$$D^2_\Sigma(a,b) = [a-b]^\top (\Sigma)^{-1} [a-b] \tag{3.10}$$

with vectors a and b, and covariance Σ. In the optimization, the thesis jointly optimizes the transformations and structure,

$$\epsilon_k\left(\theta^{M,i}, M\right) = \sum_{p=1}^{4} D^2_{\Sigma^i}\left(y^i_{k(p)}, T^{M,i} \cdot \hat{y}^M_{k(p)}\right), \tag{3.11}$$

$$f(\boldsymbol{\theta}, M) = \sum_{i=1}^{N}\left[\sum_{k=1}^{K} \mu^i_k \cdot \epsilon_k\left(\theta^{M,i}, M\right)\right], \tag{3.12}$$

and initialize all Σ^i as identity. An iterative procedure is used to calculate the diagonal elements of the noise covariances. Using the result of the first optimization, the noise covariances are recalculated and updated, after which the optimization of $f(\boldsymbol{\theta}, M)$ is repeated. This process is continued until all variances have converged. Note that to determine a unique solution, one transformation $T^{M,i}$ must be fixed.

This probabilistic formulation has the potential advantage that it avoids having heterogeneous error functions (pixel versus Euclidean). Instead, a homogeneous error function is used that comprises of the sum of squared Mahalanobis distances. Furthermore, it provides the option to include prior knowledge on board and sensor poses, however this direction is not pursued here. The disadvantages are also twofold. First, the optimization is more complex and therefore it takes more time. Second, the loop closure constraint is not explicitly enforced.

3.3.5. POSE ESTIMATION OF BODY REFERENCE FRAME

To estimate the pose of the body reference frame of the robot, minimal three 3D reference points on the exterior of the robot are required. To determine the set of 3D points

during calibration, an external sensor must be used which can detect these reference points, and the calibration target at multiple locations. This sensor should have a high resolution and large field of view to accurately locate both the 3D reference points on the exterior as well as the calibration target. From the shared detected calibration targets, the transformation from the external sensor to the robot sensor can be found, similar to *relative calibration*. After the robot reference frame is determined in the external sensor frame too, the sought transformations between the sensor and the robot frame can be computed directly, as illustrated in Figure 3.5.

To localize the 3D body reference points within the external sensor point cloud, two general approaches can be taken:

1. *Human labeling*: The locations of the set of the 3D reference points can be manually labeled in the sensor data. These locations can be obtained by manually labeling each individual 3D reference point in the point cloud. Alternatively, multiple points can be labeled on a visible part of the robot's exterior with a specific geometric shape (e.g circular shape). After that, a geometrical shape can be fitted on the set of labeled points.

2. *Markers*: The locations of the 3D reference points can be extracted by placing physical markers that the external sensor can easily detect. This is less laborious than labeling the data afterwards, but the accuracy depends fully on how precise the markers could be placed when the calibration procedure was performed.

In practice, a lidar laser scanner is used to construct a point cloud model of the body, and either select the 3D reference points in this point cloud, or use markers that the scanner can accurately detect.

3.4. EXPERIMENTS

To evaluate the performance of the tool, sensor data of lidar, camera and radar is recorded with the Toyota Prius vehicle, which is equipped with:

- a Velodyne HDL-64E lidar (on roof)
- a Continental ARS430 radar (behind front bumper)
- a stereo camera 2× UI-3060CP Rev. 2 (behind windscreen)

For the experimental validation, the sensor set of the vehicle is calibrated with the calibration target in the vehicle's garage. The calibration target is positioned in front of the car at 30 different locations within approximately 5 meters. From these 30 calibration board locations, 29 locations were within the field of view of all three sensors (lidar, the stereo camera and the radar). See Figure 3.6a for the output of the calibration tool, where the detected calibration target locations for all three sensors are shown in the lidar reference frame. For *absolute calibration*, a Leica P40 laser scanner is used as the external sensor, see Figure 3.6b. The P40 is a high resolution laser scanner which is able to localize itself in the environment using multiple black-white markers on the walls and floor. The Leica scanner was placed at several positions around the car, and using the markers

and Leica software a merged point cloud of the vehicle is obtained, shown in Figure 3.6c. During calibration, the P40 is positioned next to the car such that this sensor can see both the car and 12 calibration board locations.

Three configuration (MCPE, FCPE and PSE) for *relative calibration* are evaluated on data from 29 calibration board locations. The computation time of the optimization depends on the number of sensors and the number of calibration board locations. If all 29 calibration board locations are used, the computation time is less than 1 second for the MCPE configuration, approximately 10 seconds for the FCPE configuration and approximately 5 minutes for PSE configuration on a high-end computer (with an Intel Xeon W-2123 @ 3.60GHz CPU).

In section 3.4.1, the thesis investigates the performance of the tool for *relative calibration*, and in section 3.4.2 for *absolute calibration*. Finally, in section 3.4.2, additional outdoor experiments are presented to demonstrate the impact outside the garage in the intended environment of the vehicle.

3.4.1. *Relative calibration*

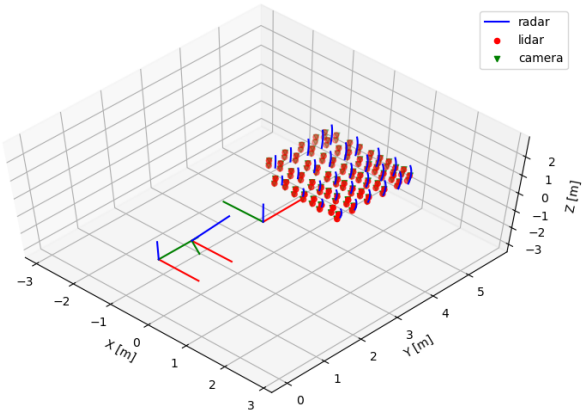
To assess calibration quality, the residual error for each pair of sensors is computed, i.e. the Euclidean distance between the measured target positions after applying the found transformation to put all measurements in the same reference frame. The root mean squared error (RMSE) of all pairwise transformations is reported, namely lidar to stereo camera ($l2c$), lidar to radar ($l2r$), stereo camera to radar ($c2r$). In the following sections, this thesis compares the *relative calibration* approaches to a baseline calibration method, and assesses the choice of reference sensors, the number of target locations, and sensitivity to additional noise.

COMPARISON TO BASELINE METHOD

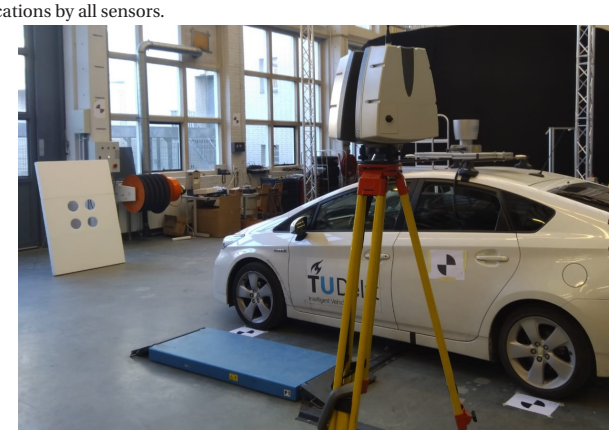
First, a comparison is made with the single-target method of Guindel et al. [13] that only calibrates a lidar to stereo camera pair. For the MCPE implementation when using sensor data of a single target location and the single target method of Guindel, the calibration is performed for all 29 calibration board locations and the mean and standard deviation of the RMSE are provided in Table 3.2. It can be seen that both single target implementations provide a similar result. In addition, the benefit of using multiple calibration board locations is investigated. Table 3.2 shows that the $l2c$ RMSE reduces from 39 mm to 15 mm.

Table 3.2: Comparison with baseline method.

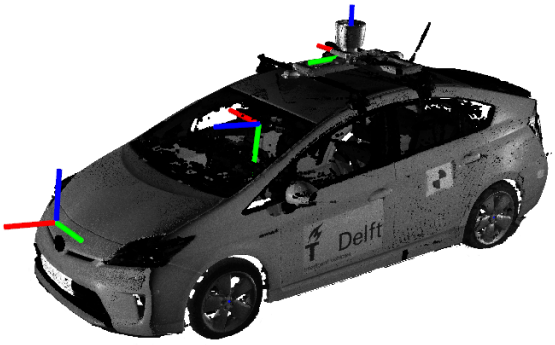
lidar to stereo method	**# boards**	**RMSE [mm]**
Guindel et al. [13]	single board	39.3 ± 10.4
MCPE	single board	39.3 ± 10.4
MCPE	all boards (29)	15.3
FCPE	all boards (29)	15.3
PSE	all boards (29)	15.3



(a) Measured target locations by all sensors.



(b) Absolute calibration setup for a vehicle.



(c) Leica point cloud of vehicle.

Figure 3.6: a Output of the calibration tool. All sensors poses and all detections of the calibration board are plotted in the lidar reference frame. b Absolute calibration setup, using an external Leica laser scanner, for a vehicle with a lidar, stereo camera, and radar. c Merged point cloud from the Leica scanner, with the calibrated coordinate systems of the three sensors after *absolute calibration* using the *Human labeling* approach.

Table 3.3: Median of the RMSE [mm] for 200 combinations of 10 calibration board locations.

	RMSE l2c [mm]	RMSE l2r [mm]	RMSE c2r [mm]
MCPE(camera)	16.0±0.3	20.5±0.5	27.7±0.8
MCPE(lidar)	16.0±0.3	20.4±0.5	27.6±0.7
MCPE(radar)	16.3±0.4	20.4±0.5	27.7±0.8
PSE	16.1±0.3	18.3±1.7	24.0±1.2
FCPE	16.0±0.3	15.0±0.6	22.3±0.9

CHOICE OF MCPE REFERENCE SENSOR

Next, an experiment is performed to investigate if the choice for reference sensor of the MCPE configuration influences its results. 10 calibration board locations are randomly picked 200 times and the sensors are calibrated. Table 3.3 shows the median RMSE for MCPE with all three reference sensors and for FCPE and PSE. The Table shows that all choices (lidar, camera and radar) give similar RMSE, however selecting the radar as reference sensors results in two links that contain radar measurements. Since radar data is 2D (range and angle) having two links with radar data might result in less accurate results, therefore the lidar with a FOV of 360° is used as reference sensor from now on. Furthermore, the RMSE for the sensor pairs $l2r$ and $c2r$ shows that configurations FCPE and PSE perform better than the MCPE configuration.

DEPENDENCE ON THE NUMBER OF CALIBRATION BOARD LOCATIONS

To understand the impact of the number of calibration board locations, K, number K is varied from 3 to 29 locations. For each value of K, 100 sets of K randomly selected locations are used to calibrate the sensor setup. Figure 3.7 shows the median and median absolute deviation of the RMSE over all 100 sets. Both FCPE and PSE show smaller RMSE than MCPE for the $l2r$ transform. The RMSE for $l2c$ and $l2r$ transforms for FCPE and PSE configurations have converged to ≤ 2 cm if more than 10 calibration board locations are used. The configuration FCPE shows the best performance for $l2r$, since the RMSE is smaller than 1.5 cm when using all 29 board locations.

SENSITIVITY TO OBSERVATION NOISE

This thesis also compares the robustness of the three configurations under additional measurement noise for a sensor, and wonders how it affects the other sensor pairs. Zero-mean Gaussian noise $\mathcal{N}(0, \sigma^2 I_3)$ is added to the 3D measurements of the lidar detections. The median and median absolute deviation of the RMSE for various values of σ are plotted in Figure 3.8, and it can be seen that the RMSE of sensor pairs with lidar increase as a result, though the $c2r$ errors for both FCPE and PSE remain fairly constant as more noise is added. Furthermore, the RMSE for $l2c$ and $l2r$ remain lower than the RMSE $c2r$ for most of values of σ.

3.4.2. ABSOLUTE CALIBRATION

For *absolute calibration*, the additional transformation between the Velodyne lidar coordinate frame and the vehicle's body reference frame is estimated using the external Leica

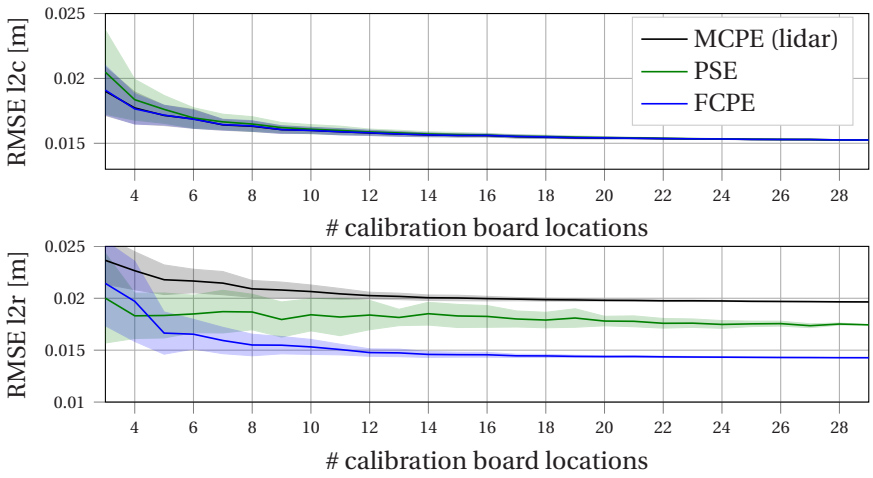


Figure 3.7: The median and median absolute deviation of the RMSE on 100 board locations for varying number of calibration board locations (K).

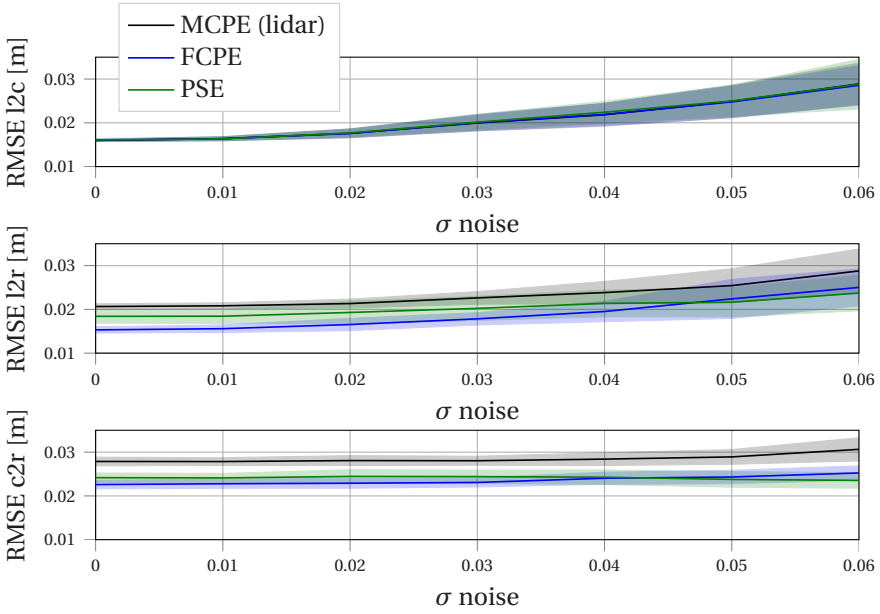


Figure 3.8: RMSE error as function of Gaussian observation noise $\mathcal{N}(0,\sigma^2 I_3)$ added to the lidar observations. The plotted median and median absolute deviation are based on 100 random combinations of 10 calibration board locations.

Table 3.4: Comparison of the standard deviations of the wheel center location for the manual labeling approaches.

	σ_x [mm]	σ_y [mm]	σ_z [mm]
Single point	1.3	1.7	1.9
3D circle fit	0.8	0.5	1.3

laser scanner. This means that the transformation between the Velodyne and the body reference frame $T^{B,i}$ (see Figure 3.5) needs to be assessed. For the Toyota Prius vehicle, the origin vehicle's body reference frame is at the center of the rear axle *projected onto the ground*, the X-axis is pointing forward, the Y-axis is pointing to the left rear wheel and the Z-axis is perpendicular to the ground (pointing upwards). Hence, to determine the pose of the body reference frame, the location of the wheel centers and ground plane must be determined.

In section 3.3, several practical approaches are discussed to determine the location of the 3D reference points, namely *Markers* and *Human labeling*, which are implemented as follows:

1. As a first *Human labeling* approach, each wheel center location is manually labeled by selecting a single point in the Leica point cloud (see Figure 3.6c). To project those locations to the ground, the normal vector and distance to the ground is found by fitting a planar model on the lower part of the point cloud.

2. Another *Human labeling* approach is to manually select N points on the rim of the wheel, and fit a 3D circle through those N points to determine the wheel center.

3. For the *Markers* approach, four Leica markers are positioned next to the wheels on the ground (see markers in Figure 3.6b) below the axles.

This section will compare the robustness of the labeling options over multiple repetitions, and compare the rotational errors of the approaches with respect to the ground normal.

ROBUSTNESS OF MANUAL LABELING

First, the two manual labeling approaches are compared by labeling the left rear wheel of the car 10 times. Table 3.4 shows the standard deviations in X, Y and Z positions in Leica reference frame. The results shows that labeling the wheel centers using multiple points ($N = 10$) on the rim and fitting a 3D circle provides slightly better results than labeling wheel centers using a single 3D point. Despite that the differences between the labeling approaches are small, the 3D circle fit on the rim is used to determine the wheel centers from now on.

More importantly, the observation is made that the standard deviation between multiple annotations is in the order of millimeters. This thesis concludes that this is sufficiently robust given the operating scale and physical size of the vehicle.

ROTATION ERROR AROUND X-AXIS AND Y-AXIS COMBINED

Now, the error in angle between the estimated and expected Z-axis is quantified. The rotational error around the vertical Z-axis will be assessed later in section 3.4.3.

Since the Z-axis of the body reference frame is perpendicular to the ground, it is expected that the observed normal vector of the ground *in the vehicle sensors* is aligned with the body's Z-axis. Sensor data was recorded in a large garage space at the same time as the absolute calibration was performed, meaning that the state of the suspension and state of the tires is unchanged. The assumption is made that the ground within 6 meters of the vehicle center is flat. The ground normal vector is estimated in the point cloud of the vehicle's Velodyne by segmenting the planar ground floor, using a maximum distance tolerance of 2.5 cm, and use the calibration to transform it to the body reference frame. The angular error θ between the observed normal vector n_{obs} and the expected normal vector $\mathbf{n_{exp}} = [0, 0, 1]$ is

$$\theta = \arccos\left(\frac{\mathbf{n_{obs}} \cdot \mathbf{n_{exp}}}{\|\mathbf{n_{obs}}\| \, \|\mathbf{n_{exp}}\|}\right). \tag{3.13}$$

Initially when the sensors were positioned based on manual adjustments, the angle θ was 0.13°, and after calibration the angle θ has decreased to 0.07° using the *Markers* approach, and 0.02° using the *Human labeling* approach.

3.4.3. OUTDOOR EXPERIMENTS

Finally, the thesis reports on additional experiments performed outside the garage at two outdoor locations. These enable us to asses the calibration impact on multi-modal perception in realistic environments, and at larger distances than possible in the garage to highlight the reduced rotational errors.

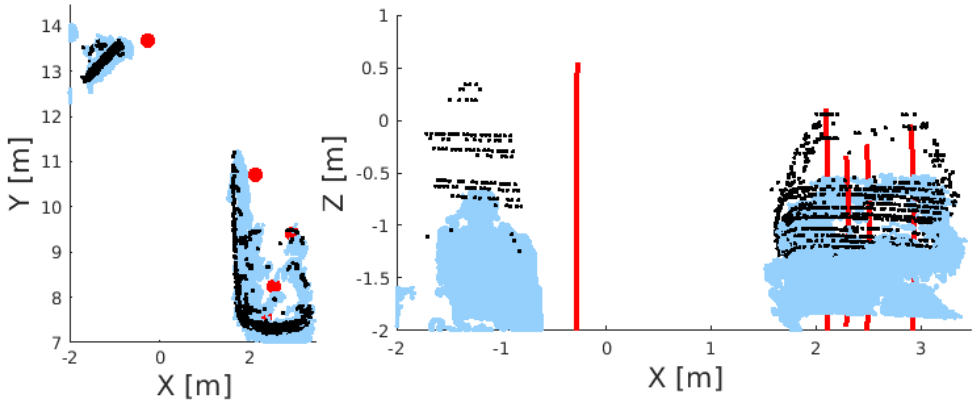
LOCATION 1: QUALITATIVE ASSESSMENT

The thesis first qualitatively demonstrates the overall spatial and rotational accuracy of *relative calibration* for all vehicle sensors in an urban outdoor scene with obstacles at 7 to 14 meter distance, see Figure 3.9. Before calibration, with initial manually set sensor poses, the data from lidar and the stereo camera have a mismatch in the Z direction, and the radar detection on the person on the left does not match the measurements from the other two sensors. After calibration the data from all sensors are well aligned, even though the used calibration targets were only placed at a few meters in front of the vehicle.

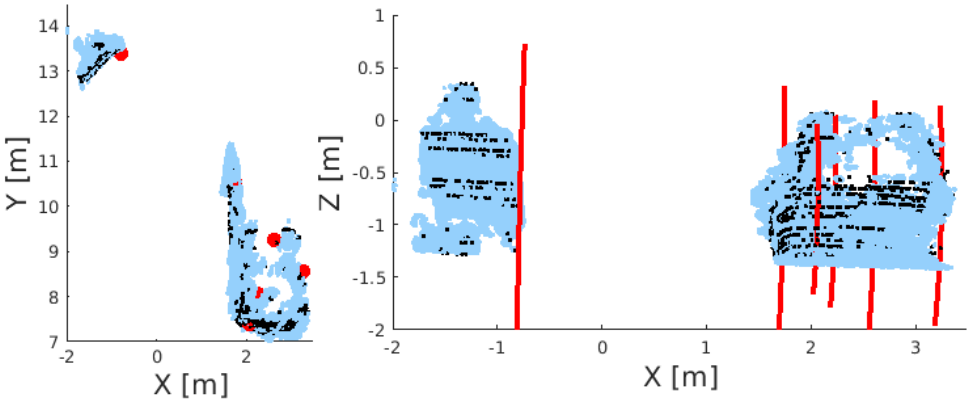
LOCATION 2: ROTATION ERROR AROUND VERTICAL Z-AXIS

To assess the rotation error around the vehicle's vertical axis for *absolute calibration*, the apparent lateral drift of static objects is measured in the sensor frame while the vehicle is moving straight forward, i.e. along the X-axis of the body reference frame. On a well calibrated setup, the expectation is that the measured lateral position of static objects, when transformed to the vehicle's body reference frame, is the same at the first and last measurement, see Figure 3.10. Therefore, the lateral positions of eight street light poles distributed along the right side of an empty 240 meter long straight road is measured in the vehicle's lidar. The poles are extracted from the point cloud by clustering the lidar

(a) Zoomed in camera image of outdoor scene

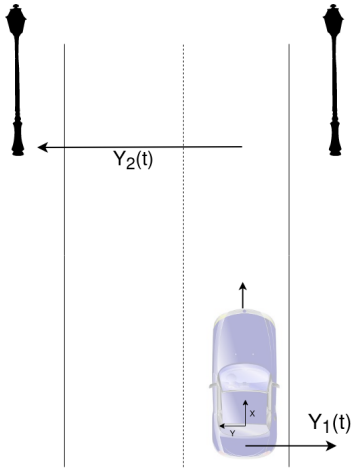


(b) Before calibration

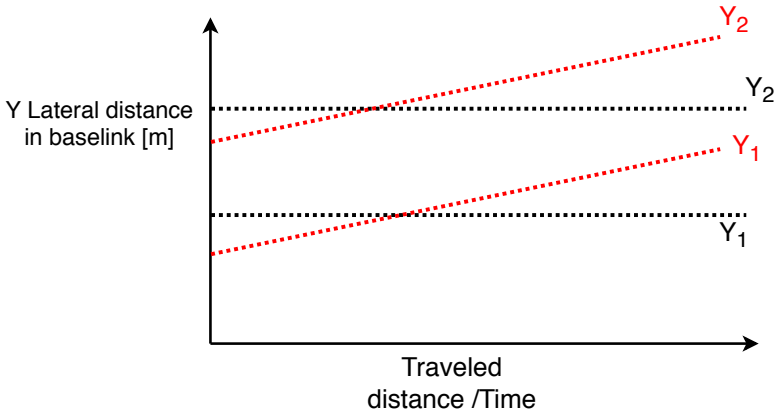


(c) After calibration

Figure 3.9: a Image of the recorded scene to test the calibration. There is a parked car ~ $6m$ in front of the sensor setup, and a person with a checkerboard at ~ $13m$. b The lidar (black) and stereo (blue) point cloud, and radar detections (red) before extrinsic calibration (based on manual adjustments). c The sensor data after extrinsic calibration. Radar detections are drawn as arcs since the elevation angle is not measured.



(a) Vehicle driving straight on road with street lights.



(b) Expected lateral position of street lights.

Figure 3.10: Experimental setup to assess the rotational error of *absolute calibration* in real-world setting over larger distances. a The vehicle drives in an approximately straight line on a long straight road with streetlights. b The angle of inclination α (slope) of the lateral position of the light in the vehicle's reference frame is expected to be near zero over the whole drive if the sensors are properly calibrated (black lines). For a bad calibration (red lines) the results would show a systematic lateral drift.

points [37]. The car drives with a maximum speed of 5.4 m/s over the road marking line (closed road), and each pole is measured for about 30 meters. To compensate for measurement errors, small deviations of the straight trajectory, and outliers at the start and end, for each pole a line through all measured positions is fitted. A pole's amount of lateral shift (ΔY) over the longitudinal range that it is observed (ΔX) allows us to compute the angular error $\alpha = \arctan(\Delta Y / \Delta X)$ of the lidar w.r.t. the body reference frame. While small deviations in the car's actual velocity can affect the number of measured positions for each street light in Figure 3.11, e.g. driving faster would result in fewer measured positions for each street light, a similar angle estimates is expected as the speed only impacts the number of points that are used for line fitting.

The measured positions of the street lights in the body frame are shown in Figure 3.11. The observation is made that the slopes for *Human labeling* are the smallest compared to the other cases. Overall, the reported median α angles in the graph captions confirm that the error has decreased from more than 0.95° to 0.33° for *Markers* and 0.02° for *Human labeling*.

3.5. DISCUSSION

Both the FCPE and the PSE configuration showed better results than the MCPE configuration (see Table 3.3 and Figure 3.7). This was expected since the FCPE configuration includes all error terms between sensors in the optimization and the PSE configuration uses a probabilistic model to simultaneously estimate the calibration board poses and the sensor poses. It was found that the FCPE configuration shows the best results on our sensor setup which consists of a lidar, a stereo camera and a radar. Furthermore, the experiments showed that the proposed method that uses sensor data of multiple calibration board locations outperforms the single target method of Guindel et al. [13]. With more than ten calibration board locations, the median RMSE is < 2 cm for lidar to camera, approximately 2 cm for lidar to radar and approximately 2.5 cm for camera to radar. For the MCPE configuration with fast computation time, the radar does not seem to be a good choice as reference sensor, since it results in having two links with 2D radar measurements (range and angle).

The PSE configuration simultaneously estimates the calibration board poses and sensor poses. The noise covariances are estimated iteratively using sensor data of all calibration board locations, however the noise covariances might not be constant for all calibration board locations as the radar observation noise is usually larger at the edges of the field of view. It is assumed that the observations are samples from a probabilistic model with zero-mean Gaussian noise and that for every sensor the (2D/3D Euclidean) measurements are uncorrelated (i.e. all off-diagonal entries of the observation covariance matrices are equal to zero). In computation of the root mean squared error, the errors in the various dimensions (e.g. X,Y,Z) are treated equally (i.e. identity weights). These identity weights are also used in optimization of Euclidean error terms in the MCPE and the FCPE configuration. However in case of the PSE configuration, the total error term in the optimizer is based on the squared Mahalanobis distance, which means that the inverse covariance matrices are used as weights (i.e. different weights for the various dimensions). This means that in case of the PSE configuration, the total error is internally optimized using the inverse covariance matrices as weights, however when the RMSE is

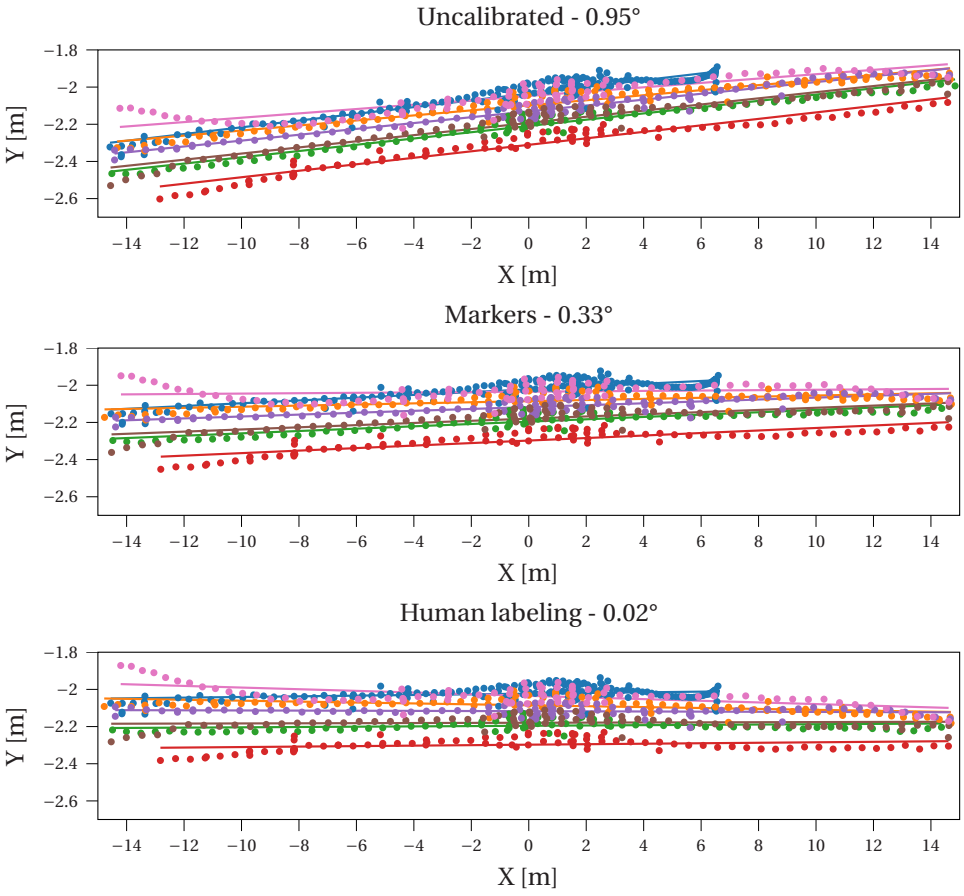


Figure 3.11: Locations of all streetlights in vehicle body reference frame while driving on a straight road. The value in the title represents the median α angle for each method.

computed then identity weights are used. This could explain why the PSE configuration performs worse than the FCPE configuration.

Furthermore, some practical considerations are important for users. The calibration board design consists of four key points for lidar and camera and one key point for radar (e.g. trihedral corner reflector). The number of key points for every sensor affects the optimization. In the FCPE configuration, the error term consists of all pairwise errors and the total error for a single calibration board locations consist of four error terms for lidar to camera and one error term for the other two links. In addition, all error terms for sensor pairs with a radar are 2D Euclidean errors, whereas lidar to camera terms are 3D Euclidean errors. This means that the error in the FCPE configuration is dominated by the lidar to camera errors, since it has four 3D Euclidean errors for every calibration board location. Furthermore, there are multiple loop closure constraints for the FCPE configuration for $N > 3$ sensors. The number of constraints (loop constraints) increases with the number of sensors in the FCPE configuration. Therefore, the optimizer needs to deal with a increasing number of constraints. This might influence the performance of this configuration. In addition, the PSE configuration requires the measurement noise covariances for all sensors, therefore these are estimated in an iterative manner. In practice, this means that the computation time is significantly affected by the number of calibration board locations.

For calibration with respect to the body reference frame, a circle fitting approach was used on the rims of the wheels to determine its center, which makes this calibration approach suitable for robots with visible wheels. In absence of visible wheels, the users should use other 3D reference points to determine the pose of a body reference frame. The main difference between the approach *Markers* and the approach *Human labeling* can be found in the rotation error around the vertical axis, which can be explained by the fact that accurate marker placement is challenging for the former method. Moreover, the accuracy completely depends on how well the markers were placed during the calibration procedure. In case of inaccurate marker placement, the calibration procedure needs be performed again. When the *Human labeling* needs to be performed again, the point cloud model of the car (including the wheel center locations) can be reused. In that case, the transformation between the point cloud model and the current scan of the external sensor can be estimated using point set registration techniques (e.g. ICP) to determine the wheel center locations in the reference frame of the Leica. In addition, the absolute calibration of the lidar sensor was evaluated. For the lidar sensor, the method *Human labeling* using *3D circle fitting* showed most accurate results, namely a median angle of 0.02° around the vertical Z-axis. To provide insights on how orientation errors affect position estimates at a larger distance, the displacement error due to rotation errors ϵ for objects located at distance d_{obj} can be computed using: $\Delta = \sin(\epsilon) \cdot d_{obj}$. Initially when the sensors were positioned based on manual adjustments a median angle error of 0.95°results in a displacement error of approximately 50 cm for an object at 30 meters. After calibration, the median angle reduces from approximately 1° to 0.02° with a factor 50, therefore the displacement error decreased with a factor 50 assuming small-angle approximation ($\sin(\epsilon) \approx \epsilon$ where ϵ is in radians).

3.6. CONCLUSION

An open-source extrinsic calibration tool to jointly calibrate sensor setups consisting of lidar, camera and radar sensors was presented. The tool offers three configurations to estimate the sensor poses from simultaneous detections of multiple calibration board locations. Important factors like configuration choice, dependency on the number of calibration board locations and choice for the reference sensor are investigated using a real multi-modal sensor setup that consists of a lidar, a stereo camera and a radar. The experiments show that all configurations can provide good calibration results, though *fully connected pose estimation* showed the best performance. When ten calibration board locations are used, the median RMSE is less than 2 cm for lidar to camera, approximately 2 cm for lidar to radar and approximately 2.5 cm for camera to radar. This chapter's findings highlight the importance of calibrating multiple sensors modalities jointly, rather than separately for each pair.

In addition, two approaches were described to calibrate the sensors to the body reference frame using an external laser scanner, a process referred to as *absolute calibration*. To measure the body frame pose of a vehicle in the external point cloud, it was found that the best approach was to manually annotate several points on each wheel, and perform geometric shape fitting on the wheels and ground plane. For the lidar sensor, a low horizontal error w.r.t. the perceived ground plane normal $< 0.2°$ was achieved, an outdoor driving experiment showed a rotation error around the vertical axis of $0.02°$ an order of magnitude smaller than the alternatives.

By sharing the ROS compatible calibration tool, and detailing the approach and findings, other researchers are facilitated that need to regularly calibrate such multi-modal sensor setups.

REFERENCES

[1] J. Domhof, J. F. Kooij, and D. M. Gavrila, *A joint extrinsic calibration tool for radar, camera and lidar,* IEEE Transactions on Intelligent Vehicles **6**, 571 (2021).

[2] S. Sivaraman and M. M. Trivedi, *Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis,* IEEE Transactions on Intelligent Transportation Systems **14**, 1773 (2013).

[3] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, *Survey of pedestrian detection for advanced driver assistance systems,* IEEE Transactions on Pattern Analysis and Machine Intelligence **32**, 1239 (2010).

[4] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision* (Cambridge university press, 2003).

[5] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, *3D LIDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization,* The International Journal of Robotics Research **31**, 452 (2012).

[6] R. Szeliski, *Computer vision: algorithms and applications* (Springer Science & Business Media, 2010).

[7] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, *Multisensor data fusion: A review of the state-of-the-art,* Information Fusion **14**, 28 (2013).

[8] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, *Automatic extrinsic calibration of vision and lidar by maximizing mutual information,* Journal of Field Robotics **32**, 696 (2015).

[9] N. Schneider, F. Piewak, C. Stiller, and U. Franke, *RegNet: Multimodal sensor registration using deep neural networks,* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2017) pp. 1803–1810.

[10] J. Levinson and S. Thrun, *Automatic online calibration of cameras and lasers.* in *Robotics: Science and Systems*, Vol. 2 (2013).

[11] Z. Zhang, *A flexible new technique for camera calibration,* IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000).

[12] J. Peršić, I. Marković, and I. Petrović, *Extrinsic 6DoF calibration of a radar–lidar–camera system enhanced by radar cross section estimates evaluation,* Robotics and Autonomous Systems **114**, 217 (2019).

[13] C. Guindel, J. Beltrán, D. Martín, and F. García, *Automatic extrinsic calibration for lidar-stereo vehicle sensor setups,* in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2017) pp. 1–6.

[14] P. J. Besl and N. D. McKay, *A method for registration of 3-D shapes,* IEEE Transactions on Pattern Analysis and Machine Intelligence **14**, 239 (1992).

[15] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster, *Automatic camera and range sensor calibration using a single shot,* in *2012 IEEE International Conference on Robotics and Automation* (IEEE, 2012) pp. 3936–3943.

[16] M. Velas, M. Španěl, Z. Materna, and A. Herout, *Calibration of RGB camera with velodyne lidar,* Comm. Papers Proc. International Conference on Computer Graphics, Visualization and Computer Vision (WSCG) , 135 (2014).

[17] A. Dhall, K. Chelani, V. Radhakrishnan, and K. M. Krishna, *LiDAR-Camera Calibration using 3D-3D Point correspondences,* arXiv preprint arXiv:1705.09785 (2017).

[18] Q. Zhang and R. Pless, *Extrinsic calibration of a camera and laser range finder (improves camera calibration),* in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, Vol. 3 (IEEE, 2004) pp. 2301–2306.

[19] H. Alismail, L. D. Baker, and B. Browning, *Automatic calibration of a range sensor and camera system,* in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission* (IEEE, 2012) pp. 286–292.

[20] F. Vasconcelos, J. P. Barreto, and U. Nunes, *A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder,* IEEE Transactions on Pattern Analysis and Machine Intelligence **34**, 2097 (2012).

[21] C.-Y. Chen and H.-J. Chien, *Geometric calibration of a multi-layer LiDAR system and image sensors using plane-based implicit laser parameters for textured 3-D depth reconstruction,* Journal of Visual Communication and Image Representation **25**, 659 (2014).

[22] X. Gong, Y. Lin, and J. Liu, *3D LIDAR-camera extrinsic calibration using an arbitrary trihedron,* Sensors **13**, 1902 (2013).

[23] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, *Obstacle detection using millimeter-wave radar and its visualization on image sequence,* in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Vol. 3 (IEEE, 2004) pp. 342–345.

[24] D. Gao, J. Duan, X. Yang, and B. Zheng, *A method of spatial calibration for camera and radar,* in *2010 8th World Congress on Intelligent Control and Automation* (IEEE, 2010) pp. 6211–6215.

[25] T. Wang, N. Zheng, J. Xin, and Z. Ma, *Integrating millimeter wave radar with a monocular vision sensor for on-road obstacle detection applications,* Sensors **11**, 8992 (2011).

[26] G. E. Natour, O. Ait-Aider, R. Rouveure, F. Berry, and P. Faure, *Toward 3D reconstruction of outdoor scenes using an MMW radar and a monocular vision sensor,* Sensors **15**, 25937 (2015).

[27] *Apollo an open autonomous driving platform,* https://github.com/ApolloAuto/apollo, accessed: 2019-07.

[28] S. Sim, J. Sock, and K. Kwak, *Indirect correspondence-based robust extrinsic calibration of lidar and camera,* Sensors **16**, 933 (2016).

[29] Z. Pusztai, I. Eichhardt, and L. Hajder, *Accurate calibration of multi-lidar-multi-camera systems,* Sensors **18**, 2139 (2018).

[30] J. L. Owens, P. R. Osteen, and K. Daniilidis, *MSG-cal: Multi-sensor graph-based calibration,* in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2015) pp. 3660–3667.

[31] H. Durrant-Whyte and T. Bailey, *Simultaneous localization and mapping: part I,* IEEE Robotics & Automation Magazine **13**, 99 (2006).

[32] D. Scaramuzza and F. Fraundorfer, *Visual odometry [tutorial],* IEEE Robotics & Automation Magazine **18**, 80 (2011).

[33] J. Peršić, I. Marković, and I. Petrović, *Extrinsic 6DoF calibration of 3D lidar and radar,* in *2017 European Conference on Mobile Robots (ECMR)* (IEEE, 2017) pp. 1–6.

[34] M. A. Fischler and R. C. Bolles, *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,* Communications of the ACM **24**, 381 (1981).

[35] E. Jones, T. Oliphant, P. Peterson, *et al.*, *SciPy: Open source scientific tools for Python,* (2001–), [Online; accessed September 2018].

[36] W. Kabsch, *A solution for the best rotation to relate two sets of vectors,* Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography **32**, 922 (1976).

[37] I. Bogoslavskyi and C. Stachniss, *Efficient online segmentation for sparse 3D laser scans,* PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science , 1 (2017).

4

OBJECT TRACKING

This chapter addresses the multi-sensor tracking filter that deals with time-varying and state dependent conditions. The filter comprises of a estimator for the probability of detection, and a tracking filter that utilizes the estimated detection probabilities. In this way, the detection probability of each sensor is continuously estimated, to adapt to changing weather and illumination conditions. Experiments in a controlled environment with artificial fog and experiments with the vehicle prototype in various weather and illumination conditions have been performed. The experiments showed that the proposed adaptive tracking filter outperforms the tracking filter with fixed hyper-parameters for the detection probabilities.

This chapter is based on the paper [1].

4.1. INTRODUCTION

To ensure safety, an intelligent vehicle must be aware of the objects in its surroundings. Mainly three sensors are used for object tracking, namely (monocular/stereo) vision, radar and lidar [2]. The object tracker provides state estimates (position, velocity, etc.) for objects such as road users to the path planning algorithm in order to plan a safe trajectory. Many state-of-the-art object tracking approaches assume perfect weather and illumination conditions; however each sensor sensing modality has their drawbacks. For example, the main drawback of vision is that it is sensitive to changes in weather and illumination conditions [3], and lidar is more sensitive to precipitation than radar [3]. Since the intelligent vehicle is a safety critical system, the sensing quality should be monitored during highly automated driving to avoid safety-critical situations. On the basis of the estimated sensing quality, the system should decide if it can continuously operate.

Multi-object filtering approaches require a parameterization of the detection probability and clutter level, however such models are not always available. Incorrect parameterization of these two variables lead to missing objects and/or spurious objects in the object tracker. Furthermore, these parameters depend on the weather and illumination conditions, which can change while driving.

There are three challenges considering adverse weather conditions. First, the probability of detection and clutter density are usually considered time invariant (fixed). These parameters are tuned based on a training data set, however these values change when for instance weather conditions change (e.g. visibility in fog).

In other words, when the weather and illumination conditions in the test set are deviating from the training set, the performance of the object tracker will be affected. Second, the probability of detection of a sensor might be dependent on the state of the object. For instance, visibility in fog depends on distance. However it is hard to parameterize the probability of detection as function of the object's state for all weather and illumination conditions. Third, the perception system should be able to inform when the sensors working are not working properly. This means that there is a need for a measure for the quality of sensing of the system. One aspect to quantify the 'health' of the sensors is to have an estimate of the probability of detection for each sensor.

Approaches that use a single sensor for object tracking are not able to deal with sensor or detector failures. In case of total sensor failure, a single sensor object tracker encounters an ambiguity, as it can reason in two ways: either the probability of detection is low or there are no objects in the field of view. In case of some sensor or detector failure due to adverse weather conditions (e.g, reduced visibility due to fog) this ambiguity exists. The same ambiguity exists when, for instance, the image detector is no longer able to detect objects at a certain distance. Therefore, this thesis motivates for having two sensing modalities for object tracking is to deal with sensor or detector failures.

In this chapter, a novel multi-sensor approach that dynamically estimates the detection probability for each sensor is presented, which uses these estimates within a multi-sensor tracking filter to improve tracking performance in challenging weather and illumination conditions.

4.2. RELATED WORK

Multi-object tracking filters combine Kalman and particle filters with data association techniques. Popular approaches include Multiple Hypothesis Tracking (MHT) [4] and Joint Probabilistic Data Association (JPDA) [5, 6]. While MHT builds all track hypothesis by propagating the data association hypothesis over time, JPDA approximates the state update by an weighted measurement update based on all measurement to track association probabilities. However, the number of association hypotheses might be intractable for real-time object tracking. The Random Finite Set (RFS) approach [7] uses an alternative way to model the collection of unknown targets, namely using a random finite set; both the elements in the set and the number of elements are a random variable. Commonly used filters are: Probability Hypothesis Density (PHD)[8], Cardinalized Probability Hypothesis Density (CPHD)[9] and the labeled multi-Bernoulli filter (LMB) filter [10]. Common implementations of multi-sensor tracking filters in automotive applications use CPHD tracking filters [11] and Labeled Multi-Bernoulli (LMB) tracking filters [12, 13] as a tracking framework.

To deal with an unknown clutter rate and detection probability, Mahler et al. [14] proposed three filters, namely a CPHD filter that can handle an unknown clutter rate, a CPHD filter that can deal with an unknown detection probability, and a CPHD filter that deals with an unknown clutter and detection probability. In the work of [15], the authors

propose a bootstrap filter consisting of a clutter estimator and a tracking filter to deal with an unknown clutter density. A robust multi-Bernoulli particle filtering approach has been proposed that deal with non-linear dynamics in [16]. Furthermore, Rezatofighi et al. [17] use a bootstrap filter with a clutter rate and detection probability estimator for multi-target tracking of cells in microscopic images. The clutter rate and detection probabilities from the estimator are bootstrapped onto a tracking filter that outputs estimates for the objects. In this way, the authors can deal with time-varying clutter and a detection profile. In [18], the authors simultaneously estimate an unknown clutter rate and detection probability for pedestrian tracking using a single camera. In addition, adaptive filtering approaches are proposed that estimate a time-varying detection probability for radar [19] or sonar [20] sensors. A state dependent probability of detection model has been derived offline in [21] for indoor pedestrian tracking with lidar and vision. The probability of detection model depends on the distance from the object, the occlusions, and the sensor field of view; however, this model is unable to deal with the time-varying probability of detection.

Furthermore, existing work bench marked the performance of automotive sensors in adverse weather conditions. For example, in the work of [22], the authors show that all state-of-the-art laser scanners perform poorly in dense fog conditions. Experiments in the fog chamber showed that the maximum viewing distance is reduced to a fraction of the clear-weather viewing distance. In [23], lidar sensors were bench marked in snow and icy conditions. A methodology to test and evaluate imaging sensors was proposed in [24]. Based on experiments in a fog chamber, the authors concluded that (standard) cameras suffer from loss of contrast that deteriorates the visibility of edges.

Nowadays, deep learning techniques show very promising results in object detection [25, 26] and object tracking [27–29]. To train these models, data sets are required with a large corpus of training data. For adverse weather conditions, the number of data sets is limited, because adverse weather conditions are relative rare. A dedicated multi-modal data set in adverse weather conditions was recorded by [30]. The authors use the clear weather part of the data set to train a deep learning neural network for object detection using sensor data of radar, lidar, camera and gated camera, and use the adverse weather data as the test set. Due to lack of data, researchers move to the usage of synthetic data, e.g. [31–33].

Alternative approaches to deal with adverse weather conditions focus on weather detection and weather removal techniques. For example, Pavlic et al. [34] and Hautiére et al. [35] developed a fog detector. Zhang et al. developed a classifier for sunny, rainy, hazy, and snowy conditions [36]. In addition, image enhancement techniques [37, 38] can be used to remove effects of adverse weather conditions. The work of [39] survey techniques to mitigate the effect of rain in images and analyse existing deep learning approaches for object detection in rainy weather conditions. Hassaballah et al. [40] use image restoration technique to enhance visibility in images, and use a deep learning neural network for object detection in combination of a PHD tracking filter. Image enhancement methods are bench marked in dense fog in [41]. Running all these algorithms continuously in parallel for each sensor might be computationally intractable. Furthermore, these methods cannot deduce the sensing quality, because it is not evident how these conditions impact the sensor observations (i.e. object detectors), which are used

for object tracking.

In summary, related work does not address object tracking with multiple sensing modalities when the detection probabilities are potentially time-varying and state-dependent due to weather and illumination conditions.

4.2.1. CONTRIBUTION

The main contribution of this chapter is threefold. First, a novel multi-modal tracking filter is proposed that is able to deal with adverse weather conditions. Second, controlled experiments in a fog simulator have been performed with static targets to investigate the tracking performance (cardinality and localization errors) in slowly changing conditions. Third, a data set is collected with the TU Delft vehicle prototype that contains recorded scenarios with multiple dynamic objects. This data set with lidar, camera, and radar data is used to investigate the performance of the proposed tracking filter in various weather and illumination conditions.

4.3. METHODOLOGY

Robust perception cannot be achieved by using a single sensing modality. There is a need for having an additional sensing modality. An additional sensing modality is beneficial for estimating the detection probabilities of the sensor, because of three reasons. First, each of the three sensing modalities (lidar, radar and vision) use a different part of electromagnetic spectrum which means that the weather conditions (rain, fog, snow, etc.) do not affect these modalities in the same way. Second, in case of sensor failure (e.g. something is covering the camera lens), an ambiguity exists in the single-sensor case. The system might reason that there are no objects or that the probability of detection is very low. Third, the probability of detection of a sensor might depend on the state of the object in case of adverse weather conditions. For instance, visibility in fog depends on the distance as objects further away are affected more by the fog.

Based on these three reasons, this thesis proposes to use at least two sensing modalities for object tracking. The single sensor bootstrap filters in [15, 17] served as inspiration for this work. Experiments in [15] showed better performance for the bootstrap filter than the filter dealing with an unknown clutter rate. Similarly, the proposed bootstrap filter outperformed the filter with unknown detection probability and clutter rate [17]. Instead of a single sensor bootstrap tracking filter, this thesis uses a multi-sensor bootstrap tracking filter that estimates the detection probability for each sensor to deal with adverse weather and illumination conditions. The multi-sensor bootstrap filter comprises of an estimator for the detection probabilities and a multi-sensor tracking filter. Figure 4.1 shows the schematic of the proposed multi-sensor bootstrap filter. The estimator provides the detection probabilities estimates for each object to the multi-sensor tracking filter, which uses the detection probabilities estimates for improved tracking.

The original p_D-CPHD filter, which is the CPHD filter that can deal with unknown detection probability, is summarized in section 4.3.1. In section 4.3.2, the proposed multi-sensor p_D-CPHD filter is discussed, which is the estimator for the probability of detection (true positive rate). The last section, section 4.3.3, addresses how the estimator and the tracking filter are integrated in the proposed multi-sensor bootstrap filter.

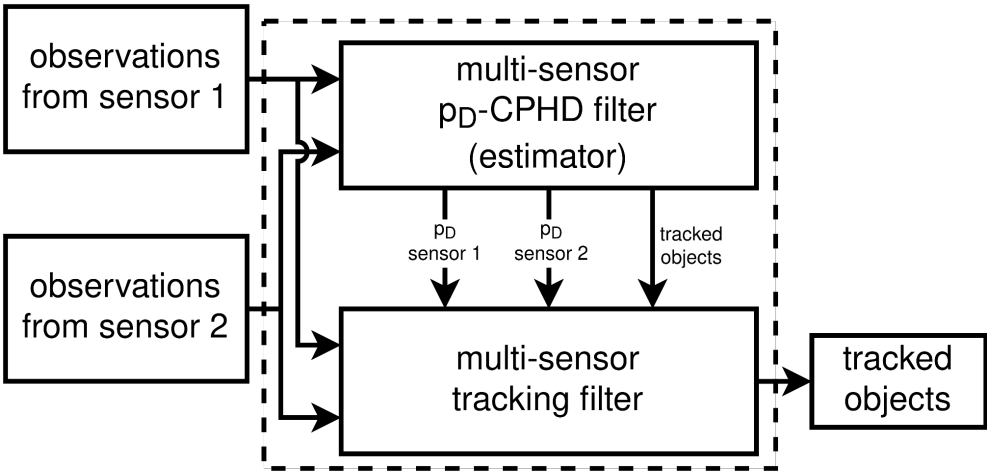


Figure 4.1: Schematic of the multi-sensor bootstrap tracking filter (4.3.3), which consists of two components, namely the estimator (4.3.2) and the tracking filter. The estimator aims at estimating the probability of detection for each sensor, and the tracking filter is using these estimates to track objects.

Both the estimator and the tracking filter receive sensor data from more than one sensor. Most sensors have a different frame rate and others are not able to be triggered (i.e. synchronized); therefore, observations are obtained asynchronously. Therefore, the estimator and the tracking filter should process the observation from a sensor one after the other; the sensor data fusion takes place by a sequential measurement update [42]. This is also called the iterated-corrector method [43]. This way, the detection probability for the various sensors can be easier taken into account compared to parallel updating.

4.3.1. ORIGINAL p_D-CPHD FILTER

This section summarizes the p_D-CPHD filter from Mahler et al. [14], which is the CPHD filter that deals with an unknown detection probability.

The state $\underline{x} = [x, a]$ consists of a kinematic state x and an augmented part a that represents the detection probability between 0 and 1. The state $\underline{x}$ is also called the augmented state, as it consists of the kinematic state and the augmented detection probability. The random finite set with augmented states at time k is represented by $X_k = \{\underline{x}_{k,1}, \dots \underline{x}_{k,N(k)}\}$, where $N(k)$ indicates the number of objects. The filter estimates the states using observations $Z_k = \{\underline{z}_{k,1}, \dots \underline{z}_{k,M(k)}\}$, where $M(k)$ denotes the number of observations at time k.

Modeling of the kinematic state The probability of an object's kinematic state x is modeled by a Gaussian distribution $\mathcal{N}(x; m, P)$, where m denotes the mean and P its covariance, The kinematic state follow a linear Gaussian dynamical model, hence the

predicted kinematic state equals

$$m_{k|k-1} = F_{k-1} m_{k-1} \tag{4.1}$$
$$P_{k|k-1} = Q_{k-1} + F_{k-1} P_{k-1} F_{k-1}^T \tag{4.2}$$

where F_{k-1} represents the state transition matrix and Q_{k-1} is the process noise matrix at time $k-1$. The subscript $k|k-1$ denotes that the estimate of a certain parameter hold for time k, given the observations up to and including time $k-1$.

The estimate of the kinematic state that is updated by observation z is equal to

$$m_k(z) = m_{k|k-1} + K_k \left(z - H_k m_{k|k-1}\right) \tag{4.3}$$
$$P_k = (I - K_k H_k)\, P_{k|k-1} \tag{4.4}$$

where H_k is the observation matrix. For brevity, the subscript k is used to denote $k|k$. Furthermore, K_k represents the Kalman gain at time k

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \tag{4.5}$$

with R_k denoting the observation noise covariance.

Modeling of the detection probability The state vector is augmented with a probability of detection a, which is modeled by a Beta distribution $\beta(a; s, t)$ which is described by two parameters, namely $s > 0$ and $t > 0$. As the probability of detection is part of the state, it is estimated by the filter for each time k, hence $\beta(a; s_k, t_k)$. From the properties of the Beta distribution, the expected value of $\beta(a; s_k, t_k)$ is

$$\mu_{\beta,k} = \frac{s_k}{s_k + t_k} \tag{4.6}$$

and its variance equals

$$\sigma_{\beta,k}^2 = \frac{s_k \cdot t_k}{(s_k + t_k)^2 (s_k + t_k + 1)}. \tag{4.7}$$

Prediction for the augmented part of the state follows a preserved mean and a dilated variance, i.e. $\mu_{\beta,k|k-1} = \mu_{\beta,k-1}$ and $\sigma_{\beta,k|k-1}^2 = k_\beta \cdot \sigma_{\beta,k-1}^2$, where $\mu_{\beta,k-1}$ and $\sigma_{\beta,k-1}^2$ are the expected value and the variance at the previous time instance $k-1$, respectively. Furthermore, k_β is a hyper parameters > 0 that scales the variance. To increase the predicted variance, a factor of $k_\beta > 1$ is chosen. Then the prediction for parameters s_k and t_k boils down to

$$s_{k|k-1} = \left(\frac{\mu_{\beta,k|k-1}(1 - \mu_{\beta,k|k-1})}{\sigma_{\beta,k-1}^2} - 1 \right) \mu_{\beta,k|k-1} \tag{4.8}$$
$$t_{k|k-1} = \left(\frac{\mu_{\beta,k|k-1}(1 - \mu_{\beta,k|k-1})}{\sigma_{\beta,k-1}^2} - 1 \right) (1 - \mu_{\beta,k|k-1}). \tag{4.9}$$

Implementation This section details the closed-form implementation of the filter equations. To estimate the augmented states X_k, the filter aims at jointly propagating an intensity distribution and a cardinality distribution. The intensity distribution $\underline{v}(x,a)$ describes the density of the expected number of targets per unit volume, at x, a, [44]. This means that when $\underline{v}(x,a)$ is higher, then it is more likely that there is a target at x, a, [43]. When $\underline{v}(x,a)$ is integrated, $\int_S \int_0^1 v(x,a)dadx$, then the expected number of targets in region S is obtained. The cardinality distribution is a probabilistic representation of the estimated number of objects in the set. Thus, for each n, the cardinality distribution contains the probability that there are n targets in the environment. The cardinality distribution is a probability mass function of the number of targets, which is propagated by considering existing targets and appearing targets as well as false alarms and the measurement set. To avoid propagating infinite terms, the cardinality distribution is truncated at N_{max}, with N_{max} significantly greater than the number of targets in the environment [44].

If the posterior intensity $\underline{v}_{k-1}$ and posterior cardinality distribution ρ_{k-1} at time $k-1$ are known and the $\underline{v}_{k-1}$ is a Beta-Gaussian mixture [14] with J_{k-1} components, each representing a target hypothesis, that have weight $w_{k-1}^{(i)}$

$$\underline{v}_{k-1}(x,a) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} \beta\left(a; s_{k-1}^{(i)}, t_{k-1}^{(i)}\right) \times \mathcal{N}\left(x; m_{k-1}^{(i)}, P_{k-1}^{(i)}\right). \tag{4.10}$$

The predicted intensity and cardinality distribution are adapted by considering the probability that previously tracked targets continue to exist (surviving targets) and that new targets appear, also known as 'births'. The surviving RFS and birth RFS are assumed to be independent. Then the predicted intensity is given by

$$\underline{v}_{k|k-1}(x,a) = p_{S,k} \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} \beta\left(a; s_{S,k|k-1}^{(i)}, t_{S,k|k-1}^{(i)}\right) \times \mathcal{N}\left(x; m_{S,k|k-1}^{(i)}, P_{S,k|k-1}^{(i)}\right) + \gamma_k(x,a) \tag{4.11}$$

where subscript S denotes the surviving states, $p_{S,k}$ is the probability of survival and $\gamma_k(x,a)$ is the intensity of the birth targets, represented by a Beta-Gaussian mixture. The predicted cardinality distribution $\rho_{k|k-1}(n)$ involves a convolution of the birth and surviving targets, see [14].

Given the predicted intensity $\underline{v}_{k|k-1}(x,a)$ and the predicted cardinality distribution $\rho_{k|k-1}(n)$, the updated intensity is also a Beta-Gaussian mixture

$$\begin{aligned}\underline{v}_k(x,a) = &\sum_{j=1}^{J_{k|k-1}} w_{M,k}^{(j)} \cdot \beta\left(a; s_{k|k-1}^{(j)}, t_{k|k-1}^{(j)} + 1\right) \\ &\times \mathcal{N}\left(x; m_{k|k-1}^{(j)}, P_{k|k-1}^{(j)}\right) + \sum_{z \in Z_k} \sum_{j=1}^{J_{k|k-1}} w_{D,k}^{(j)}(z) \\ &\times \beta\left(a; s_{k|k-1}^{(j)} + 1, t_{k|k-1}^{(j)}\right) \mathcal{N}\left(x; m_k^{(j)}(z), P_k^{(j)}\right).\end{aligned} \tag{4.12}$$

where $w_{M,k}^{(j)}$ is the weight after a missed detection and $w_{D,k}^{(j)}(z)$ is the weight that is updated with an observation (detection). The first term of equation 4.12 contains the missed term, which means that the weight ($w_{M,k}^{(j)}$) are updated considering a missed detection; for the beta distribution t is incremented by one and the kinematic state is predicted using equation 4.1. The second term represents the term that is updated with a detection. Thus, the weights are updated based on the measurement likelihood, clutter intensity and a correction factor considering a detection. Furthermore, the kinematic states are updated according to the Kalman filter update step (equations 4.3 to 4.5) and for the detection probability a detection is considered; increment s of the Beta distribution by one. The updated cardinality $\rho_k(n)$ is a function of the clutter cardinality, the measurements, the expected detection probability (similar to equation 4.6), the predicted weights and the predicted cardinality distribution. For the exact equations for the updated cardinality $\rho_k(n)$, weights $w_{M,k}^{(j)}$ and weights $w_{D,k}^{(j)}(z)$, the reader is referred to the original work, [14].

The expected number of targets can be estimated in two ways; first option is to compute the mean of the weights $\hat{N}_k = \sum_{j=1}^{J_k} w_k^{(j)}$, whereas the second option involves taking the mode of $\rho_k(n)$, namely $\hat{N}_k = \arg\max_n \rho_k(n)$.

To avoid increase of the number mixture components, component pruning, merging and capping are required. Component pruning involves removing mixture components that have a weight smaller than a threshold T'. Component merging aims at merging mixture components that are similar. In the work of [14], two mixture components are merged if the similarity measure is below predefined threshold S'. As a measure of similarity, the authors chose a metric that is inspired by the Hellinger distance [45], which quantifies the distance between two probability distributions. It is defined in the range between zero and one, where zero indicates that two distributions are the same. For a set with mixture components that are similar $I = \{i : d_{ij} < S'\}$, the set I is replaced by a single Beta-Gaussian mixture component that is approximated by a summed weight $\sum_{i\in I} w_k^{(i)}$ and a weighted average for the augmented and kinematic state, e.g. $m_k^{(j)} = \frac{1}{\sum_{i\in I} w_k^{(i)}} \sum_{i\in I} w_k^{(i)} m_k^{(i)}$. Finally, if the number of mixture components is greater than J_{max}, then component capping reduces the number of mixture components by taking the J_{max} components with the highest weight.

4.3.2. MULTI-SENSOR p_D-CPHD FILTER

Instead of having one detection probability a, now the multi-sensor case is considered. This means that apart from the kinematic state, multiple augmented variables needs to be considered, which are denoted as $a^{(n_s)}$, and represent the detection probability for sensor n_s. The assumption is made the detection probabilities are independent of each other. From now on, two sensors ($N_s = 2$) are considered, hence the set of sensor indices equals $1, 2$. To include the probability of detection for the second sensor, equation 4.10

is extended:

$$\begin{aligned}\underline{v}_{k-1}(x, a^{(1)}, a^{(2)}) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} \beta\left(a^{(1)}; s_{k-1}^{(1,i)}, t_{k-1}^{(1,i)}\right) \\ \times \beta\left(a^{(2)}; s_{k-1}^{(2,i)}, t_{k-1}^{(2,i)}\right) \\ \times \mathcal{N}\left(x; m_{k-1}^{(i)}, P_{k-1}^{(i)}\right)\end{aligned} \tag{4.13}$$

This means that for the predicted intensity, both β distributions are predicted using equations 4.8 and 4.9.

To denote from which sensor, sensor data is received, two variables are introduced, namely n_z and $n_{\neg z}$. Variable n_z indicates the sensor for which sensor data Z^{n_z} is received at time k, and $n_{\neg z}$ to denote the other sensor. This means that at each time k, either $n_z = 1$ and $n_{\neg z} = 2$ when sensor data from sensor 1 is received or $n_z = 2$ and $n_{\neg z} = 1$ when sensor data from sensor 2 is received. The updated intensity equals

$$\begin{aligned}\underline{v}_k(x, a^{(n_z)}, a^{(n_{\neg z})}) = \\ \sum_{j=1}^{J_{k|k-1}} w_{M,k}^{(j)}(n_z) \times \beta\left(a^{(n_z)}; s_{k|k-1}^{(n_z,j)}, t_{k|k-1}^{(n_z,j)} + 1\right) \\ \times \beta\left(a^{(n_{\neg z})} a; s_{k|k-1}^{(n_{\neg z},j)}, t_{k|k-1}^{(n_{\neg z},j)}\right) \times \mathcal{N}\left(x; m_{k|k-1}^{(j)}, P_{k|k-1}^{(j)}\right) \\ + \sum_{z \in Z_k} \sum_{j=1}^{J_{k|k-1}} w_{D,k}^{(j)}(z) \times \beta\left(a^{(n_z)}; s_{k|k-1}^{(n_z,j)} + 1, t_{k|k-1}^{(n_z,j)}\right) \\ \times \beta\left(a^{(n_{\neg z})}; s_{k|k-1}^{(n_{\neg z},j)}, t_{k|k-1}^{(n_{\neg z},j)}\right) \times \mathcal{N}\left(x; m_k^{(j)}(z), P_k^{(j)}\right).\end{aligned} \tag{4.14}$$

The intensity and cardinality distribution are updated similarly as in section 4.3.1 if the measurement model of sensor n_z, clutter density of sensor n_z, the expected detection probability of sensor n_z are selected. The kinematic state x and the augmented part $a^{(n_z)}$ are updated using sensor data of sensor n_z, which means that the observation model of sensor n_z is used for the Kalman filter update step and that only β part for sensor n_z is updated. Pruning, merging and capping techniques are used to prevent unbounded growth of the number of components. Pruning and capping techniques work similarly as for the p_D-CPHD filter, however for merging there is a difference. As the augmented state now includes additional variables, namely the detection probability for each sensor, the similarity measure for merging the components (after the update of equation 4.14) should be extended. To accommodate multiple augmented variables, the similarity measure should take into account the β distributions for each of the augmented

variables. Similarly to [14], a measure inspired by the Hellinger distance is used

$$d_{ij} = \sqrt{1 - d_{i,j}^{\beta,1} \cdot d_{i,j}^{\beta,2} \cdot d_{i,j}^{\mathcal{N}}} \tag{4.15}$$

$$d_{i,j}^{\beta} = \frac{B\left(\frac{s_k^{(i)}+s_k^{(j)}}{2}, \frac{t_k^{(i)}+t_k^{(j)}}{2}\right)}{\sqrt{B\left(s_k^{(i)}, t_k^{(i)}\right) B\left(s_k^{(j)}, t_k^{(j)}\right)}} \tag{4.16}$$

$$d_{i,j}^{\mathcal{N}} = \frac{\det(P_k^{(i)})^{1/4} \det(P_k^{(j)})^{1/4}}{\det(P)^{1/2}}$$

$$\exp\left\{-\frac{1}{8}(m_k^{(i)} - m_k^{(j)})^T (P)^{-1} (m_k^{(i)} - m_k^{(j)})\right\} \tag{4.17}$$

with $P = \frac{P_k^{(i)}+P_k^{(j)}}{2}$ and where B is the Beta function. In equation 4.15, $d_{i,j}^{\beta,1}$ is calculated between the two β distributions for sensor 1, $d_{i,j}^{\beta,2}$ for sensor 2, and finally $d_{i,j}^{\mathcal{N}}$ is calculated for components of i and j, which are based on the Hellinger distance for multivariate normal distribution and Beta distribution, see the Appendix B.

4.3.3. BOOTSTRAP FILTERS

Section 4.3.2 addressed the estimator for the probability of detection for each sensor. As Figure 4.1 is showing, the bootstrap filter is composed of the estimator and a tracking filter. This section explains how the estimator and the tracking filter are integrated.

The tracking filter requires a parametrization of several parameters. One of these parameters is the probability of detection p_D for each sensor. To benefit from the estimator, the tracking filter needs to be adapted to use the estimated detection probabilities from the estimator. For that, this thesis distinguishes two approaches how the estimated probability of detection is used inside the multi-sensor tracking filter. The first approach is to consider *time-varying* detection probabilities, which means that the probability of detection (p_D) for sensor n_s is constant in the field of view and it depends on time k, $p_{D,k}^{(n_s)}$. The second approach considers the detection probabilities to be *time-varying and state dependent*, i.e. the detection probability also depends on the location in the field of view.

Time varying The probability of detection is considered constant in the field of view, however, it can vary over time. In that case, a weighted average of estimated probability of detection is used to parameterize the multi-sensor tracking filter.

$$p_{D,k}^{(n_s)} = \frac{1}{\sum_{i=1}^{J_k} w_k^{(i)}} \sum_{i=1}^{J_k} w_k^{(i)} \mu_{\beta,k}^{(n_s,i)} \tag{4.18}$$

where $\mu_{\beta,k}^{(n_s,i)}$ is the mean of the Beta distribution for sensor n_s and mixture component i at time k. The mean is computed using as equation 4.6.

Time varying and state dependent Now consider that the probability of detection depends on the state of the object, e.g. a lower detection probability when the object is further from the sensor. The p_D-CPHD filter estimates the probability of detection for each object in the environment, and the detection probability estimate for each component of the p_D-CPHD filter are shared with the tracking filter along with the weight, kinematic state and (kinematic) state covariance. The tracking filter requires an estimate for the detection probability for each object, for which the estimate in the neighborhood of the object's state is used, i.e. the closest component from the p_D-CPHD filter. To find the closest component, a distance measure is calculated between the kinematic state of the object in the tracking filter and between the components of the p_D-CPHD filter. Similarly to the component merging within the p_D-CPHD filter, the distance measure is inspired by the Hellinger distance. Distance is calculated between the kinematic states of i and j:

$$d_{ij} = \sqrt{1 - d_{i,j}^{\mathcal{N}}} \tag{4.19}$$

where $d_{i,j}^{\mathcal{N}}$ is defined in equation 4.17.

For each object i in the tracking filter, which is described by mean $m^{(i)}$ and covariance $P_k^{(i)}$, the distance $d_{i,j}$ from each $m^{(j)}, P^{(j)}$ in the p_D-CPHD filter is computed. Next, the component with minimum distance is found: $\hat{j} = \text{argmin}_j \; d_{i,j}$. The probability of detection $p_{D,k}^{n_s,\hat{j}}$ is assigned to the object i if the distance $d_{i,\hat{j}}$ is smaller than the threshold D', otherwise the estimate of the previous time is kept. To avoid tracking objects with very low detection probabilities, pruning is performed, requiring a minimum detection probability.

When the detection probability is considered to be time-varying and independent of the object's state, the filter is called a bootstrap-filter, following the naming as used in the single sensor bootstrap filters of [15] and [17]. In that case, the probability of detection for all object at time k is equal for all objects. Since, CPHD and LMB filters are considered here as tracking filters, this thesis calls these bootstrap-CPHD and bootstrap-LMB, respectively.

Furthermore, in the case where the probability of detection is time-varying and state dependent, the name local-bootstrap-filter is used, as the filter finds locally the closest component for retrieving the estimate for the probability of detection. Now, the filters are called local-bootstrap-CPHD and local-bootstrap-LMB.

4.4. EXPERIMENTS

The first section describes the experiment in a controlled environment, and the second section elaborates on the experiment with the TU Delft vehicle prototype (a Toyota Prius) with multiple dynamic objects in various weather and illumination conditions. To evaluate performance, the Optimal Sub Pattern Assignment (OSPA) metric is used, which is a commonly used metric to evaluate the performance of multi object filters [46]. The OSPA metric consists of two error components, namely the localization error and the cardinality error. The localization error quantifies how precise the state estimates are, whereas the cardinality error indicates how accurate it estimates the number of targets in the scene. The proposed filter variations from section 4.3.3 are compared with two

baselines, namely CPHD filter and LMB filter are used that both use a fixed detection probability for the sensors.

4.4.1. EXPERIMENTS IN FOG SIMULATOR

In the rain & fog simulator from Cerema (www.cerema.fr), experiments were conducted to investigate the influence of adverse weather on radar and camera in a controlled environment. Very dense fog, with a visibly range of 50 m, was injected in the environment (see Figure 4.3). The visibility range slowly increase, because the particles slowly dissipate. To simulate that the fog is slowly appearing (see Figure 4.2) in this experiment, the recorded sensor data and visibility data (by means of a transmissometer) has been reversed.

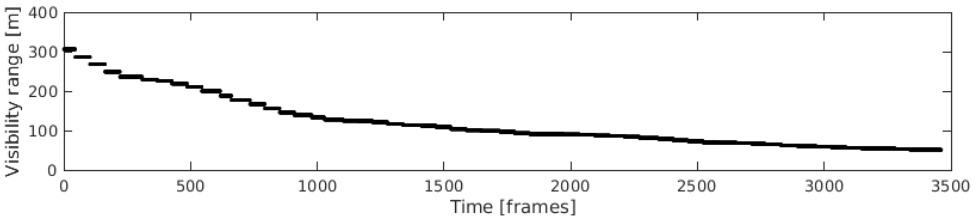


Figure 4.2: Measured visibilty range [m] by the transmissometer as function of time in frames.

The sensor setup consists of a radar and a monocular camera. The Continental ARS309-2 radar runs at 15 Hz and the Melexis MLX75411 HDR camera with a high dynamic range of 154 dB runs at 30 Hz. In this experiment, the camera images are temporally aligned with the radar frames, resulting in sensor data at a rate of 15 Hz. The recording consists of 3461 radar frames (corresponding to a duration of ≈ 4 minutes). In total, 19 radar frames (0.5% of total) were missing due to the CAN to USB converter.

A trihedral corner reflector is a common radar target because of its distinctive Radar Cross-Section (RCS) value (i.e. reflectivity). To have a common target for radar and vision, a vision-based detector is needed for the trihedral corner reflector, therefore a linear Support Vector Machine (SVM) using Histogram Oriented Gradient (HOG) features is trained on a data set with images in normal conditions (no fog or precipitation).

The ground truth of the stationary targets (see Figure 4.3) is obtained by a combination of manually measured 3D locations and annotated pixel locations in the image. The initial manually measured 3D locations are refined by minimizing the reprojection pixel error with the annotated pixel locations.

The object's kinematic state vector $x_k = [x, y, v_x, v_y]^T$ consists of the positions (x and y) and the velocities in x and y (v_x and v_y). As motion model a Discrete White Noise Acceleration (DWNA) model is used. The radar observation vector consists range (r), azimuth angle (α) and Doppler velocity ($\dot{r}$). The polar radar observations can be converted to Cartesian coordinates when the following conditions is valid [42]: $\frac{r \cdot \sigma_\alpha^2}{\sigma_r} < 0.4$ where σ_α and σ_r are the standard deviations of observation noise in azimuth and range, respectively. This method is called Converted Measurement Kalman Filter (CMKF) [42]. In this experiment, the standard deviation for the observation noise equals $\sigma_r = 0.25$ m, $\sigma_\alpha = 0.017$ rad and $\sigma_{\dot{r}} = 0.14$ m/s. For the vision-based detector, the measurement vec-

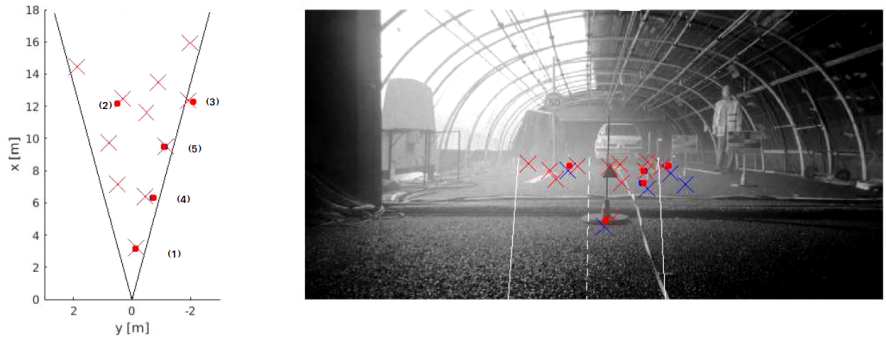


Figure 4.3: Experiment setup at the rain/fog simulator. The left figure shows the top view of the environment and the right image shows the camera view. The red crosses in both the top view and image represent the radar detection. Blue crosses represent the bottom center of the bounding box of the trihedral detector. The red dots represent the multi-sensor local-bootstrap-LMB tracks. The trihedral corner reflectors are numbered in the top view.

tor consists of the bottom center of the 2D bounding box (u, v). In case of a nonlinear observation model, the nonlinear observation model is linearized around the current mean estimate and covariance estimate, and an extended Kalman filter update is used. Standard deviations for the camera model are 6.15 pixels and 10 pixels for σ_u and σ_v, respectively. Both observations u and v are used for gating, and only u is used in the Kalman filter update.

For the filters, the maximum number of components is set to 100, the pruning thresholds equals 10^{-5}, the maximum cardinality equals 10, and a value of 0.5 is used as merging threshold. In addition, the variance of the Beta distribution is increased with factor $k_\beta = 1.1$ in the prediction step. Furthermore, gating is used with a gate probability of 0.99. The Gaussian mixture births are located at the expected locations of the trihedral corner reflectors with a weight of 0.03. Furthermore based on observations, the Poisson average rate of uniform clutter per scan is assumed to be constant and equal to 5 for the radar and equal to 1 for the camera. The OSPA metric with parameters $p = 2$ and $c = 1$ m is used to quantify the performance. For assigning the detection probability in the local bootstrap filters, a distance threshold D' of 0.3 is used (see section 4.3.3). As a prior for the birth model in the CPHD filter, a prior for $\beta(a, s, t)$ with $s = 9$ and $t = 1$ for the radar detector and the camera detector. As a baseline, the multi-sensor implementation of the CPHD and LMB filters are used. Both filters use fixed detection probabilities; the detection probability for radar is equal to 0.99, and that for the camera detector is 0.95.

In table 4.1, the average OSPA is summarized. The OSPA distance is decomposed into localization and cardinality errors in table 4.1 in order to compare the performance of the various filters. Furthermore, the rows in the table are sorted based on the OSPA distance column, with the best performing filter at the top (local-bootstrap-LMB). It can be seen that the local-bootstrap-LMB filter outperforms the other filters, because the cardinality (i.e. the estimated number of objects) is more accurate. In terms of localization accuracy, the LMB filters show similar localization errors and for the CPHD filters the local-bootstrap-CPHD performs better than the bootstrap-CPHD and CPHD filter.

Table 4.1: Average OSPA in m for all filters (lower is better). Bold numbers indicate best performance for each column.

	OSPA Distance	OSPA Localisation	OSPA Cardinality
local-bootstrap-LMB (proposed)	**0.322**	0.303	**0.037**
bootstrap-LMB (proposed)	0.469	0.304	0.306
local-bootstrap-CPHD (proposed)	0.500	0.348	0.296
LMB (baseline)	0.511	**0.299**	0.381
bootstrap-CPHD (proposed)	0.623	0.482	0.345
CPHD (baseline)	0.667	0.447	0.480

Figure 4.4 demonstrates the detection probability of the third trihedral corner reflector (see Figure 4.3) estimated by the local-bootstrap-LMB filter. At the beginning of the scenario, the probability of detection for radar and camera is both larger than 0.8, whereas at the end of the scenario the detection probability for camera dropped to values smaller than 0.3 due to the reduced visibility.

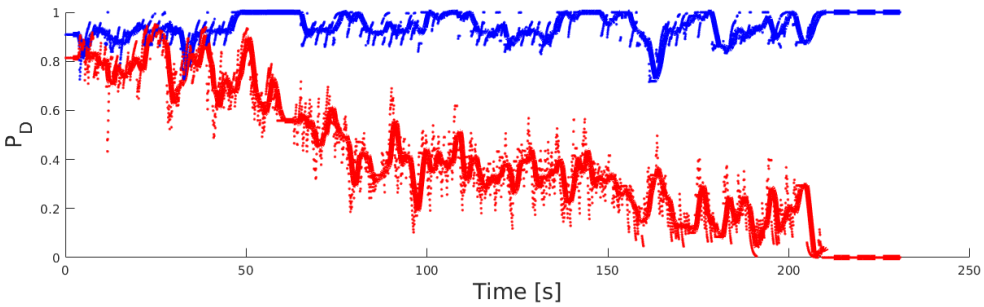


Figure 4.4: The probability of detection of trihedral corner reflector three in the multi-sensor local-bootstrap-LMB filter. The blue dots represent the p_D for the radar, and the red dots represent the p_D for the camera. Furthermore, the solid lines present the averaged probability of detection of a period of 2 seconds.

4.4.2. RECORDED DATA WITH PRIUS

In the fog simulator experiments, it was shown that the best performing filter was the local-bootstrap-LMB filter, therefore this thesis now evaluates the performance of the proposed local-bootstrap-LMB tracker on recorded data of the prototype vehicle. The sensors of the prototype vehicle consist of a Velodyne HDL-64E lidar (on the roof), a Continental ARS430 radar (behind the front bumper), and a camera (UI-3060CP Rev. 2) which is mounted behind the windscreen at the height of the rear view mirror. These sensors are calibrated extrinsically using [47]. The vehicle is equipped with an Inertial Measurement Unit (IMU) and GPS receiver for egomotion estimation and localization.

The prototype vehicle is parked on a straight road, with a pavement on the right-hand side of the car. Four different staged scenarios with three or four pedestrians are performed and recorded, where the crossing pedestrians are at a distances of 15 m and 30 m, and the longitudinal moving pedestrians walk up to 100 m from the car. To obtain

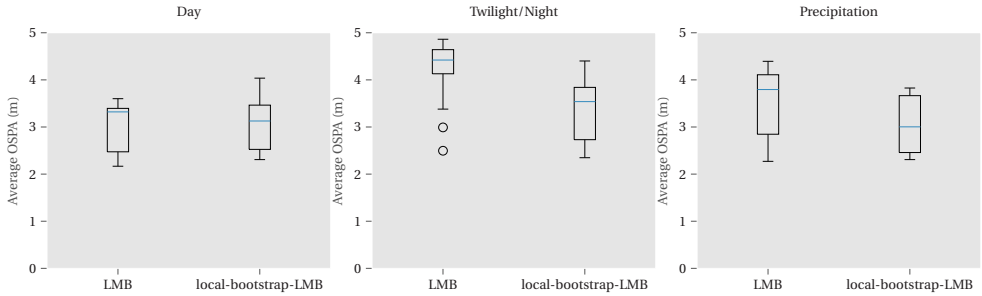


Figure 4.5: Box plots with average OSPA distance results. For the proposed filter, the results are obtained using a uniform distribution (i.e. $\beta(a,1,1)$) for the initialization of the unknown detection probabilities in the birth model of the p_D-CPHD filter.

ground truth annotations, each pedestrian wears on his head a wearable Emlid Reach M+ with a Tallysman multi-GNSS antenna. For each reading of the GPS devices, the GPS coordinate is transformed to vehicle coordinate frame using the vehicle's GPS and IMU data. Then, the distance to the points in the lidar point cloud is computed, and the closest lidar point is assigned as ground truth location, if the distance is smaller than 2 m.

In total, 36 recordings were recorded in various weather conditions (with/without precipitation) and illumination conditions; 6 scenarios of those recordings were recorded when there was precipitation, 24 scenarios were recorded in twilight/night conditions, i.e. after sunset. Figure 4.6 shows the camera images for each recorded scenario.

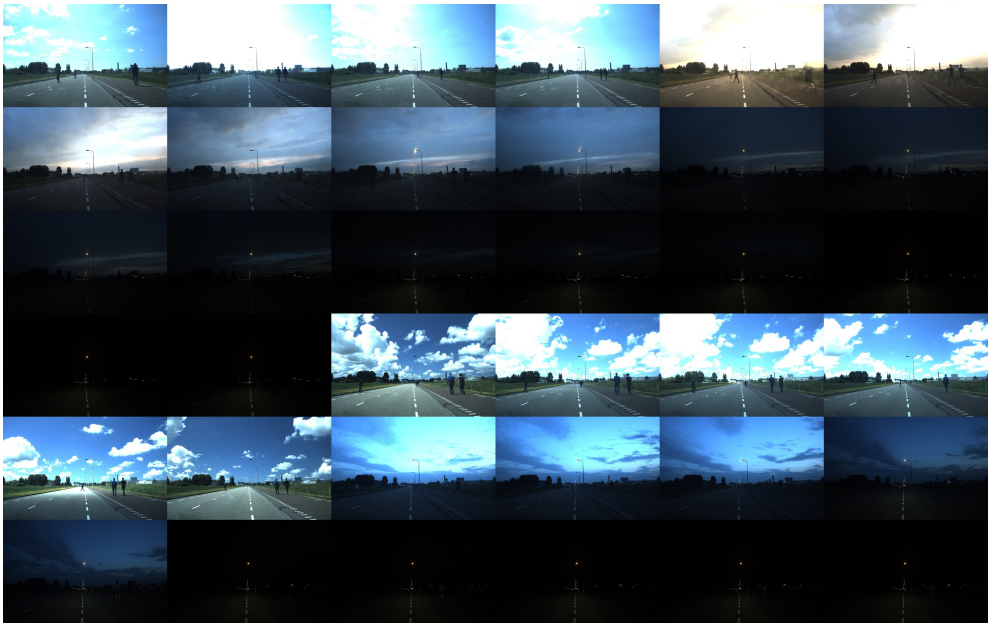


Figure 4.6: Camera images showing the conditions in which the scenarios were recorded.

To the detect the moving pedestrians in the radar and camera data, a detector is required for moving road users. To detect road users in the camera images, the Single Shot Multibox Detector (SSD) is used [26]. As measurements for the camera detector, the bottom center of 2D bounding box are fed into the object tracker(s). The moving target detector for radar data consist of three three pre-processing steps. Radar detections are filtered for detections with a Radar Cross Section (RCS) value smaller than -20 dBm^2 and, in addition, only detections that have a minimum absolute Doppler velocity value of 0.02 m/s are kept. After that, DBSCAN [48] is used to extract the clusters using a distance threshold of 0.7 and with minimum number of points set to 1. For each cluster, the mean of the cluster is computed and the mean is converted to range and azimuth which serve as measurements for the tracker.

Within the tracking filters, the state vector of the object is represented by the 3D positions (x, y, z), and the 3D velocities (v_x, v_y, v_z). As motion model, a Discrete White Noise Acceleration (DWNA) model is used. The non-linear observation models that object's states to the observations space of the radar and the SSD detector, are approximated using the unscented transformation, i.e. an Unscented Kalman Filter is used [49]. For the radar sensor, the standard deviation of the observation noise is considered to be 0.6 m for the range and 2° for the azimuth angle. Furthermore, a pinhole camera model is used as an observation model with a standard deviation of the observation noise equal to 10 and 15 pixels for the horizontal pixel location and the vertical pixel location, respectively. To initialize new tracks, a birth approach is required. For the p_D-CPHD filter, the birth approach from [50] is used and the adaptive birth procedure from [10] is used for initialization for the LMB filters. New tracks are initialized based on radar observations as the detections from the SSD detector (based on the monocular camera) lack depth information. The maximum cardinality is set to 20 objects. Furthermore, a pruning threshold of 10^{-4} is used, and the Gaussian components of each track are merged on the basis of a threshold of 1. The tracks in the local-bootstrap-LMB filter that have a probability of detection > 0.5 for one of the sensors are kept and the tracks with a maximum probability of detection < 0.5 are pruned. For the p_D-CPHD filter, a maximum number of 100 components is used as the capping threshold. In the LMB and local-bootstrap-LMB filter, the number components for each track is limited to 20. A distance threshold D' of 0.3 is used, for assigning the detection probabilities in the local-bootstrap-LMB filter.

In this experiment, this thesis investigates the performance of the LMB filter, which showed better performance than the CPHD in section 4.4.1, that is optimized offline for daylight conditions without precipitation with the local-bootstrap-LMB filter that online estimates the probability of detection. The adaptive tracking filter is expected to be more robust in various weather and illumination conditions than the tracking filter with fixed hyper-parameters for the probability of detection and clutter rate for both sensors. Two scenarios (see Appendix C) that were recorded in daylight conditions without precipitation are used as training dataset for finding the optimal hyper parameters. These optimal parameters for the detection probabilities and clutter rates are found using the minimization of the summed OSPA distance in all frames of both scenarios using a derivative-free global optimization routine (adapted version of [51]) from the dlib toolkit [52]. For the camera detector, a detection probability of 0.73 with a clutter rate of 1.9 were found as optimal parameters. For the radar detector, the detection probability

equals 0.6 and the clutter rate equals 4.9. The optimal clutter rates are known to both the LMB and local-bootstrap-LMB filters.

For each scenario, the average OSPA is calculated across all frames with parameters $p = 2$ and $c = 5.0$ m to obtain a single metric for each scenario. This thesis considers that the conditions are rainy when during the recording of the scenario rain was observed and when the hourly weather classification data of the nearby Rotterdam weather station of the Royal Dutch Meteorological Institute (KNMI) reports rain.

Figure 4.5 shows the result for the LMB and local-bootstrap-LMB filter in three conditions. For the p_D-CPHD filter, a uniform distribution for the detection probabilities is used in the birth model, i.e. $\beta(a,1,1)$. The label *Day* means that the recording was made in daylight conditions (including precipitation) and the label *Twilight/Night* indicates that the recording was made after sunset. Finally, *Precipitation* describes if the sensor data was recorded in rainy conditions, regardless of illumination conditions. In *Day* conditions, both filter show similar performance as both median values are very close. In *Twilight/Night* scenarios, the LMB filter shows a decrease in performance compared to *Day* conditions, as the median value of the average OSPA distances increases to 4.4 m. The local-bootstrap-LMB filter shows better results than the LMB filter, as the median value is approximate 3.5 m and the whisker boundaries are lower. Two outliers are identified in *Twilight/Night* conditions from which the outlier with the lowest OSPA distance can be explained as it is recorded shortly after sunset.

To investigate why the local-bootstrap-LMB performs better for *Twilight/Night* and *Precipitation* conditions, the cardinality component of the OSPA metric in these two conditions is inspected. In Twilight/Night conditions, the median value of the average OSPA cardinality errors equals 4.4 m for the LMB filter, and 2.7 for the local-bootstrap-LMB filter. In *Precipitation*, the median values are 3.7 m and 2.4 m for the LMB filter and the local-bootstrap-LMB filter, respectively. The lower median values for the local-bootstrap-LMB filter show that the cardinality estimates are better, meaning that this filter is better at estimating the number of objects.

4.5. DISCUSSION

In section 4.4.2, a uniform distribution is used to initialize the detection probabilities in the p_D-CPHD filter. Even with a uniform distribution, the local-bootstrap-LMB filter showed better performance than the LMB filter with fixed detection probabilities. However if more optimal values for the initialization of the detection probabilities are used, improved results can be expected. For instance, the radar sensor is more robust than the camera in adverse weather conditions, therefore in the birth model a Beta distribution with higher mean and lower standard deviation could be used.

The optimal settings found in parameter optimization for detection probability and clutter rate for the radar detector are equal to 0.6 and 4.9 per frame, respectively. False positives detections could result in tracked objects if the false positive is persistent. For instance, undesired detections due to multi path reflections or detection due to moving vegetation could result in a tracked objects. The local-bootstrap-LMB filter could be more sensitive to persistent false positives as the filter estimates a low detection probability for camera and it keeps these tracked objects alive. Instead, the LMB filter with fixed detection probabilities suppresses these objects as it expects camera detections. To

reduce the number of false positives, the minimum number of points in DBSCAN clustering could increased, however this might result in missed detections at large distances. The minimum absolute value of the Doppler velocity threshold could be increased to decrease the number of false positives; however, this might lead to a decrease in the detection probability. Furthermore, more advanced signal processing techniques could be used to reduce false positives in the radar point cloud. Finally, a state-of-the-art detector for moving objects (e.g. [53]) could be used, as these typically have a higher detection probability and less false positives.

For the camera detector, a detection probability of 0.73 with a clutter rate of 1.9 per frame were found as optimal parameters. The detection probability is low; however, objects appear to be quite small in the image when the pedestrian is moving up to 100 m from the camera. Moreover, it is the recommended to investigate why the found optimal clutter rate is high. A possible explanation could be that a pinhole camera model is not optimal as observation model the camera detector or that the observation noise was set too low. Furthermore, the challenge with using a monocular camera is that there is a lack of depth information. This means that 3D points that are along the same ray have similar pixel coordinates, which might result in that objects further away are associated with a bounding box that belongs to a closer object. One way to deal with that is to add a second camera in a stereo camera setup and use stereo imaging to infer depth or use the bounding box size to improve the association.

In adverse weather conditions, detecting road users in camera images is more challenging, therefore the number of false negatives increases. For instance, the images in *Twilight/Night conditions* (Figure 4.6) show that for a human it is hard to detect the road users in the images. To reduce the effect of the increased number of false negatives, the local-bootstrap-LMB filter estimates the detection probabilities, which are subsequently used by the tracking filter to improve the cardinality estimates and thereby the tracking performance. Apart from affected detection probabilities, also the number of false positives could be dependent on the weather and illumination conditions, therefore adding an estimator for clutter rate to the bootstrap filter of Figure 4.1 might be a direction for future research. In addition, it should be investigated if the observation noise is depending on weather conditions. Furthermore, in experiments in a controlled environment, it has been shown that the detection probability for both sensors can be estimated for an object. These estimated detection probabilities could be used to assess the quality of the sensing. For instance, in Figure 4.4 the detection probability for the camera detector drops below 0.3 in dense fog, which could indicate that this camera detector is not working as expected.

4.6. CONCLUSION

In this chapter, a tracking filter that deals with affected detection performance due to adverse weather conditions was proposed. The experiment in a fog simulator showed that local-bootstrap-LMB filter performs better than the local-bootstrap-CPHD filter according to the OSPA metric. Furthermore, the experiment recorded with the prototype vehicle showed that the proposed local-bootstrap-LMB filter outperforms the LMB with fixed detection probabilities. For instance, the median value of the average OSPA distances reduces from 4.4 m to 3.5 m in *Twilight/Night* conditions. As a consequence

of online estimation of the detection probabilities, the cardinality estimates in case of adverse weather conditions improved with respect to the LMB filter that uses fixed parameters which were optimized on daylight recordings. Future work involves adding an estimator for the clutter rates to the bootstrap filter in order to estimate the number of false positives for each sensor.

REFERENCES

[1] J. Domhof, J. F. Kooij, and D. M. Gavrila, *Adaptive multi-sensor object tracking filter in adverse weather conditions,* submitted to IEEE Transactions on Intelligent Vehicles (2022).

[2] E. Marti, M. A. de Miguel, F. Garcia, and J. Perez, *A review of sensor technologies for perception in automated driving,* IEEE Intelligent Transportation Systems Magazine **11**, 94 (2019).

[3] S. Sivaraman and M. M. Trivedi, *Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis,* IEEE Transactions on Intelligent Transportation Systems **14**, 1773 (2013).

[4] S. S. Blackman and R. Popoli, *Design and analysis of modern tracking systems* (Artech House Publishers, 1999).

[5] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe, *Multi-target tracking using joint probabilistic data association,* in *1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes* (IEEE, 1980) pp. 807–812.

[6] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, *Sonar tracking of multiple targets using joint probabilistic data association,* IEEE journal of Oceanic Engineering **8**, 173 (1983).

[7] R. P. Mahler, *Statistical multisource-multitarget information fusion* (Artech House, Inc., 2007).

[8] R. P. S. Mahler, *Multitarget bayes filtering via first-order multitarget moments,* IEEE Transactions on Aerospace and Electronic Systems **39**, 1152 (2003).

[9] R. Mahler, *PHD filters of higher order in target number,* IEEE Transactions on Aerospace and Electronic Systems **43** (2007).

[10] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, *The labeled multi-Bernoulli filter,* IEEE Transactions on Signal Processing **62**, 3246 (2014).

[11] L. Lamard, R. Chapuis, and J. P. Boyer, *Multi target tracking with CPHD filter based on asynchronous sensors,* in *2013 16th International Conference on Information Fusion (FUSION)* (IEEE, 2013) pp. 892–898.

[12] F. Kunz, D. Nuss, J. Wiest, H. Deusch, S. Reuter, F. Gritschneder, A. Scheel, M. Stübler, M. Bach, P. Hatzelmann, *et al.*, *Autonomous driving at Ulm university: A modular, robust, and sensor-independent fusion approach,* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2015) pp. 666–673.

[13] A. Scheel, S. Reuter, and K. Dietmayer, *Vehicle tracking using extended object methods: An approach for fusing radar and laser,* in *IEEE International Conference on Robotics and Automation (ICRA), 2017* (IEEE, 2017) pp. 231–238.

[14] R. P. Mahler, B.-T. Vo, and B.-N. Vo, *CPHD filtering with unknown clutter rate and detection profile,* IEEE Transactions on Signal Processing **59**, 3497 (2011).

[15] M. Beard, B.-T. Vo, and B.-N. Vo, *Multitarget filtering with unknown clutter density using a bootstrap GMCPHD filter,* IEEE Signal Processing Letters **20**, 323 (2013).

[16] B.-T. Vo, B.-N. Vo, R. Hoseinnezhad, and R. P. Mahler, *Robust multi-Bernoulli filtering,* IEEE Journal of Selected Topics in Signal Processing **7**, 399 (2013).

[17] S. H. Rezatofighi, S. Gould, B. T. Vo, B.-N. Vo, K. Mele, and R. Hartley, *Multi-target tracking with time-varying clutter rate and detection profile: Application to time-lapse cell microscopy sequences,* IEEE Transactions on Medical Imaging **34**, 1336 (2015).

[18] Y. Punchihewa, B.-T. Vo, B.-N. Vo, and D. Y. Kim, *Multiple object tracking in unknown backgrounds with labeled random finite sets,* arXiv preprint arXiv:1706.01584 (2017).

[19] G. Papa, P. Braca, S. Horn, S. Marano, V. Matta, and P. Willett, *Multisensor adaptive Bayesian tracking under time-varying target detection probability,* IEEE Transactions on Aerospace and Electronic Systems **52**, 2193 (2016).

[20] F. Meyer, P. Braca, F. Hlawatsch, M. Micheli, and K. D. LePage, *Scalable adaptive multitarget tracking using multiple sensors,* in *2016 IEEE Globecom Workshops (GC Wkshps)* (IEEE, 2016) pp. 1–6.

[21] J. Pallauf, J. Wagner, and F. P. León, *Evaluation of state-dependent pedestrian tracking based on finite sets,* IEEE Transactions on Instrumentation and Measurement **64**, 1276 (2015).

[22] M. Bijelic, T. Gruber, and W. Ritter, *A benchmark for lidar sensors in fog: Is detection breaking down?* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2018) pp. 760–767.

[23] M. Kutila, P. Pyykönen, M. Jokela, T. Gruber, M. Bijelic, and W. Ritter, *Benchmarking automotive lidar performance in arctic conditions,* in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2020) pp. 1–8.

[24] M. Bijelic, T. Gruber, and W. Ritter, *Benchmarking image sensors under adverse weather conditions for autonomous driving,* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2018) pp. 1773–1779.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *You only look once: Unified, real-time object detection,* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) pp. 779–788.

[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, *SSD: Single shot multibox detector,* in *European Conference on Computer Vision* (Springer, 2016) pp. 21–37.

[27] T. Meinhardt, A. Kirillov, L. Leal-Taixe, and C. Feichtenhofer, *Trackformer: Multi-object tracking with transformers,* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022) pp. 8844–8854.

[28] D. Frossard and R. Urtasun, *End-to-end learning of multi-sensor 3D tracking by detection,* in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2018) pp. 635–642.

[29] G. Brasó and L. Leal-Taixé, *Learning a neural solver for multiple object tracking,* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020) pp. 6247–6257.

[30] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, *Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather,* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020) pp. 11682–11692.

[31] M. Hahner, C. Sakaridis, D. Dai, and L. Van Gool, *Fog simulation on real lidar point clouds for 3D object detection in adverse weather,* in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021) pp. 15283–15292.

[32] M. J. Mirza, C. Buerkle, J. Jarquin, M. Opitz, F. Oboril, K.-U. Scholl, and H. Bischof, *Robustness of object detectors in degrading weather conditions,* in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* (IEEE, 2021) pp. 2719–2724.

[33] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, *Virtual worlds as proxy for multi-object tracking analysis,* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) pp. 4340–4349.

[34] M. Pavlic, G. Rigoll, and S. Ilic, *Classification of images in fog and fog-free scenes for use in vehicles,* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2013) pp. 481–486.

[35] N. Hautiére, J.-P. Tarel, J. Lavenant, and D. Aubert, *Automatic fog detection and estimation of visibility distance through use of an onboard camera,* Machine Vision and Applications **17**, 8 (2006).

[36] Z. Zhang, H. Ma, H. Fu, and C. Zhang, *Scene-free multi-class weather classification on single images,* Neurocomputing **207**, 365 (2016).

[37] S. G. Narasimhan and S. K. Nayar, *Contrast restoration of weather degraded images,* IEEE Transactions on Pattern Analysis and Machine Intelligence **25**, 713 (2003).

[38] K. Garg and S. K. Nayar, *Vision and rain,* International Journal of Computer Vision **75**, 3 (2007).

[39] M. Hnewa and H. Radha, *Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques,* IEEE Signal Processing Magazine **38**, 53 (2020).

[40] M. Hassaballah, M. A. Kenk, K. Muhammad, and S. Minaee, *Vehicle detection and tracking in adverse weather using a deep learning framework,* IEEE Transactions on Intelligent Transportation Systems **22**, 4230 (2020).

[41] M. Bijelic, P. Kysela, T. Gruber, W. Ritter, and K. Dietmayer, *Recovering the unseen: Benchmarking the generalization of enhancement methods to real world data in heavy fog.* in *CVPR Workshops* (2019) pp. 11–21.

[42] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and data fusion* (YBS publishing, 2011).

[43] R. Mahler, *'Statistics 102' for multisource-multitarget detection and tracking,* IEEE Journal of Selected Topics in Signal Processing **7**, 376 (2013).

[44] B.-T. Vo, B.-N. Vo, and A. Cantoni, *Analytic implementations of the cardinalized probability hypothesis density filter,* IEEE Transactions on Signal Processing **55**, 3553 (2007).

[45] E. Hellinger, *Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen.* Journal für die reine und angewandte Mathematik **1909**, 210 (1909).

[46] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, *A consistent metric for performance evaluation of multi-object filters,* IEEE Transactions on Signal Processing **56**, 3447 (2008).

[47] J. Domhof, J. F. Kooij, and D. M. Gavrila, *A joint extrinsic calibration tool for radar, camera and lidar,* IEEE Transactions on Intelligent Vehicles **6**, 571 (2021).

[48] T. N. Tran, K. Drab, and M. Daszykowski, *Revised DBSCAN algorithm to cluster data with dense adjacent clusters,* Chemometrics and Intelligent Laboratory Systems **120**, 92 (2013).

[49] E. A. Wan and R. Van Der Merwe, *The unscented kalman filter for nonlinear estimation,* in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)* (IEEE, 2000) pp. 153–158.

[50] S. Reuter, D. Meissner, and K. Dietmayer, *Multi-object tracking at intersections using the cardinalized probability hypothesis density filter,* in *2012 15th International IEEE Conference on Intelligent Transportation Systems* (IEEE, 2012) pp. 1172–1177.

[51] C. Malherbe and N. Vayatis, *Global optimization of Lipschitz functions,* in *International Conference on Machine Learning* (PMLR, 2017) pp. 2314–2323.

[52] D. E. King, *Dlib-ml: A machine learning toolkit,* Journal of Machine Learning Research **10**, 1755 (2009).

[53] A. Palffy, J. Dong, J. F. Kooij, and D. M. Gavrila, *CNN based road user detection using the 3D radar cube,* IEEE Robotics and Automation Letters **5**, 1263 (2020).

5

CONCLUSION

This chapter summarizes the main technical findings of each chapter, then turns to more general discussion of the thesis research, including future work.

5.1. SENSOR SELECTION

To select the sensor setup for a certain automated driving application, the tracking performance of a candidate sensor setup needs to be estimated in an early design phase. It is important to know if a certain sensor setup can comply with the system requirements. Therefore, tracking performance limits of a sensor setup are required in an early (design) phase. Chapter 2 presented a framework to predict tracking performance limits using the Cramér-Rao lower bound (CRLB) for multi-sensor setups. Numerical studies on an example sensor set consisting of state-of-the art automotive sensors showed that in close range stereo vision performs well, and accurate positional x and y estimates can be found for the lidar sensor. Fusing data from radar and lidar showed the best performance in highway environments and any two sensor combination provides a CRLB σ_x and a CRLB σ_y of less than 0.1 m within an observation time of 0.5 seconds in urban environments. The framework can be used to design a redundant perception system, since the impact of a sensor failure on the tracking performance can be predicted. Comparing the performance limits to the system requirements, early decisions can be made if the perception system is still able to continuously operate or not.

5.2. EXTRINSIC SENSOR CALIBRATION

Chapter 3 addressed the problem of joint extrinsic sensor calibration for lidar, camera and radar sensors. Three configurations to optimize the set of sensors were identified and evaluated. Minimally Connected Pose Estimation (MCPE) computes the sensor-to-sensor transformations with respect to a single reference sensor. Fully Connected Pose Estimation (FCPE) optimizes the transformations between all sensor pairs using a constraint that forces loop closure. Finally, Pose and Structure Estimation (PSE) jointly estimates sensor poses as well as the calibration board poses. Both the FCPE and the PSE

configuration showed better performance than the MCPE configuration, since the former configuration includes all pairwise error terms in combination with a loop closure constraint and the latter configuration simultaneously estimates the calibration board poses and sensor poses. The experiments with the TU Delft prototype vehicle with lidar, camera and stereo camera suggest that the FCPE configuration performs better than the PSE configuration. The median RMSE is less than 2 cm for lidar to camera, approximately 2 cm for lidar to radar and approximately 2.5 cm for camera to radar when ten calibration boards are used.

Furthermore, the sensors are calibrated with respect to the body reference frame of the robot in this chapter. For that, two requirements are identified, namely the need for an external sensor and a set of 3D reference points. Two methods are compared to determine the pose of the body reference frame in an external sensor. For *absolute calibration*, the experiments suggest that the method *Human labeling* using *geometrical shape fitting* provides more accurate results than the method *Markers*. The main difference between the method *Markers* and the method *Human labeling* can be found in the rotation error around the vertical axis, which can be explained by the fact that accurate marker placement is challenging for the *Markers* method.

The software is available as an open-source extrinsic calibration tool with bindings to Robot Operating System (ROS). It consists of three configurations to estimate the sensor poses from simultaneous detections of multiple calibration board locations.

5.3. OBJECT TRACKING

Chapter 4 of this thesis addresses object tracking in challenging weather and illumination conditions, for which the multi-modal bootstrap filter is proposed, comprising of a Cardinalized Probability Hypothesis Density (CPHD) filter that can deal with an unknown detection probability (p_D-CPHD filter) [1] that serves as an estimator of the probability of detection and a tracking filter. Experiments in a controlled environment with artificial generated fog and experiments that were recorded with the TU Delft prototype vehicle showed that according to the Optimal Sub Pattern Assignment (OSPA) [2] metric, the proposed filter outperforms tracking filters with fixed detection probabilities. Experiments with the TU Delft prototype vehicle illustrated that estimating the detection probabilities and using these estimates in the tracking filter improves the tracking for scenarios recorded in adverse illumination (twilight/night) and precipitation conditions. Improved cardinality estimates in case of adverse weather conditions were found with respect to the Labeled Multi-Bernoulli (LMB) [3] filter using fixed parameters that were optimized on daylight recordings. Furthermore, the detection probability for radar and camera was visualized for a single target in a controlled environment. These estimated detection probabilities could serve a measure for the sensing quality as it indicates the quality of data entering the fusion center.

5.4. FINAL REMARKS

The aim of the thesis is to develop methods and algorithms for developing a robust perception system that is able to deal with adverse weather and illumination conditions.

A systematic approach to evaluate the tracking performance limits for sensor setups

are computed in chapter 2, for static objects in front of the vehicle observed by different sensor setups. Future work should address the problem of moving objects, and thereby the observation noise and the observation matrices vary with time as these matrices depend on the object's position.

Robust perception cannot be achieved by using a single sensing modality, as in that case there is a common mode of failure. Therefore, there is a need for having an additional sensing modality. Each of the three sensing modalities (lidar, radar and vision) use a different part of electromagnetic spectrum, therefore environmental conditions (rain, fog, snow, etc.) do not affect these sensor in the same way. The use of multiple sensing modalities reduces the probability of a common mode of failure due to adverse weather and illumination conditions. For redundancy purposes, the perception system should comprise of at least two sensors that completely observe the 3D object and have an overlapping field of view (FoV). If one of the sensors fails the system can still operate using the other sensor. In case of sensor failure, the system can observe that one of the sensors is detecting the object and the other sensor not. Therefore for tracking moving objects, a sensor setup with two sensors including a monocular camera is not recommended as a monocular camera is not able to measure depth directly, although depth from monocular images [4, 5] might be accurate enough in the future.

The calibration tool proposed in chapter 3 uses a styrofoam calibration target with four holes, and including a single trihedral corner reflector as target. To improve the detectability of the calibration target for a monocular camera, the calibration board design can be improved to include fiducial markers [6].

The thesis addresses the problem of joint extrinsic calibration of camera, lidar and radar sensors in an offline setting, which is important for vehicles leaving the factory. However, while being in operation, there is a need for an online calibration method to verify if the sensor pose is still correct, or it has been misaligned. Misalignment could happen after servicing, due to vibrations or by impact. Classical methods for online calibration of lidar and camera (e.g. [7, 8]), and more recently online extrinsic calibration is performed using deep learning techniques [9, 10]. Future work should involve online extrinsic and intrinsic calibration for multi-sensor setups consisting of lidar, camera and radar sensors. Online calibration results should be monitored, to confirm that the sensors are not misaligned. Without well calibrated sensors, there is no common coordinate frame in which the sensor data can be expressed and fused.

In chapter 4, a sensor setup consisting of a radar and a camera is used for object tracking in adverse weather conditions. The challenge with the radar - monocular camera setup is that the monocular camera does not provide depth estimates. Without these depth measurements, the monocular camera cannot initialize any new objects, unless the object's size is used to infer the depth. Due to the lack of depth estimates, the state of the object cannot be estimated solely on the monocular camera, as the state is not fully observable. Furthermore, the lack of depth accurate information makes it also harder to monitor the health of the radar sensor, because the object cannot be tracked based on monocular camera only. In case of malfunctioning radar sensor, the tracking filter is not able to track based on camera only, and the estimator of the probability of detection cannot estimate the detection probabilities. In order to improve the monitoring of the sensing quality of the radar, the monocular camera could be replaced with a stereo cam-

era or a lidar. Alternatively, deep learning techniques can be applied to infer depth from monocular vision [4, 5].

To assess the reliability of all the perception sensors in real time, one parameter to monitor is the detection probability of each of the sensors. Apart from the detection probability, there are other parameters that needs to be monitored that can serve as a measure for the sensing quality. In other words, these parameters indicate if the perception system works as expected. Parameters such as the observation noise, clutter and frame rate of each of the individual sensors need be monitored on a regular basis to determine the sensors still works as expected. Furthermore, this thesis addressed the problem of a time-varying and state dependent detection probability for object tracking. Other parameters, e.g. clutter and observation noise, could also be time-varying and state dependent, therefore similar techniques need to be developed with the aim to create a robust perception system dealing adverse weather and illumination conditions.

Deep learning techniques are nowadays also used for object tracking [11–13]. Training these models require large data sets for training and testing, and especially data in adverse weather conditions are more challenging to record, as these are more rare. To train and test deep learning models for object tracking, the research community requires more specialized data sets on adverse weather conditions, similar to [14]. Alternatively, techniques to 'simulate' adverse weather conditions on radar, lidar and camera data in existing datasets are needed. For instance, [15] simulate fog on lidar point clouds, and [16] render rain in camera images.

While important challenges remain, this thesis serves as a stepping stone to more robust perception using heterogeneous sensors, contributing to safer intelligent vehicles in adverse weather and illumination conditions.

REFERENCES

[1] R. P. Mahler, B.-T. Vo, and B.-N. Vo, *CPHD filtering with unknown clutter rate and detection profile,* IEEE Transactions on Signal Processing **59**, 3497 (2011).

[2] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, *A consistent metric for performance evaluation of multi-object filters,* IEEE Transactions on Signal Processing **56**, 3447 (2008).

[3] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, *The labeled multi-Bernoulli filter,* IEEE Transactions on Signal Processing **62**, 3246 (2014).

[4] Y. Ming, X. Meng, C. Fan, and H. Yu, *Deep learning for monocular depth estimation: A review,* Neurocomputing **438**, 14 (2021).

[5] F. Khan, S. Salahuddin, and H. Javidnia, *Deep learning-based monocular depth estimation methods—a state-of-the-art review,* Sensors **20**, 2272 (2020).

[6] J. Beltrán, C. Guindel, F. García, *et al.*, *Automatic extrinsic calibration method for lidar and camera sensor setups,* IEEE Transactions on Intelligent Transportation Systems (2022).

[7] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, *Automatic extrinsic calibration of vision and lidar by maximizing mutual information,* Journal of Field Robotics **32**, 696 (2015).

[8] J. Levinson and S. Thrun, *Automatic online calibration of cameras and lasers.* in *Robotics: Science and Systems*, Vol. 2 (2013).

[9] N. Schneider, F. Piewak, C. Stiller, and U. Franke, *RegNet: Multimodal sensor registration using deep neural networks,* in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2017) pp. 1803–1810.

[10] G. Iyer, R. K. Ram, J. K. Murthy, and K. M. Krishna, *Calibnet: Geometrically supervised extrinsic calibration using 3D spatial transformer networks,* in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2018) pp. 1110–1117.

[11] T. Meinhardt, A. Kirillov, L. Leal-Taixe, and C. Feichtenhofer, *Trackformer: Multi-object tracking with transformers,* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022) pp. 8844–8854.

[12] D. Frossard and R. Urtasun, *End-to-end learning of multi-sensor 3D tracking by detection,* in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2018) pp. 635–642.

[13] G. Brasó and L. Leal-Taixé, *Learning a neural solver for multiple object tracking,* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020) pp. 6247–6257.

[14] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, *Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather,* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020) pp. 11682–11692.

[15] M. Hahner, C. Sakaridis, D. Dai, and L. Van Gool, *Fog simulation on real lidar point clouds for 3D object detection in adverse weather,* in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021) pp. 15283–15292.

[16] S. S. Halder, J.-F. Lalonde, and R. de Charette, *Physics-based rendering for improving robustness to rain,* in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019) pp. 10203–10212.

ACKNOWLEDGEMENTS

My PhD period was a challenging period in my life. Fortunately, I have had help from many people, therefore this chapter is devoted to thank some of them in more detail.

First, I would like to thank Prof. Dr. ir. Pieter Jonker and Dr. ir. Riender Happee for giving me the opportunity to do a PhD. When I started as a PhD student, only few people were working on automated driving in our group and faculty. At the time, our group was working on equipping the car with automotive sensors to develop a research platform for automated driving. Furthermore, I am thankful the valuable lessons and experiences that I gained by participating in the Wepods project.

In September 2017, Prof. Dr. Dariu M. Gavrila and Dr. Julian F.P. Kooij officially took over the supervision from Pieter and Riender. I would like to thank you both for your scientific guidance. During the time we worked together, you have asked me countless detailed questions, which helped me to improve my research. Looking back, I think it was a missed opportunity that it was not possible to work with you as supervisors from the start of my PhD, since the scientific output would be have been higher. I am grateful for the time we have worked together. Thank you Prof. Dr. Dariu M. Gavrila, for your application driven research approach, which is important since it results in identifying and understanding all important factors related to the automated driving platform. Thank you for organizing and facilitating the necessities and circumstances that are required for doing research on intelligent vehicles. The visits to Daimler R&D were interesting and informative, especially the appendix of these visits is something that I will never forget: skiing and Kuhhandel. Dr. Julian F.P. Kooij thanks for your expertise and knowledge. Your guidance helped me to improve my research skills. In particular, your clear and constructive feedback was of great importance for me. I appreciate your patient feedback, my paper writings skills improved significantly.

I would like to thank all committee members for the time and effort they devoted to review my thesis.

Furthermore, I would like to thank the colleagues from the radar group that were part of the S4Drive project: Rossiza Gourova, Dr. Oleg Krasnov, Prof. dr. Alexander Yarovoy.

In the first two years of my PhD, I shared an office with the two Greek guys, namely Dimitrios Kotiadis and Athanasios Tasoglou. Thank you Dimitrios for your optimism and your 'don't worry' mentality, because it was something that I really needed in the beginning of my PhD. Athanasios, I am extremely grateful with your help. You were always available for advice on any level: mental, technical, etc.. You always believed in me and stimulated me to finish my PhD, more than once. Many times, you made me think about situations more optimistically. Athanasios and Dimitrios, I look forward to work with you guys again!

Also, I would like to thank my former colleagues at the university. I am really grateful for colleagues that I shared an office with, namely Ewoud Pool and András Pállfy. It was a lot of fun to share an office with you. Together with Thomas Hehn, we were the

'perception & modeling' PhD students in the Intelligent Vehicles group. Ewoud, András and Thomas, thanks for the interesting on-topic and off-topic discussions and knowledge sharing moments. Furthermore, without the work of the technicians/engineers, Tom Dalhuisen, Frank Everdij, Ronald Ensing, doing research is impossible, therefore I am very grateful for your help.

In addition, I would also like to thank my fellow PhD students and all other colleagues in the Intelligent Vehicles group, including Floris Gaisser, Jork Stapel, Tugrul Irmak, Yanggu Zheng, Laura Ferranti and Barys Shyrokau. In addition, I would like to thank the following people from Daimler R&D: Markus Braun, Markus Roth, Sebastian Krebs, Christian Münch, Christoph Rist and Fabian Flohr. Thanks for the brainstorm sessions and workshops that we have had together. Especially, the skiing and Kuhhandel workshops were the best. Furthermore, I would like to thank all master students I have worked with, including Randy Stakelbeek, Jan Wymenga, Ronald Immerzeel, Xinyu Gao, Georgios Katsaounis and Jiaao Dong. I really enjoyed the many discussions.

Also, I would like to thank my 2getthere colleagues, especially Dimitrios and Sjoerd.

Friends and family have had an important contribution to the completion of my PhD. First, I would like to thank the members of SOG: Christ 'Akky' Akkermans, Frerik 'Fredje' Andriessen, Tijmen Bregt, Jurjen C. Kamphuis, Bastiaan F. Lagaune, Salwan Al Jaberi, Philip Raaphorst en Jeroen A. Siebers. Except for Tijmen en Fredje, we all met at outdoor sport association, S.B.V. Slopend. Jeroen Siebers for the countless coffees at the Korvezeestraat/Rotterdamseweg. Second, I would like to thank my friends Niels, Leon and Daniel for the outdoor activities we did together; climbing, survival and running, etc. Thanks Jorn for your levelheaded view on life since we met at Aerospace engineering. Although we do not meet as often as in Delft, it still feels the same as we meet.

I would like to thank my in-laws for their understanding and support: Rong & Eva, Hester & Sicco, Jasmijn, Oscar & Benthe. I am very lucky with you as my in-laws. Hester and Rong thanks for (bi-)weekly taking care of Voske, because it resulted that I could focus on my PhD since I knew she was in good hands.

Pap & Mam thank you for teaching me the important things in live that let to the person that I am. You gave me the foundations and building blocks to develop myself so that I was able to move, study and do my PhD in Delft. Furthermore, I would like to thank my brothers and my sister: Marijn, Elske & Arnoud.

Finally and most importantly, I would like to thank my partner Merel, our daughter Voske and son Dasse. Without your love, patience and support, it would not have been possible to get through this intense period. Merel, thanks for your everlasting resilience in the difficult moments during my PhD time. Voske & Dasse thanks for making me realize the truly important things in life. Seeing you grow up makes me so proud.

A

CRLB

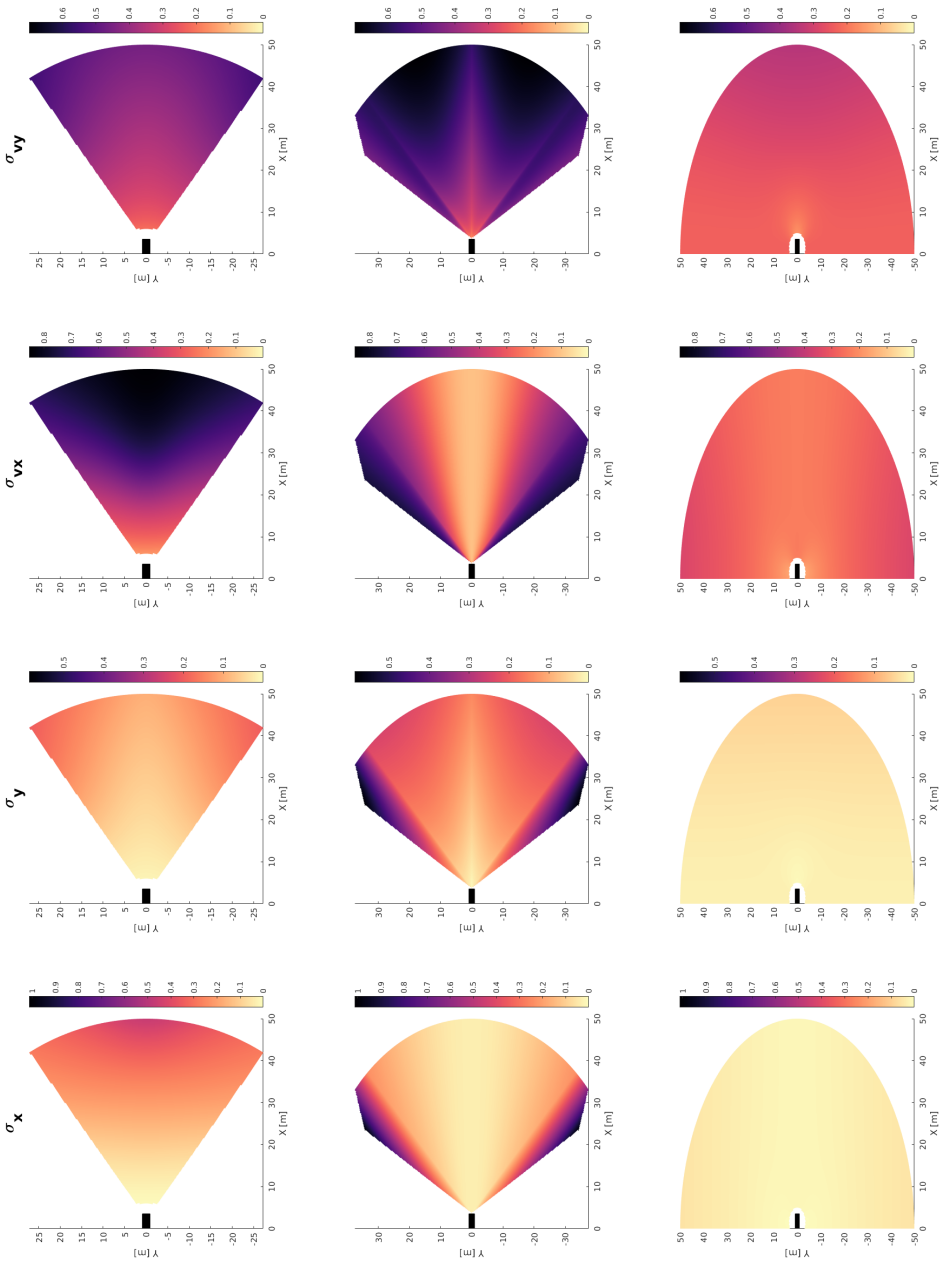


Figure A.1: Cramér-Rao lower bound results for single sensor object tracking with, from top to bottom, stereo camera (S), radar (R) (near range sensor) and lidar (L), respectively. From left to right the standard deviation in x, y, v_x and v_y is plotted. The units in the colourbar are [m] for σ_x and σ_y, and [m/s] for σ_{vx} and σ_{vy}. These results are obtained after a tracking (observation) time of 0.5 seconds. The lower the value in the plot the better the predicted tracking performance limit.

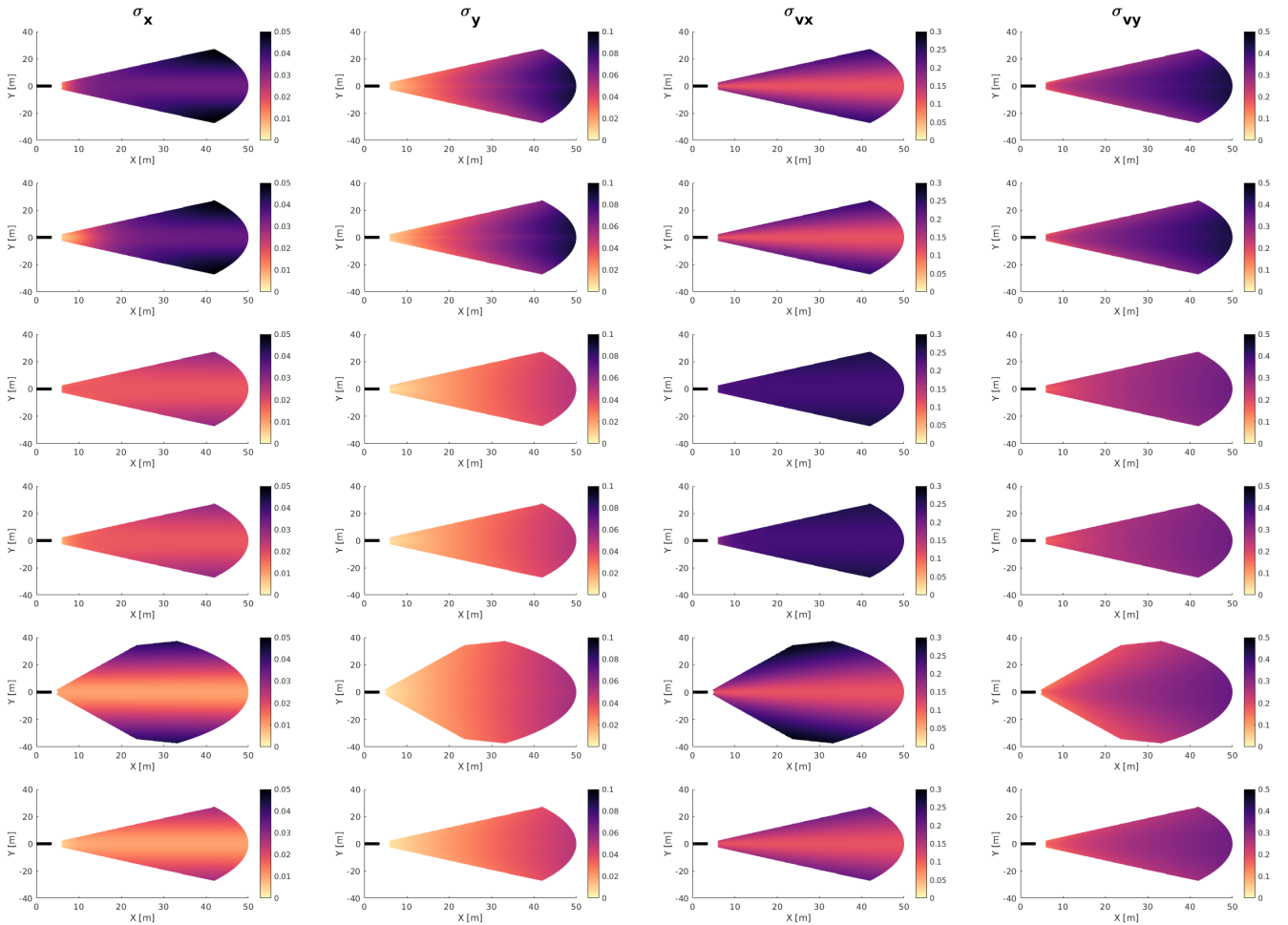


Figure A.2: Cramér-Rao lower bound results for multi-sensor object tracking with from top to bottom radar + monocular camera (R+C), radar + stereo camera (R+S), lidar + monocular camera (L+C), lidar + stereo camera (L+S), lidar + radar (L+R) and lidar + radar + stereo camera (L+R+S). From left to right the standard deviation in x, y, v_x and v_y is plotted. These results are obtained after a tracking time of 0.5 seconds.

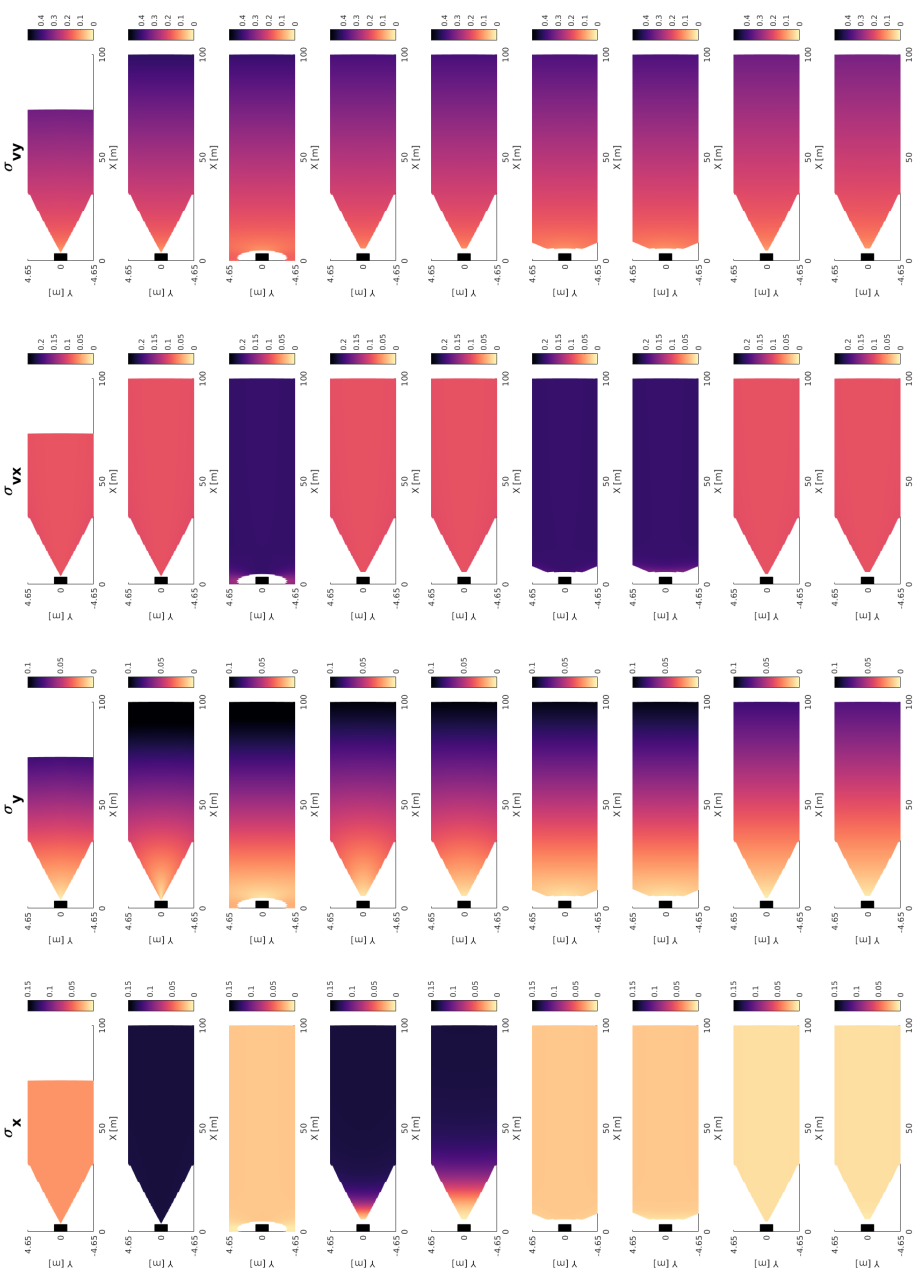


Figure A.3: The tracking performance limits for different configurations of lidar, radar and vision are visualised. From top to bottom: radar near range + radar far range ($R_{NR} + R_{FR}$), radar far range (R_{FR}), lidar (L), radar far range + monocular camera ($R_{FR} + C$), radar far range + stereo vision ($R_{FR} + S$), lidar + camera ($L + C$), lidar + stereo ($L + S$), lidar + radar far range ($L + R_{FR}$), lidar + radar far range + stereo vision ($L + R_{FR} + S$). These results are obtained after a tracking time of 0.5 seconds.

B

HELLINGER DISTANCE

The Hellinger distance between two multivariate normal distributions equals $\sqrt{1-d_{i,j}^{\mathcal{N}}}$ [1], where $d_{i,j}^{\mathcal{N}}$ is defined

$$d_{i,j}^{\mathcal{N}} = \frac{\det(P_k^{(i)})^{1/4}\det(P_k^{(j)})^{1/4}}{\det(P)^{1/2}} \exp\left\{-\frac{1}{8}(m_k^{(i)}-m_k^{(j)})^T(P)^{-1}(m_k^{(i)}-m_k^{(j)})\right\} \tag{B.1}$$

with $P = \frac{P_k^{(i)}+P_k^{(j)}}{2}$. The Hellinger distance between two Beta distributions equals $\sqrt{1-d_{i,j}^{\beta}}$ (e.g. [2]), where $d_{i,j}^{\beta}$ equals

$$d_{i,j}^{\beta} = \frac{B\left(\frac{s_k^{(i)}+s_k^{(j)}}{2}, \frac{t_k^{(i)}+t_k^{(j)}}{2}\right)}{\sqrt{B\left(s_k^{(i)},t_k^{(i)}\right)B\left(s_k^{(j)},t_k^{(j)}\right)}}. \tag{B.2}$$

REFERENCES

[1] L. Pardo, *Statistical inference based on divergence measures* (CRC press, 2018).

[2] W. Ji, S. M. AbouRizk, O. R. Zaïane, and Y. Li, *Complexity analysis approach for prefabricated construction products using uncertain data clustering,* Journal of Construction Engineering and Management **144**, 04018063 (2018).

C

EXPERIMENTAL RESULTS

Table C.1: Figure with average OSPA results for both filters on all recordings. For the local-bootstrap-LMB filter, the results are obtained using a uniform prior for the both detection probabilities in the birth model of in the p_D-CPHD filter. The two recordings with an asterisk (*) are the recordings used for tuning of the hyper parameters (i.e. the probability of detection and the false alarm rate). The values in bold indicate the best filter for each recorded scenario.

Recording	Precipitation	Light	LMB	local-bootstrap-LMB
2019-5-13 15:32*	-	day	**2.2**	2.4
2019-5-13 15:36*	-	day	**2.3**	2.5
2019-5-13 15:42	-	day	**2.7**	2.7
2019-5-13 15:49	-	day	**3.6**	4.0
2019-6-19 21:59	Rain	day	**2.3**	**2.3**
2019-6-19 22:02	Rain	day	**2.5**	**2.5**
2019-6-19 22:06	Rain	twilight/night	3.8	**3.5**
2019-6-19 22:10	Rain	twilight/night	4.4	**3.8**
2019-6-19 22:13	Rain	twilight/night	3.8	**3.7**
2019-6-19 22:17	Rain	twilight/night	4.2	**2.4**
2019-6-19 22:29	-	twilight/night	4.3	**3.9**
2019-6-19 22:32	-	twilight/night	4.4	**3.7**
2019-6-19 22:34	-	twilight/night	4.4	**3.3**
2019-6-19 22:40	-	twilight/night	4.7	**4.2**
2019-6-19 22:48	-	twilight/night	4.6	**3.9**
2019-6-19 22:50	-	twilight/night	4.6	**3.6**
2019-6-19 22:56	-	twilight/night	4.8	**4.4**
2019-6-19 23:29	-	twilight/night	4.6	**4.0**
2019-6-19 23:31	-	twilight/night	4.7	**4.1**
2019-6-19 23:34	-	twilight/night	4.9	**3.8**
2019-6-21 12:24	-	day	**3.3**	3.4
2019-6-21 12:26	-	day	**3.4**	3.6
2019-6-21 12:27	-	day	**3.3**	3.4
2019-6-21 12:33	-	day	**3.3**	3.4
2019-6-21 12:44	-	day	3.4	**2.8**
2019-6-21 12:46	-	day	**3.4**	3.5
2019-6-24 22:20	-	twilight/night	**2.5**	2.7
2019-6-24 22:24	-	twilight/night	3.9	**2.9**
2019-6-24 22:26	-	twilight/night	4.4	**3.7**
2019-6-24 22:36	-	twilight/night	4.4	**3.5**
2019-6-24 22:38	-	twilight/night	4.4	**2.6**
2019-6-24 23:14	-	twilight/night	3.0	**2.7**
2019-6-24 23:16	-	twilight/night	3.4	**2.7**
2019-6-24 23:21	-	twilight/night	4.7	**3.4**
2019-6-24 23:23	-	twilight/night	4.7	**2.3**
2019-6-24 23:30	-	twilight/night	4.6	**2.7**

Curriculum Vitæ

Joris Ferdinandus Maria Domhof

28-11-1989	Born in Winterswijk, The Netherlands

Education

2008–2011	Bachelor Aerospace Engineering (cum laude) Delft University of Technology
2011–2015	Master Mechanical Engineering Delft University of Technology
2015 – present	PhD Researcher Delft University of Technology *Promotor:* Prof. dr. D.M. Gavrila

List of Publications

5. J. Domhof, J. F. P. Kooij, D. M. Gavrila, *Adaptive Multi-Sensor Object Tracking in Adverse Weather Conditions*, in preparation for IEEE Transactions on Intelligent Vehicles.

   Author contributions: J. Domhof created proposed filter, performed data acquisition and performed experiments. J.F.P. Kooij and D. M. Gavrila provided guidance and supervision.

4. J. Domhof, J. F. P. Kooij, D. M. Gavrila, *An Joint Extrinsic Calibration Tool for Radar, Camera and Lidar*, IEEE Transactions on Intelligent Vehicles (T-IV), vol. 6, nr. 3, pp. 571-582, 2021.

   Author contributions: J. Domhof created calibration tool, performed data acquisition and performed experiments. J.F.P. Kooij provided guidance, supervision and contributed to writing. D. M. Gavrila provided guidance and supervision.

3. J. Domhof, J. F. P. Kooij, D. M. Gavrila, *An Extrinsic Calibration Tool for Radar, Camera and Lidar*, International Conference on Robotics and Automation (ICRA), (IEEE, 2019).

   Author contributions: J. Domhof created calibration tool, performed data acquisition and performed experiments. J.F.P. Kooij provided guidance, supervision and contributed to writing. D. M. Gavrila provided guidance and supervision.

2. J. Domhof, R. Happee and P. Jonker, *Robust multi-sensor bootstrap tracking filter for quality of service estimation*, 20th International Conference on Information Fusion (Fusion), (IEEE, 2019).

   Author contributions: J. Domhof created proposed filter, performed data acquisition and performed experiments. R. Happee contributed to writing. P. Jonker provided supervision.

1. J. Domhof, R. Happee and P. Jonker, *Multi-sensor object tracking performance limits by the Cramér-Rao lower bound*, 20th International Conference on Information Fusion (Fusion), (IEEE, 2019).

   Author contributions: J. Domhof created proposed methodology and performed simulations. R. Happee contributed to writing. P. Jonker provided supervision.