Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation UVEK
Département fédéral de l'environnement, des transports, de l'énergie et de la communication DETEC
Dipartimento federale dell'ambiente, dei trasporti, dell'energia e delle comunicazioni DATEC

Bundesamt für Strassen
Office fédéral des routes
Ufficio federale delle Strade

Automated Driving Sensor Testing Vehicle

Automated Driving Sensor Testfahrzeug

Véhicule d'essai du capteur AD

Empa, Automotive Powertrain Technologies Laboratory
Christian Hohl
Dejan Milojevic
Dr. Miriam Elser
Jannik Zraggen
Nikola Vulin

Forschungsprojekt ASTRA 2019/004 auf Antrag des Bundesamtes für Strassen (ASTRA)

Juli 2021 **1703**

Der Inhalt dieses Berichtes verpflichtet nur den (die) vom Bundesamt für Strassen unterstützten Autor(en). Dies gilt nicht für das Formular 3 "Projektabschluss", welches die Meinung der Begleitkommission darstellt und deshalb nur diese verpflichtet.
Bezug: Schweizerischer Verband der Strassen- und Verkehrsfachleute (VSS)

Le contenu de ce rapport n'engage que les auteurs ayant obtenu l'appui de l'Office fédéral des routes. Cela ne s'applique pas au formulaire 3 « Clôture du projet », qui représente l'avis de la commission de suivi et qui n'engage que cette dernière.
Diffusion : Association suisse des professionnels de la route et des transports (VSS)

La responsabilità per il contenuto di questo rapporto spetta unicamente agli autori sostenuti dall'Ufficio federale delle strade. Tale indicazione non si applica al modulo 3 "conclusione del progetto", che esprime l'opinione della commissione d'accompagnamento e di cui risponde solo quest'ultima.
Ordinazione: Associazione svizzera dei professionisti della strada e dei trasporti (VSS)

The content of this report engages only the author(s) supported by the Federal Roads Office. This does not apply to Form 3 'Project Conclusion' which presents the view of the monitoring committee.
Distribution: Swiss Association of Road and Transportation Experts (VSS)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation UVEK
Département fédéral de l'environnement, des transports, de l'énergie et de la communication DETEC
Dipartimento federale dell'ambiente, dei trasporti, dell'energia e delle comunicazioni DATEC

Bundesamt für Strassen
Office fédéral des routes
Ufficio federale delle Strade

Automated Driving Sensor Testing Vehicle

Automated Driving Sensor Testfahrzeug

Véhicule d'essai du capteur AD

Empa, Automotive Powertrain Technologies Laboratory
Christian Hohl
Dejan Milojevic
Dr. Miriam Elser
Jannik Zraggen
Nikola Vulin

Forschungsprojekt ASTRA 2019/004 auf Antrag des Bundesamtes für Strassen (ASTRA)

Juli 2021 **1703**

Impressum

Forschungsstelle und Projektteam

Projektleitung
Christian Hohl

Mitglieder
Dejan Milojevic
Dr. Miriam Elser
Mathias Huber
Jannik Zraggen
Nikola Vulin

Begleitkommission

Präsident
Bernhard Gerster

Mitglieder
Sven Britschgi
Markus Deubelin
Anton Keller
Heinz Berger

KO-Finanzierung des Forschungsprojekts

SCCER Mobility / Innosuisse
Begleitgruppe aus Wirtschaft
Lexus / Toyota Schweiz

Antragsteller

Bundesamt für Strassen (ASTRA)

Bezugsquelle

Das Dokument kann kostenlos von http://www.mobilityplatform.ch heruntergeladen werden.

Contents

Impressum **4**
Zusammenfassung **7**
Résumé **21**
Summary **33**

1 Introduction **35**
1.1 Motivation 35
1.2 Contribution 35
1.3 Organization 36

2 Testing Approaches for AD-Systems **37**
2.1 Function-Based 37
2.2 Real-World Testing 37
2.3 Shadow Mode 38
2.4 Simulation Testing 38
2.5 Staged Introduction 38
2.6 Formal Verification 38
2.7 Scenario-Based Testing 38

3 Sensing and Perception in Automated Driving **41**
3.1 Sensors 41
3.1.1 Radar 41
3.1.2 Lidar 44
3.1.3 Camera 46
3.2 Perception Functionalities 47
3.2.1 Detection 47
3.2.2 Classification 48
3.2.3 Distance Estimation 49
3.2.4 Tracking 49
3.2.5 Localization 49

4 Methodology **51**
4.1 Function-Scenario-Based Testing 51
4.2 Performance Metrics 52
4.3 Evaluation and Data Pipeline 52

5 Field Testing Setup and Preparatory Work **53**
5.1 Sensor Calibration 53
5.2 Test track 54
5.3 Track Mapping 55
5.4 Test objects 57
5.5 Weather Conditions 58

6 Function-Scenario-Based Testing **63**
6.1 Object Detection and Distance Estimation 63
6.1.1 Sensors, Algorithms and Black Boxes 63
6.1.2 Implementation and first insights 64
6.1.3 Walking Pedestrian Scenario 65
6.1.4 Mixed Urban Scenario 86
6.1.5 Lidar Targets 104
6.2 Localization 120
6.2.1 Lidar localization in urban environment 120

7 Conclusion 135
7.1 Outlook 137

Appendix 139
Nomenclature 143
References 145
Project Completion 149

Zusammenfassung

Dieses Dokument dokumentiert die Aktivitäten und Ergebnisse aus dem Projekt "Automated Driving - Sensor Testing Vehicle". Das Projekt wurde gestartet, um Verfahren zur Sicherheitsvalidierung für automatisierte Fahren zu verstehen und aufzubauen. Sensoren spielen eine entscheidende Rolle im Automated Driving (AD)-Stack, da sie die relevanten Informationen aus der Umgebung liefern. Das Projekt konzentrierte sich auf die Evaluierung von AD-Sensoren mit Feldtests während des realen Fahrens auf einer Teststrecke und den Vergleich der Ergebnisse mit Daten aus experimentellen Labortests, die am METAS durchgeführt wurden. Zu diesem Zweck wurde ein Lexus RX 450h Fahrzeug mit einem handelsüblichen Satz der wichtigsten AD-Sensoren, d.h. Lidar, Kamera und Radar, ausgestattet. Weiterhin wurde auf dem Gelände des Empa-Campus in Dübendorf eine Teststrecke mit Testobjekten aufgebaut.

Für die Bewertung der Leistungsfähigkeit der Sensoren wurde der so genannte funktions-szenariobasierte Testansatz verwendet. Bei diesem Verfahren wird eine bestimmte Wahrnehmungsfunktion im AD, z.B. die Objekterkennung, in einem entworfenen realen Fahrszenario getestet. Dieser Testansatz ermöglicht sowohl die Wiederholbarkeit von Feldtests als auch ein genaues Verständnis des Szenarios. Mit verschiedenen Szenarien haben wir die Objekterkennung, die Distanzschätzung, Tracking und Lokalisierungsleistung der Mobileye-Kamera, des Delphi ESR-Radars und des Velodyne VLP-16-Lidars getestet. Da es sich bei einigen der Sensoren, z. B. bei der Mobileye-Kamera, um Black-Box-Systeme handelt, haben wir neue Bewertungsspezifikationen entwickelt, z. B. die Tracking-Ratio, welche angibt, wie empfindlich bzw. robust die Objekttrackingfunktionalität über die Zeit ist. Das Design unserer Szenarien ermöglichte es uns, präzise Informationen über die Sensorleistung zu erhalten, z. B. die Falsch-Negativ-Rate der Objekterkennung bei unterschiedlichen Objektabständen. Die Feldtests wurden unter verschiedenen Wetterbedingungen und mit verschmutzten Sensoren durchgeführt.

Unsere Analysen ergaben, dass die Mobileye-Kamera bei der Objekterkennung eine hohe Wetterabhängigkeit aufweist und grosse Leistungseinbussen bei verschneiten und regnerischen Wetterbedingungen zeigt. Ebenfalls signifikant beeinflusst der Objekttyp, -grösse und -entfernung die Erkennungsleistung der Mobileye Kamera. Fahrzeuge wurden insgesamt leichter erkannt als Fussgänger und grössere (erwachsene) Fussgänger leichter als kleinere (Kinder) Fussgänger. Generell verschlechterte sich die Erkennungsleistung stark bei grossen Entfernungen, aber auch bei sehr geringen Entfernungen, was wir auf Bewegungsunschärfe-Effekte durch das sich bewegende Testfahrzeug zurückführten. Die Entfernungsschätzungen der Mobileye-Kamera erwiesen sich unter allen getesteten Bedingungen als sehr schlecht, mit einer deutlichen Tendenz zur Unterschätzung der wahren Entfernung. Das ESR-Radar zeigte unter allen Umgebungsbedingungen eine gute Erkennungsleistung und lieferte sehr genaue Distanzmessungen. Wir beobachteten jedoch, dass die Objekttrackingfunktionalität recht empfindlich ist, da der Sensor oft Tracks verliert und/oder das gleiche Objekte mehrfach verfolgt. Der Velodyne-Lidar zeigte unter allen getesteten Umgebungsbedingungen eine gute Lokalisierungsleistung. Selbst ein teilweise blockiertes Sichtfeld durch angesammelten Schnee auf der Sensoroberfläche führte nicht zu einem Lokalisierungsfehler. Unsere Feldtests und Laborexperimente zeigten jedoch, dass dunkler gefärbte Oberflächen vom Lidar nicht gesehen werden, wenn der Einfallswinkel zwischen dem Laserstrahl und der Oberfläche nicht senkrecht ist, was ein grosses Problem für die Objekterkennung darstellt.

Testverfahren

Im ersten Kapitel werden die verschiedenen Ansätze zur Validierung der Sicherheit von Automatisierten Fahrzeugen beschrieben. Hierbei werden die einzelnen Verfahren kurz vorgestellt und die spezifischen Vor- und Nachteile aufgezeigt.

In diesem Projekt wird der funktions-szenariobasierte Ansatz verwendet. Dabei besteht die Idee darin, automatisiertes Fahren in typische einzelne Verkehrssituationen zu zerlegen. Man versucht also, die unendliche Menge möglicher Fahrsituationen auf eine endliche

Menge von zu testenden Szenarien zu reduzieren. Diese endliche Menge von Tests kann entweder in der Simulation, auf einer geschlossenen Strecke oder auf der Strasse durchgeführt werden. Der Vorteil dieses Ansatzes ist, dass wir uns auf kritische Situationen konzentrieren können, ohne Testzeit auf die einfachen Aufgaben zu verwenden. Die grösste Herausforderung dieses Ansatzes ist, dass er davon ausgeht, dass das Automated Vehicle (AV), wenn es ein bestimmtes Szenario erfolgreich durchläuft, auch andere ähnliche Szenarien bestehen wird.

Sensoren und Wahrnehmungsfunktionen für Automatisiertes Fahren.

Im zweiten Kapitel werden die wichtigsten AD-Sensoren und Wahrnehmungsfunktionen vorgestellt. Die verschiedenen Sensortechnologien (Kamera, Radar, Lidar...) werden verwendet, um Echtzeitinformationen über die Umgebung des Fahrzeugs zu erfassen. Durch Datenverarbeitung können die Sensoren Objekte in der Umgebung des Fahrzeugs erkennen und verfolgen. Die von verschiedenen Sensoren erfassten Daten werden durch Datenfusion kombiniert und in Echtzeit verarbeitet, um eine umfassende Darstellung der Realität zu erstellen. Diese wird von Algorithmen verwendet, um die Situation zu analysieren, mögliche Trajektorien zu planen und die zu ergreifenden Massnahmen (Lenken, Bremsen...) zu bestimmen.

Unser AD-Testfahrzeug ist mit einem kommerziell erhältlichen Delphi-Radarsystem ESR 2.5 ausgestattet, welches in der vorderen Stossstange installiert ist. Das ESR 2.5 ist ein 76,5 GHz Solid-State Radar mit elektronisch gelenktem Strahl und Puls-Doppler-Signalverarbeitung. Das Radar ermöglicht die gleichzeitige Messung eines breiten Sichtfeldes im mittleren Bereich und eines engen Sichtfeldes im Fernbereich. Auf dem Dachträger des Fahrzeug ist ein Velodyne VLP-16-A LiDAR Puck montiert, ein 16-Kanal-Spinningl-Lidar. Der Velodyne LiDAR Puck hat eine Reichweite von 100 m mit einer Genauigkeit von 3 cm. Er bietet ein volles horizontales Sichtfeld von 360° und ein vertikales Sichtfeld von ± 15°. Als Kamerasystem ist eine Mobileye 630 hinter der Windschutzscheibe angebracht. Diese enthält den EyeQ2 Image Processing Chip mit den Bildverarbeitungsalgorithmen von Mobileye. Die Algorithmen versuchen, Fahrspuren, Verkehrsschilder, Fahrzeuge, Radfahrer und Fussgänger zu erkennen und den Abstand zu diesen Objekten zu messen.

Objekterkennung & Klassifikation

Die Erkennung und Klassifikation von Objekten ist eine entscheidende Funktionalität für AVs. Für AD können Objekte aus Kamerabildern, Radardaten und aus Lidar-Punktwolken erkannt werden. Typische Objekte, die erkannt werden müssen, sind nicht nur Fahrzeuge und Fussgänger, sondern auch andere Strassenhindernisse wie Bäume, Fahrspuren, Geschwindigkeitsschilder, Ampeln und vieles mehr. Aus Kamerabildern wird ein Objekt in der Regel entweder durch Deep-Learning-Methoden oder klassische Computer-Vision-Methoden auf Basis von Kantenerkennung erkannt. Die Objekterkennung in Echtzeit, insbesondere aus Kamerabildern, ist ein sehr aktives Forschungsthema sowohl in der akademischen als auch in der industriellen Forschung, und in den letzten zehn Jahren wurden mit verbesserten Algorithmen und erhöhter Rechenleistung grosse Fortschritte erzielt.

Tracking

Tracking ist der Prozess der Verfolgung von bewegten Objekten. In vielen Anwendungen, z. B. bei der Trajektorienplanung, ist es oft notwendig, die Position und Geschwindigkeit jedes Objekts mit einer festen Aktualisierungsrate bereitzustellen. Die Tracks werden mit einem Filter gewonnen, der Objekterkennungs- und Lokalisierungsdaten verwendet. Beim Entwurf eines Filters gibt es viele Herausforderungen, wie verrauschte Sensordaten, fehlende oder falsche Erkennungen und unsynchronisierte Sensordaten. Ausserdem kann nicht jeder Sensor alle erforderlichen Daten für die Anwendung liefern. So kann z. B. die Geschwindigkeit des Objekts nicht aus einem einzigen Kamerabild bestimmt werden, weshalb ein Tracking-Filter diese Daten in diesem Fall schätzen muss.

Lokalisierung

Wie menschliche Fahrer in herkömmlichen Autos müssen auch AD-Fahrzeuge ihre absolute Position kennen, um zum Zielort zu navigieren. Für diesen Navigationszweck ist eine GNSS-Position mit einer Genauigkeit in der Grössenordnung von ein paar Metern ausreichend. Darüber hinaus verlassen sich menschliche Fahrer bei der Wegverfolgung auf eine visuelle Lokalisierung, d. h. auf das Erkennen von Fahrspuren und Strassenbegrenzungen und der relativen Position in Bezug auf diese. Ebenso verlassen sich viele AD-Fahrzeuge auf eine sehr genaue absolute Lokalisierung im Zentimeterbereich und verwenden ähnlich genaue detaillierte Strassenkarten zum Navigieren. Für eine solch genaue Lokalisierung ist die Lidar-Lokalisierung ein sehr vielversprechender Ansatz, insbesondere für urbane Umgebungen. Sie bietet eine Genauigkeit im geforderten cm-Bereich, ähnlich wie ein Differential- oder RTK-GNSS. Zusätzlich leidet die Lidar-Lokalisierung nicht unter Signalverlusten oder Mehrwegproblemen wie GNSS und kann daher Redundanz bieten oder als Backup fungieren, wenn GNSS-Positionierungsmethoden verloren gehen oder ausfallen. Die Lidar-Lokalisierung funktioniert, indem die aktuelle Punktwolke mit einer meist vordefinierten 3D-Karte verglichen wird und so die Position in der Karte gefunden wird, an der die aktuelle Punktwolke am besten zur Karte passt.

Methodik

Um die verschiedenen Sensoren und Wahrnehmungsfunktionen eines AD-Fahrzeugs unter realen Fahrbedingungen bewerten zu können, benötigen wir einige realistische, aber dennoch wiederholbare Experimente. Die Wiederholbarkeit der Experimente ermöglicht es uns, die Leistung der Sensoren bei unterschiedlichen Umgebungsbedingungen zu vergleichen und die verschiedenen Sensoren untereinander zu vergleichen. Um einen Test auswerten zu können muss immer definiert werden, wie die Performance eines Sensors beziehungsweise einer Funktion gemessen und evaluiert wird. Beispielsweise wird für die Detektionsperformance die False-Negative-Rate verwendet. Um eine False-Negative Rate zu berechnen,muss für jeden Testrun zuerst die Ground-Truth bestimmt werden, das heisst es muss festgelegt werden, wo sich das Objekt zu welchem Zeitschritt befindet. Ein Schritt in der Datenauswertung welcher oftmals manuell geschieht.

Testaufbau und Vorbereitungsarbeiten

Im Kapitel Testaufbau und Vorbereitungsarbeiten werden die Arbeiten dokumentiert, die notwendig waren, um die späteren Versuche korrekt durchführen zu können. In einem ersten Schritt mussten die verschiedene verwendeten Sensoren kalibriert werden. Hierfür wurde für jeden Sensor ein Koordinatensystem definiert, die genaue Position und Ausrichtung der einzelnen Sensorkoordinatensysteme zueinander vermessen und ein mathematisches Modell implementiert.

Um unsere Testszenarien wiederholbar zu gestalten, wurde auf dem Empa Campus in Dübendorf eine Teststrecke erstellt. In einem ersten Schritt wurde das Layout der Teststrecke im CAD gezeichnet und auf die Erfüllung der Normen überprüft. Anschliessend konnten die Strassenmarkierungen auf dem Empa Campus aufgebracht werden. Abschliessend wurde die Teststrecke mit dem Lidar kartographiert und eine präzise 3D-Karte der Testumgebung erstellt.

Zusätzlich zur Teststrecke wurden für die einzelnen Szenarien verschiedene Testobjekte hergestellt. Einerseits wurden mehre Schaufensterpuppen erworben und mittels Farbe und Kleidern zu Fussgängern umgestalten, welche auf der Teststrecke aufgestellt werden können. Andererseits wurden drei grosse Holztafeln auf Paletten gebaut und mit weisser, grauer beziehungsweise schwarzer Farbe angestrichen, welche uns als Lidar Testziele dienen.

Im letzten Abschnitt des Kapitels werden die verschiedenen Wetter und Umweltbedingungen definiert, bei welchen die Sensoren getestet werden. Für unsere Versuche sind dies Sonne, Regen, Nebel, Schneefall und Verschmutzung.

Funktions-szenariobasierter Testansatz

Im letzten Kapitel werden die Durchführung, die Datenaufbereitung, Analysen und Ergebnisse unserer Tests festgehalten. Das Kapitel ist in zwei Abschnitte unterteilt, wobei der erste Abschnitt die drei verschiedenen Szenarien zum Testen der Objekterkennung und Distanzschätzung und der zweite Abschnitt das Szenario zum Test der Lokalisierungsfunktion beinhaltet.

Walking Pedestrian Szenario

Beim Walking Pedestrian Szenario testen wir die AD Sensoren Mobileye Kamera und Radar bezüglich der Erkennung, Lokalisierung und Positionsbestimmung eines gehenden Fussgängers. Hierbei startet der Fussgänger 5 m vom stehenden AD-Fahrzeug entfernt und läuft von diesem weg, wobei er alle 5 m für 10 s anhält. Sobald er die Markierung bei 50 m erreicht hat, dreht sich der Fussgänger um und geht wieder in 5-m-Schritten auf das Auto zu. Bei den Stillstandspositionen kennen wir die exakte Entfernung des Fussgängers und können daher die Leistung der Entfernungsschätzung genau bewerten.

Als Performance-Metriken verwenden wir den Fehler der Distanzmessung bzw. Schätzung, die Anzahl der für den Fussgänger erzeugten Tracks, die durchschnittliche Dauer der Tracks, die Tracking Ratio, das heisst wieviel Prozent der Zeit der Fussgänger getrackt wurde und, im Falle der Kamera, ob der Fussgängers korrekt als Fussgänger klassifiziert wurde.

Das Test-Szenario haben wir unter den verschiedenen definierten Wetter/Umweltbedingen durchgeführt und mehrfach wiederholt. Unsere Testergebnisse zeigen, dass die Entfernungsschätzungen des Mobileye in diesem Szenario unter allen Umgebungsbedingungen sehr schlecht sind. Die Entfernung wird generell stark unterschätzt. Der absolute Fehler und seine Variabilität zwischen den Testwiederholungen scheint bei höheren Entfernungen zuzunehmen, insbesondere bei nebligen Bedingungen.

Betrachtet man die Tracking-Ergebnisse aus allen Testläufen, so sieht man, dass bei guten Lichtverhältnissen die Tracking-Ratio des Mobileye nahe bei 100 % liegt. Im Gegensatz dazu erreicht die Tracking-Ratio bei schlechten Lichtverhältnissen den Wert Null, was aufgrund der Mobileye-Spezifikationen zu erwarten war. Interessanter ist, dass wir bei schlechten Wetterbedingungen eine schlechtere Tracking-Stabilität haben, d.h. der Track geht häufiger verloren, die Anzahl der erzeugten Tracks steigt und die durchschnittliche Trackdauer sinkt. Dies zeigt, dass die Tracking-Stabilität der Mobileye-Kamera deutlich von guten Lichtverhältnissen, wie sie bei sonnigem Wetter vorliegen, profitiert.

Im Gegensatz dazu ist die Entfernungsschätzung des Delphi ESR-Radars für dieses Szenario viel präziser (Fehler innerhalb unserer Positionierungs-/Ground Truth Präzision) und wir sehen weder einen Trend mit zunehmender Entfernung noch Auswirkungen der ungünstigen Umgebungsbedingungen. Obwohl in den technischen Datenblättern des Delphi ESR-Radars angegeben wird, dass der Sensor nur Tracks meldet, sehen wir in unseren Tests viele sehr kurze Tracks, die eher wie Einzeldetektionen als gefilterte Tracks aussehen. Die Tracking Ratio für lange Tracks (d. h. Tracks, die länger als 1 s sind) variiert zwischen 20 % und 85 %, wobei die Mehrheit der Tests eine Verfolgungsrate von etwa 45 % aufweist. Das bedeutet, dass das Radar den Fussgänger die meiste Zeit über nicht gut oder zumindest nicht kontinuierlich verfolgen kann. Auch in Bezug auf die Trackingsleistung konnten wir keine signifikanten Auswirkungen der Umgebungsbedingungen beobachten. Unsere Erkenntnisse aus den Praxistests mit dem Radar stimmen insgesamt gut mit den Ergebnissen aus den Labormessungen am METAS überein, wo ebenfalls eine hohe Anzahl von Stör-Tracks und Phantomobjekten gemessen wurde. Darüber hinaus zeigten unsere Praxistests, dass die Radarleistung stark von Gebäuden und anderen grossen Objekten entlang der Strasse beeinflusst wird. Tatsächlich scheinen Objekte mit einem grossen Radarquerschnitt, wie die T-förmigen Balken an unserem Laborgebäude oder der Metallcontainer neben unserer Teststrecke, das Radar bei der Verfolgung des Fussgängers zu stören, was die Ursache für den beobachteten Trackverlust des Fussgängers in der Nähe solcher Objekte zu sein scheint.

Vergleicht man die Ergebnisse beider Sensoren, kann man zu dem Schluss kommen, dass weder die Mobileye-Kamera noch das Delphi ESR-Radar bei individueller Betrachtung eine akzeptable Leistung für die Fussgängererkennung und Entfernungsschätzung erbringen, da beide Sensoren unter bestimmten Umgebungsbedingungen deutliche und reproduzierbare Schwächen aufweisen. Insbesondere ist die Mobileye-Entfernungsschätzung insgesamt sehr schlecht und die Verfolgungsleistung nimmt bei ungünstigen Wetterbedingungen ab, während das Radar, das präzise Entfernungsschätzungen liefert, Probleme bei der Tracking-Leistung zeigt, wenn sich Gebäude und andere Objekte entlang der Strasse befinden. Mit unserem Fussgängerszenario konnten wir also zeigen, dass ein einzelnes Sensorsystem allein nicht zuverlässig genug ist, um unter allen Bedingungen zu funktionieren, und dass jeder Sensor seine eigenen spezifischen Stärken und Schwächen hat. Daher würde ein AD-System, um sicher zu sein, stark von der Kombination mehrerer Sensortechnologien und der Sensorfusion profitieren, um Fussgänger zuverlässig zu erkennen und zu verfolgen.

Mixed Urban Szenario

In diesem Szenario testen wir die AD Sensoren Mobileye Kamera und Radar bezüglich der Erkennung, Lokalisierung und Positionsbestimmung von Fussgängern und Fahrzeugen. Dabei wollen wir die genauen Positionen der getesteten Hindernisse relativ zum AD-Fahrzeug kennen, damit wir die Performanceänderungen bei unterschiedlichen Abständen, Winkeln oder Fahrzeuggeschwindigkeiten analysieren können. Zu diesem Zweck haben wir dieses Mixed Urban-Szenario entwickelt, das eine städtische Umgebung darstellen soll. Die Objekte in diesem Szenario sind stationär und wir kennen ihre genaue Position. Im Szenario fährt das AD-Fahrzeug mit einer konstanten Geschwindigkeit auf der vorderen Geraden unserer Teststrecken. Vier Schaufensterpuppen (zwei weibliche, eine männliche und ein Kind) sind entlang der vorderen Geraden installiert. Am Ende der vorderen Geraden befindet sich ein stationärer Transporter (Iveco Daily).

Zusätzlich zu den bereits im Walking Pedestrian Szenario verwendeten Metriken evaluieren wir die False Negative Rate der Sensoren. Wir möchten herausfinden, wie hoch die FNR bei unterschiedlichen Objektabständen, Objekttypen, Objektgrössen oder Umgebungsbedingungen ist. Diese Dimensionen geben uns ein besseres Verständnis, warum bestimmte Sensoren für gewisse AD-Aufgaben besser sind.

Unsere Ergebnisse zeigen, dass die Leistung der Mobileye-Kamera bei der Objekterkennung stark von der Objektart/-grösse und den Wetterbedingungen abhängt. Bei der Fussgängererkennung beobachten wir die beste Leistung (d. h. eine niedrigere mittlere FNR-Kurve für alle Entfernungen) für sonnige und neblige Wetterbedingungen. Wie in der Literatur berichtet, wird erwartet, dass die Leistung der Mobileye-Kamera bei sonnigen Bedingungen am besten ist, da der passive Kamerasensor von den guten Lichtverhältnissen profitiert und die Bilder nicht durch Wassertropfen, Schneeflocken oder Nebel beeinträchtigt werden. Die ähnliche Leistung, welche für die nebligen Bedingungen wie für die sonnigen Bedingungen erzielt wurde, war eher unerwartet. Auch beim "Walking Pedestrian" Szenario hatte der Nebel keinen Einfluss auf die Tracking-Leistung von Mobileye, Die Experimente für die Szenarien "Walking Pedestrian" und "Mixed Urban" wurden jeweils am gleichen Tag mit einem Zeitunterschied von ein paar Minuten durchgeführt, was bedeutet, dass die Nebeldichte mehr oder weniger gleich war. Wir gehen davon aus, dass die relativ gute Leistung des Sensors für die nebligen Bedingungen darauf zurückzuführen sein könnte, dass der Nebel während unserer Experimente nicht dicht genug war, um einen Einfluss auf die Kamerabilder und die Objekterkennung für Entfernungen bis zu 50 m zu haben. Im Gegensatz dazu wurden für alle Entfernungen höhere FNRs für regnerische und verschneite Bedingungen ermittelt, wie aufgrund des zusätzlichen Rauschens, das durch die Tropfen/Flocken im Kamerabild entsteht erwartet wurde.

In Bezug auf die Gesamtleistung zeigt das Mobileye eine eher schwache Fussgängererkennungsleistung. Interessanterweise zeigen alle Kurven eine Parabelform mit einem globalen Minimum meist zwischen 15 und 25 m. Der Grund für steigende FNRs mit der Entfernung ist, dass die Kamera allmählich weniger Informationen von den Objekten erfasst, d.h. je höher die Entfernung, desto geringer die Menge an Pixeln des

Objekts im Bild. Die geringere Leistung bei kleineren Entfernungen ist weniger intuitiv, da nähere Objekte mehr Pixel haben sollten. Wir vermuten, dass der Grund mit der höheren tangentialen Relativgeschwindigkeit zwischen der Kamera und dem Objekt zusammenhängen könnte, welche zu Bewegungsunschärfe in den Kamerabildern führt. Da wir leider keinen Zugriff auf die von der Mobileye-Kamera erzeugten Rohbilder haben, können wir das Auftreten der Bewegungsunschärfe jedoch nicht beweisen.

Neben den Wettereffekten haben wir auch Unterschiede bei der Erkennung gleicher Objekttypen mit unterschiedlichen Grössen festgestellt. Hierzu vergleichen wir die FNR-Kurven für die Erkennung von erwachsenen und kindlichen Fussgängern bei sonnigen und verschneiten Wetterbedingungen. Dabei zeigt sich, dass die Leistung für die Erkennung von erwachsenen Fussgängern im Vergleich zur Erkennung von Kinderfussgängern deutlich besser ist. Auch hier könnten mehrere Faktoren eine wichtige Rolle spielen. Aus der Kameraperspektive ist ein Kind kleiner als ein Erwachsener, und folglich erscheint es auch kleiner im Bild. Das bedeutet, dass es weniger Pixel / Informationen über das Objekt gibt.

Bei der Fahrzeugerkennung zeigt die Mobileye Kamera bei sonnigen Bedingungen die beste Performance mit einer FNR-Kurve von nahe 0, wobei wir bei Nebel wiederum eine nahezu gleiche Performance feststellen. Bei Regenbedingungen zeigt sich eine grössere Unsicherheit und die FNR steigt zwischen 30 und 40m auf 0,6 an. Dies zeigt, dass Wassertropfen vom Regen einen Einfluss auf die Fahrzeugerkennung haben und dass der Einfluss bei grösseren Entfernungen zunimmt. Die schlechteste Leistung zeigt sich bei der Fahrzeugerkennung unter Schneebedingungen. Es scheint, dass Schneeflocken einen höheren Einfluss haben als Regentropfen.

Beim Vergleich zwischen verschiedenen Objekttypen zeigt sich, dass die Kamera für Fahrzeuge eine bessere Performance aufweist als für Fussgänger. Für diesen Effekt könnte es mehrere Erklärungen geben. Zum einen ist der Iveco Daily viel grösser als ein Fussgänger und daher enthalten die Kamerabilder mehr nützliche Informationen für seine Erkennung. Ausserdem sind grössere Objekte möglicherweise weniger empfindlich gegenüber Sensorrauschen oder Umgebungsrauschen, das durch Regen oder Nebel entsteht. Ausserdem hat die Bewegungsunschärfe der Kamera, die mit der Bewegung des AD-Fahrzeugs zusammenhängt, bei kleineren Objekten einen grösseren Einfluss. Schliesslich könnten die Fahrzeuge leichter maschinenverständliche Formen haben. In der Zukunft planen wir, Experimente auch mit kleineren Fahrzeugen durchzuführen, um eine bessere Verteilung der möglichen Fahrzeuge zu haben

Schliesslich haben wir die Genauigkeit der Abstandsschätzung der Mobileye-Kamera für die Fussgänger- und Fahrzeugerkennung analysiert. Die Fussgänger werden meist zu nah erkannt (positiver Fehler) und die Streuung des absoluten Fehlers nimmt mit der Entfernung zu. Im Fall des Kinderfussgängers beobachten wir jedoch, dass die Genauigkeit der Entfernungsschätzung bei Entfernungen über 20 m sehr schlecht ist und dass der Kinderfussgänger meist zu weit entfernt erkannt wird, wahrscheinlich, weil er kleiner ist als das Mobileye-Fussgängermodell. Beim Iveco Daily tendiert der absolute Fehler dazu, bei grösseren Entfernungen linear anzusteigen wobei der Fehler bei regnerischen und verschneiten Wetterbedingungen stärker zunimmt. Ausserdem nimmt die Variabilität des absoluten Fehlers bei grösseren Entfernungen deutlich zu. Der absolute Fehler ist meist positiv, was damit zusammenhängen könnte, dass der Iveco Daily ein ziemlich grosses Fahrzeug ist (höchstwahrscheinlich grösser als die Fahrzeuge, die für die Anpassung des Mobileye-Modells verwendet wurden) und der Abstandsschätzungs-algorithmus daher das Fahrzeug als näher schätzt als es ist.

Im Gegensatz zur Mobileye Kamera zeigt das ESR-Radar eine recht wetterunabhängige Leistung bei der Fahrzeugerkennung. Insbesondere ist die mittlere FNR-Kurve fast immer Null, d.h. das Fahrzeug wurde zu jeder Zeit erkannt. Nur bei verschneiten Wetterbedingungen sehen wir eine etwas höhere Unsicherheit in der FNR-Kurve für grössere Entfernungen (von 50 bis 70 m). Die Wetterbeständigkeit liegt daran, dass Radare aktive Sensoren sind und ihre Wellenlänge gross genug ist, um nicht von Schneeflocken oder Regentropfen reflektiert zu werden.

Interessanterweise entdeckten wir einige Einbrüche in der Radarerkennungsleistung, wenn der Sensor kontaminiert ist. Die FNR-Kurve für die Fahrzeugerkennung mit verunreinigtem Radar zeigt, dass die FNR bei Entfernungen zwischen 40 und 50 m von 0 auf 0,9 ansteigt, bevor sie wieder abfällt. Der Grund für diesen Effekt ist uns nicht klar. Möglicherweise haben Schmutz, Blätter und Reifenabrieb, die für diese Tests auf der Sensoroberfläche abgelagert werden, einen Einfluss auf die Radarreflexion. Ausserdem könnte es sein, dass bei Entfernungen zwischen 40 und 60 m ein Blatt das FoV des Radars ganz oder teilweise genau in Richtung des Fahrzeugs verdeckte. Wir planen, diese Experimente in Zukunft zu wiederholen, um die Auswirkungen der Verschmutzung weiter zu untersuchen.

Schliesslich haben wir die Leistung der Mobileye-Kamera und das ESR-Radars zur Fahrzeugerkennung miteinander verglichen. Dabei ist der grösste Unterschied bei der härtesten Umgebungsbedingung zu sehen, nämlich bei Schnee. Die FNR-Kurve der Mobileye-Kamera bei Schnee ist viel höher als die des ESR-Radars. Das bedeutet, dass das ESR-Radar zuverlässiger und witterungsbeständiger für die Fahrzeugerkennung ist wobei der Unterschied ziemlich gross ist.

Zusammenfassend, konnten wir mit unserem Mixed Urban Szenario Erkenntnisse auf verschiedenen Ebenen gewinnen. Die Mobileye-Kamera zeigte im Allgemeinen eine unzuverlässige Fussgänger-Erkennungsleistung und wäre daher als einzelner Sensor nicht geeignet. Das ESR-Radar hat eine recht zuverlässige Fahrzeugerkennung bei allen Wetterbedingungen. Der Nachteil des ESR-Radars ist jedoch seine Sensitivität. Ausserdem zeigte das Delphi-Radar Leistungseinbussen, wenn es verschmutzt war. Jedoch sind weitere Messungen erforderlich, um genauere Schlussfolgerungen und Bewertungen über beide Sensoren zu machen. Auf der Ebene der allgemeinen Entdeckungen zeigen unsere Ergebnisse, dass der Radarsensor wetterbeständig ist, während der Kamerasensor Leistungsänderungen bei verschiedenen Wetterbedingungen zeigte. Beide Entdeckungen wurden aus der Literatur erwartet und konnten mit unseren Auswertungen bestätigt werden. Schliesslich haben wir auch gezeigt, dass es nicht ausreicht, ein allgemeines statistisches Mass wie die mittlere durchschnittliche Genauigkeit auf einem Testdatensatz zu haben. Zu statistischen Massen wie FNR müssen weitere Dimensionen hinzugefügt werden. Wir haben zum Beispiel gezeigt, wie sich die FNR mit unterschiedlichen Abständen zum Testobjekt ändert. Dies ist eine relevante Information, wenn wir über Sensorfusion, Zustandsschätzung und Planung nachdenken. Dimensionen wie unterschiedliche Objektgrössen haben ebenfalls einen Einfluss auf die Objekterkennungsleistung. Mit unserem Vergleich zwischen Erwachsenen- und Kinder-Fussgängererkennung haben wir gezeigt, wie sich die Erkennungsleistung für zwei Objekte mit gleicher Form, aber unterschiedlicher Grösse ändert. Wetterbedingungen und Verschmutzung sind ebenfalls relevante Grössen. Viele weitere Dimensionen könnten hinzugefügt werden, wie z. B. die relative Orientierung zu den Objekten, die Dynamik der Objekte oder die relative Geschwindigkeit zwischen dem AD-Fahrzeug und dem Hindernis. Aus der parabolischen Form der FNR-Kurven lässt sich ableiten, dass die Bewegung des AD-Fahrzeugs die Sensorleistung beeinflusst. Zusätzliche Dimensionen könnten auf der Sensorebene liegen, wie z. B. unterschiedliche Belichtungszeiten für die Kamerasensoren. All diese zusätzlichen Dimensionen, die sich auf die Leistung der Sensoren auswirken, sind für die Sicherheitsbewertung von AD-Fahrzeugen, die Sensorauswahl, die Sensorfusion, das AD-Fahrzeug-Co-Design oder den Reglerentwurf für AD-Funktionen relevant.

Lidar Targets

In diesem Szenario testen wir spezifisch den Lidar Sensor bezüglich seiner Eignung zur Erkennung und Lokalisierung von Objekten. Um den Einfluss des Wetters auf die Lidar-Detektionsleistung zu testen, haben wir die in verschiedenen Farben lackierten Holz- Ziele auf der Teststrecke installiert. Das Szenario ist so aufgebaut, dass wir auf der Gegengerade der Teststrecke mit dem AD-Fahrzeug aus einer Entfernung von 70m bis auf 5m auf die Lidar-Ziele zufahren.

Im Gegensatz zur Mobileye-Kamera und dem Delphi-Radar mit ihren eingebauten Erkennungsalgorithmen, liefert das Lidar nur Rohdaten. Um die gleichen Performance-Metriken verwenden zu können, müssten wir also die Lidar-Rohdaten mit einem

entsprechenden Erkennungsalgorithmus verarbeiten. Da unser Ziel jedoch nur der Test des Sensors ist, haben wir andere Metriken definiert. Die gewählten Metriken, Anzahl Punkte pro Target, Reflektivität und Intensität, spiegeln die Informationen wider, die in den Lidar-Detektionsalgorithmen verwendet werden und erlauben uns daher, die potenzielle Lidar-Detektionsleistung nur auf Basis der Rohdaten zu analysieren und zu bewerten.

Unsere Ergebnisse aus den Feldtests, den Tests in unserem Laborgebäude und den Labortests am METAS zeigen, dass die Farbe des Objekts und der Einfallswinkel zwischen dem Laser und der Oberfläche des Objekts einen starken Einfluss auf die Leistung des Lidars haben. Insgesamt verhalten sich die weissen und grauen Oberflächen wie vom Funktionsprinzip des Sensors erwartet, da die gemessene Reflektivität der weissen Oberfläche durchweg höher ist als die der grauen und die Variabilität zwischen aufeinanderfolgenden Durchläufen/Testbedingungen meist gering ist. Während jedoch erwartet wird, dass die Reflektivität unabhängig von der Entfernung zum Objekt ist, sehen wir eine kontinuierliche Abnahme der Reflektivität von den weissen und grauen Targets für Entfernungen unter 30 m. Dies könnte mit der Sättigung der Intensitätsmessung bei kurzen Entfernungen zusammenhängen. Im Gegensatz dazu haben wir keine umfassende Erklärung für das seltsame Verhalten der schwarzen Fläche, für die die Reflektivität eine viel höhere Variabilität und einen völlig anderen Trend mit der Entfernung zeigt als die der weissen und grauen Flächen. Die schwarze Fläche führt auch zu unerwarteten Ergebnissen in Bezug auf die Anzahl der Punkte, die bei kleinen Entfernungen deutlich abfallen, anstatt exponentiell anzusteigen, wie für die weissen und grauen Ziele beobachtet und von der Theorie erwartet. Dies hängt höchstwahrscheinlich mit der Tatsache zusammen, dass das Target bei geringen Entfernungen möglicherweise nicht perfekt senkrecht zum Laser steht, was, wie die Tests am METAS und in unserem Workshop gezeigt haben, die Erkennung schwarzer Oberflächen noch schwieriger machen kann. Noch schlimmer wird es, wenn die Oberfläche glänzend statt matt oder halbmatt ist, wie in den Ergebnissen von METAS zu sehen ist.

Unsere Ergebnisse aus dem Szenario "Lidar Targets" zeigen, dass die Reflektivitätsmessung bei schlechten Wetterbedingungen wie Regen, Nebel oder Schnee insgesamt verrauschter ist und eine höhere Variabilität zwischen aufeinanderfolgenden Messungen aufweist als bei sonnigen Bedingungen. Im Gegensatz dazu scheint die Verschmutzung der Lidar-Oberfläche die gemessene Reflektivität nicht signifikant zu beeinflussen. Auf der anderen Seite scheint die Anzahl der Punkte, die jedes Ziel erreichen, weder von den unterschiedlichen Wetterbedingungen noch von der Sensorverschmutzung beeinflusst zu werden.

Grundsätzlich konnten wir beobachten, dass die Lidar-Messungen nicht dramatisch von ungünstigen Wetterbedingungen beeinflusst werden. Allerdings stellt die Tatsache, dass dunkle Objekte bei nicht senkrechten Einfallswinkeln und bei grösseren Entfernungen nicht genügend Energie reflektieren, ein grosses Problem für den Einsatz des Lidars zur Objekterkennung dar.

Lidar Localization

Im Gegensatz zu den drei anderen Szenarien geht es bei diesem Szenario nicht um die Objekterkennung, sondern um die Lokalisierung des Fahrzeuges mittels Lidar-Sensor. Das Szenario ist so definiert, dass wir bei verschiedene Wetter- und Verschmutzungsbedingungen jeweils mehrere Runden auf der Teststrecke fahren um die Lidar-Lokalisierungsfunktionalität zu bewerten.

Idealerweise würde man die Differenz zwischen der wahren und der geschätzten Position als Metrik für die Lokalisierung verwenden, wie z. B. den quadratischen Mittelwert oder den maximalen Distanzfehler. Die grösste Herausforderung hierbei ist, wie man die wahre Position erhält, da die von unserem xSens GPS-Empfänger gelieferte GPS-Position nicht genau genug ist. Das andere Problem ist, dass man für die Lidar-Lokalisierung eine Punktwolkenkarte benötigt, die genau genug georeferenziert ist, d.h. mit weniger Fehler als die Lokalisierung selbst. Um diese beiden Probleme zu überwinden, verwenden wir anstelle der Analyse der Lokalisierungsgenauigkeit Metriken, die aus dem

Lokalisierungsalgorithmus berechnet werden, um zu beurteilen, wie schnell und robust der Lokalisierungsalgorithmus konvergiert.

Die Ergebnisse aus den Lidar-Lokalisierungstests unter verschiedenen Wetterbedingungen und Verschmutzungen deuten darauf hin, dass aus reiner Lokalisierungssicht die Auswirkungen von ungünstigen Wetterbedingungen auf das Lidar nicht dramatisch sind. Insgesamt sehen wir keine grossen Schwankungen innerhalb der Ergebnisse und keine wesentliche Verschlechterung der Lokalisierungsleistung bei Regen, Nebel oder Schnee im Vergleich zu sonnig. Selbst wenn ein volles Viertel des Sichtfeldes des Lidars durch den anhaftenden Schnee blockiert war, konnte der Lokalisierungsalgorithmus immer konvergieren. Die gewonnen statistischen Daten unterstützen diese Schlussfolgerung, wo wir, ausser im Fall von Schneefall, keinen signifikanten Anstieg der Anzahl der Iterationen und der Punktzahl bei sich verschlechternden Bedingungen sehen. Der einzige deutlich sichtbare Anstieg der Punktzahl ist bei Schneefall, aber auch in diesem Fall änderte sich die Anzahl der Iterationen bis zur Konvergenz nicht.

Unsere Ergebnisse zeigen also, dass Algorithmus und Sensor nicht immer getrennt voneinander betrachtet werden können. Tatsächlich hatte der Lokalisierungsalgorithmus in allen getesteten Fällen eine ähnliche Leistung, auch wenn sich die reine Sensorleistung deutlich unterschied, wie die Punktwolkenvisualisierungen zeigen.

Bislang haben wir nur die Lokalisierung in einer strukturierten städtischen Umgebung untersucht, in der es viele strukturierte Objekte im FoV gibt. Es könnten weitere Dimensionen hinzugefügt und verglichen werden, wie z. B. das Fahren an Orten, an denen die 3D-Karte weniger Merkmale zeigt, z. B. Autobahnen, Landstrassen usw. Allerdings ist die GNSS-Lokalisierungsleistung in diesen Fällen weniger problematisch, da die Sichtlinie zu den Satelliten weniger gestört ist. Aber alles in allem werden diese zusätzlichen Überlegungen für eine zuverlässige Lokalisierung von AD-Fahrzeugen in allen Fahrumgebungen relevant sein.

Fazit

In diesem Projekt haben wir gezeigt, wie AD-Sensoren mit funktions-szenariobasierten Methoden getestet werden können. Das funktions-szenariobasierten Testen bietet die Vorteile von AD-Tests in der realen Welt, d. h. die Sensoren werden während des realen Fahrens getestet und nicht durch mathematische Modelle der Sensoren, wie in Simulationen. Ausserdem werden die Sensoren in der realen Welt gemessen, anstatt wie bei Vehicle-in-the-Loop- oder Hardware-in-the-Loop-Testsystemen gefälscht zu werden. Da die Szenarien gut definiert sind, sind die Tests wiederholbar und die relevanten Testspezifikationen wie Wetterbedingungen oder Objektabstand können extrahiert werden. Zudem werden die AD-Wahrnehmungsfunktionen (z. B. Objekterkennung oder Objektverfolgung) separat getestet, so dass Schwachstellen und Fehlerursachen besser erkannt werden können. Die Kombination mit Laborexperimenten ermöglicht es, Ergebnisse aus Feldtests in Laborumgebungen zu verifizieren und umgekehrt. Neben dem Testverfahren haben wir auch einige sensorspezifische Ergebnisse für die Mobileye-Kamera, das Delphi ESR-Radar und das Velodyne VLP-16-Lidar vorgestellt.

Von allen getesteten Sensoren wird die Mobileye-Kamera eindeutig am stärksten von den Umgebungsbedingungen beeinflusst. Tatsächlich zeigen unsere Ergebnisse deutlich, dass sowohl die Objekterkennung als auch die Objektverfolgungsfunktionalität der Mobileye-Kamera durch ungünstige Wetterbedingungen wie Regen, Nebel oder Schnee negativ beeinflusst werden. Die Ergebnisse aus dem Walking Pedestrian-Szenario zeigen, dass die Kamera bei guten Lichtverhältnissen eine Tracking-Quote nahe 100 % hat, während bei schlechten Lichtverhältnissen die Tracking-Quote Null erreicht. Darüber hinaus führen ungünstige Wetterbedingungen zu einer schlechteren Tracking-Stabilität, d. h. der Track geht häufiger verloren, die Anzahl der erzeugten Tracks steigt und die durchschnittliche Trackdauer sinkt. Ähnliche Erkenntnisse ergaben die Analysen der Tests im Urban Mixed-Szenario hinsichtlich der Objekterkennungsleistung, die zeigten, dass die FNR der Mobileye-Kamera für die Fahrzeug- und Fussgängererkennung bei verschneiten und regnerischen Wetterbedingungen signifikant ansteigt. Insgesamt zeigen unsere

Ergebnisse, dass die Abnahme der Beleuchtungstärke und das Rauschen, welches in den Kamerabildern durch Regentropfen, Schneeflocken und Nebel erzeugt wird, die Objekterkennungs- und Verfolgungsfunktionalitäten der Mobileye-Kamera negativ beeinflussen.

Die Tests in den Szenarien "Walking Pedestrian" und "Urban Mixed" wurden auch zur Analyse der Entfernungsschätzungsfunktionalität der Mobileye-Kamera verwendet. In beiden Szenarien führte die Mobileye-Kamera unter allen Umgebungsbedingungen zu sehr schlechten Abstandsschätzungen. Insgesamt neigt die Mobileye-Kamera dazu, die Entfernungen zu den Objekten zu unterschätzen wobei der absolute Fehler linear mit der Entfernung ansteight. Nur für den Kinderfussgänger überschätzt die Mobileye-Kamera die Entfernung bei grossen Entfernungen, wahrscheinlich, weil sie statt eines Kindes einen weiter entfernt stehenden erwachsenen Fussgänger sieht.

Im Urban Mixed-Szenario haben wir auch den Effekt von verschiedenen Objektgrössen und -typen untersucht. Unsere Ergebnisse zeigen, dass die Mobileye-Kamera Fahrzeuge besser erkennt als Fussgänger, und dass unter den Fussgängern erwachsene Fussgänger leichter erkannt werden als Kinder. Diese Ergebnisse hängen höchstwahrscheinlich mit den unterschiedlichen Objektgrössen zusammen, da grössere Objekte durch mehr Pixel in den Kamerabildern repräsentiert werden, was ihre Erkennung erleichtert. Aber auch Unterschiede in den Mobileye-Modellen für verschiedene Objekttypen könnten eine Rolle spielen (z. B. könnten Fahrzeuge eine charakteristischere Form haben als Fussgänger).

Schliesslich haben wir im "Mixed Urban"-Szenario auch die Veränderungen im FNR der Mobileye-Kamera für die Fussgänger- und Fahrzeugerkennung in Abhängigkeit von den Objektabständen betrachtet. Während das Fahrzeug bei allen getesteten Entfernungen richtig erkannt wird (d.h. FNR nahe Null), verschlechtert sich die Erkennung von Fussgängern bei grösseren Entfernungen, da die Anzahl der Pixel des Objekts auf dem Bild abnimmt. Ausserdem steigt der FNR für die Fussgängererkennung bei Entfernungen unter 10 m, was wir auf die Bewegungsunschärfe der Kamerabilder aufgrund der hohen relativen Tangentialgeschwindigkeit zwischen dem sich bewegenden AD-Fahrzeug und dem statischen Objekt zurückführen.

Die Szenarien "Walking Pedestrian" und "Mixed Urban" wurden ebenfalls verwendet, um die Objekterkennungs- und -trackingleistung des Delphi ESR-Radars zu analysieren. Die Tests im Walking Pedestrian-Szenario zeigten, dass das Radarsystem eine geringere Trackstabilität und Verfolgungsrate als die Mobileye-Kamera hat. Zusätzlich haben wir beobachtet, dass die Leistung des ESR-Radars stark von der Strassenumgebung beeinflusst wird. Tatsächlich beobachteten wir bei Vorhandensein bestimmter Fassadengeometrien, metallischer Container und Zäune eine hohe Anzahl von Trackverlusten und Lücken in den Tracks des gehenden Fussgängers. Andererseits liefert das ESR-Radar bei Vorhandensein eines Tracks eine sehr präzise Entfernungsmessung unabhängig von den Wetterbedingungen und der Objektentfernung. Aufgrund des grossen Einflusses der Umgebung ist das Radar jedoch nicht zuverlässig genug für Fussgängererkennung.

Von unseren Tests im Szenario "Mixed Urban" haben wir nur die Erkennungsleistung des Radars für Fahrzeuge analysiert, da die Schaufensterpuppen einen anderen Radarquerschnitt als der menschliche Körper haben und daher zu keinen aussagekräftigen Ergebnissen geführt hätten. Wie im Szenario "Walking Pedestrian" lieferte das Radar eine sehr genaue Entfernungsmessung, unabhängig von den Wetterbedingungen und der Entfernung. Im Fall des Fahrzeugs waren auch die Erkennung und Verfolgung des ESR-Radars sehr gut, was zu sehr niedrigen Falsch-Negativ-Raten für alle Entfernungen und Wetterbedingungen führte (FNR immer unter 10 % und sehr nahe an 0 % für Objektentfernungen zwischen 5 und 70 m). Nur im Fall des kontaminierten Sensors beobachteten wir einen Abfall der Radarerfassungsleistung zwischen 40 und 50 m Entfernung. Der Grund für diesen Effekt ist uns bisher nicht klar. Wahrscheinlich haben Schmutz, Laub und Reifenabrieb auf der Radar Oberfläche einen gewissen Einfluss auf die Radarreflexion, aber weitere Tests werden nötig sein, um dies zu bestätigen.

Einer der grössten Nachteile des ESR-Radars, den unsere Tests aufgedeckt haben, ist seine hohe Empfindlichkeit, die zu Geistererfassungen (Falsch Positive) und Reflexionen führt, wie sowohl in unseren Feldversuchen als auch bei den Labormessungen am METAS festgestellt wurde. Für den Einsatz in einem AD-Fahrzeug müssen diese Geisterstracks dann sorgfältig behandelt und herausgefiltert werden, was zu einer höheren Komplexität führt und potenziell die Fehlerrate erhöht.

Generell zeigen unsere Ergebnisse, dass bei separater Betrachtung weder die Mobileye-Kamera noch das Delphi ESR-Radar eine akzeptable Leistung bei der Fussgängererkennung aufweisen, da jeder Sensor bei bestimmten Umgebungs- oder Testbedingungen deutliche und reproduzierbare Schwächen zeigt. Die Leistung von Mobileye nimmt bei ungünstigen Wetterbedingungen ab, während das Radar Probleme zeigt, wenn sich Gebäude und andere Objekte entlang der Strasse befinden. Daher würde ein AD-System, um sicher zu sein, stark von mehreren Sensoren und Sensorfusion profitieren, um Fussgänger zuverlässig zu erkennen und zu verfolgen.

Wir untersuchten auch die Leistung der Erkennungs- und Lokalisierungsfunktionalität des Velodyne VLP-16-A Lidarsensors. Für die Bewertung der Erkennungsfunktionalität verwendeten wir unser Lidar-Target-Szenario. Die Ergebnisse dieser Praxistests zeigten uns die Anzahl der von den Testobjekten reflektierten Punkte unter verschiedenen Umweltbedingungen. Interessanterweise zeigten unsere Tests grosse Leistungsprobleme des Lidars bei dunkel gefärbten Oberflächen, wie z.B. schwarz und dunkelblau. Insbesondere werden diese dunkleren Oberflächen vom Lidar nicht erkannt, wenn der Einfallswinkel zwischen dem Laserstrahl und der Oberfläche nicht senkrecht ist. Diese Erkenntnisse aus unseren Praxistests wurden auch durch die Ergebnisse der Messungen am METAS bestätigt. Die Tatsache, dass dunkle Objekte bei nicht ganz senkrechten Winkeln keine Energie reflektieren, stellt ein Problem für den Einsatz des Lidars zur Objekterkennung dar. Für die anderen getesteten Farben (grau und weiss) fanden wir insgesamt vernünftigere Ergebnisse, bei denen die Anzahl der Punkte auf beiden Zielen immer sehr nahe an der theoretischen Anzahl für alle Entfernungen und Wetterbedingungen lag. Für beide Farben wird die gemessene Reflektivität nur geringfügig durch ungünstige Wetterbedingungen beeinflusst (zumindest nicht so stark wie bei der Mobileye-Kamera). Bei sonnigen Bedingungen bleibt die Reflektivität über alle Entfernungen hinweg ziemlich konstant, mit einer Abnahme bei Entfernungen unter 10 m. Die Abnahme unter 10 m könnte mit der Erkennungssättigung und einem begrenzten numerischen Bereich zusammenhängen. Wir sehen auch eine erhöhte Anzahl von Ausreissern in den Reflektivitätswerten bei schlechteren Bedingungen und insgesamt etwas mehr Variation.

Die Ergebnisse aus dem Lidar-Lokalisierungsszenario zeigen, dass der Lokalisierungsalgorithmus recht robust gegenüber Sensorrauschen und Beeinträchtigungen durch ungünstige Umgebungsbedingungen ist. Obwohl bestimmte ungünstige Wetterbedingungen die Lidar-Punktwolken deutlich verschlechtern, konnten wir für keine der getesteten Bedingungen, einschliesslich Sonne, Regen, Nebel, Schnee und Verschmutzung, eine signifikante Verschlechterung der Lokalisierungsleistung beobachten. Tatsächlich konvergierte der Lokalisierungsalgorithmus immer erfolgreich, selbst wenn ein volles Viertel des Sichtfeldes des Lidars durch anhaftenden Schnee blockiert war. Insgesamt zeigen unsere Ergebnisse, dass für eine reine Lokalisierung die Leistung des Lidars selbst nicht der limitierende Faktor ist.

Im Rahmen dieses Projekts haben wir die Leistung verschiedener AD-Sensoren unter realen Fahrbedingungen untersucht. Die Fragen, die sich nun stellen, sind: Wie können wir diese Ergebnisse für die AD-Validierung nutzen? Wie hängt die endgültige AD-Leistung mit der Leistung der Sensoren zusammen? Und welche Auswirkungen haben die Sensorfusion und die Regelalgorithmen? Spezifikationen wie die FNR-Kurven, die wir aus unseren Tests erhalten haben, können verwendet werden, um probabilistische Sensormodelle zu erstellen, die dann in Simulationen verwendet werden können. Dabei hätte das Sensormodell die gleiche Unsicherheit wie die reale Hard- und Software, wodurch die Simulation näher an der Realität wäre. Sobald präzise und realistische Spezifikationen in Tests in der realen Welt ermittelt wurden, kann die Simulation zur Validierung vieler weiterer Szenarien verwendet werden und damit den Validierungsprozess beschleunigen.

Darüber hinaus können die Leistungsanalysen einzelner Sensoren dabei helfen, bessere Sensorfusionsmodelle und Zustandsschätzungsmodelle für AD zu erstellen, da sie wichtige Erkenntnisse über die Genauigkeit und Unsicherheit der Sensoren liefern. Auch ohne direktes Testen des gesamten AD-Stacks können einige allgemeine Schlussfolgerungen direkt aus den Sensortestergebnissen gezogen werden, z. B. wäre unser Testfahrzeug nicht in der Lage, bei Schneefall nachts sicher und effizient zu fahren, wenn nur die aktuell installierten Sensoren verwendet würden.

Ausblick

Die Infrastruktur und das erworbene Wissen aus diesem Projekt eröffnen viele Möglichkeiten und neue Forschungsthemen im Bereich der AD-Prüfung. Um die gewonnenen Ressourcen optimal zu nutzen, möchten wir unsere funktions-szenariobasierten Sensortestaktivitäten mit zusätzlichen Sensoren, Wahrnehmungsalgorithmen und Szenarien fortsetzen.

Zunächst wollen wir einige Tests wiederholen oder erweitern, um einige der in diesem Projekt entdeckten Probleme in der Sensorleistung weiter zu untersuchen. Zum Beispiel soll der Test mit dem kontaminierten Radar wiederholt werden, um die beobachtete erhöhte FNR für die Fahrzeugerkennung zu verstehen. Auch die zweiten lokalen Minima in den FNR-Kurven für die Fussgängererkennung mit der Mobileye-Kamera bei nebligen und verschneiten Wetterbedingungen sollen weiter untersucht werden. Ausserdem planen wir, die Sensoren auch bei dichterem Nebel zu testen, um zu sehen, ob/ wann die Leistung der Mobileye-Kamera abfällt. Weiterhin wollen wir die Experimente zur Fahrzeugerkennung mit einem kleineren Fahrzeug wiederholen, um Unterschiede zum Iveco Daily zu sehen. Das merkwürdige Verhalten des Lidars bei kurzen Entfernungen und dunklen Oberflächen soll ebenfalls weiter analysiert werden, um herauszufinden, ob es mit einem Sensorfehler zusammenhängt oder ob dies bei allen Lidars der Fall ist. Für die Lokalisierungsleistung wollen wir den Grad der Sensordegradation (in Form von Umgebungsrauschen) bestimmen, der notwendig ist, damit der Lokalisierungsalgorithmus nicht konvergiert.

Ausserdem wollen wir neue Kameras testen und verschiedene Open-Source-Wahrnehmungsalgorithmen anwenden. Dies würde uns erlauben, sowohl die rohen Kamerabilder als auch die Arbeitsweise der Algorithmen zu analysieren. Neben konventionellen RGB-Kameras planen wir, auch ereignisbasierte und Infrarot-Kameras zu testen. Wir wollen auch Open-Source-3D-Objekt-Erkennungsalgorithmen verwenden, um die Leistung des Velodyne VLP-16-Lidars für die Fussgänger- und Fahrzeugerkennung zu evaluieren. In diesem Zusammenhang wäre es auch interessant, ein Lidar mit mehr als 16 Kanälen zu testen.

In Bezug auf die Umgebungsbedingungen wollen wir Methoden entwickeln, um die Wetter- und Verschmutzungsbedingungen mit Zahlen zu quantifizieren, um z.B. die Veränderungen in der Erfassungsleistung bei unterschiedlichen Regenfällen oder Verschmutzungen zu analysieren. Neben dem Wetter- und Verschmutzungseinfluss wollen wir auch die Alterungs- und Abnutzungseffekte von AD-Sensoren untersuchen, sowie die Interferenzen zwischen ähnlichen Sensoren, die auf zwei verschiedenen AD-Fahrzeugen montiert sind.

Unsere Ergebnisse aus den funktions-szenariobasierten Tests haben gezeigt, dass viele Parameter die Erfassungsleistung beeinflussen können: die Fahrzeuggeschwindigkeit, der Objektabstand, die Objektgrösse, die Wetterbedingungen. Daher wollen wir unsere Tests aus dem Mixed Urban-Szenario mit unterschiedlichen Fahrzeuggeschwindigkeiten und möglicherweise auch mit dynamischen Objekten wiederholen. Diese zusätzlichen Tests würden uns helfen, alle relevanten Faktoren, die die Leistung der Sensoren beeinflussen, zu bestimmen und eine Sensitivitätsanalyse ihrer Relevanz für die AD-Wahrnehmungsleistung durchzuführen. Die detaillierten Analysen der Leistung der einzelnen Sensoren können dann verwendet werden, um Methoden der Sensorfusion zu analysieren und um festzustellen, wie gut verschiedene Sensorkombinationen unter verschiedenen Bedingungen funktionieren. Schliesslich können wir durch die Verwendung

der aus unseren Praxistests abgeleiteten Sensormodelle in Simulationen die Erfassungsleistung mit der AD-Leistung verknüpfen.

Résumé

Ce document rend compte des activités et des résultats du projet "Véhicule d'essai de capteurs AD". Le projet a été lancé pour comprendre et construire des procédures de validation de la sécurité pour la conduite automatisée. Les capteurs jouent un rôle crucial dans le système de conduite automatisée, car ils fournissent des informations pertinentes sur l'environnement. Le projet s'est concentré sur l'évaluation des capteurs AD par des essais sur le terrain lors d'une conduite réelle sur une piste d'essai et sur la comparaison des résultats avec les données des essais expérimentaux en laboratoire réalisés à METAS. À cette fin, un véhicule Lexus RX 450h a été équipé d'un ensemble de capteurs AD disponibles dans le commerce parmi les plus importants, à savoir un lidar, une caméra et un radar. En outre, une piste d'essai avec des objets d'essai a été mise en place sur le campus de l'Empa à Dübendorf.

L'approche de test basée sur des scénarios de fonctions a été utilisée pour évaluer les performances des capteurs. Dans cette procédure, une fonction de perception spécifique de l'AD, par exemple la reconnaissance des objets, est testée dans un scénario de conduite réel. Cette approche permet à la fois la répétabilité des tests sur le terrain et une compréhension précise du scénario. À l'aide de différents scénarios, nous avons testé les performances de détection d'objets, d'estimation de la distance, de suivi et de localisation de la caméra Mobileye, du radar Delphi ESR et du lidar Velodyne VLP-16. Étant donné que certains des capteurs, comme la caméra Mobileye, sont des systèmes à boîte noire, nous avons développé de nouvelles spécifications d'évaluation, comme le ratio de suivi, qui indique la sensibilité ou la robustesse de la fonctionnalité de suivi des objets dans le temps. La conception de nos scénarios nous a permis d'obtenir des informations précises sur les performances des capteurs, par exemple le taux de faux négatifs de la détection d'objets à différentes distances d'objets. Des essais sur le terrain ont été réalisés dans différentes conditions météorologiques et avec des capteurs sales.

Notre analyse a montré que la détection d'objets de la caméra Mobileye dépend fortement des conditions météorologiques, avec des dégradations importantes des performances dans des conditions de neige et de pluie. Le type, la taille et la distance de l'objet ont également influencé de manière significative les performances de détection de la caméra Mobileye. Dans l'ensemble, les véhicules ont été détectés plus facilement que les piétons et les piétons de grande taille (adultes) plus facilement que les piétons de petite taille (enfants). En général, les performances de détection se sont fortement dégradées à longue distance, mais aussi à très courte distance, ce que nous avons attribué aux effets de flou de mouvement causés par le véhicule d'essai en mouvement. Les estimations de distances de la caméra Mobileye se sont avérées très faibles dans toutes les conditions testées, avec une nette tendance à sous-estimer la distance réelle. Le radar ESR a montré de bonnes performances de détection dans toutes les conditions environnementales et a fourni des mesures de distance très précises. Cependant, nous avons observé que la fonctionnalité de suivi des objets est assez sensible, car le capteur perd souvent la trace et/ou suit le même objet plusieurs fois. Le lidar Velodyne a montré de bonnes performances de localisation dans toutes les conditions environnementales testées. Même un champ de vision partiellement bloqué en raison de la neige accumulée sur la surface du capteur n'a pas entraîné d'erreur de localisation. Cependant, nos tests sur le terrain et nos expériences en laboratoire ont montré que les surfaces de couleur plus foncée ne sont pas vues par le lidar si l'angle d'incidence entre le faisceau laser et la surface n'est pas perpendiculaire, ce qui constitue un problème majeur pour la détection des objets.

Méthodes d'essai

Dans le premier chapitre, les différentes approches pour valider la sécurité des véhicules automatisés sont décrites. Les différentes procédures sont brièvement présentées et les avantages et inconvénients spécifiques sont soulignés.

Dans ce projet, l'approche basée sur les scénarios de fonctions est utilisée. L'idée est de décomposer la conduite automatisée en situations de trafic individuelles typiques. Ainsi,

on essaie de réduire l'ensemble infini de situations de conduite possibles à un ensemble fini de scénarios à tester. Cet ensemble fini de tests peut être réalisé en simulation, sur une piste fermée ou sur la route. L'avantage de cette approche est que nous pouvons nous concentrer sur les situations critiques sans consacrer de temps de test aux tâches simples. Le plus grand défi de cette approche est qu'elle suppose que si le véhicule automatisé (AV) passe avec succès un certain scénario, il passera également d'autres scénarios similaires.

Capteurs et fonctions de perception pour la conduite automatisée.

Dans le deuxième chapitre, les capteurs AD et les fonctions de perception les plus importants sont présentés. Les différentes technologies de capteurs (caméra, radar, lidar...) sont utilisées pour collecter des informations en temps réel sur l'environnement du véhicule. Grâce au traitement des données, les capteurs peuvent détecter et suivre des objets dans l'environnement du véhicule. Les données recueillies par différents capteurs sont combinées par fusion de données et traitées en temps réel pour créer une représentation complète de la réalité. Celle-ci est utilisée par des algorithmes pour analyser la situation, planifier les trajectoires possibles et déterminer les actions à entreprendre (direction, freinage...).

Notre véhicule d'essai AD est équipé d'un système radar Delphi ESR 2.5 disponible sur le marché, qui est installé dans le pare-chocs avant. L'ESR 2.5 est un radar à semi-conducteurs de 76,5 GHz avec un faisceau orienté électroniquement et un traitement du signal Doppler pulsé. Le radar permet de mesurer simultanément un champ de vision large à moyenne portée et un champ de vision étroit à longue portée. Monté sur la galerie de toit du véhicule, un Velodyne VLP-16-A Li-DAR Puck, un lidar rotatif à 16 canaux. Le Velodyne LiDAR Puck a une portée de 100 m avec une précision de 3 cm. Il offre un champ de vision horizontal complet de 360° et un champ de vision vertical de ± 15°. Le système de caméra est un Mobileye 630 monté derrière le pare-brise. Il contient la puce de traitement d'image EyeQ2 avec les algorithmes de traitement d'image de Mobileye. Les algorithmes tentent de détecter les voies, les panneaux de signalisation, les véhicules, les cyclistes et les piétons et de mesurer la distance par rapport à ces objets.

Reconnaissance et classification des objets

La détection et la classification des objets sont des fonctionnalités essentielles pour les systèmes audiovisuels. Pour l'AD, les objets peuvent être détectés à partir d'images de caméras, de données radar et de nuages de points lidar. Les objets typiques qui doivent être détectés sont non seulement les véhicules et les piétons, mais aussi d'autres obstacles routiers tels que les arbres, les voies, les panneaux de vitesse, les feux de signalisation et bien d'autres encore. À partir d'images de caméra, un objet est généralement détecté à l'aide de méthodes d'apprentissage profond ou de méthodes classiques de vision par ordinateur basées sur la détection des bords. La reconnaissance d'objets en temps réel, notamment à partir d'images de caméras, est un sujet de recherche très actif, tant dans le milieu universitaire que dans l'industrie. De grands progrès ont été réalisés au cours de la dernière décennie grâce à l'amélioration des algorithmes et à l'augmentation de la puissance de calcul.

Suivi

Le suivi est le processus qui consiste à suivre des objets en mouvement. Dans de nombreuses applications, telles que la planification de trajectoires, il est souvent nécessaire de fournir la position et la vitesse de chaque objet à une fréquence de mise à jour fixe. Les traces sont obtenues avec un filtre qui utilise les données de détection et de localisation des objets. La conception d'un filtre pose de nombreux défis, tels que des données de capteur bruités, des détections manquantes ou fausses, et des données de capteur non synchronisées. En outre, tous les capteurs ne peuvent pas fournir toutes les données requises pour l'application. Par exemple, la vitesse de l'objet ne peut pas être déterminée à partir d'une seule image de caméra. Dans ce cas, un filtre de suivi doit estimer cette donnée.

Localisation

À l'instar des conducteurs humains dans les voitures conventionnelles, les véhicules AD doivent connaître leur position absolue afin de se diriger vers leur destination. Pour cette navigation, une position GNSS d'une précision de l'ordre de quelques mètres est suffisante. En outre, les conducteurs humains s'appuient sur la localisation visuelle, c'est-à-dire la reconnaissance des limites des voies et de la route et la position relative par rapport à celles-ci, pour suivre leur chemin. De même, de nombreux véhicules AD s'appuient sur une localisation absolue très précise au niveau centimétrique et utilisent des cartes routières détaillées tout aussi précises pour naviguer. Pour une localisation aussi précise, la localisation par lidar est une approche très prometteuse, notamment pour les environnements urbains. Il fournit une précision dans la gamme de cm requise, similaire à celle d'un GNSS différentiel ou RTK. En outre, la localisation lidar ne souffre pas de perte de signal ou de problèmes de trajets multiples comme le GNSS et peut donc fournir une redondance ou servir de sauvegarde lorsque les méthodes de positionnement GNSS sont perdues ou défaillantes. La localisation Lidar fonctionne en comparant le nuage de points actuel avec une carte 3D généralement prédéfinie pour trouver la position sur la carte où le nuage de points actuel correspond le mieux à la carte.

Méthodologie

Afin d'évaluer les différents capteurs et les fonctions perceptives d'un véhicule AD dans des conditions de conduite réelles, nous avons besoin d'expériences réalistes, mais reproductibles. La répétabilité des expériences nous permet de comparer les performances des capteurs dans différentes conditions environnementales et de comparer les différents capteurs entre eux. Afin d'évaluer un test, il est toujours nécessaire de définir comment la performance d'un capteur ou d'une fonction est mesurée et évaluée. Par exemple, le taux de faux négatifs est utilisé pour les performances de détection. Afin de calculer le taux de faux négatifs, la réalité du terrain doit d'abord être déterminée pour chaque essai, c'est-à-dire qu'il faut préciser où se trouve l'objet à quel incrément de temps. Il s'agit d'une étape de l'évaluation des données qui est souvent effectuée manuellement.

Mise en place des tests et travaux préparatoires

Dans le chapitre préparation des tests et travaux préparatoires, les travaux nécessaires à la réalisation correcte des tests ultérieurs sont documentés. Dans un premier temps, les différents capteurs utilisés ont dû être calibrés. À cette fin, un système de coordonnées a été défini pour chaque capteur, la position exacte et l'alignement des systèmes de coordonnées des différents capteurs les uns par rapport aux autres ont été mesurés et un modèle mathématique a été mis en œuvre.

Afin de rendre nos scénarios d'essai reproductibles, une piste d'essai a été créée sur le campus de l'Empa à Dübendorf. Dans un premier temps, le tracé de la piste d'essai a été dessiné en CAO et sa conformité aux normes a été vérifiée. Les marquages routiers ont ensuite été appliqués sur le campus de l'Empa. Enfin, la piste d'essai a été cartographiée avec Lidar et une carte 3D précise de l'environnement d'essai a été créée.

En plus de la piste d'essai, divers objets d'essai ont été produits pour les différents scénarios. D'une part, plusieurs mannequins ont été achetés et transformés en piétons au moyen de peinture et de vêtements, qui peuvent être placés sur la piste d'essai. D'autre part, trois grandes planches de bois ont été construites sur des palettes et peintes avec de la peinture blanche, grise et noire, respectivement, qui nous servent de cibles de test Lidar.

Dans la dernière section du chapitre, nous définissons les différentes conditions météorologiques et environnementales dans lesquelles les capteurs seront testés. Pour nos expériences, il s'agit du soleil, de la pluie, du brouillard, des chutes de neige et de la l'encrassement.

Approche de test basée sur des scénarios fonctionnels

Dans le dernier chapitre, l'exécution, le traitement des données, les analyses et les résultats de nos tests sont consignés. Le chapitre est divisé en deux sections, la première

contient les trois différents scénarios pour tester la détection d'objets et l'estimation de la distance et la deuxième section contient le scénario pour tester la fonction de localisation.

Scénario pour les piétons

Dans le scénario de la marche d'un piéton, nous testons les capteurs AD - caméra et radar Mobileye - pour la détection, la localisation et le positionnement d'un piéton. Ici, le piéton part à 5 m du véhicule AD stationnaire et s'en éloigne en s'arrêtant pendant 10 s tous les 5 m. Dès qu'il atteint le marqueur à 50 m, le piéton s'arrête, fait demi-tour et se dirige à nouveau vers la voiture en faisant des étapes de 5 m. Pour les positions arrêtées, nous connaissons la distance exacte du piéton et pouvons donc évaluer avec précision les performances de l'estimation de la distance.

Comme mesures de performance, nous utilisons l'erreur de mesure ou d'estimation de la distance, le nombre de traces générées pour le piéton, la durée moyenne des traces, le ratio de suivi, c'est-à-dire le pourcentage du temps où le piéton a été suivi et, dans le cas de la caméra, si le piéton a été correctement classé comme piéton.

Nous avons réalisé le scénario de test dans différentes conditions météorologiques/environnementales définies et l'avons répété plusieurs fois. Les résultats de nos tests montrent que les estimations de distance du Mobileye dans ce scénario sont très faibles dans toutes les conditions environnementales. En général, la distance est gravement sous-estimée. L'erreur absolue et sa variabilité entre les répétitions du test semblent augmenter à des distances plus élevées, notamment dans des conditions de brouillard.

En examinant les résultats de suivi de tous les essais, nous constatons que dans de bonnes conditions d'éclairage, le taux de suivi du Mobileye est proche de 100 %. En revanche, dans des conditions de faible luminosité, le taux de suivi atteint zéro, ce qui était prévisible d'après les spécifications de Mobileye. Plus intéressant encore, dans de mauvaises conditions météorologiques, la stabilité du suivi est moins bonne, c'est-à-dire que la piste est perdue plus souvent, le nombre de pistes générées augmente et la durée moyenne des pistes diminue. Cela montre que la stabilité du suivi de la caméra Mobileye bénéficie clairement de bonnes conditions d'éclairage, telles que celles présentes par temps ensoleillé.

En revanche, l'estimation de la portée du radar Delphi ESR pour ce scénario est beaucoup plus précise (erreur dans la précision de notre positionnement/réalité au sol) et nous ne voyons ni une tendance à l'augmentation de la distance ni les effets de conditions environnementales défavorables. Bien que les fiches techniques du radar Delphi ESR indiquent que le capteur ne rapporte que des traces, dans nos tests nous voyons beaucoup de traces très courtes qui ressemblent plus à des détections uniques qu'à des traces filtrées. Le taux de suivi pour les longues pistes (c'est-à-dire les pistes de plus de 1 s) varie de 20 % à 85 %, la majorité des tests montrant un taux de suivi d'environ 45 %. Cela signifie que le radar ne suit pas bien le piéton, ou du moins pas en continu, la plupart du temps. Nous n'avons pas non plus observé d'impact significatif des conditions environnementales sur les performances de suivi. Dans l'ensemble, les résultats des essais pratiques avec le radar concordent bien avec les résultats des mesures en laboratoire à METAS, où un grand nombre de traces parasites et d'objets fantômes ont également été mesurés. En outre, nos essais sur le terrain ont montré que les performances du radar sont fortement influencées par les bâtiments et autres objets de grande taille situés le long de la route. En effet, les objets ayant une grande section transversale radar, tels que les poutres en forme de T sur notre bâtiment de laboratoire ou le conteneur métallique à côté de notre piste d'essai, semblent interférer avec le radar lors du suivi du piéton, ce qui semble être la cause de la perte de trajectoire observée du piéton à proximité de tels objets.

En comparant les résultats des deux capteurs, nous pouvons conclure que ni la caméra Mobileye ni le radar Delphi ESR n'offrent des performances acceptables pour la détection des piétons et l'estimation de la distance lorsqu'ils sont considérés individuellement, car les deux capteurs présentent des faiblesses importantes et reproductibles dans certaines conditions environnementales. En particulier, l'estimation de la distance par Mobileye est

très mauvaise dans l'ensemble et les performances de suivi diminuent dans des conditions météorologiques défavorables, tandis que le radar, qui fournit des estimations précises de la distance, présente des problèmes de performances de suivi lorsqu'il y a des bâtiments et d'autres objets le long de la route. Ainsi, avec notre scénario de piéton, nous avons pu montrer qu'un seul système de capteurs n'est pas suffisamment fiable pour fonctionner dans toutes les conditions, et que chaque capteur a ses forces et ses faiblesses spécifiques. Par conséquent, pour être sûr, un système AD bénéficierait grandement de la combinaison de plusieurs technologies de capteurs et de la fusion de capteurs pour détecter et suivre les piétons de manière fiable.

Scénario urbain mixte

Dans ce scénario, nous testons les capteurs AD Mobileye camera et radar en ce qui concerne la détection, la localisation et le positionnement des piétons et des véhicules. Nous voulons connaître les positions exactes des obstacles testés par rapport au véhicule AD afin de pouvoir analyser les changements de performance à différentes distances, angles ou vitesses du véhicule. À cette fin, nous avons développé ce scénario urbain mixte pour représenter un environnement urbain. Les objets dans ce scénario sont stationnaires et nous connaissons leur position exacte. Dans le scénario, le véhicule AD se déplace à une vitesse constante sur la ligne droite avant de nos pistes d'essai. Quatre mannequins (deux femmes, un homme et un enfant) sont installés le long de la ligne droite avant. Au bout de la ligne droite avant se trouve une camionnette immobile (Iveco Daily).

En plus des mesures déjà utilisées dans le scénario Walking Pedestrian, nous évaluons le taux de faux négatifs des capteurs. Nous voulons savoir à quel point le FNR est élevé pour différentes distances d'objets, types d'objets, tailles d'objets ou conditions environnementales. Ces dimensions nous permettent de mieux comprendre pourquoi certains capteurs sont meilleurs pour certaines tâches AD.

Nos résultats montrent que les performances de la caméra Mobileye en matière de détection d'objets dépendent fortement du type/de la taille de l'objet et des conditions météorologiques. Pour la détection des piétons, nous observons les meilleures performances (c'est-à-dire une courbe FNR moyenne plus basse pour toutes les distances) pour les conditions météorologiques ensoleillées et brumeuses. Comme indiqué dans la littérature, les performances de la caméra Mobileye devraient être meilleures dans des conditions ensoleillées, car le capteur de la caméra passive bénéficie de bonnes conditions d'éclairage et les images ne sont pas affectées par les gouttelettes d'eau, les flocons de neige ou le brouillard. La performance similaire obtenue pour les conditions de brouillard comme pour les conditions ensoleillées était plutôt inattendue. De même, pour le scénario "piéton", le brouillard n'a pas affecté les performances de suivi de Mobileye.Les expériences pour les scénarios "piéton" et "urbain mixte" ont été menées le même jour avec une différence de temps de quelques minutes, ce qui signifie que la densité du brouillard était plus ou moins la même. Nous supposons que la performance relativement bonne du capteur pour les conditions de brouillard peut être due au fait que le brouillard n'était pas assez dense pendant nos expériences pour avoir un impact sur les images de la caméra et la détection des objets pour des distances allant jusqu'à 50m. En revanche, des FNR plus élevés ont été obtenus pour toutes les distances dans des conditions pluvieuses et neigeuses, comme prévu en raison du bruit supplémentaire introduit par les gouttes/flocons dans l'image de la caméra.

En termes de performances globales, le Mobileye présente des performances de détection des piétons plutôt faibles. Il est intéressant de noter que toutes les courbes montrent une forme parabolique avec un minimum global principalement entre 15 et 25 m. La raison pour laquelle les FNR augmentent avec la distance est que la caméra capture progressivement moins d'informations des objets, c'est-à-dire que plus la distance est élevée, plus le nombre de pixels de l'objet dans l'image est faible. Les performances inférieures à de plus petites distances sont moins intuitives, puisque les objets plus proches devraient avoir plus de pixels. Nous pensons que la raison pourrait être liée à la vitesse tangentielle relative plus élevée entre la caméra et l'objet, ce qui entraîne un flou de mouvement dans les images de la caméra. Cependant, comme nous n'avons pas accès aux images brutes produites par la caméra Mobileye, nous ne pouvons pas prouver l'occurrence du flou de mouvement.

Outre les effets de la météo, nous avons également constaté des différences dans la détection des mêmes types d'objets de tailles différentes. À cette fin, nous comparons les courbes FNR pour la détection des piétons adultes et enfants dans des conditions météorologiques ensoleillées et neigeuses. Cela montre que les performances de la détection des piétons adultes sont nettement meilleures que celles de la détection des piétons enfants. Là encore, plusieurs facteurs peuvent jouer un rôle important. Du point de vue de la caméra, un enfant est plus petit qu'un adulte, et il apparaît donc plus petit dans l'image. Cela signifie qu'il y a moins de pixels / d'informations sur l'objet. Pour la détection des véhicules, la caméra Mobileye montre les meilleures performances dans des conditions ensoleillées avec une courbe FNR proche de 0. Dans le brouillard, nous observons à nouveau presque les mêmes performances. Dans des conditions pluvieuses, une plus grande incertitude est montrée et le FNR augmente à 0,6 entre 30 et 40m. Cela montre que les gouttelettes d'eau provenant de la pluie ont un impact sur la détection des véhicules et que cet impact augmente à des distances plus longues. La plus mauvaise performance est montrée pour la détection de véhicules dans des conditions de neige. Il semble que les flocons de neige aient une plus grande influence que les gouttes de pluie.

La comparaison entre les différents types d'objets montre que la caméra est plus performante pour les véhicules que pour les piétons. Il pourrait y avoir plusieurs explications à cet effet. Premièrement, l'Iveco Daily est beaucoup plus grand qu'un piéton et les images de la caméra contiennent donc plus d'informations utiles pour sa détection. En outre, les objets plus grands peuvent être moins sensibles au bruit du capteur ou au bruit ambiant causé par la pluie ou le brouillard. En outre, le flou de mouvement de la caméra lié au déplacement du véhicule AD a un impact plus important sur les petits objets. Enfin, les véhicules pourraient avoir des formes plus facilement compréhensibles par les machines. À l'avenir, nous prévoyons de mener des expériences avec des véhicules plus petits également afin d'avoir une meilleure répartition des véhicules possibles.

Enfin, nous avons analysé la précision de l'estimation de la distance de la caméra Mobileye pour la détection des piétons et des véhicules. Les piétons sont généralement détectés trop près (erreur positive) et la dispersion de l'erreur absolue augmente avec la distance. Cependant, dans le cas de l'enfant piéton, nous observons que la précision de l'estimation de la distance est très faible à des distances supérieures à 20 m et que l'enfant piéton est généralement détecté trop loin, probablement parce qu'il est plus petit que le modèle de piéton Mobileye. Pour l'Iveco Daily, l'erreur absolue tend à augmenter linéairement à des distances plus grandes, l'erreur augmentant davantage dans des conditions météorologiques pluvieuses et neigeuses. En outre, la variabilité de l'erreur absolue augmente de manière significative à des distances plus longues. L'erreur absolue est principalement positive, ce qui pourrait être lié au fait que l'Iveco Daily est un véhicule assez grand (très probablement plus grand que les véhicules utilisés pour l'ajustement du modèle Mobileye) et que l'algorithme d'estimation de la distance estime donc que le véhicule est plus proche qu'il ne l'est.

Contrairement à la caméra Mobileye, le radar ESR présente des performances de détection de véhicules relativement indépendantes de la météo. En particulier, la courbe moyenne du FNR est presque toujours nulle, ce qui signifie que le véhicule a été détecté à tout moment. Ce n'est que par temps de neige que l'on constate une incertitude légèrement plus élevée dans la courbe FNR pour les distances plus longues (de 50 à 70 m). La résistance aux intempéries est due au fait que les radars sont des capteurs actifs et que leur longueur d'onde est suffisamment grande pour ne pas être réfléchie par les flocons de neige ou les gouttes de pluie.

Il est intéressant de noter que nous avons découvert une baisse des performances de détection radar lorsque le capteur est encrassé. La courbe du FNR pour la détection de véhicules avec un radar encrassé montre que le FNR augmente de 0 à 0,9 à des distances comprises entre 40 et 50 m avant de baisser à nouveau. La raison de cet effet n'est pas claire pour nous. Il est possible que la saleté, les feuilles et l'usure des pneus déposés sur la surface du capteur pour ces tests aient un effet sur la réflectivité du radar. En outre, il se pourrait qu'à des distances comprises entre 40 et 60 m, une feuille obscurcisse tout ou partie du FoV du radar précisément dans la direction du véhicule. Nous prévoyons de répéter ces expériences à l'avenir pour étudier plus en détail les effets de la contamination.

Enfin, nous avons comparé les performances de la caméra Mobileye et du radar ESR pour la détection des véhicules. La plus grande différence est observée dans les conditions environnementales les plus difficiles, à savoir la neige. La courbe FNR de la caméra Mobileye dans la neige est beaucoup plus élevée que celle du radar ESR. Cela signifie que le radar ESR est plus fiable et plus résistant aux intempéries pour la détection des véhicules, mais la différence est assez importante.

En résumé, nous avons pu obtenir des informations à différents niveaux grâce à notre scénario urbain mixte. La caméra Mobileye a généralement montré des performances peu fiables en matière de détection des piétons et ne conviendrait donc pas comme capteur unique. Le radar ESR a une détection des véhicules assez fiable dans toutes les conditions météorologiques. Cependant, l'inconvénient du radar ESR est sa sensibilité. En outre, le radar Delphi a montré une dégradation des performances lorsqu'il était sale. Cependant, des mesures supplémentaires sont nécessaires pour tirer des conclusions et des évaluations plus précises sur les deux capteurs. Au niveau des découvertes générales, nos résultats montrent que le capteur radar est résistant aux intempéries, tandis que le capteur cameras a montré des changements de performance dans différentes conditions météorologiques. Ces deux découvertes étaient attendues dans la littérature et ont pu être confirmées par nos évaluations. Enfin, nous avons également montré qu'il ne suffit pas de disposer d'une mesure statistique générale telle que la précision moyenne sur un ensemble de données de test. D'autres dimensions doivent être ajoutées aux mesures statistiques telles que le FNR. Par exemple, nous avons montré comment le FNR varie en fonction de la distance par rapport à l'objet testé. Il s'agit d'une information pertinente lorsque nous pensons à la fusion de capteurs, à l'estimation d'état et à la planification. Les dimensions telles que les différentes tailles des objets ont également un impact sur les performances de détection des objets. Avec notre comparaison entre la détection des piétons adultes et enfants, nous avons montré comment les performances de détection changent pour deux objets ayant la même forme mais des tailles différentes. Les conditions météorologiques et l'encrassement sont également des variables pertinentes. De nombreuses autres dimensions pourraient être ajoutées, comme l'orientation relative par rapport aux objets, la dynamique des objets ou la vitesse relative entre le véhicule AD et l'obstacle. D'après la forme parabolique des courbes FNR, on peut en déduire que le mouvement du véhicule AD affecte les performances du capteur. Des dimensions supplémentaires pourraient se situer au niveau du capteur, comme des temps d'exposition différents pour les capteurs de la caméra. Toutes ces dimensions supplémentaires affectant les performances des capteurs sont pertinentes pour l'évaluation de la sécurité des véhicules AD, la sélection des capteurs, la fusion des capteurs, la co-conception des véhicules AD ou la conception des contrôleurs pour les fonctions AD.

Cibles Lidar

Dans ce scénario, nous testons spécifiquement le capteur lidar en termes d'aptitude à détecter et à localiser des objets. Pour tester l'effet de la météo sur les performances de détection du lidar, nous avons installé des cibles en bois peintes de différentes couleurs sur la piste d'essai. Le scénario est établi de telle sorte que nous conduisons le véhicule AD sur la ligne droite arrière de la piste d'essai à partir d'une distance de 70 m jusqu'à 5 m des cibles lidar.

Contrairement à la caméra Mobileye et au radar Delphi, dont les algorithmes de détection sont intégrés, le lidar ne fournit que des données brutes. Ainsi, afin d'utiliser les mêmes mesures de performance, nous devrions traiter les données lidar brutes avec un algorithme de détection approprié. Cependant, comme notre objectif est uniquement de tester le capteur, nous avons défini d'autres métriques. Les paramètres choisis, le nombre de points par cible, la réflectivité et l'intensité, reflètent les informations utilisées dans les algorithmes de détection lidar et nous permettent donc d'analyser et d'évaluer les performances potentielles de détection lidar sur la base des données brutes uniquement. Les résultats des essais sur le terrain, les essais dans notre bâtiment de laboratoire et les essais en laboratoire à METAS montrent que la couleur de l'objet et l'angle d'incidence entre le laser et la surface de l'objet ont une forte influence sur les performances du lidar. Dans l'ensemble, les surfaces blanches et grises se comportent comme prévu d'après le principe de fonctionnement du capteur, puisque la réflectivité mesurée de la surface blanche est

systématiquement supérieure à celle de la surface grise et que la variabilité entre les essais successifs est généralement faible. Cependant, alors que l'on s'attend à ce que la réflectivité soit indépendante de la distance à l'objet, nous observons une diminution continue de la réflectivité des cibles blanches et grises pour des distances inférieures à 30 m. Cela peut être lié à la saturation de la surface blanche. Cela pourrait être lié à la saturation de la mesure d'intensité à courte distance. En revanche, nous n'avons pas d'explication complète pour le comportement étrange de la zone noire, pour laquelle la réflectivité montre une variabilité beaucoup plus élevée et une tendance complètement différente avec la distance que celle des zones blanches et grises. La surface noire conduit également à des résultats inattendus en termes de nombre de points, qui chute significativement à de petites distances au lieu d'augmenter exponentiellement comme observé pour les cibles blanches et grises et attendu par la théorie. Ceci est très probablement lié au fait que la cible peut ne pas être parfaitement perpendiculaire au laser à de petites distances, ce qui, comme l'ont montré les tests effectués à METAS et dans notre atelier, peut rendre la détection de surfaces noires encore plus difficile. La situation est encore pire si la surface est brillante au lieu d'être mate ou semi-mate, comme le montrent les résultats de METAS.

Nos résultats du scénario "Cibles Lidar" montrent que les mesures de réflectivité sont globalement plus bruitées dans de mauvaises conditions météorologiques telles que la pluie, le brouillard ou la neige, et présentent une plus grande variabilité entre les mesures successives que dans des conditions ensoleillées. En revanche, la pollution de la surface du lidar ne semble pas affecter de manière significative la réflectivité mesurée. En revanche, le nombre de points atteignant chaque cible ne semble pas être affecté par les différentes conditions météorologiques ni par la pollution du capteur.

En général, nous avons observé que les mesures lidar ne sont pas dramatiquement affectées par des conditions météorologiques défavorables. Cependant, le fait que les objets sombres ne reflètent pas suffisamment d'énergie à des angles d'incidence non perpendiculaires et à de plus grandes distances pose un problème majeur pour l'utilisation du lidar pour la détection d'objets.

Localisation Lidar

Contrairement aux trois autres scénarios, celui-ci ne concerne pas la détection d'objets, mais la localisation du véhicule à l'aide d'un capteur lidar. Le scénario est défini de manière à ce que nous fassions plusieurs tours sur la piste d'essai dans différentes conditions météorologiques et de pollution afin d'évaluer la fonctionnalité de localisation du lidar.

Idéalement, nous utiliserions la différence entre la position réelle et la position estimée comme métrique pour la localisation, comme l'erreur quadratique moyenne ou l'erreur maximale de distance. Le plus grand défi ici est de savoir comment obtenir la position réelle, puisque la position GPS fournie par notre récepteur GPS xSens n'est pas assez précise. L'autre problème est que pour la localisation lidar, il faut une carte de nuages de points qui soit géoréférencée avec suffisamment de précision, c'est-à-dire avec moins d'erreur que la localisation elle-même. Pour surmonter ces deux problèmes, au lieu d'analyser la précision de la localisation, nous utilisons des métriques calculées à partir de l'algorithme de localisation pour évaluer la rapidité et la robustesse de la convergence de l'algorithme de localization.

Les résultats des essais de localisation du lidar dans différentes conditions météorologiques et de pollution indiquent que, du point de vue de la localisation pure, les effets des mauvaises conditions météorologiques sur le lidar ne sont pas dramatiques. Dans l'ensemble, nous ne constatons pas de grandes variations dans les résultats et pas de dégradation significative de la performance de lo- calisation sous la pluie, le brouillard ou la neige par rapport au soleil. Même lorsqu'un quart complet du champ de vision du lidar était bloqué par de la neige adhérente, l'algorithme de localisation a toujours pu converger. Les données statistiques obtenues confirment cette conclusion où, sauf dans le cas de chutes de neige, nous ne constatons pas d'augmentation significative du nombre d'itérations et du score lorsque les conditions se dégradent. La seule augmentation

clairement visible du score est dans le cas des chutes de neige, mais même dans ce cas, le nombre d'itérations n'a pas changé jusqu'à la convergence.

Ainsi, nos résultats montrent que l'algorithme et le capteur ne peuvent pas toujours être considérés séparément. En fait, l'algorithme de localisation a eu des performances similaires dans tous les cas testés, même si les performances des capteurs purs différaient de manière significative, comme le montrent les visualisations des nuages de points.

Jusqu'à présent, nous n'avons étudié que la localisation dans un environnement urbain structuré où il y a de nombreux objets structurés dans le FoV. D'autres dimensions pourraient être ajoutées et comparées, comme la conduite dans des endroits où la carte 3D montre moins d'éléments, par exemple les autoroutes, les routes rurales, etc. Cependant, les performances de localisation GNSS sont moins problématiques dans ces cas, car la ligne de visée vers les satellites est moins perturbée. Mais dans l'ensemble, ces considérations supplémentaires seront pertinentes pour une localisation fiable des véhicules AD dans tous les environnements de conduite.

Conclusion

Dans ce projet, nous avons montré comment les capteurs AD peuvent être testés en utilisant des méthodes basées sur des scénarios de fonctions. Les tests basés sur des scénarios fonctionnels offrent les avantages des tests AD dans le monde réel, c'est-à-dire que les capteurs sont testés pendant la conduite réelle et non par des modèles mathématiques des capteurs, comme dans les simulations. En outre, les capteurs sont mesurés dans le monde réel au lieu d'être truqués comme dans les systèmes d'essai "vehicle-in-the-loop" ou "hardware-in-the-loop". Comme les scénarios sont bien définis, les tests sont reproductibles et les spécifications pertinentes des tests, telles que les conditions météorologiques ou la distance des objets, peuvent être extraites. En outre, les fonctions de perception AD (par exemple, la détection ou le suivi d'objets) sont testées séparément, de sorte que les points faibles et les causes d'erreurs peuvent être mieux identifiés. La combinaison avec des expériences en laboratoire permet de vérifier les résultats des essais sur le terrain dans des environnements de laboratoire et vice versa. En plus de la procédure d'essai, nous avons également présenté certains résultats spécifiques aux capteurs pour la caméra Mobileye, le radar Delphi ESR et le lidar Velodyne VLP-16.

De tous les capteurs testés, la caméra Mobileye est clairement la plus affectée par les conditions environnementales. En fait, nos résultats montrent clairement que les fonctions de détection et de suivi des objets de la caméra Mobileye sont négativement affectées par des conditions météorologiques défavorables telles que la pluie, le brouillard ou la neige. Les résultats du scénario du piéton marchant montrent que la caméra a un taux de suivi proche de 100 % dans de bonnes conditions de luminosité, tandis que dans des conditions de faible luminosité, le taux de suivi atteint zéro. De plus, des conditions météorologiques défavorables entraînent une moins bonne stabilité du suivi, c'est-à-dire que la piste est perdue plus souvent, le nombre de pistes générées augmente et la durée moyenne des pistes diminue. Des résultats similaires ont été obtenus à partir des analyses des essais dans le scénario mixte urbain en ce qui concerne les performances de détection des objets, qui ont montré que le FNR de la caméra Mobileye pour la détection des véhicules et des piétons augmente de manière significative dans des conditions météorologiques neigeuses et pluvieuses. Dans l'ensemble, nos résultats montrent que la diminution de l'intensité de l'éclairage et le bruit généré dans les images de la caméra par les gouttes de pluie, les flocons de neige et le brouillard affectent négativement les fonctionnalités de détection et de suivi des objets de la caméra Mo-bileye.

Les tests dans les scénarios Walking Pedestrian et Urban Mixed ont également été utilisés pour analyser la fonctionnalité d'estimation de la distance de la caméra Mobileye. Dans les deux scénarios, la caméra Mobileye a produit de très mauvaises estimations de la distance dans toutes les conditions environnementales. Dans l'ensemble, la caméra Mobileye a tendance à sous-estimer la distance des objets, l'erreur absolue augmentant linéairement avec la distance. Ce n'est que pour l'enfant piéton que la caméra Mobileye surestime la distance à grande distance, probablement parce qu'elle voit un piéton adulte plus éloigné qu'un enfant.

Dans le scénario urbain mixte, nous avons également étudié l'effet des différentes tailles et types d'objets. Nos résultats montrent que la caméra Mobileye détecte mieux les véhicules que les piétons, et que parmi les piétons, les adultes sont plus facilement détectés que les enfants. Ces résultats sont très probablement liés aux différences de taille des objets, car les objets plus grands sont représentés par un plus grand nombre de pixels dans les images des caméras, ce qui facilite leur détection. Cependant, les différences entre les modèles Mobileye pour différents types d'objets peuvent également jouer un rôle (par exemple, les véhicules peuvent avoir une forme plus caractéristique que les piétons).

Enfin, dans le scénario "urbain mixte", nous avons également considéré les changements dans le FNR de la caméra Mobileye pour la détection des piétons et des véhicules en fonction de la distance des objets. Alors que le véhicule est correctement détecté à toutes les distances testées (c'est-à-dire que le FNR est proche de zéro), la détection des piétons se détériore à des distances plus longues, car le nombre de pixels de l'objet sur l'image diminue. De plus, le FNR pour la détection des piétons augmente à des distances inférieures à 10 m, ce que nous attribuons au flou de mouvement des images de la caméra dû à la vitesse tangentielle relative élevée entre le véhicule AD en mouvement et l'objet statique.

Les scénarios Walking Pedestrian et Mixed Urban ont également été utilisés pour analyser les performances de détection et de suivi des objets du radar ISR Delphi. Les tests dans le scénario "Walking Pedestrian" ont montré que le système radar a une stabilité de trajectoire et un taux de suivi inférieurs à ceux de la caméra Mobileye. En outre, nous avons observé que les performances du radar ESR sont fortement affectées par l'environnement routier. En effet, en présence de certaines géométries de façades, de conteneurs métalliques et de clôtures, nous avons observé un nombre élevé de pertes de traces et de lacunes dans les traces du piéton qui marche. En revanche, en présence d'une piste, le radar ESR fournit une mesure de distance très précise, indépendante des conditions météorologiques et de la distance de l'objet. Cependant, en raison de la grande influence de l'environnement, le radar n'est pas suffisamment fiable pour la détection des piétons.

Lors de nos tests dans le scénario "urbain mixte", nous n'avons analysé que les performances de détection du radar pour les véhicules, car les mannequins ont une section transversale radar différente de celle du corps humain et n'auraient donc pas donné de résultats significatifs. Comme dans le scénario "Walking Pedestrian", le radar a fourni une mesure très précise de la distance, quelles que soient les conditions météorologiques et la distance. Dans le cas du véhicule, la détection et le suivi du radar ESR ont également été très bons, ce qui s'est traduit par des taux de faux négatifs très faibles pour toutes les distances et conditions météorologiques (FNR toujours inférieur à 10 % et très proche de 0 % pour les distances entre les objets comprises entre 5 et 70 m). Ce n'est que dans le cas du capteur encrassé que nous avons observé une baisse des performances de détection radar entre 40 et 50 m de distance. La raison de cet effet n'est pas encore claire pour nous. Il est probable que la saleté, les feuilles et l'abrasion des pneus sur la surface du radar ont une certaine influence sur la réflectivité du radar, mais des tests supplémentaires seront nécessaires pour le confirmer.

L'un des inconvénients majeurs du radar ESR révélé par nos tests est sa grande sensibilité, qui conduit à des détections fantômes (faux positifs) et à des réflexions, comme cela a été constaté tant dans nos tests sur le terrain que dans les mesures en laboratoire à METAS. Pour être utilisées dans un véhicule AD, ces pistes fantômes doivent ensuite être soigneusement traitées et filtrées, ce qui accroît la complexité et augmente potentiellement le taux d'erreur.

En général, nos résultats montrent que lorsqu'ils sont considérés séparément, ni la caméra Mobileye ni le radar Delphi ESR ne présentent des performances acceptables en matière de détection des piétons, car chaque capteur présente des faiblesses significatives et reproductibles dans certaines conditions environnementales ou de test. Les performances de Mobileye se dégradent dans des conditions météorologiques défavorables, tandis que le radar présente des problèmes lorsque des bâtiments et d'autres objets se trouvent le long de la route. Par conséquent, pour être sûr, un système de DA bénéficierait

grandement de capteurs multiples et de la fusion de capteurs pour détecter et suivre les piétons de manière fiable.

Nous avons également étudié les performances de la fonctionnalité de détection et de localisation du capteur lidar VLP-16-A de Velodyne. Nous avons utilisé notre scénario de cible Lidar pour évaluer la fonctionnalité de détection. Les résultats de ces tests en conditions réelles nous ont montré le nombre de points réfléctés par les objets testés dans différentes conditions environnementales. Il est intéressant de noter que nos tests ont révélé de gros problèmes de performance du lidar sur les surfaces de couleur sombre, comme le noir et le bleu foncé. En particulier, ces surfaces plus sombres ne sont pas détectées par le lidar si l'angle d'incidence entre le faisceau laser et la surface n'est pas perpendiculaire. Les résultats de nos essais pratiques ont également été confirmés par les résultats des mesures effectuées à METAS. Le fait que les objets sombres ne réflèctent pas l'énergie lorsque l'angle d'incidence n'est pas complètement perpendiculaire pose un problème pour l'utilisation du lidar pour la détection des objets. Pour les autres couleurs testées (gris et blanc), nous avons trouvé des résultats plus raisonnables dans l'ensemble, où le nombre de points sur les deux cibles était toujours très proche du nombre théorique pour toutes les distances et conditions météorologiques. Pour les deux couleurs, la réflectivité mesurée n'est que légèrement affectée par les mauvaises conditions météorologiques (du moins pas autant que pour la caméra Mobileye). Dans des conditions ensoleillées, la réflectivité reste assez constante sur toutes les distances, avec une diminution aux distances inférieures à 10 m. La diminution en dessous de 10 m pourrait être liée à la saturation de la détection et à un domaine numérique limitée. Nous constatons également un nombre accru de valeurs aberrantes dans les valeurs de réflectivité dans de mauvaises conditions et une variation légèrement plus importante dans l'ensemble.

Les résultats du scénario de localisation lidar montrent que l'algorithme de localisation est assez robuste au bruit du capteur et à la dégradation due à des conditions environnementales défavorables. Bien que certaines conditions météorologiques défavorables dégradent considérablement les nuages de points lidar, nous n'avons pas observé de dégradation significative des performances de localisation pour aucune des conditions testées, y compris le soleil, la pluie, le brouillard, la neige et la pollution. En fait, l'algorithme de localisation a toujours convergé avec succès, même lorsqu'un quart complet du champ de vision du lidar était bloqué par la neige adhérente. Globalement, nos résultats montrent que pour la localisation pure, les performances du lidar lui-même ne sont pas le facteur limitant.

Dans ce projet, nous avons étudié les performances de différents capteurs AD dans des conditions de conduite réelles. Les questions qui se posent maintenant sont les suivantes : Comment pouvons-nous utiliser ces résultats pour la validation AD? Comment la performance finale de l'AD est-elle liée à la performance des capteurs? Et quels sont les effets de la fusion des capteurs et des algorithmes de contrôle? Les spécifications telles que les courbes FNR obtenues lors de nos tests peuvent être utilisées pour créer des modèles probabilistes de capteurs qui peuvent ensuite être utilisés dans des simulations. Ce faisant, le modèle de capteur aurait la même incertitude que le matériel et le logiciel réels, ce qui rendrait la simulation plus proche de la réalité. Une fois que des spécifications précises et réalistes ont été établies lors d'essais en conditions réelles, la simulation peut être utilisée pour valider de nombreux autres scénarios, ce qui accélère le processus de validation. En outre, l'analyse des performances des capteurs individuels peut aider à élaborer de meilleurs modèles de fusion de capteurs et d'estimation d'état pour l'AD, car elle fournit des informations importantes sur la précision et l'incertitude des capteurs. Même sans tester directement l'ensemble du bloc AD, certaines conclusions générales peuvent être tirées directement des résultats des tests des capteurs, par exemple, notre véhicule d'essai ne serait pas en mesure de conduire de manière sûre et efficace la nuit pendant les chutes de neige si seuls les capteurs actuellement installés étaient utilisés.

Perspectives

L'infrastructure et les connaissances acquises dans le cadre de ce projet ouvrent de nombreuses possibilités et de nouveaux sujets de recherche dans le domaine du test de l'AD. Afin d'utiliser au mieux les ressources acquises, nous souhaitons poursuivre nos

activités de test de capteurs basées sur des scénarios de caractéristiques avec des capteurs, des algorithmes de perception et des scénarios supplémentaires.

Tout d'abord, nous voulons répéter ou étendre certains tests afin d'étudier plus en profondeur certains des problèmes de performance des capteurs découverts dans ce projet. Par exemple, le test du radar contaminé sera répété pour comprendre l'augmentation observée du FNR pour la détection des véhicules. En outre, les seconds minima locaux des courbes FNR pour la détection des piétons avec la caméra Mobileye dans des conditions météorologiques brumeuses et neigeuses seront étudiés plus en détail. Nous prévoyons également de tester les capteurs dans un brouillard plus dense pour voir si/quand les performances de la caméra Mobileye diminuent. De plus, nous voulons répéter les expériences de détection de véhicules avec un véhicule plus petit pour voir les différences avec l'Iveco Daily. Le comportement étrange du lidar à courte distance et sur des surfaces sombres sera également analysé plus en détail pour voir s'il est lié à une erreur du capteur ou si c'est le cas pour tous les lidars. Pour les performances de localisation, nous voulons déterminer le degré de dégradation du capteur (sous forme de bruit ambiant) nécessaire pour que l'algorithme de localisation ne converge pas.

En outre, nous voulons tester de nouvelles caméras et appliquer différents algorithmes de perception à code source libre. Cela nous permettrait d'analyser à la fois les images brutes de la caméra et le fonctionnement des algorithmes. En plus des caméras RGB classiques, nous prévoyons de tester des caméras événementielles et infrarouges. Nous voulons également utiliser des algorithmes de détection d'objets 3D à code source libre pour évaluer les performances du lidar Velodyne VLP-16 pour la détection des piétons et des véhicules. Dans ce contexte, il serait également intéressant de tester un lidar avec plus de 16 canaux.

En ce qui concerne les conditions environnementales, nous voulons développer des méthodes pour quantifier les conditions météorologiques et de encrassement avec des chiffres, par exemple pour analyser les changements dans la performance de détection sous différentes conditions de pluie ou de encrassement. Outre les effets de la météo et de la encrassement, nous voulons également étudier les effets du vieillissement et de l'usure des capteurs AD, ainsi que les interférences entre des capteurs similaires montés sur deux véhicules AD différents.

Les résultats des tests basés sur des scénarios de caractéristiques ont montré que de nombreux paramètres peuvent affecter les performances de détection: vitesse du véhicule, distance de l'objet, taille de l'objet, conditions météorologiques. Par conséquent, nous aimerions répéter nos tests à partir du scénario urbain mixte avec différentes vitesses de véhicules et éventuellement avec des objets dynamiques. Ces tests supplémentaires nous aideraient à déterminer tous les facteurs pertinents affectant les performances des capteurs et à effectuer une analyse de sensibilité de leur pertinence pour les performances perceptives de l'AD. Les analyses détaillées des performances de chaque capteur peuvent ensuite être utilisées pour analyser les méthodes de fusion des capteurs et pour déterminer les performances de différentes combinaisons de capteurs dans différentes conditions. Enfin, en utilisant les modèles de capteurs dérivés de nos tests en conditions réelles dans les simulations, nous pouvons lier les performances de détection aux performances de l'AD.

Summary

This document reports the activities and results from the project "Automated Driving – Sensor Testing Vehicle". The project was launched in order to understand and build automated driving (AD) safety validation procedures. Sensors play a crucial role in the AD stack, as they provide the relevant information from the environment. The project focused on evaluating AD sensors with field tests during real-world driving on a test track and comparing the results with laboratory tests performed at METAS. For this purpose, a Lexus RX 450h vehicle was equipped with a commercial set of the main AD sensors, i.e. lidar, camera and radar. Further, a test track with test objects was built in the premises of the Empa campus in Dübendorf. For the evaluation of the sensors' performance we used the so called function-scenario-based testing approach. In this procedure, a specific perception function in AD, e.g. object detection, is tested in a designed real-world driving scenarios. This testing approach allows repeatability of field tests as well as precise understanding of the scenario. With different scenarios, we tested the object detection, distance estimation, tracking and localization performance of the Mobileye camera, Delphi ESR radar and Velodyne VLP-16 lidar. Since some of the sensors are black box systems, e.g. Mobileye camera, we developed new evaluation specifications such as the tracking ratio, which indicates how sensitive and robust the object tracking functionality is over time. The design of our scenarios allowed us to obtain precise information about the sensor performance, such as the false negative rate of object detection at different object distances. The field tests were executed under different weather conditions and with contaminated sensors. Our analyses revealed that the Mobileye camera has a high weather dependency for object detection, showing large performance drops for snowy and rainy weather conditions. Object type, size and distance also significantly influenced the Mobileye detection performance. Vehicles were overall detected easier than pedestrians, and larger (adult) pedestrians easier than smaller (child) pedestrians. Generally, the detection performance strongly deteriorated for large distances, but also for very close distances, which we attributed to motion blur effects due to the moving testing vehicle. The distance estimates of the Mobileye camera were found to be very poor under all tested conditions, with a clear tendency to underestimate the true object distance. The ESR radar showed good detection performance under all environmental conditions and provided very accurate distance measurements. However, we observed that its object tracking functionality is quite sensitive, as the sensor often loses tracks and/or tracks same objects multiple times. The Velodyne lidar showed good localization performance under all tested environmental conditions. Even a partially blocked field of view from accumulated snow on the sensor surface did not lead to a localization failure. However, our field tests and laboratory experiments revealed that darker colored surfaces are not seen by the lidar if the angle of incidence between the laser beam and the surface is anything but perpendicular, which poses a big problem for object detection

1 Introduction

1.1 Motivation

Automated driving is a heavily discussed technology on future mobility. This technology has the potential to make road traffic safer, more convenient and to make better use of vehicles. The opportunities and risks of these technologies are the subject of numerous studies.

AD-mobility is not only an engineering challenge, it stimulates and activates also other areas such as jurisprudence, ethics and socio-economics. Challenges are for instance who is liable when an AD accident happens or how will the AD technology change the mobility of individuals. Furthermore, there are also ecological questions such as how the AD technology will influence environmental pollution. Many of these questions and challenges are interdisciplinary. For example, the regularization of automated vehicles (AVs) involves engineering, jurisprudence and maybe other not so obvious disciplines such as ethics. From the jurisprudence perspective, it is hard to define new traffic laws if they do not understand how safe AD systems are. On the other hand, from the engineering perspective, it is hard to estimate what is safe enough for the society and how to prove safety. Therefore, a cooperation between these two disciplines is required in order to accomplish this particular challenge of AV regularization.

In order to understand and quantify the AD safety, testing and validation procedures are required. Thereby, the whole AD vehicle can be tested at once as a black box system or the different components can be tested separately. The separation of the components can provide more insides about the failure origins and safety requirements.

One very important component of every AV is the sensor set. It provides information from the environment such that driving control inputs can be calculated. The "AD Sensor Testing Vehicle" project addresses the development of AD sensor testing procedures and the evaluation of commercially available AD sensors.

AD systems require field tests for further development. The overall objective of this project is to provide first insights into the real-world testing of AVs and lay out the fundamentals for the definition of minimum requirements for AD sensor technology and data recording for field tests/pilot trials.

1.2 Contribution

In this report we present the results from the function-scenario-based AD sensor testing procedures developed within the "AD Sensor Testing Vehicle" project. Thereby, AD perception tasks such as object detection, distance estimation, classification, tracking and localization are tested and evaluated with field test scenarios. For this purpose, a Lexus RX450h was equipped with commercially available range and vision sensors as well as with AD drive-by-wire systems and computing units. A test track with human-like test objects was designed to realize real-world field tests at Empa in Dübendorf. Sensor performance results from the function-scenario-based AD sensor testing are presented for the Mobileye 630 camera, Velodyne VLP-16-A puck lidar and Delphi ESR 2.5 radar. The sensors' performance is evaluated for different weather conditions such as sun, fog, snow and rain as well as with contaminated sensors. Besides these field tests, the aforementioned sensors were also tested at the METAS laboratory, which allows a comparison between real-world field testing and laboratory testing.

1.3 Organization

The rest of this report is structured as follows. In Chapter 2, we give an overview of the different testing approaches used for the evaluation and verification of AD-Systems. Each testing approach is described along with its advantages and disadvantages. In chapter 3, we introduce the role of sensing and perception in AD. Thereby, we describe the main sensors used in this field along with those tested in this project, and we define the main functionalities of each sensor. In chapter 4, we describe our testing methodology, including the test scenarios, the choice of the performance metrics and the data evaluation pipeline. In chapter 5, we describe all the preparatory work we performed prior to the actual tests, including the calibration of the sensors and the set-up of the test track and test objects, and we define the different environmental conditions (weather and contamination) that were evaluated in this project. In chapter 6, we report the main results of our field tests. Each section presents the results of a single testing scenario. First the scenario and tested functionality are described, followed by the main data processing steps, the description of the chosen performance metrics and the test results. Each section is then completed with a discussion of the main results. In the last chapter we summarize all our results, findings and conclusions from the function-scenario-based testing.

2 Testing Approaches for AD-Systems

In this section, we describe the different approaches for validating the safety of an Automated Vehicle (AV). Figure 1 shows an overview of these different approaches.

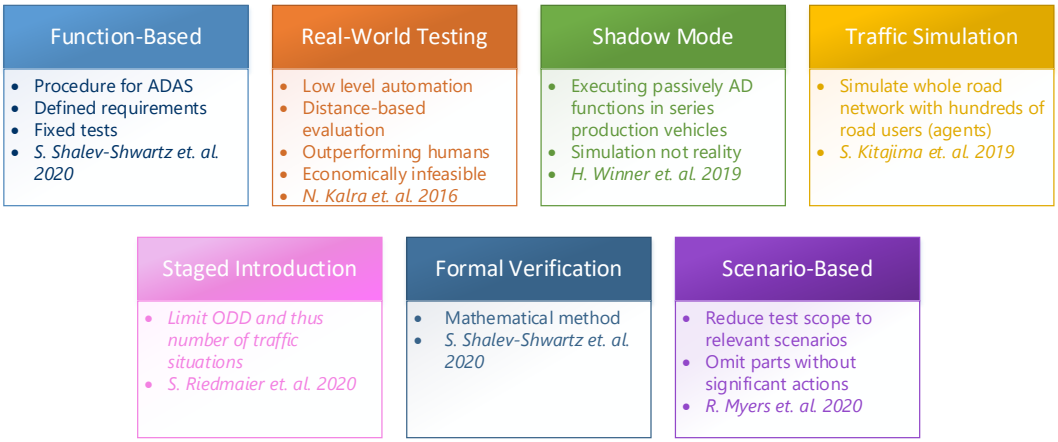

Figure 1: *Overview of the different testing approaches.*

Each testing approach has its advantages and disadvantages and a complete safety validation of an AV vehicle will probably require a mix of different testing approach.

2.1 Function-Based

This approach focuses on the function and its use case. It is often used to test Advanced Driver Assistance Systems (ADAS). The functions (ADAS) are defined and designed based on a set of requirements and tested against those [1]. Function-based testing is an approach currently used in couple of standards for ADAS.

While this approach works well for systems with a limited functionality, defining requirements for AVs that cover every possible functionality is impossible [1].

2.2 Real-World Testing

This approach relies on testing an automated vehicle on public roads, hence the name real-world testing. Often, the safety of the AV is evaluated based on the idea of outperforming the human driver with respect to safety. According to the US Bureau of Transportation Statistic, around 3 trillion miles are driven in America every year, with around 2.3 million reported injuries [2]. This corresponds to a failure rate of around 77 injuries and 1.09 fatalities per 100 million miles [2] or around 10^{-6} fatalities per hour [3]. Therefore, to show that an AV performs better than a human driver requires driving hundreds of millions of kilometers or more [2]. For example, Alphabet's Waymo has announced in January 2020, that they had driven over 20 million miles on public roads [4]. Even though these numbers look impressive, many doubt that it is enough [2], [3].

Usually an automated vehicle should perform considerably better than a human operated one, which would lead to even more test kilometers to show the required safety [3].

Additionally, it is very difficult to assess whether the driven test kilometer really cover all situations, including edge cases and rare events. For example, driving 100 km on a lonely highway on a sunny day is by no means the same as driving 100 km in a busy city during rush hour.

Another problem that remains is how to cope with software updates. One changed line of code could render the whole test useless with respect to statistical safety arguments. At latest then, proving safety trough real-world testing alone becomes economically infeasible.

2.3 Shadow Mode

In the so-called shadow mode, the AD functions run passively in the background without actively controlling the car [5]. The shadow mode works well for testing perception and sensing, since they are open loop by nature. The main drawback of this approach lies in the fact that decision making in automated driving affects the decisions of the other traffic participants and vice versa. Therefore, without executing the decisions, assessing the safety is difficult [1]. Shadow mode is widely used by car manufactures like Tesla for testing new systems or updates in a large fleet.

2.4 Simulation Testing

In contrast to real-world testing, this approach aims to validate the safety of AVs trough simulation. The basic idea is to build a virtual world that can represent any kind of urban environment and everyday scenarios.

The advantage over real-world testing is that it tackles its two biggest issues: the number of driven kilometers and the necessary variation within those kilometers. With simulation, it is possible to drive the huge number of kilometers in all situations needed to demonstrate its safety with enough statistical confidence. In addition, testing scenarios involving virtual people allow to also test accidents without injuries.

The biggest issue with safety testing trough simulation is that it is equally hard to validate the simulation itself as it is to validate the driving software directly [3]. Additionally, as with every simulation, there is an inherent need to simplify the real world. However, the required level of fidelity for safety evaluation is still a large unknown.

A big advantage of this method is that once the simulation is validated, it is possible to revalidate driving software with a manageable effort and in a highly standardized fashion, since driving millions of kilometers in simulation is much faster and cheaper than in real world.

2.5 Staged Introduction

In staged introduction, the idea is to limit the ODD (Operational Design Domain) and therefore the number of traffic situations to be tested. These limitations should then allow to assess the safety of the AV in real-world testing. Once the safety is assessed in this limited ODD with an appropriate methodology, the ODD is gradually widened. Staged introduction is an approach applied by many manufacturers and can be very useful for the introduction of level 4 AVs in selected downtown areas [1].

2.6 Formal Verification

This approach gives a mathematical proof of the correctness of the algorithm/code with respect to the formal specifications. Therefore, both the source code and an abstract model of the correct behavior of the algorithm to verify are required [3]. Given a complete specification, formal verification allows the safety assessment for the whole ODD [1].

2.7 Scenario-Based Testing

In this approach, the idea is to break down AD into typical individual traffic situations. Thus, one tries to reduce the infinite amount of possible driving situations into a finite set of scenarios to be tested.

This finite amount of tests can be done in either simulation, closed track testing or on-road testing. The advantage of this approach is that we can focus on critical situations without spending test time on the easy tasks. The main challenge of this approach is that it assumes that if the AV passes a given scenario successfully, it will pass other similar scenarios as well [3]. While humans are very good at this, it is hard to prove it for an AV.

3 Sensing and Perception in Automated Driving

Figure 2 shows a diagram of a typical AD-stack. Different sensor technologies (camera, radar, lidar…) are used to capture real-time information of the surroundings of the vehicle. Through data processing, the sensors can detect and track objects surrounding the car. The data collected from different sensors is combined through data fusion and processed in real-time to create a comprehensive representation of the reality. This is used by algorithms to analyze the situation, plan possible trajectories and determine the action to be taken (steering, braking…). In this project we focus on the first two steps of this process, sensing and perception, which are described in more detail in the following two sections.

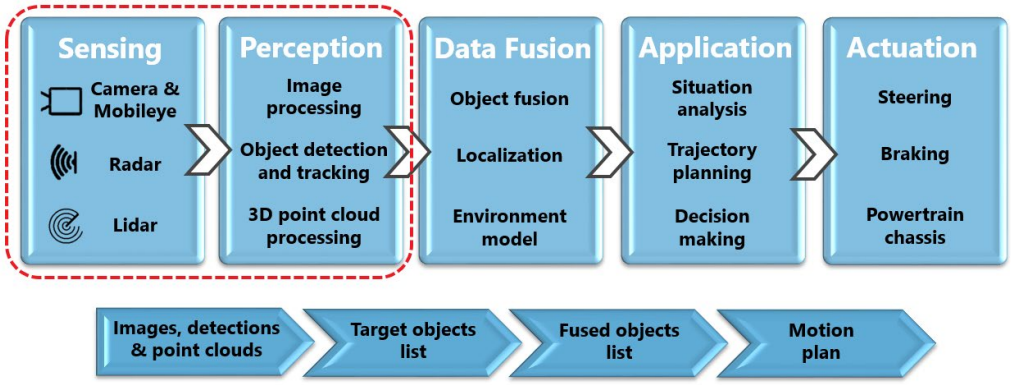

Figure 2: *AD-Stack with sensing and perception highlighted.*

3.1 Sensors

In this section, we briefly describe the main sensors used in AD and provide the specifications of the sensors installed in our AD Testing Vehicle.

3.1.1 Radar

Radar (Radio Detection and Ranging) is a remote sensing technology that uses radio waves to determine the range, angle, and/or velocity of objects. A radar system consists of a transmitter, a transmitting antenna, a receiving antenna (often the same antenna as for transmitting), a receiver and a processor.

A radar system emits radio waves into a predetermined direction. When these radio waves hit an object, some of the energy is reflected back to the transmitter (Figure 3).

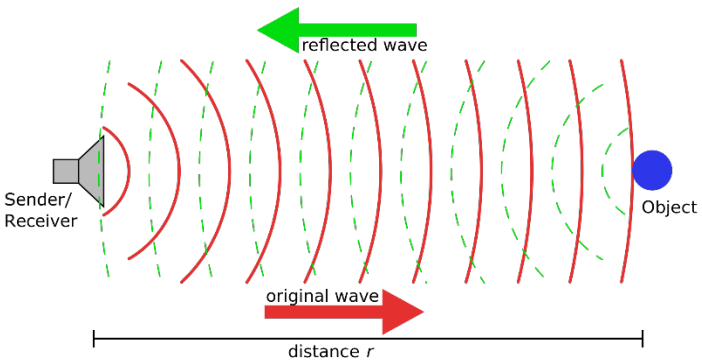

Figure 3: *Basic radar principal (Source:* **[6]***).*

The reflections are then received and processed. In order to detect an object, a certain amount of power must reach the receiver. The received power from a target at a certain distance can be described with the radar equation:

$$P_r = P_t \cdot \frac{G \cdot \sigma_{RCS} \cdot A_e}{(4\pi)^2 d^4},$$

where P_r is the received power, P_t is the transmitted power, d is the distance to the target, G is the antenna gain, σ_{RCS} is the radar cross section of the target and A_e is the effective antenna aperture.

The radar equation shows that the received power from a target declines with the 4th power of the distance to the target. Therefore, to double the detection range one needs 16 times more power.

Traditional radars use an antenna with a fixed shape and therefore with a fixed antenna pattern/radar beam. In contrast, Active Electronically Scanned Array (AESA) radars consist of multiple small solid-state transmit/receive modules (Tx/Rx modules).

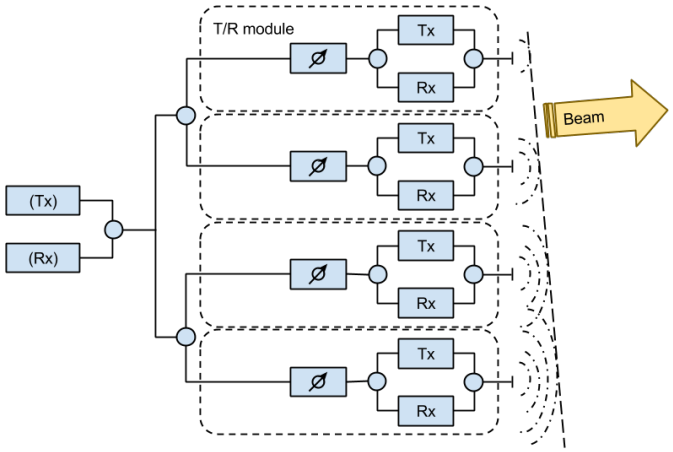

Figure 4: *Principle of an AESA radar (Source:* **[6]***).*

By actively controlling the phase and frequency of each transmitter (Tx), the radar beam can be electronically steered to point in different directions without moving the antenna (Figure 4). Instead of just one beam with a single frequency, an AESA radar can also use multiple smaller sub-beams, each potentially with a different frequency.

Similarly, adjusting the phase of each individual receiver (Rx) signal before summation, allows the steering of the reception beam. An alternative approach for the reception part is to sample each individual receiver signal with sufficiently high bandwidth and perform digital beam forming. This allows steering the reception beam in multiple directions at the same time and therefore simultaneous detection of multiple targets.

Radar sensors for automotive applications are typically divided into two categories: frequency modulated continuous-wave (FMCW) and pulse radar. They typically use frequencies of 24, 77 or 79 GHz, known as millimeter wave (MMW).

A continuous wave transmitter generates a continuous sinusoidal signal with frequency f_t, which is radiated by the antenna. Reflection from a moving object creates a sinusoidal echo, albeit with a different frequency f_r due to the Doppler effect. The frequency change depends on the relative radial velocity v_r between radar and target and the speed of light c:

$$f_r = f_t \left(\frac{1 + \frac{v_r}{c}}{1 - \frac{v_r}{c}} \right)$$

For an approaching target, the received frequency is higher than the transmitted one and vice versa. To measure the range, a radar must be able to determine the time between transmission and reception. For a continuous wave radar, this is accomplished by frequency or other modulation techniques, which allow reconstruction of the time of flight from the received signal.

On the other hand, pulse radar transmit only short pulses and wait for reflections. For a pulse radar, the range is directly calculated from the time of flight while the radial velocity of the target is calculated from the Doppler frequency shift.

Delphi ESR Radar

Our AD testing vehicle is equipped with a Delphi (electronically scanning radar) ESR 2.5 [7] radar system installed in the front bumper (Figure 5).

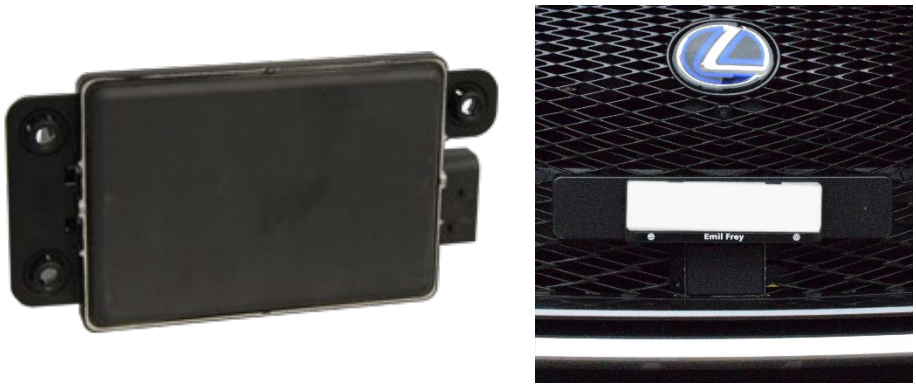

Figure 5: *Left: Delphi ESR 2.5 (Source:* **[8]***). Right: Radar fitted into the front grill below the license plate.*

The ESR 2.5 is a 76.5 GHz, solid-state single radar system with an electronically steered beam and pulse-Doppler signal processing. The ESR multimode radar provides simultaneous measurements of a wide field of view at mid-range and a narrow field of view at long-range (Figure 5).

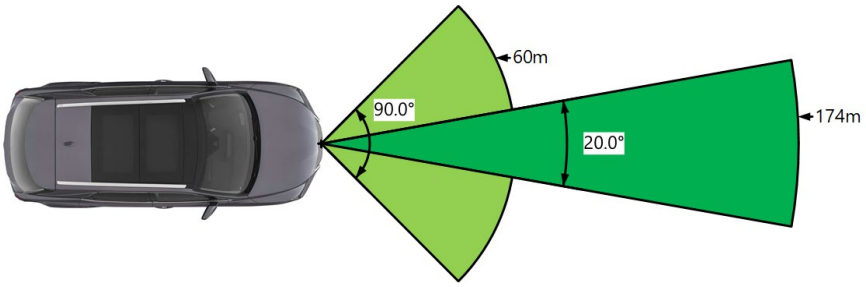

Figure 6: *Delphi ESR 2.5 field of view of.*

The long-range mode provides accurate range and range rate information about targets within the field of view. The wide range mode allows the detection of lane changing or crossing vehicles as well as pedestrian well outside the width of the vehicle. In total, the ESR radar provides information of up to 64 identified and tracked targets.

The target data from the tracking filter is provided at an update rate of 50 Hz. Target data contains range, range-rate, range-acceleration, azimuth (lateral angle), lateral relative velocity, and the track ID.

3.1.2 Lidar

Lidar (Light Detection and Ranging) is a remote sensing technology that uses laser light to measure distances. An object is illuminated with a laser pulse and the reflected light is measured.

The distance to the object is then calculated from the elapsed time between transmission and reception:

$$d = c \cdot \frac{\Delta t}{2}$$

where d is the distance to the object, Δt the elapsed time between transmission and reception and c the speed of light.

By changing the horizontal and vertical direction of the laser, a 3D-lidar can scan the surrounding space, creating a 3D point cloud. Most lidars have a much higher horizontal than vertical angular resolution, which results in the typical contour lines that can be seen in a top view visualization (see Figure 7).

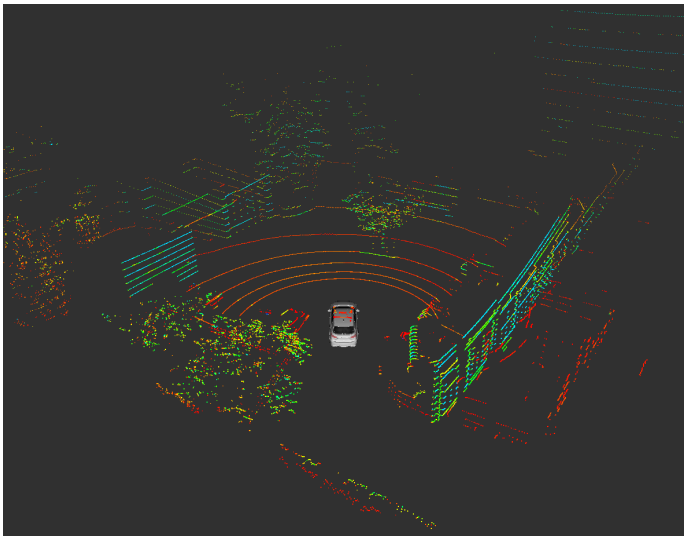

Figure 7: *Visualization of a point cloud from a 3D lidar.*

The vertical resolution of a lidar is determined by the number of channels and the vertical field of view. Each channel consists of a pair of laser emitter and detector at a fixed elevation angle relative to the horizontal sensor plane.

Today, horizontal scanning is usually accomplished by a rotating mirror, which deflects the laser beam from all channels simultaneously into the desired direction (Figure 8). The horizontal resolution is therefore determined by the rotational speed of the mirror and the number of measurements per second. Due to the complex mechanical setup, spinning lidars are quite expensive and have a relative low lifetime.

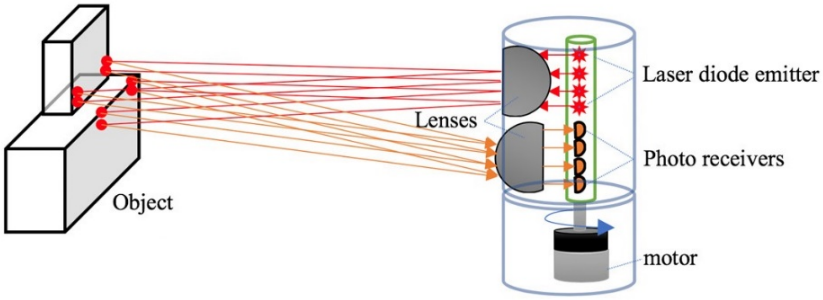

Figure 8: *Spinning 3D-lidar (Source:* **[9]***).*

To get rid of the moving mechanical parts, many companies are working today on so-called solid-state lidars. Similar to phased array radars, solid-state lidars contain many optical emitters, which can be controlled individually. By carefully adjusting the emission pattern, the resulting beam can be focused and steered into the desired direction without the need of any moving mechanical parts.

Another concept is deployed in the Micro-Electro Mechanical System (MEMS) lidars (Figure 9), where an array of micro mirrors is used to steer the laser beam into the desired direction.

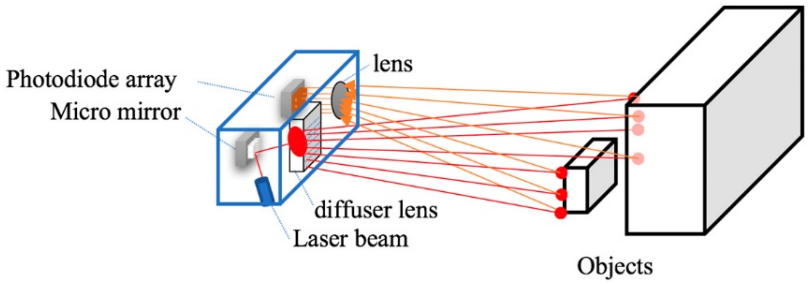

Figure 9: *MEMS lidar (Source:* **[9]***).*

Today's state-of-the-art lidar systems use laser light with a wavelength of either 905 nanometers (nm) or 1550 nm. Both wavelengths have their own advantages and disadvantages, such as different water absorption. Therefore, tradeoffs must be made with respect to performances, safety and cost.

Velodyne (VLP-16-A) LiDAR Puck

Mounted on the front bar of the roof rack of our AD Testing Vehicle is a Velodyne VLP-16-A LiDAR Puck [10], a 16 channel spinning mirror lidar (Figure 10). It uses a 905 nm wavelength, class 1 laser.

The Velodyne LiDAR Puck has a range of 100 m with an accuracy of 3cm. It provides full 360° horizontal and ± 15° vertical field of view (from the horizontal plane). The 16 channels provide a vertical resolution of 2°. The horizontal resolution depends on the rotation rate and can be chosen between 0.1° at 300 RPM (as used in our setup) and 0.4° at 1200 RPM.

Figure 10: *Left: Velodyne (VLP-16-A) LiDAR Puck (Source:* **[8]***). Right: Velodyne installed on the roof.*

3.1.3 Camera

A digital camera is an optical device that captures visible or infrared light to produce images. It consists of lenses that focus light from a scene and an imaging sensor (Figure 11) to capture that light. The imaging sensor consists of a 2D array of light sensitive elements, which capture photons and produce an electric signal that is proportional to the number of captured photons.

The two main types of imaging sensors are charge-coupled device (CCD) and active-pixel sensor/ complementary metal oxide semiconductor (CMOS). For a long time, CCD sensors were superior over CMOS with respect to image quality, but this is no longer true due to massive improvements and innovations in CMOS technology [11].

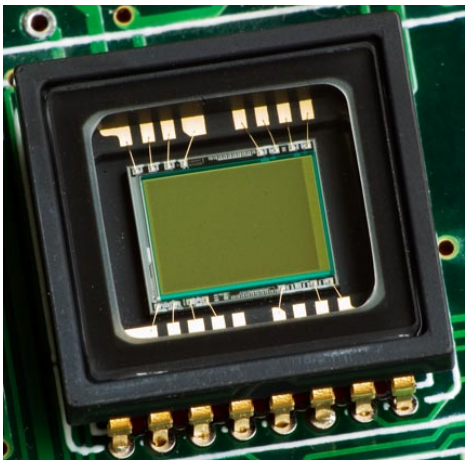

Figure 11: *CCD imaging sensor from a webcam.*

Standard imaging sensors capture light in the same visible light range as the human eye. The visible light range is further divided into three color bands (red, green and blue) and each band is detected separately. Besides RGB imaging sensors, there are also imaging sensors that detect light in the near infrared range.

For use in automated vehicles, high-resolution cameras are combined with computer vision algorithms to extract information from the generated images (Figure 12).

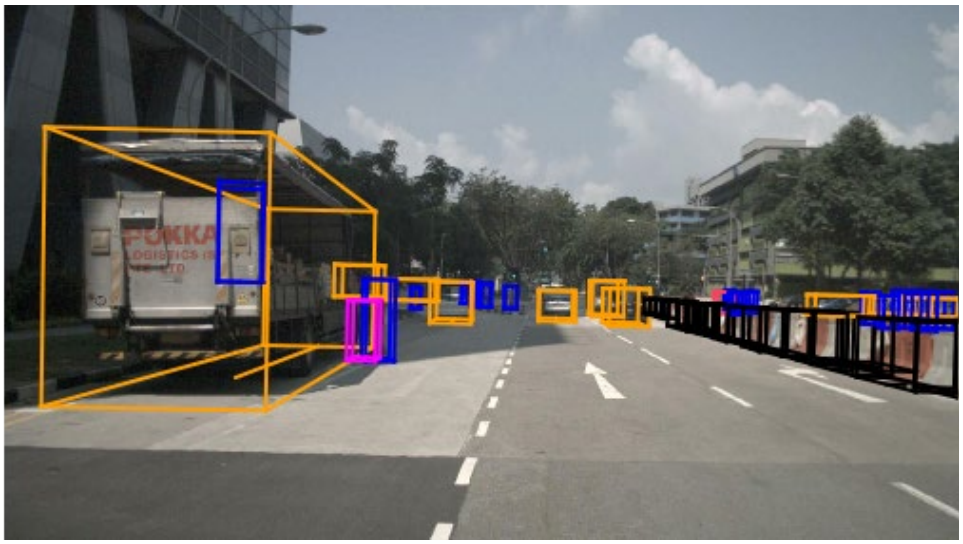

Figure 12: *RGB-Image with detections (Source:* **[12]***).*

Mobileye 630

Our AD Testing Vehicle is equipped with a Mobileye 630 camera, mounted behind the front windshield underneath the original Lexus camera (Figure 13).

Figure 13: *Left: Mobileye 630 camera (Source:* **[8]***). Right: Mobileye Camera Module installed on Lexus.*

The camera contains the EyeQ2 Image Processing Chip with Mobileye's image processing algorithms [13]. The algorithms try to detect lanes, traffic signs, vehicles, cyclists and pedestrians and measure the distance to those objects. The EyeQ2 processor has an update rate of 15 frames per second.

3.2 Perception Functionalities

In this section, we briefly describe the main functionalities associated with perception in AD.

3.2.1 Detection

The detection of objects is a crucial functionality for AVs. For AD, objects can be detected from camera images, radar data and from lidar point clouds. Typical objects that have to be detected include not only vehicles and pedestrians, but also other road obstacles such

as trees, traffic lanes, speed signs, traffic lights and many more (see Figure 12 for an example of detected objects in a camera image).

From camera images, an object is usually detected through either deep learning methods [14], [15], [16] or classical computer vision methods based on edge detections [17]. Real time object detection, especially from camera images, is a very active research topic both in academic and industrial research and a lot of progress was made in the last decade with improved algorithms and increased computational power.

Classical and machine learning methods can be applied as well for object detection from point clouds [18], [19], [20], [21].

3.2.2 Classification

Once an object is detected, it needs to be classified. Classifying means assigning a class label to each detected object. Figure 12 shows an example of an image with drawn bounding boxes around the detected objects and the classification labels are indicated with different colors for each object class.

3.2.3 Distance Estimation

Distance estimation is the process of determining distances between objects with data from passive sensors, usually from camera images where the distances are not measured directly as with radar or lidar. There are several options to perform distance estimation with a monocular camera:

- Geometry-based
  - Structure from motion (estimate 3D structure from a series of 2D images)
  - Infer distance from geometry of single image
- Deep-learning based
- Sensor-based: RGB-D sensor detects pixel level depth directly

This paper [22] by Mobileye describes their efforts in geometry-based range estimation. They mention the two approaches for geometry-based range estimation. In the first approach, the vehicle width in the image space is detected and the distance to the vehicle is computed by using pinhole-camera geometrics. Since the real vehicle width varies a lot, this can induce large error in the distance estimation. In the second approach, the bottom-line of the vehicle (point of contact width the road) is detected and the distance to the camera-horizon allows inferring the distance to the vehicle by using again pinhole-camera geometrics. This approach leads to errors at varying road inclination. Both approaches can be translated to pedestrians and bicycles.

3.2.4 Tracking

Tracking is the process of following moving objects. In many applications, for example trajectory planning, it is often necessary to provide the position, the velocity and the instance of every object at a fixed update rate. The object tracks are obtained with a filter which uses object detection and localization data. There are many challenges when designing a filter, such as noisy sensor data, missing detections, false detections and unsynchronized sensor data. Additionally, not every sensor can provide all the required data for this application. For example, the object's velocity cannot be determined from a single camera image and therefore a tracking filter has to estimate this data in this case. Tracking uses a dynamic filter to smooth out the raw detection data, estimate the missing data and extrapolate/interpolate between detections.

Within any multi-object tracking filter, i.e. a filter which can track multiple objects simultaneously, there is a crucial processing step. All the raw detections need to be correlated to the existing tracks to find the best matches, i.e. to figure out which detection belongs to which track. Once the detections are associated with existing tracks, the tracks are updated with the new data from the detections. Of course, if no suitable track can be found for a detection, a new track is initiated.

Common filter designs for object tracking include Kalman filters and their variations and particle filters among many others. The choice of the filter depends on the requirements and the available computational power.

3.2.5 Localization

As human drivers in conventional cars, AD vehicles need to know their absolute position in order to navigate to the target location. For this navigation purpose, a GNSS position with an accuracy in the order of a couple of meters is sufficient.

Moreover, human drivers rely on vision-based local localization for path following, i.e. detecting traffic lanes and street boundaries and the relative position with respect to them.

Likewise, many AD vehicles also rely on very accurate absolute localization in the centimeter range and use similar accurate detailed street maps.

For such accurate localization, lidar localization is a very promising approach, especially for urban environments. It provides an accuracy in the required cm range, similar to a differential or RTK GNSS. Additionally, lidar localization does not suffer from signal loss or multipath problems like GNSS and can therefore act as a backup and provide redundancy when GNSS positioning methods are lost or fail.

Lidar localization works by comparing the current point cloud to a usually predefined 3D map and thus finding the position in the map where the current point cloud fits best to the map [23]. There are several different algorithms available, e.g. iterative closest point (ICP) and normal distribution transform matching (NDT-matching) [24]. Recently also research in the field of deep learning based lidar localization methods is being performed [23].

For our tests, we used the NDT-matching algorithm implemented in autoware.ai [1].

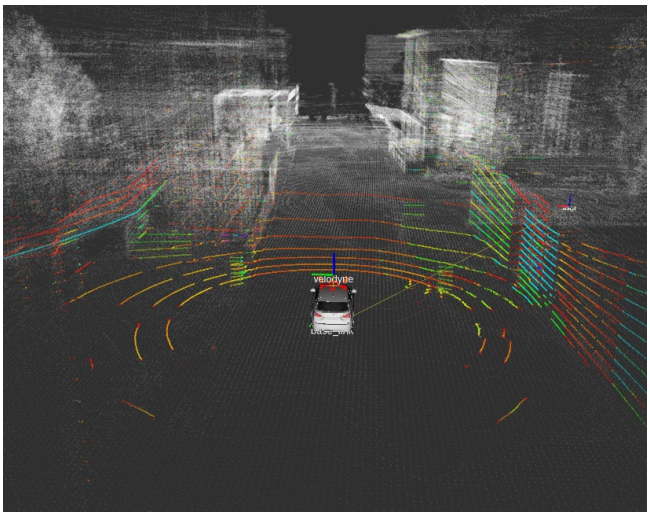

Figure 14: *Example of lidar localization with autoware.ai NDT-matching algorithm on the Empa campus in Dübendorf with the lidar point cloud in colors and the reference map in grey.*

[1] https://www.autoware.ai/

4 Methodology

In this project, we focus on the approach of function-scenario-based testing. For each test, we choose a function and define a scenario, performance metrics and the evaluation/data pipeline.

4.1 Function-Scenario-Based Testing

First, we have to define what functionality of the AV we want to test and in which scenario the functionality will be tested.

For example, if we want to test the reliability of the camera-based pedestrian detection (function), we could define a testing scenario in which the car drives along the test track while a pedestrian crosses the track in front of the vehicle.

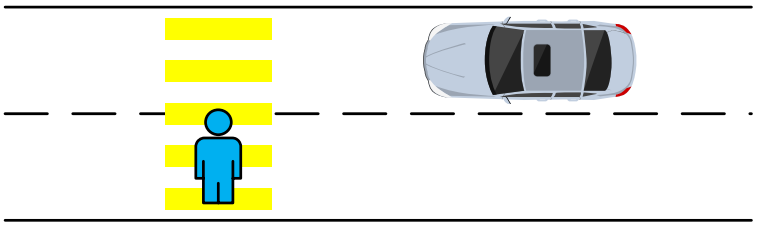

Figure 15: *Pedestrian crossing scenario.*

Figure 16 shows an example of a pedestrian on the street from the perspective of the camera with overlaid Mobileye and radar detection as well as the lidar points.

Figure 16: *Pedestrian crossing the street. Red box: Mobileye detection. Green cuboid: Radar detection. Colored horizontal points: lidar points with corresponding intensity.*

4.2 Performance Metrics

For each test, we then define performance metrics, which describe how we evaluate the performance of each sensor for a given functionality in a given scenario.

Detection and Classification Evaluation

For our pedestrian detection scenario, an example of such a performance metric would be the false negative rate, i.e. how many pedestrians the sensor (and algorithm) missed to detect in areas in which there was a pedestrian.

$$FNR = \frac{FN}{FN + TP}$$

The objet detection performance is evaluated with annotated ground truth data. Ground truth is usually the human eye. The data has to be annotated manually, which can be done with different open source programs or in our case with a self-made Matlab code. The detection usually consists of a localization marked by a bounding box and a classification, where a word describes the object. Both are evaluated by the mean average precision (mAP), where the true positive (TP) is trivial for classification and evaluated for bounding boxes by the "Intersection over Union" (IoU):

$Precision = \frac{TP}{TP+FP}$,

$Localization: TP\ if\ IoU > 0.5,\ where\ IoU = \frac{|A \cap B|}{|A \cup B|}$.

The IoU value ranges from 0.0 to 1.0, where 1.0 is a perfect detection. The threshold when a detection is correct can be set up individually, but in general a threshold of 0.5 is used.

Stability of Detection / Tracking

Another way to test the sensors is to see how stable the detection or tracking is over the complete trajectory and how fast an object is detected, i.e. to look at the reaction time of a sensor (and its algorithm).

Distance Estimation

The evaluation of the distance estimation is straightforward. We measure the distance manually and compare it with the different estimations. The performance metrics are then the absolute and relative errors.

4.3 Evaluation and Data Pipeline

As a last step, we define how we evaluate the data from a test scenario given the chosen performance metrics. This includes the definition of the signals we need to gather, as well as the procedure to extract the required information out of those signals.

For our pedestrian detection example, we need to log the output of the detection algorithm but also the true appearance of the pedestrian in the field of view of the camera. Quite often, part of the data pipeline includes some manual steps, e.g. determining the ground truth in the pedestrian scenario

5 Field Testing Setup and Preparatory Work

5.1 Sensor Calibration

In order to compare data from different sensors, the sensors need to be calibrated. This means measuring each sensors' position and orientation on the vehicle with respect to a common reference. For our vehicle, this reference is named *base_link* and is located at the rear axle (Figure 17).

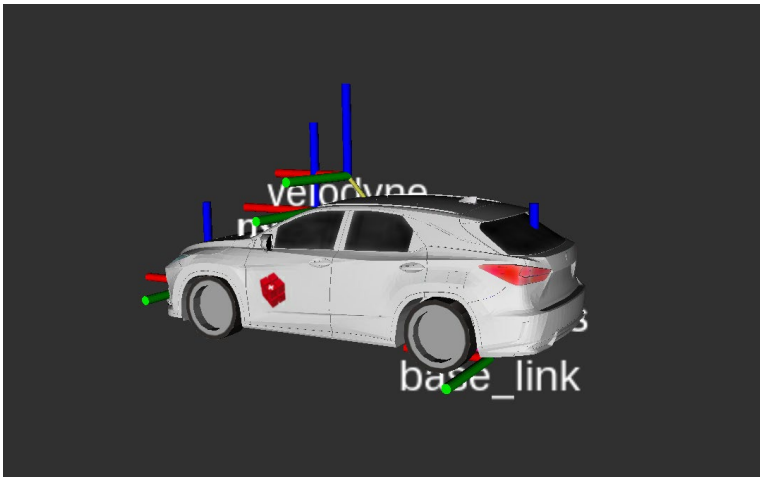

Figure 17: *Lexus rendering with sensor coordinate systems.*

Additionally, we have to estimate the lens distortion and the intrinsic parameters of our camera, which are needed to project 3D-points on a 2D camera image.

The camera intrinsic parameters are the focal length (f_x and f_y), optical center (c_x and c_y) and the skew coefficient (s) [25]. The camera intrinsic matrix K is then defined as:

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$

Since no camera lens is perfect, we also need to consider their distortion. For our camera, we use a plumb bob model [26], which is a combination of a radial distortion model and a tangential/thin prism model. Radial distortion occurs when light at the edge of the lens bends differently than at the center. The result is that straight lines aren't mapped to straight lines but as curved lines (see Figure 18).

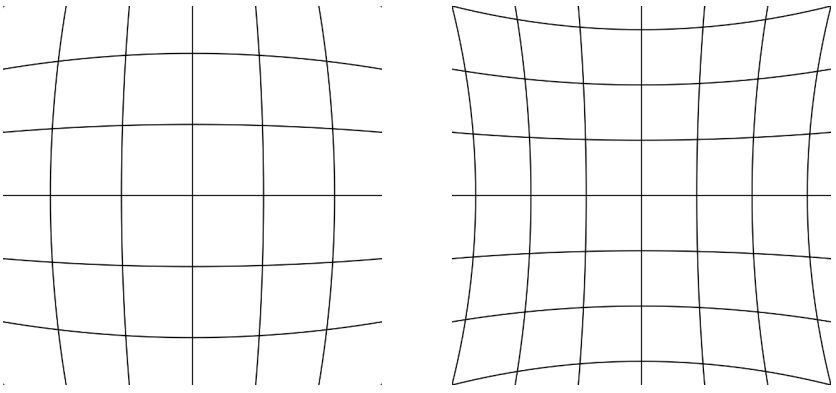

Figure 18: *Left: Barrel distortion. Right: Pincushion distortion. (sources: wikimedia.org)*

The radial distortion can be modeled as a function of the distance r from the center of the image plane.

$$\begin{bmatrix} x_{dist} \\ y_{dist} \end{bmatrix} = (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) \begin{bmatrix} x \\ y \end{bmatrix}.$$

The second effect, tangential distortion, occurs when the lens and image plane are not mounted exactly parallel.

$$\begin{bmatrix} x_{dist} \\ y_{dist} \end{bmatrix} = \begin{bmatrix} 2k_4xy + k_5(r^2 + 2x^2) \\ k_4(r^2 + 2y^2) + 2k_5xy \end{bmatrix}.$$

Combined, the plumb bob model can be expressed as [26]:

$$\begin{bmatrix} x_{dist} \\ y_{dist} \end{bmatrix} = (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2k_4xy + k_5(r^2 + 2x^2) \\ k_4(r^2 + 2y^2) + 2k_5xy \end{bmatrix}.$$

The estimation of the intrinsic (f_x, f_y, c_x, c_y, s) and the distortion $(k_1, k_2, k_3, k_4, k_5)$ parameters is done with a ROS-node[2]. The node uses a series of images from a checkerboard with known dimensions (Figure 19). For each image, the algorithm detects the corners on the checkerboard and from the known length of the squares and the fact that they are known to be straight lines, estimates the model parameters.

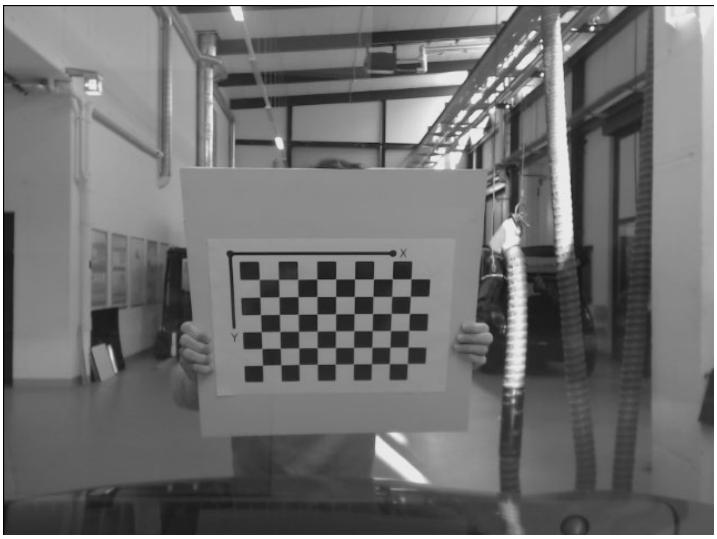

Figure 19: *Image from the camera parameter estimation process with a checkerboard.*

5.2 Test track

For the real-world function-scenario-based testing, we built a short test track at the Empa campus in Dübendorf (Figure 20). The track has an oval shape and includes two intersections and a zebra crossing (Figure 21). one of the intersections is a yield intersection (Figure 21) while the other one is a stop (Figure 21). While the back straight has only a single lane due to limited space (Figure 21), all other parts of the test track have two lanes, which allows for crossing or overtaking testing scenarios.

[2] https://www.ros.org/about-ros/

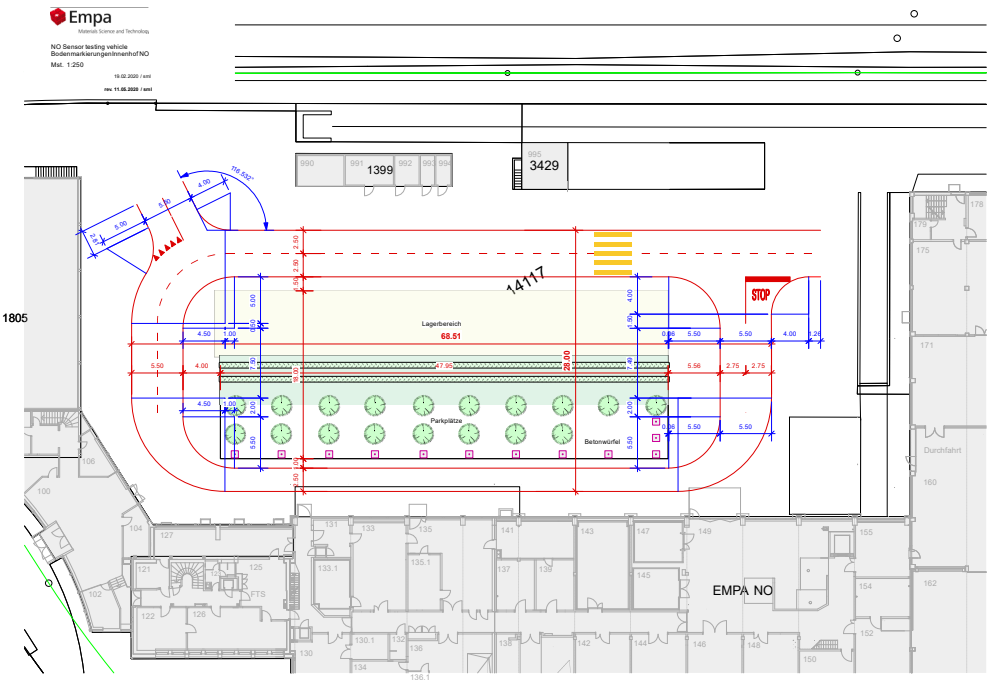

Figure 20: *Test track layout.*

Figure 21: *Uppder feft: Test track stop intersection with zebra crossing Upper right: Yield intersection. Lower left: Painting the stop marking. Lower right: Single-lane back straight*

5.3 Track Mapping

In order to use lidar localization, the test area has to be mapped in 3D. A visualization of such a 3D map of our test track is shown in Figure 22.

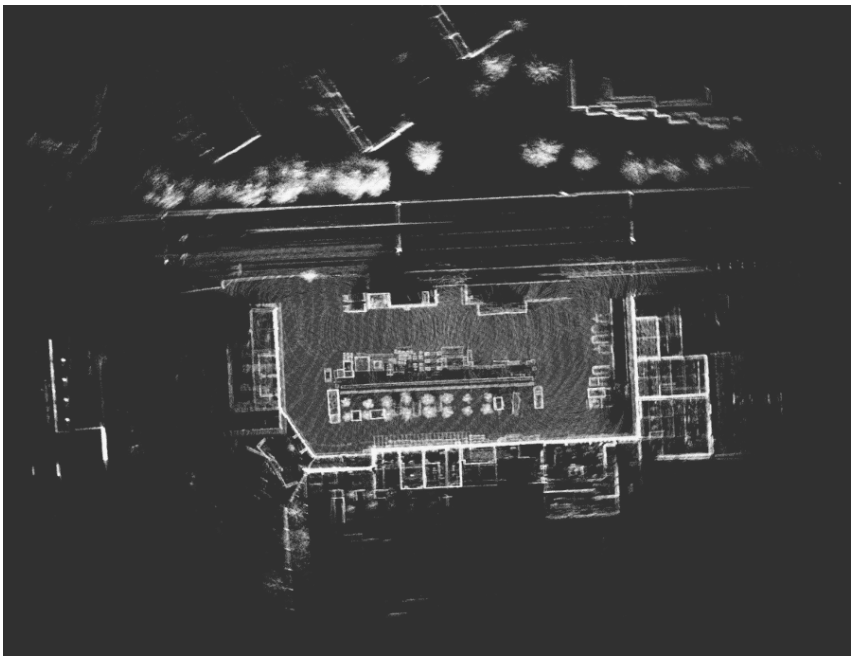

Figure 22: *Top view of the generated 3D point map from the test track.*

Such maps are generated by matching and overlying lidar data from different measurements. Figure 23 shows this process with two point clouds. Before the registration, the two point clouds are off by about 2 m. After the registration and applying the found transformation, the point clouds are overlaid, which is visible from the points representing the walls and trees around the test track.

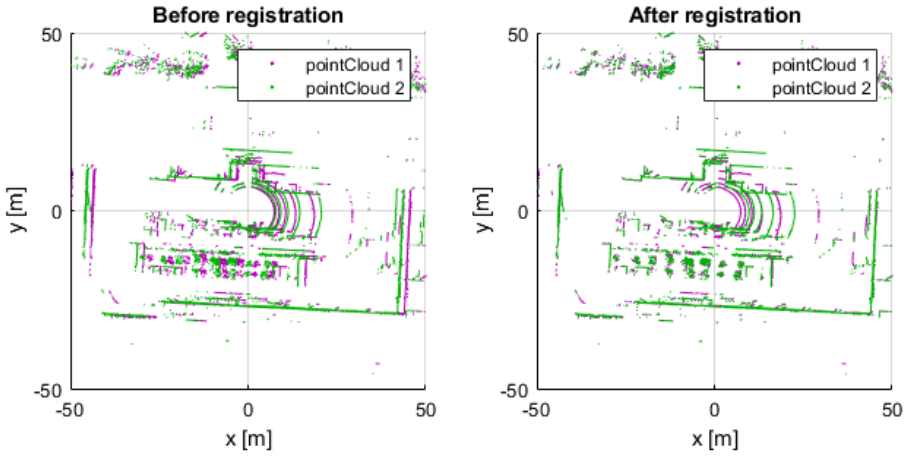

Figure 23: *Effect of point cloud registration used in 3D-Mapping.*

For mapping, this process is performed repeatedly. The mapping process starts with either a single point cloud as the current map or an existing 3D map. Each new point cloud is registered to the current map and then merged with it to form the new map (see Figure 24). Of course, additional care is required when the point clouds contain moving objects, i.e. cars, pedestrians etc. Since our test track is semi closed, for simplicity we mapped it using only cloud points that did not include any moving objects.

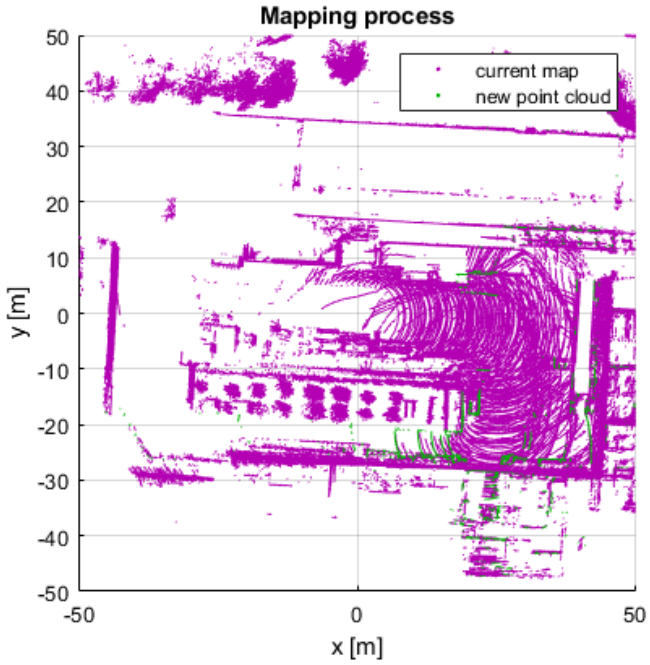

Figure 24: *Mapping process with accumulated map and the current new point cloud.*

5.4 Test objects

In order to enable repetitive tests with a fixed set-up, we built several static test objects and placed them in fixed positions on the test-track.

To represent pedestrians, we used several mannequins (including males/females and adults/child) which were painted, clothed and mounted onto sturdy metal tripods (Figure 25).

Figure 25: *Painted mannequins with clothes, mounted on tripods.*

For testing the lidar, we built three large targets. They are made out of wood panels (Figure 26) mounted onto euro palettes for stability and easy handling. Each wood panel measures 75 by 200 cm and was painted in either semi matte white, grey or black color.

Figure 26: *Wood panels for lidar targets.*

5.5 Weather Conditions

Sunshine

Sunny, i.e. bright and clear view but no glare conditions are considered to be the best or easiest conditions for sensing and perception. This includes bright illumination of the scene with usually high contrast, no reflections from water on the ground, no noise from absorptions or reflections on rain droplets or snowflakes and no blocking of the sensors' field of view.

Figure 27: *Impression of our test track in bright sunshine.*

Rain

Figure 28 shows an image taken on the Empa test track during rainfall. This image already shows some of the effects of rain, for example lower overall illumination and some light reflection on the ground. For our tests, rainy weather conditions means active rainfall, not just wet ground after rainfall.

Figure 28: *Impression of our test track in rainfall.*

Fog

Similar as for human drivers, fog also poses problems for automated vehicles. Camera images become low in contrast and objects are thus harder to detect. Fog also potentially dampens the laser light from the lidar, thus decreasing the signal to noise ratio and possibly reducing the maximum range.

Figure 29: *Impression of our test track in fog.*

Snowfall

Figure 30 shows an image taken on the Empa campus in Dübendorf during a heavy snowfall. Snowfall has multiple effects on perception, including scattering and reflection of light. For our tests, snowy weather conditions means active snowfall, not just snow on the ground.

Figure 30: *Impression of snowfall in Dübendorf.*

Contamination

For the contamination test, we artificially covered the sensors with dirt and leaves as shown in Figure 31 and Figure 32. Already before starting the test, the leaves on the lidar blew away due to the airflow, but the needles and dirt stuck to the sensor surface. For the radar, which given its position close to the ground is more prone to such contaminations, also the leaves stuck on the surface for the full length of the test. The Mobileye Camera was not considered in this test, as we assume that any potential dirt/contamination blocking its field of view can be easily and quickly be removed at any time with the windscreen wipers.

Figure 31: *Lidar covered with dirt and leaves.*

Figure 32: *Radar covered with dirt and leaves.*

6 Function-Scenario-Based Testing

In this chapter, we show results and the application of our function- scenario-based testing methodology described in Chapter 4. Thereby, we tested the Mobileye camera, VLP-16 lidar and Delphi radar (details about the sensors are given in Section 3.1). Testing scenarios, analyses, results and discussion for the object detection and distance estimation functions are described in Section 6.1. The localization function is shown in Section 6.2.

6.1 Object Detection and Distance Estimation

In this section we evaluate the object detection and distance estimation function (described in Section 3.2.1 and 3.2.3) for the Mobileye camera, the VLP-16 lidar and the Delphi radar using different field test scenarios. In Subsection 6.1.3 we present the results from our Walking Pedestrian scenario which illustrates a scenario in which a pedestrian is walking in front of the AV and the Mobileye camera and the Delphi radar need to detect and track the pedestrian. In Subsection 6.1.4 we report the results from our Mixed Urban Scenario, in which the vehicle is driving on an urban-like road with different test objects and the Mobileye camera and the Delphi radar need to detect the objects. Finally, in Subsection 6.1.5 the Lidar Targets scenario is shown, which is used to evaluate the lidar object detection capabilities. For all three scenarios, the object detection and distance estimation functions were evaluated under different environmental conditions (weather and sensor contamination).

6.1.1 Sensors, Algorithms and Black Boxes

As we have seen in Section 3.2, detection is a crucial perception function in AD. Thereby, sensors observe the environment and create raw data formats such as images or point clouds. These sensor measurements are then processed by an algorithm in order to obtain the object detections. Since the algorithm provides the final detections, it is hard to separate the algorithm from the sensor in the evaluation. For example, it is hard to say how good a specific camera is for object detection at night without considering the corresponding algorithm. Nonetheless, some conclusions can still be made from the raw sensor data, e.g. the probability that an algorithm will detect an object based on a lidar point cloud increases with the number of lidar points lying on the object.

For the distance estimation, we need to differentiate between active and passive (range or vision) sensors. The lidar and radar sensors are active sensors and their measurements are actually ranges. The accuracy of the measurements depends on the chosen hardware of the sensor and physics. On the other hand, camera sensors are passive sensors, which provide images from the environment. Since an image does not directly provide depth or distance information, this needs to be estimated with an algorithm. Same as for detection, the sensor and algorithm cannot be separated. In stereo vision, the distance estimation depends on the camera calibrations.

Then there are sensors which provide tracks as described in Subsection 3.2.4. The tracks are filtered and processed detection and distance estimation data. The Mobileye camera and the Delphi radar provide directly tracks and we cannot access the raw sensor data such as images or the detections. Instead we get tracks, which are processed detections with additional information such as location of the object, velocity of the object and instance identification of the object. Such systems are called black box systems, since we do not know how the sensor tracked an object, e.g. what image did the camera create, which objects were detected and which detections were matched to each track. Since such black box systems might be the standard for coming sensors, our experiments can give first insights on how to test such end-to-end black boxes.

The lidar and the RGB-camera can also be used for object detection but they need to be merged with additional external detection software. Autoware.ai[3] provides algorithms to extract information from both sensors individually or by sensor fusion. However, testing such algorithms was out of the scope of this project. Therefore, we focused on the testing of the main detection sensors, i.e. the Mobileye camera and the Delphi radar. Further, we look also in the possible usage and capabilities of the Velodyne lidar for object detection with our Lidar Targets scenario.

6.1.2 Implementation and first insights

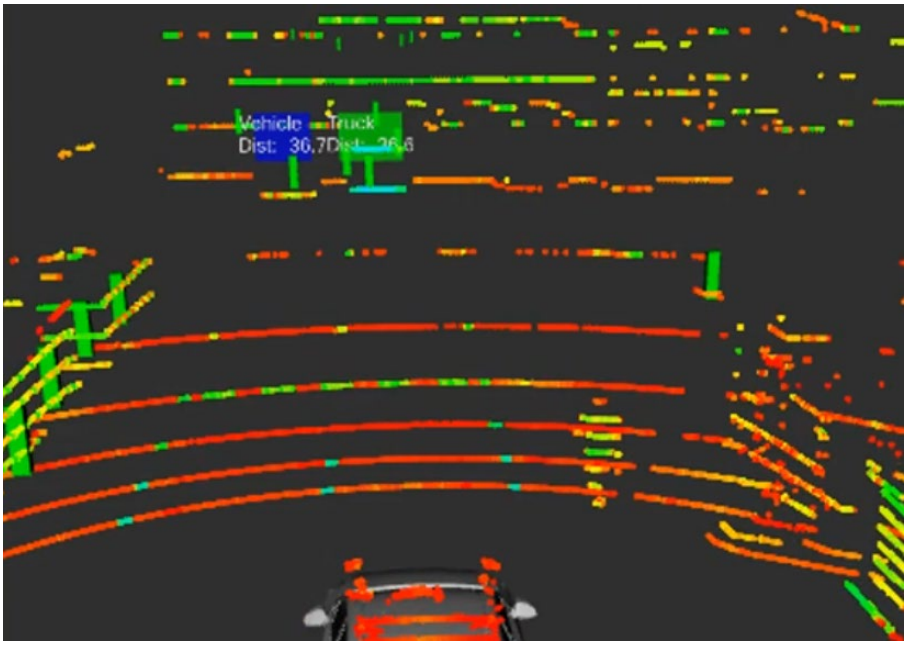

Figure 33: *Example for multiple radar tracks for a single object. The truck appears as three radar tracks.*

During the implementation and installation of the Mobileye sensor and the Delphi radar, we collected several first insights.

In Figure 33 we can see which functionalities or measurements are covered by the different sensors: The Mobileye returns a bounding box with distance estimation and classification of the object. Unfortunately, the ROS driver for the Mobileye camera does not allow visualizing the sensor data as necessary. The bounding boxes of the object detection are vertically centered on the lane detection. Therefore, we can either visualize the lanes correctly on the ground and have the bounding boxes inside the ground or we visualize the bounding boxes on the correct height and have the lanes visualized at the level of the drivers eyes. We are currently working on an own solution for this.

In contrast, the radar works as specified by the manufacturer. It mostly detects the object and estimates the correct distance to those objects. One downside of this system is that the detections from the radar are not classified and that the radar marks bigger objects with several markers (see Figure 33). Additionally, it is difficult to differentiate between a fix object beside the road (e.g. a tree, pole, etc.) and a standing or even slow moving pedestrian. Hence, the sensor is suited for collision avoidance, but not for the perception of the environment. Further, we observed another difficulty: since the radar is mounted at the height of the license plate, it concentrates on the detection of objects near the ground. For pedestrians and bicycles this leads to problems in detection, since most of the reflection area is farther away from the ground.

[3] https://www.autoware.ai/

6.1.3 Walking Pedestrian Scenario

Scenario Description

The setup of the scenario is shown in Figure 34. The pedestrian starts 5 meters away from the stationary AD-vehicle and walks away from it stopping for 10 s every 5 m. Once it reaches the mark at 50 m the pedestrian turns around and walks back towards the car, again in 5m steps. At the standstill positions, we know the exact distance and can therefore precisely evaluate the distance estimate performance. Figure 35 shows an image taken at the 30m position.

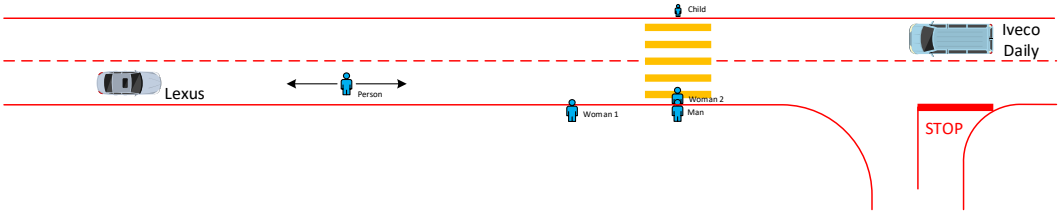

Figure 34: *Walking pedestrian scenario setup.*

Figure 35: *Picture of a pedestrian in front of the AD-vehicle (29.09.2020 around 09:00 a.m.).*

Data Processing and Analysis

In the Walking Pedestrian Scenario, we evaluate the performance of the Mobileye Camera and the Delphi radar with respect to pedestrian detection in different environmental conditions.

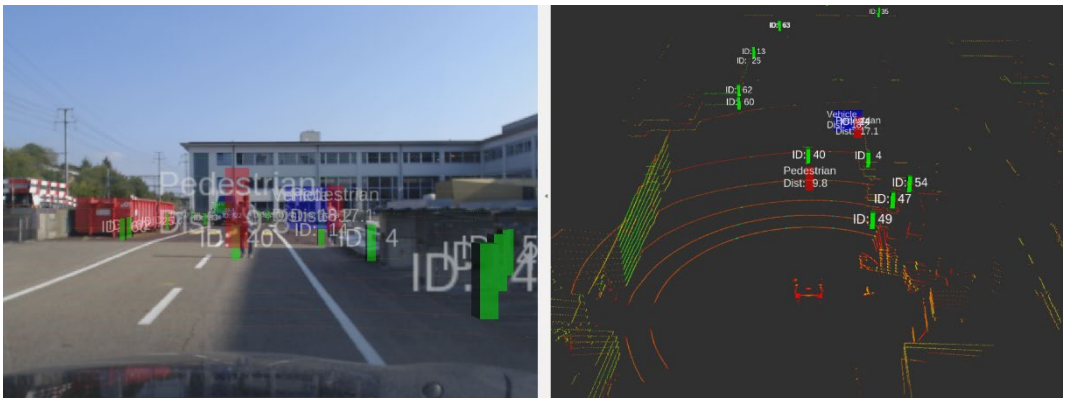

Figure 36: *Visualization of data from a Walking Pedestrian scenario (log 09 from 30.9.2020)*

As described above, the Mobileye camera and the ESR Radar provide tracks of all the detected objects. To evaluate multiple object tracking algorithms previous studies have used the performance metric "Multiple Object Tracking Accuracy" (MOTA) [27]. However, because we are dealing with black box systems, we cannot use this metric in our case. In fact, the Mobileye camera does not output all the single detections, which would be needed to evaluate the false positives and false negatives. Instead, we only receive the object tractions. A traction consists of a detection, classification and position estimate of an object, where a unique tracking identification number (ID) is assigned to the object. The sensor tries to identify the same object in successive time steps and thereby to assign the same tracking ID to each object. If the sensor loses or changes the track ID of an object, we have no way to know what caused it. The same holds for the radar, for which the lack of the object classification function makes it even harder to define false positives and false negatives. Thus, we had to come up with different metrics to evaluate the detection and distance estimates of the two black box sensors (Mobileye camera and ESR Radar). The metrics that we used to evaluate the sensors in the Walking Pedestrian scenario are summarized in Table 1 and described more in detail in the next subsections. Since the ESR Radar does not classify its tracks, we cannot evaluate this metric in this case. All other used metrics apply to both sensors.

Table 1: *Performance metrics for Walking Pedestrian Scenario.*

Metric	Mobileye Camera	Delphi ESR Radar
Distance	X	X
Number of tracks	X	X
Average track duration	X	X
Tracking Ratio	X	X
Classification	X	-

Distance

The evaluation of the distance estimate is straightforward: we know the distance at the standstill positions and compare it with the distance estimates from the sensors. The performance metrics consist of the absolute and relative errors.

Number of tracks, average track duration and tracking ratio

In order to evaluate the stability of the detection and tracking algorithms we analyze the number of tracks, the average track duration and the tracking ratio for each test run.

First, we compile the number of tracks (n_{tracks}) and the duration of the tracks (t_{track}) created in each test run for the target object.

From this we can calculate the total tracking time, which is given by the sum of the track durations over all tracks:

$$totalTrackingTime = \sum_{n_{tracks}} t_{track}$$

We then calculate the average track duration as the ratio between the total tracking time and the number of tracks. The more tracks there are, the smaller this number becomes.

$$averageTrackDuration = \frac{totalTrackingTime}{n_{tracks}}$$

Finally, we calculate the tracking ratio (TR) as the ratio between the total tracking time and the total test time (given by the difference between the start and end times of the test run). The TR gives a measure of what percentage of the time the pedestrian was tracked, or, in other words, how often we have a false negative (i.e. no track).

$$totalTime = tEnd - tStart$$

$$TR = \frac{totalTrackingTime}{totalTime}$$

Classification

Classification only applies to the Mobileye camera. In the Walking Pedestrian scenario this metric is trivial to evaluate. The sensors' classification is plotted, then checked manually for each track and was found to be always correct for the Walking Pedestrian scenario.

Data Processing

In order to analyze the pedestrian tracks, we first have to filter out all the other tracks, i.e. we have to identify and separate all the tracks belonging to the walking pedestrian. After manual annotation, all further processing steps can be performed based on the track IDs.

For the radar, since the position estimate of the tracks, i.e. distance and angle, is accurate, we can identify the pedestrian tracks filtering on a narrow region of interest. Figure 37 shows all radars tracks from a test run and the pedestrian tracks within the Region of Interest (i.e. attributed to the walking pedestrian) marked in blue.

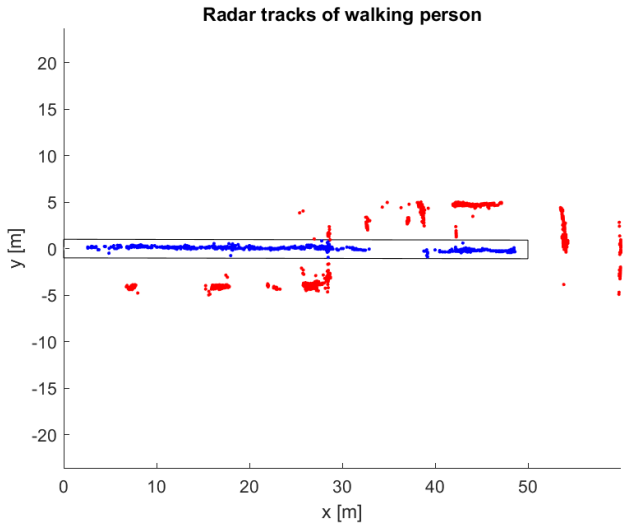

Figure 37: *Radar filtering based on region of interest (log 06 from 28.9.2020).*

For the Mobileye camera, because the distance estimate is much worse but the number of tracks is low, we mostly manually mark all tracks (or track IDs) that belong to the walking pedestrian. Figure 38 shows all Mobileye tracks and their track ID from the same test run; the blue marks correspond to the pedestrian.

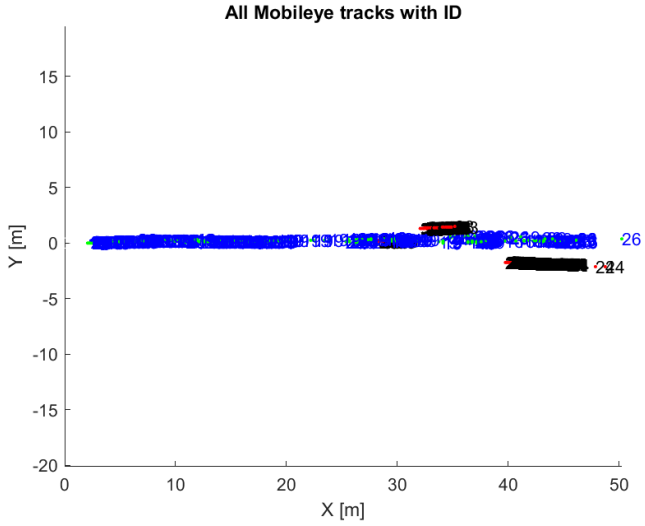

Figure 38: *All Mobileye tracks with their track ID (log 06 from 28.9.2020).*

Results

In this section we present the results obtained from the Walking Pedestrian scenario, focusing on the tracking performance and the accuracy of the distance estimates for different environmental conditions, first for the Mobileye camera end then for the ESR radar.

Mobileye

In this section, we show the results for the Mobileye camera in the Walking Pedestrian scenario for the different environmental conditions. For each condition, we only show figures from one test run but describe also the interesting observations and differences from the other test runs in the same conditions.

Sunshine

Figure 39 shows the results from a selected test run on a sunny day. As can be seen, the detection stability is good, as the pedestrian always has the same track ID. As shown in Table 2 this is always the case for sunny conditions. We also see that for distances above 15 m there is quite often a significant overshoot in the distance estimation when the pedestrian stops walking, but once the pedestrian stands still the distance estimate stabilizes (Figure 39). Regarding the precision, we find that the distance is generally underestimated (see Figure 44 and Figure 45 for plots of the error distribution).

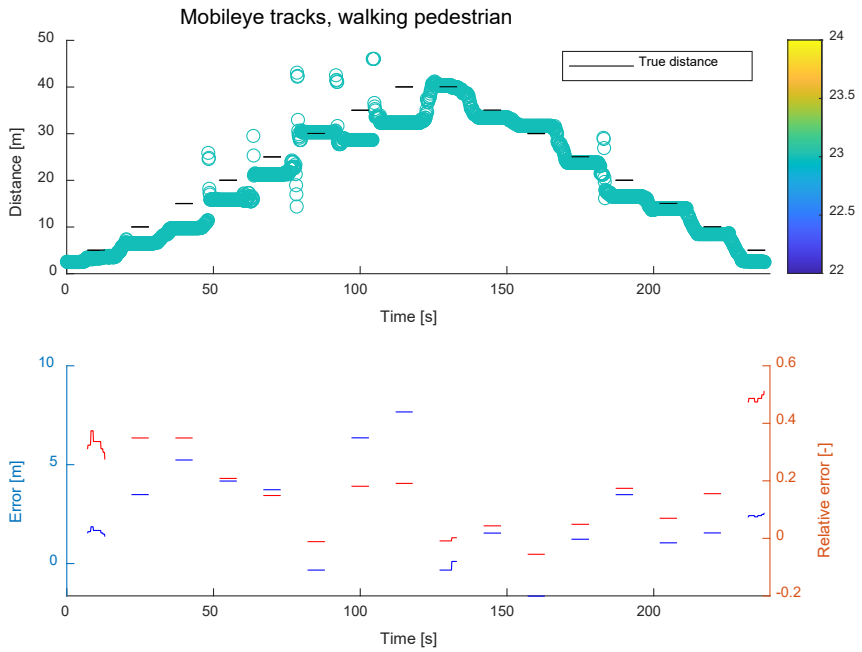

Figure 39: *Distance estimation results from a single test run on a sunny day (log 09 from 30.09.2020).*

Rain

Figure 40 shows an example from a selected test run on a rainy day. The tracking performance remains quite good, with an accuracy similar to the test runs in sunny conditions, but the tracking stability is worse (see also Statistical Results). The Mobileye camera loses the pedestrian twice, resulting in a total of three tracks for the pedestrian in this test run.

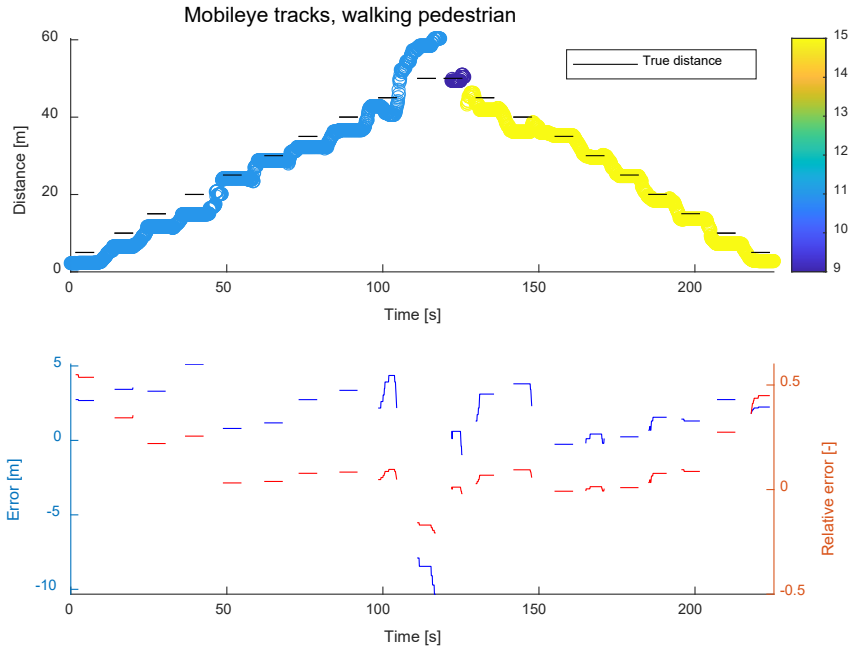

Figure 40: *Distance estimation results from a test run from a rainy day (log 09 from 29.9.2020).*

Fog

Figure 41 shows an example from a test run on a foggy day. Even though there is a continuous track, the tracking performance is really bad (error up to 30 m, relative error up to 60 %). Up to 25 m true distance, the distance is consistently underestimated by 20 to 50 %. After the 25 m marker, the estimated distances decrease abruptly and are underestimated by around 60 % until the pedestrian is back to the 15 m marker at the end of the test.

In the other test run performed under foggy conditions, the distance estimation was much better, with a consistent underestimation of around 30 %.

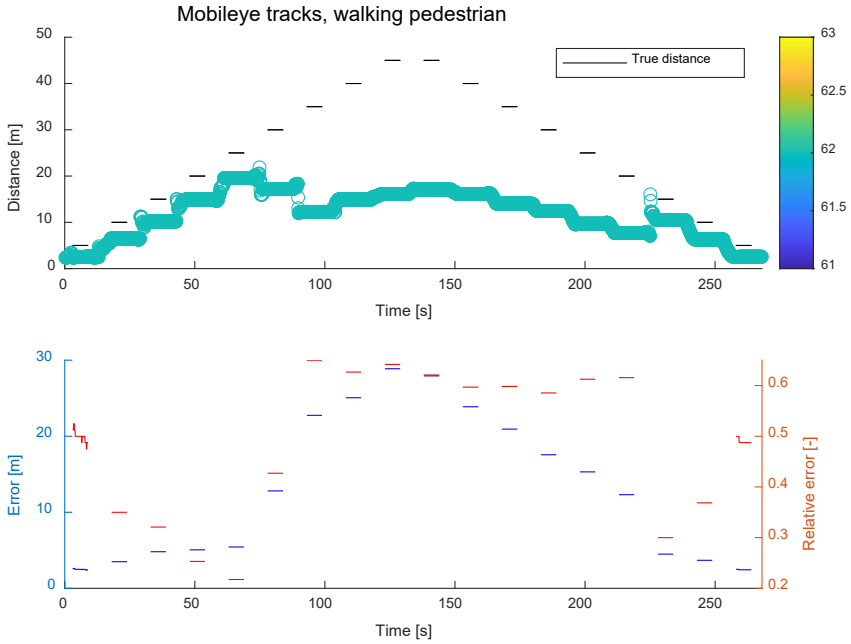

Figure 41: *Distance estimation results from a test run on a foggy day (log 16 from 18.11.2020).*

Snow

During our test runs in snowy condition, we recognized a significant increase in the number of tracks created for the pedestrian, which translates into many track losses and short tracks (see Figure 42 and Table 2). Since all test runs in snowy conditions showed the same behavior, we can conclude that the snow and lighting conditions during snowfall were the cause for the bad tracking performance. Regarding the distance estimate accuracy, we do not see any worsening of the generally bad performance observed in the other conditions.

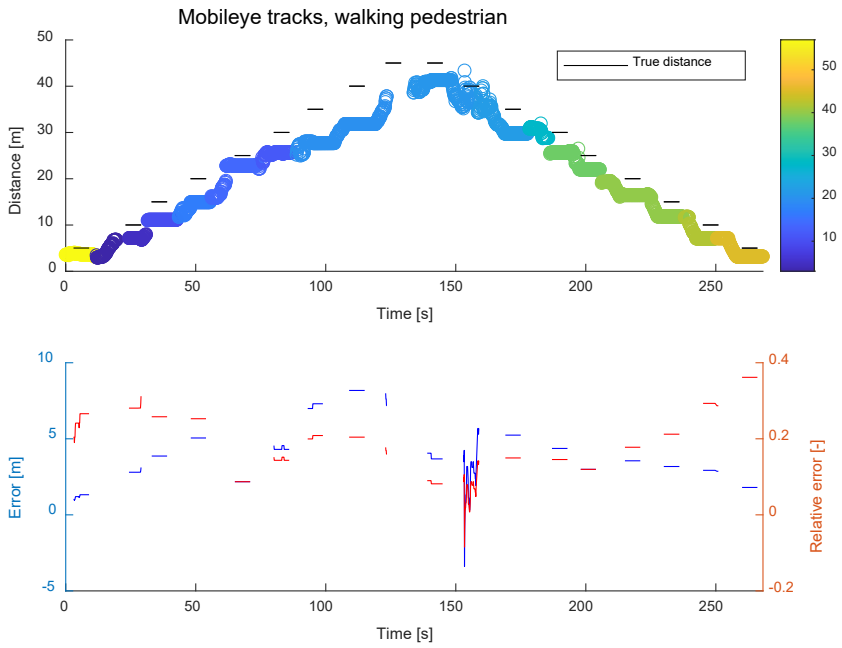

Figure 42: *Distance estimation results from a test run on a snowy morning (log 16 from 1.12.2020).*

Sunrise

Figure 43 shows the results of a test run in difficult environmental conditions: slight rain combined with bad illumination just after sunrise. We can see that under these conditions the Mobileye camera does not detect the pedestrian at all.

After looking at the raw data, we saw that in this early morning test run the Mobileye camera did manage to detect the Iveco Daily, but not the walking pedestrian, nor the mannequins positioned around the track. This check ensures us that the Mobileye camera was working properly but could just not detect the pedestrian under such light conditions.

According to the Mobileye Data sheet [13], pedestrian detection is only possible during daylight. The question that then arises is whether shortly after sunrise counts as daylight or not.

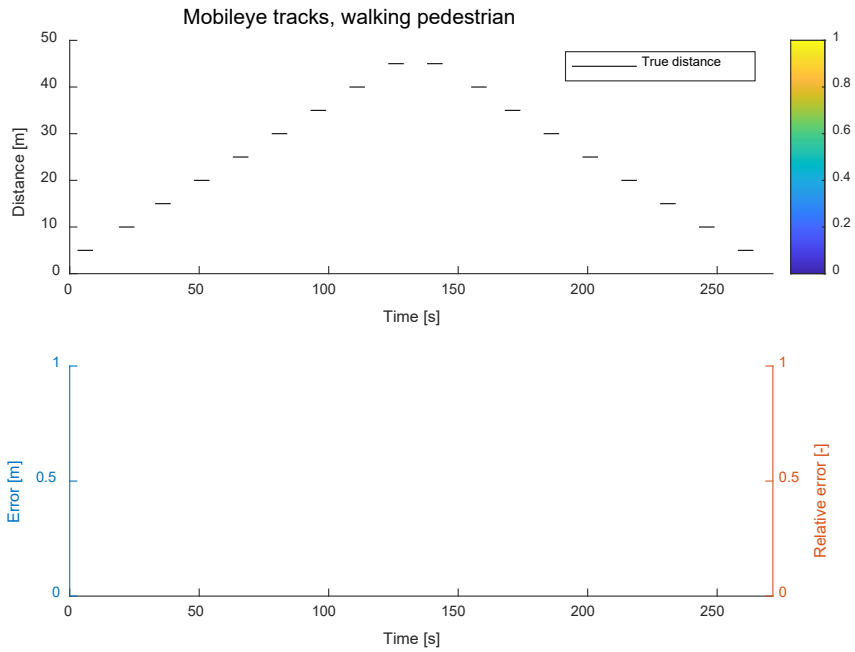

Figure 43: *Distance estimation results from a test run at early morning (log 13 from 16.10.2020)*

Contamination

The Mobileye is mounted behind the windscreen and can be cleaned continuously with the windscreen wiper and fluid if necessary. Therefore, exterior contamination is not a real problem for the Mobileye.

Statistical Results

Table 2 shows the performance metrics for the single test runs. During bright and dry weather conditions, we have a continuous single track resulting in a tracking ratio close to 100 %. We also see that in the early morning and light rain condition the pedestrian is not detected at all. The number of tracks increases in rainy conditions and even more so for snowfall. In both snowfall test runs we had 15 tracks, which results in an average track duration of around 16 s and a slightly lower tracking ratio than in dry weather conditions. Surprisingly, fog does not seem to influence the tracking stability (1 long track in both test runs), but, as discussed above, the distance estimate is much worse in this case.

Table 2: *Mobileye results from all test runs.*

Date	Condition	#Track-Ids #Tracks	Average Track Duration [s]	Tracking Ratio [%]
28.08.2020	Sunny, with few clouds	3	65.6	88.8
30.09.2020	Sunny afternoon	1	234.2	100
30.09.2020	Sunny afternoon	1	237.9	99.3
15.10.2020	Bright morning with few clouds	3	132.8.7	98.2
15.10.2020	Bright morning with few clouds	2	127.8	95.9
18.10.2020	Sunny afternoon	1	209.3	98.3
18.10.2020	Sunny afternoon	1	272.2	99.4
29.09.2020	Rainy, dim light	3	73.1	94.5
29.09.2020	Rainy, dim light	3	73.7	96.2
29.09.2020	Rainy, dim light	3	73.6	98.0
16.10.2020	Light rain, early morning	0	-	0
16.10.2020	Light rain, early morning	0	-	0
16.10.2020	Light rain, early morning	0	-	0

16.10.2020	Light rain, early morning	0	-	0
19.10.2020	Light rain	9	33.32	98.2
18.10.2020	Foggy morning	1	267.6	99.9
18.10.2020	Foggy morning	1	259.1	100
01.12.2020	Snowfall	15	16.7	93.3
01.12.2020	Snowfall	15	16.2	90.8

The following figures show the distributions of the absolute errors (Figure 44) and relative errors (Figure 45) of the distance estimate for the different environmental conditions as a function of the distance to the pedestrian. The distributions include the results from all the test runs reported in Table 2 for a given condition. It can be seen that for all conditions the absolute error increases for higher distances, especially for foggy conditions. Moreover, the variability of the error among different tests also tends to increase in all cases with the distance. The trends for the relative errors are a bit less clear, although it seems that the relative error tends to decrease for increasing distances for most conditions.

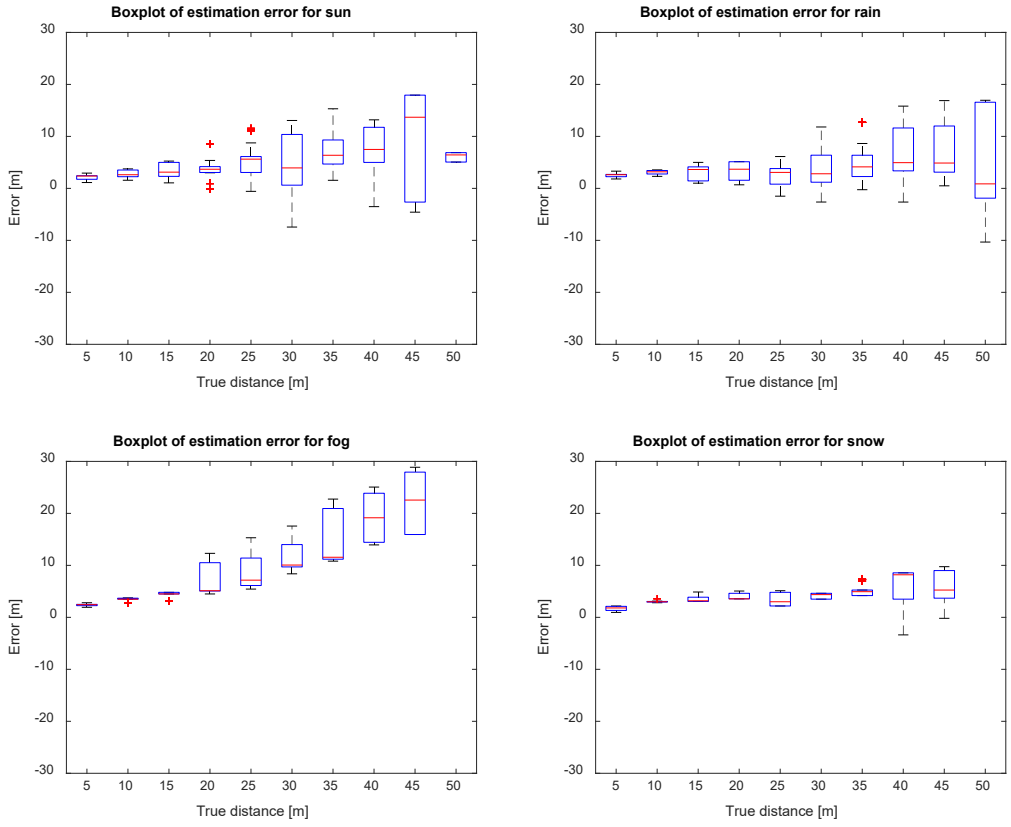

Figure 44: *Box plots of distance errors for different weather conditions of Mobileye camera (upper left: sun, upper right: rain, lower left: fog, lower right: snow).*

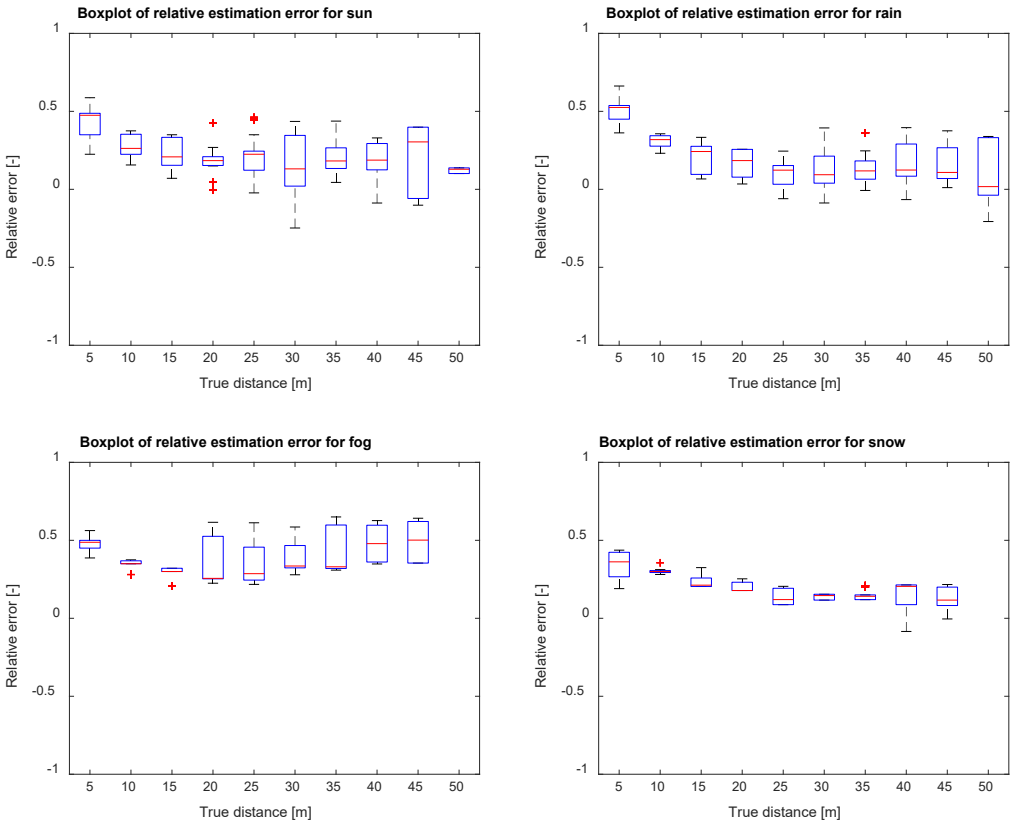

Figure 45: *Box plots of relative distance errors for different weather conditions of Mobileye camera (upper left: sun, upper right: rain, lower left: fog, lower right: snow).*

ESR Radar

In this section, we report the results from the radar. For each condition, we only show figures from one test run but describe also the interesting observations and differences from the other test runs in the same conditions.

Sunshine

Figure 46 shows all radar tracks corresponding to the walking pedestrian from a test run on a sunny day with the colors indicating different track IDs. As reported in Table 3, we observe over 100 short tracks and 13 tracks with a duration longer than 1 second. Moreover, there are two distinct gaps in the tracks at 60 s and at 160 s, corresponding to a distance between the radar and the pedestrian of 20 m respectively 30 m as shown in Figure 46. We assume that the cause for these two gaps are material changes of the road from asphalt to reinforced concrete. These two gaps are a repeating phenomenon in our measurements which is further discussed later on in this section.

Figure 46 also shows the accuracy of the distance estimate. As shown, the distance estimate from the radar is very consistent when the pedestrian is standing at a still position and shows almost no over and undershoots when the pedestrian is walking. The difference between true and measured distance (± 1 m) is well within the accuracy of the pedestrian ground truth positioning and size, i.e. the annotation and object size tolerance. Since the distance accuracy and uncertainty could be much better evaluated at the METAS laboratories [28], here we focus more on the repeatability of the pedestrian detection itself.

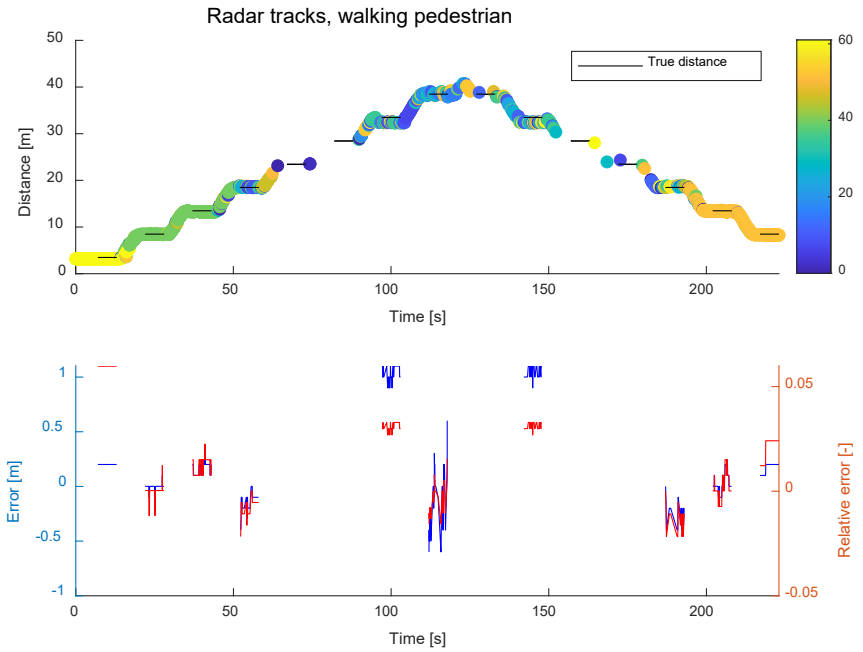

Figure 46: *Stationary distance errors from a test run on a sunny day (log 09 from 30.09.2020).*

Rain

Overall, the results from rainy conditions resemble the ones of the sunny conditions. Figure 47 shows all tracks corresponding to the walking pedestrian from a test run on a rainy day. As for sunshine conditions, we see a few longer tracks and many spurious tracks which only last a few time steps. In addition, the distance accuracy and consistency is again within the annotation and object size tolerance.

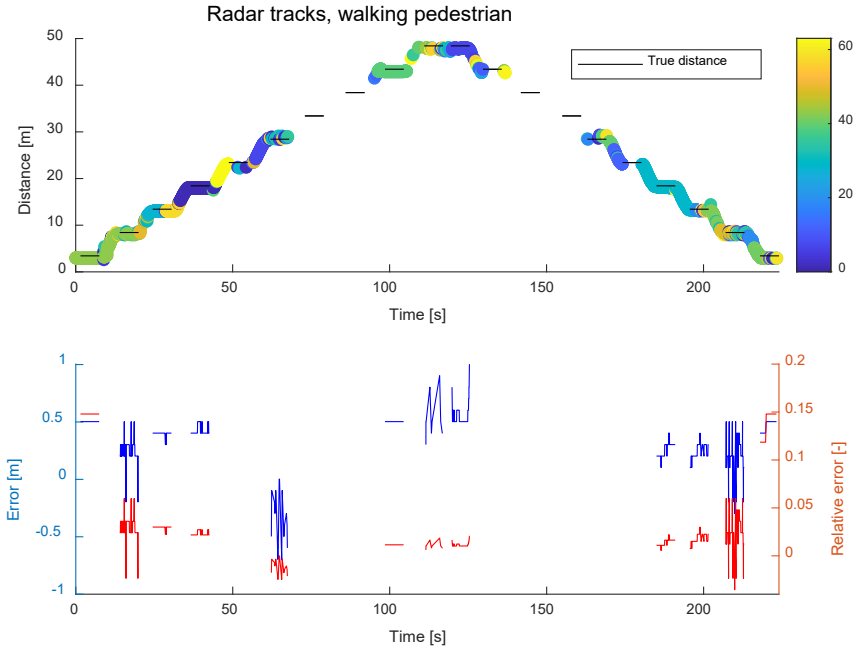

Figure 47: *Distance errors from a test run on a rainy day (log 08 from 29.9.2020).*

Fog

Figure 48 shows the results of a test run during fog. Overall, we see the same performance as in the sunny and rainy conditions. As shown also in Table 3, the number of spurious tracks slightly decreases in the fog conditions, while the tracking ratio remains about the same (around 45 %) as in the other test cases.

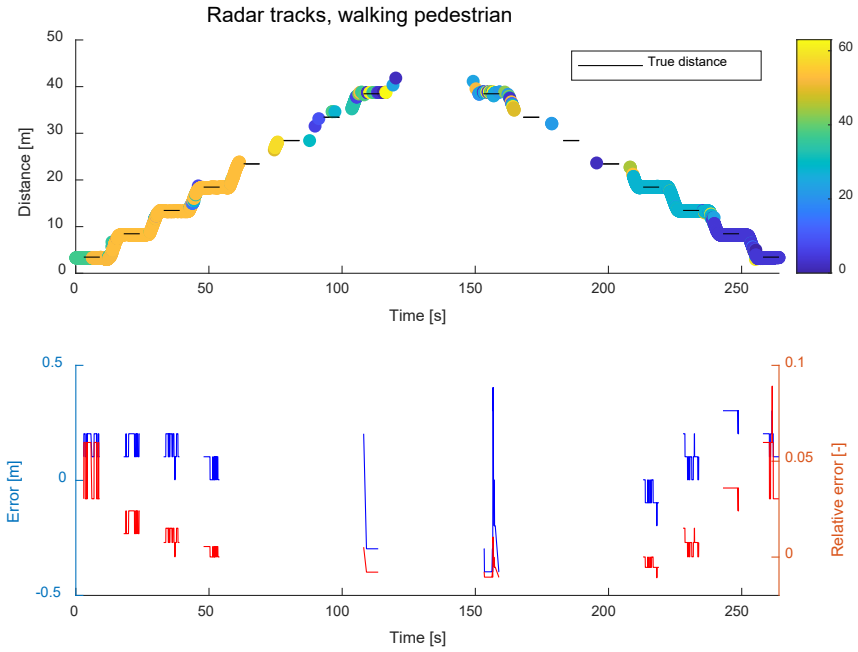

Figure 48: Distance errors from a test run on a foggy day (log 16 from 18.11.2020).

Snow

In the case of snowfall we do not see any significant differences in the detection performance compared to the aforementioned environmental conditions (see Figure 49). Also number of spurious tracks and the error of the distance estimate remain similar to the other cases. However, it is important to note that since we did our tests on our semi-closed, low-speed test track, the radar, despite being mounted on the front bumper, did not catch up a lot of snow. This might not always be the case; when driving behind other vehicles in busy streets, or stopping for longer times at a traffic light, the amount of snow and dirt from the road covering the radar surface might be substantially higher, resulting in a reduced detection performance.

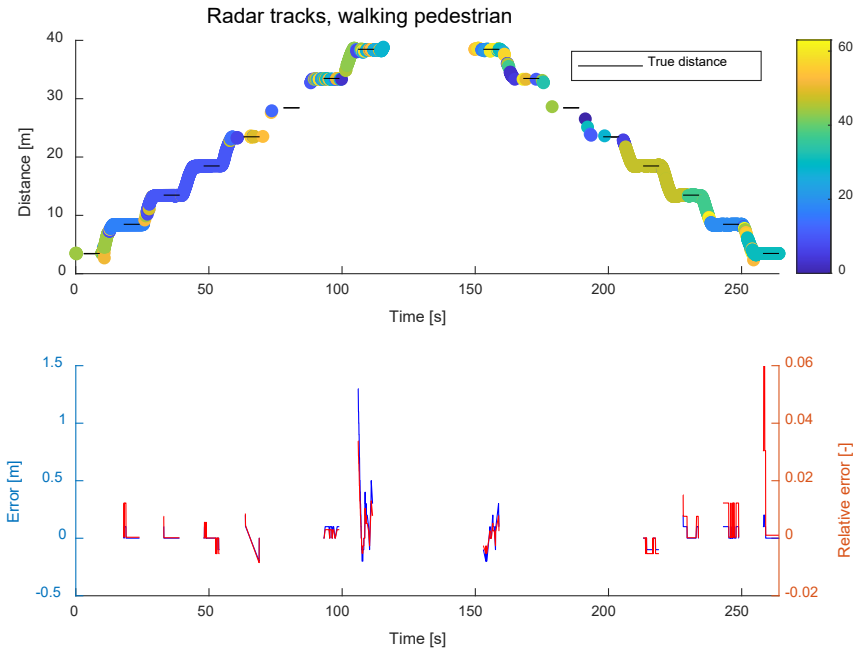

Figure 49: *Distance errors from a test run on a snowfall day (log 16 from 1.12.2020)*

Contamination

Figure 50 shows that during the test with external contamination (leaves and dirt), the radar tracks the pedestrian surprisingly well compared to all other tests. In fact, for most of the time, there is a stable, longer track and only a few spurious tracks. This results in a tracking ratio of 85% when considering only the long tracks (Table 3). The distance estimate accuracy remains very high and well within the positioning accuracy. For the time being we do not have an explanation for the observed improvement of the tracking stability and how/whether it could be linked to the contamination. Additional tests will be necessary to clarify this.

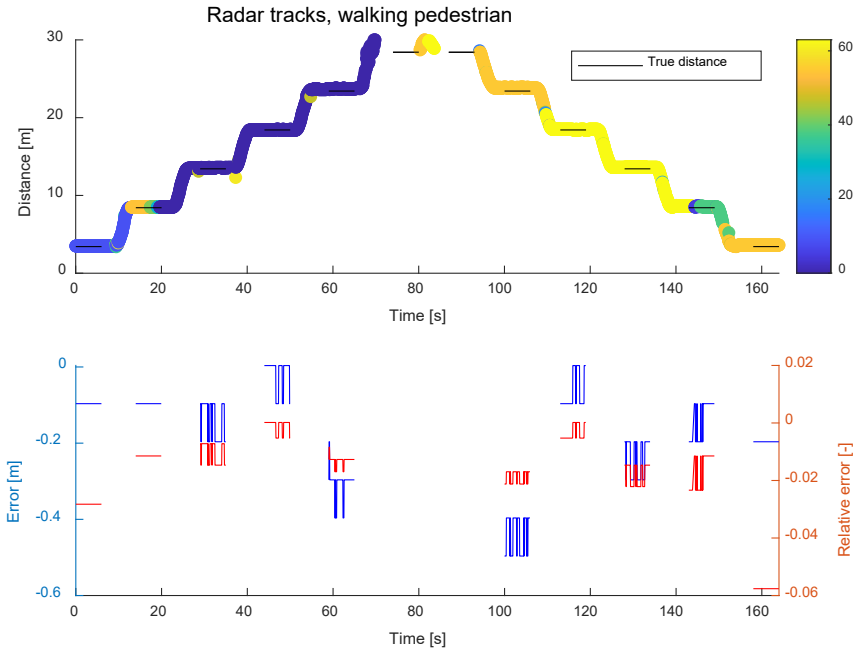

Figure 50: *Distance errors from a test run with contaminated sensor (log 14 from 27.11.2020).*

Statistical Results

Table 3 shows the statistics for each test run. Given the large amount of very short tracks, in this case we also report the number, average track duration and the tracking ratio of the "long tracks", which we define as those being longer than 1 s. We see that for each test there are usually over 10 long tracks and more than 100 tracks in total. The average track duration is generally below 1 s when all tracks are considered, and mostly remains below 10 s for the long tracks. The tracking ratio is around 50 % for the long tracks and does not significantly increase when all tracks are considered. Interestingly enough, we sometimes observe slightly lower tracking ratios, both for the total and long tracks, in sunny conditions. Except from this effect and the increased tracking performance observed during the test with the contaminated sensor, we cannot see any significant links between the weather conditions and the detection and tracking performance of the radar, which was expected from its working principle.

Table 3: **Radar results from all test runs.**

Date	Condition	# Tracks	Average Track Duration [s]	Tracking Ratio [%]	# Long Tracks	Average Long Track Duration	Long Tracking Ratio [%]
28.08.2020	Sunny with clouds	269	0.34	41.7	28	1.51	19.2
30.09.2020	Sunny afternoon	200	0.60	52.1	19	5.77	47.3
30.09.2020	Sunny afternoon	224	0.59	55.4	13	8.88	48.5
15.10.2020	Bright morning with few clouds	188	0.82	57.2	17	8.60	54.3
15.10.2020	Bright morning with few clouds	152	0.98	56.3	20	6.81	51.6
18.11.2020	Sunny sky	94	0.77	26.9	3	23.0	25.7
18.11.2020	Sunny sky	155	0.87	50.0	7	18.32	47.5
29.09.2020	Rainy, dim light	135	0.96	56.2	21	5.88	53.6
29.09.2020	Rainy, dim light	159	0.72	49.2	14	7.40	45.3
29.09.2020	Rainy, dim light	148	0.72	47.8	17	5.89	44.7
16.10.2020	Early morning, light rain	209	0.92	75.9	22	8.24	71.3
16.10.2020	Early morning, light rain	222	0.58	48.8	22	5.19	43.2
19.10.2020	Light rain	118	1.21	53.1	10	12.93	48.2
18.10.2020	Foggy morning	110	1.18	49	8	15.29	46.3
18.10.2020	Foggy morning	104	1.29	51.9	13	9.73	49.0
01.12.2020	Snowfall	137	0.92	47.2	15	7.72	43.5
01.12.2020	Snowfall	124	0.98	45.9	10	10.7	40.5
27.10.2020	Contamination	35	4.02	85.8	11	12.72	85.3

The following figures show the distributions of the absolute errors (Figure 51) and relative errors (Figure 52) of the distance measurements of the ESR radar for the different environmental conditions as a function of the distance to the pedestrian. The distributions include the results from all the test runs reported in Table 2 for a given condition. We see no clear trend in the absolute error with distance or weather condition. The absolute error is mostly below half a meter and there are only a few outliers with errors of 1 to 2 m. We also observe that there are often no results at distances around 20 m. This issue is discussed later in this section in more detail.

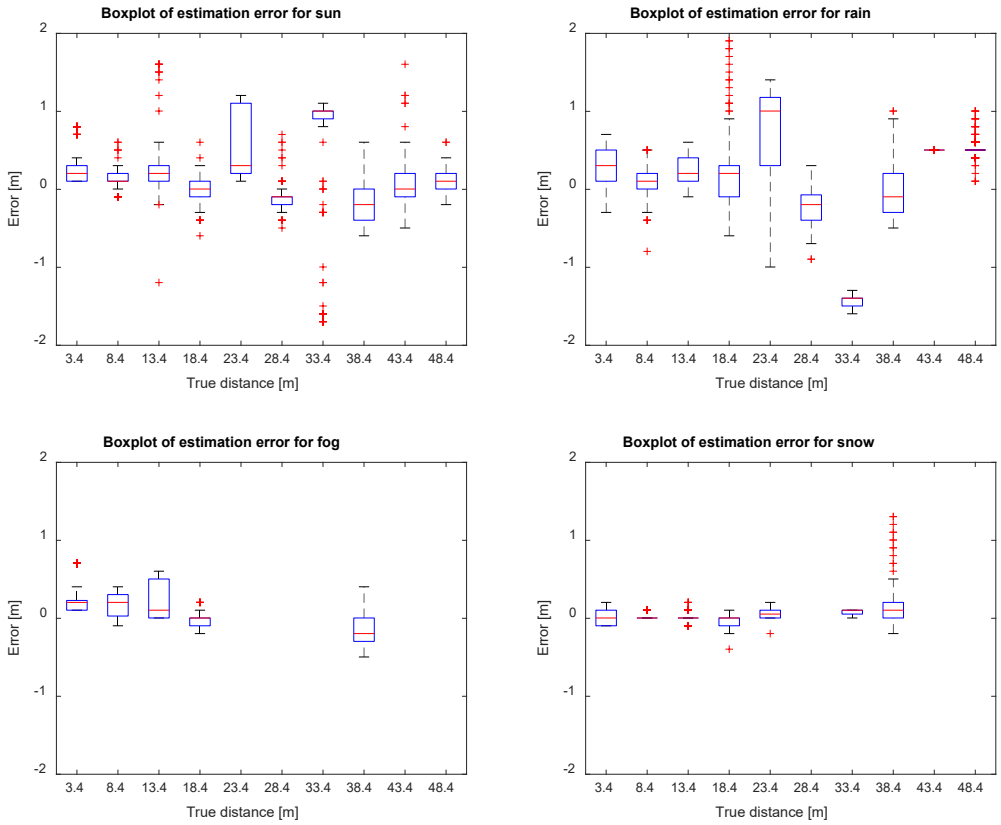

Figure 51: *Box plots of distance errors for different weather conditions of the ESR radar (upper left: sun, upper right: rain, lower left: fog, lower right: snow).*

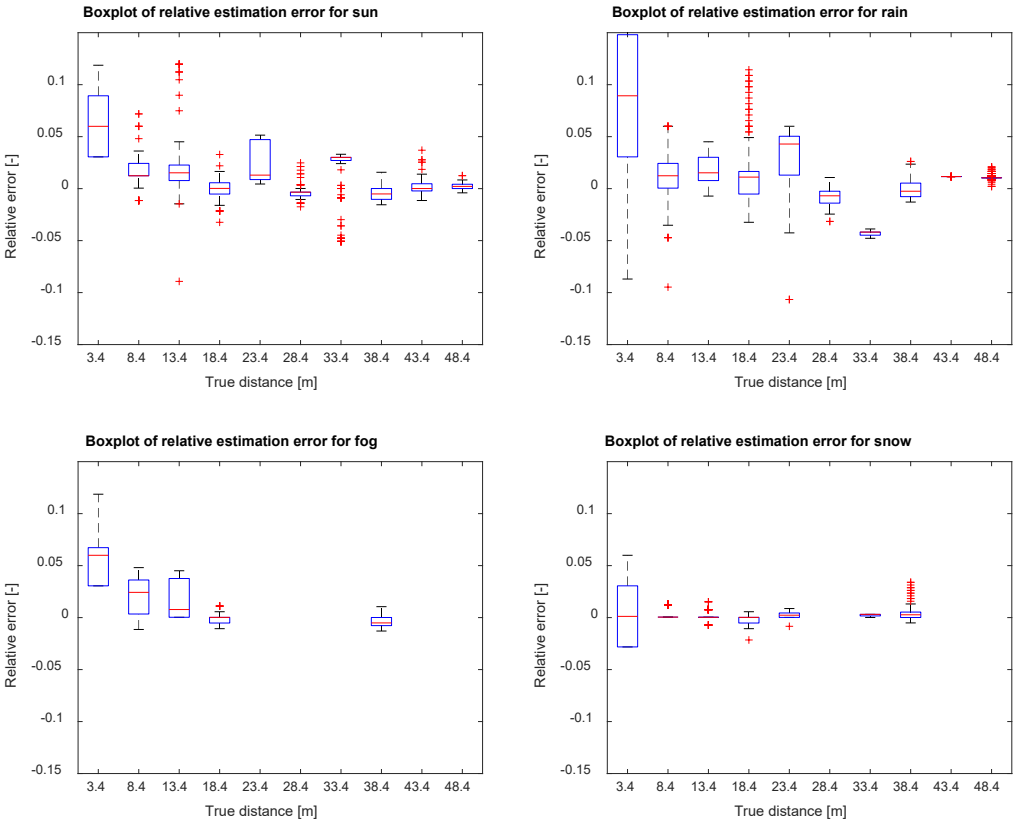

Figure 52: *Box plots of relative distance errors for different weather conditions of the ESR radar (upper left: sun, upper right: rain, lower left: fog, lower right: snow).*

Comparison with laboratory measurements

In our tests, the measured distance accuracy is well within the positioning accuracy of the pedestrian. For comparison, Figure 53 shows the results of the distance linearity measurements that were performed at METAS. In the laboratory measurements they observed a rather consistent overestimation of the distance estimate that increases from about 0.85 m at a distance of 10 m to 1.35 m at 170 m. This effect was not seen in the test-track measurements, where the distance estimate absolute error generally varied between ± 1 m (up to ± 2 m for some outliers) and no clear over/underestimation tendency was observed for reference distances varying between 2 and 50 m.

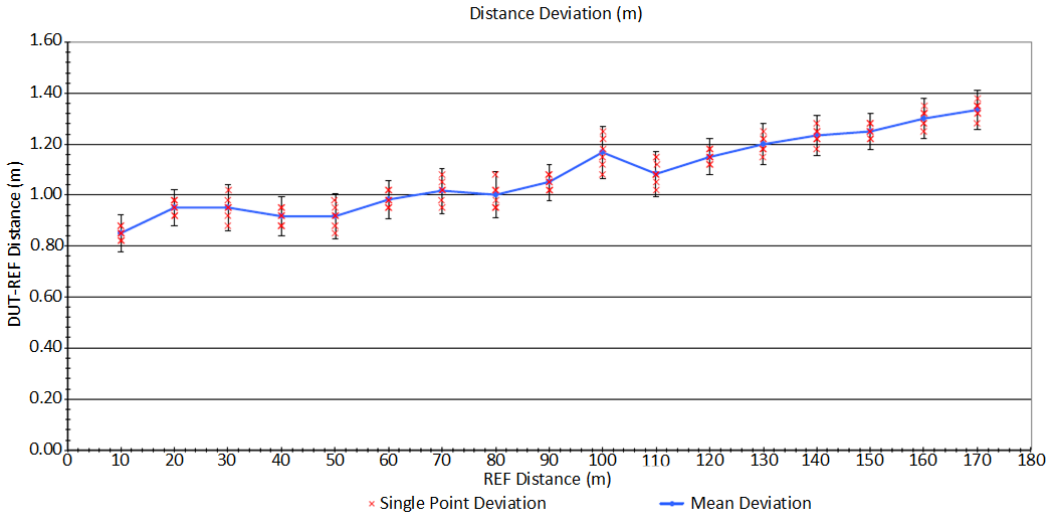

Figure 53: *Distance linearity evaluation at METAS (Source: METAS* **[28]***).*

As previously mentioned, we saw many spurious tracks of short duration in our real-world test runs with the radar. This matches the findings from the laboratory measurements at METAS, where they reported many issues with reflections, ghost targets and false detections.

Furthermore, they observed that the ESR radar is very sensitive, i.e. it filters out less detections compared to other similar radars they had previously tested. This was even more apparent when they performed tests with dynamic objects. For the same object, the sensor often indicated several different Track-IDs with short track-trajectories and sometimes also phantom target detections [28]. For example, for objects moving with a relatively low speed (up to 15 m/s), and mostly when moving away from the sensor, the radar detected a second phantom object moving parallel to the real object with double its speed [28]. According to METAS, this is a known phenomenon in radar technology, but radars designed for speed measurements normally filter out such phantom targets [28].

Figure 54 shows the results of the speed linearity measurements after filtering out the spurious and phantom targets. The laboratory experiments showed that the maximum measurable speed for an object approaching the sensor (direction approaching) is limited to 81.92 m/s, instead of the 100 m/s reported in the specifications. For objects moving away from the sensor (direction departing), speeds between 37 m/s and 40 m/s were always indicated as approximately 24 m/s (see Figure 55), which they think is caused by a radar firmware bug [28]. The laboratory test results also show that the specified speed accuracy of +-0.12 m/s seems to be rather utopian [28].

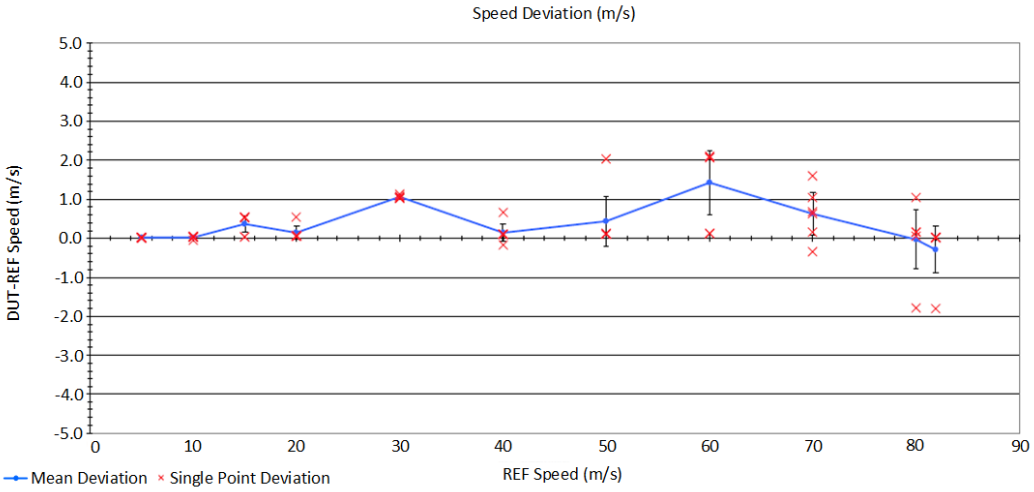

Figure 54: *Speed linearity evaluation at METAS for direction approaching (Source: METAS* **[28]***).*

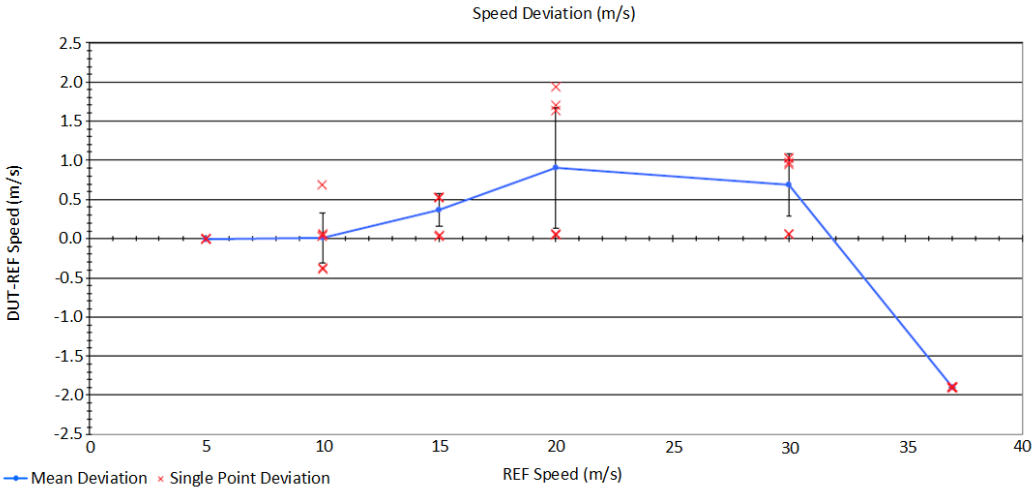

Figure 55: *Speed linearity evaluation at METAS for direction departing (Source: METAS* **[28]***).*

Road and building influence

In our tests for the Walking Pedestrian scenario we often noted more pronounced gaps in the tracks when the pedestrian was between 20 m and 35 m away from the vehicle. Our first conjecture was that the gap might be caused by a change in the road surface from asphalt to iron plates, which were installed to reinforce the surface, as our test-track area was previously used for storing construction machines. However, plotting all tracks from the Walking Pedestrian tests over an aerial image of our test track revealed that the tracks gap rather correlates to a narrowing of the test track caused by a metal garbage container and only loosely to the change in the road surface from asphalt to cast iron (see Figure 56). Note that the decrease of the number of tracks at the two ends of the test area is related to spatially shifted test runs and does not represent gaps in the data.

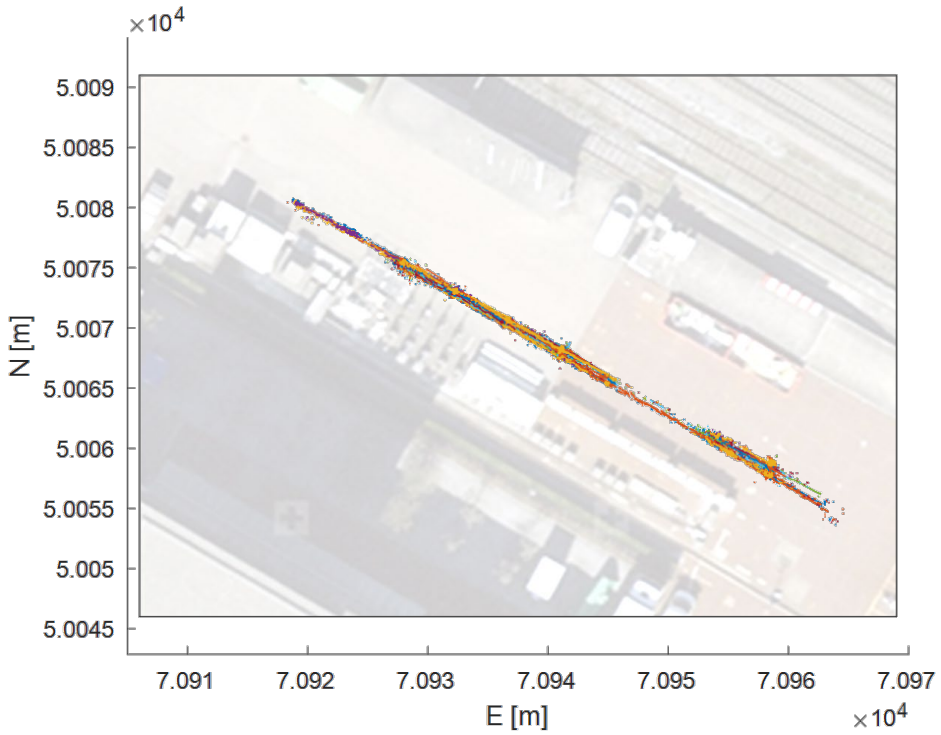

Figure 56: *Overlay of all tracks on an aerial image of test track section used for the waking pedestrian scenario (Source of aerial image: swisstopo.ch)*

In order to investigate our conjecture about the effects from the road surface and the road sides further, we repeated the Walking Pedestrian scenario in front of our laboratory, where the road is asphalt only (see Figure 57).

Figure 57: *Road in front of our laboratory (Source of aerial image: swisstopo.ch).*

We repeated our Walking Pedestrian scenario five times in this new setting. Again, we observed large gaps in the tracks at several different distances. Plotting all tracks together (Figure 59), we see clear gaps between 13 m and 27 m and between 38 m and 46 m. Moreover, in the region between 29 m and 35 m we observe a decrease in the number of tracks and a high number of track ID changes. These gaps occur in both walking directions (see also Figure 58), which indicates that they are most probably related to the environment. Because the road surface and the environment on the right side of the car are fairly homogeneous, we hypothesize that the gaps could rather be linked to changes in

the walls of the laboratory building on the left side of the car. The laboratory building has a large glass window with vertical T-shaped steel bars (see Figure 60) that starts exactly at the 13 m marker. These steel bars with a 90° angle can act as a retroreflector for radar waves, potentially overpowering the radar echo of the pedestrian, and could therefore cause the loss of the pedestrian track for distances above 13 m. The second gap (between 29 m and 35 m) occurs at the end of the laboratory building. Finally, the region with higher track losses corresponds to the garage door of the building which is around one meter set back from the wall. This set back again creates an area with 90° angles.

We also excluded the possibility of the missing detections being related to a saturation of the radar with respect to the number of tracks. According to the specifications [7], the radar can track a maximum of 64 targets simultaneously. Figure 61 shows the number of tracks during one test run, and shows that the radar never tracked more than 46 objects. If we consider all five test runs, the maximum simultaneously tracked number of objects was 52 with an average of 39.

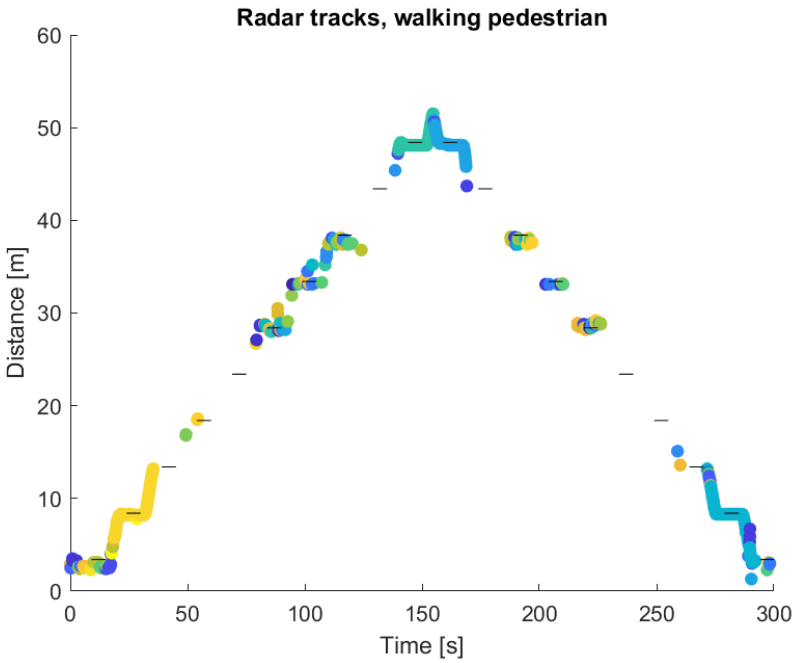

Figure 58: *Repetition of the Walking Pedestrian scenario at a different location (log 04 from 04.03.2021).*

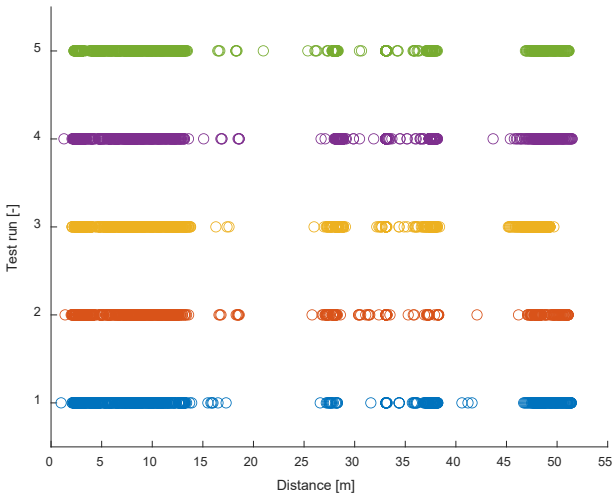

Figure 59: *Track distances plot with all repetitions of the Walking Pedestrian scenario at a different location.*

Figure 60: *Laboratory building on the left side of our repetition test track with windows and garage door. The yellow tape strips on the asphalt are the distances marks for the pedestrian.*

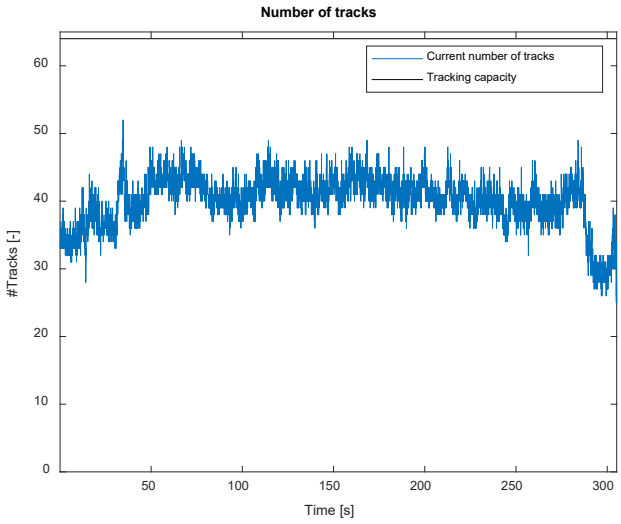

Figure 61: *Number of objects currently tracked by the radar (log 04 from 04.03.2021)*

Discussion

Our test results show that the distance estimates from the Mobileye in the Walking Pedestrian Scenario are very poor in all environmental conditions. As shown in Figure 44 and Figure 45, the distance is generally strongly underestimated (maximum error 28.8 m, maximum relative error 0.63, mean error 5.55 m, mean relative error 0.24) The absolute error and its variability among test repetitions seems to increase for higher distances, especially for foggy conditions.

If we look at the tracking results from all test runs (Table 3), we see that under good lighting conditions the tracking ratio of the Mobileye is close to 100 %. On the contrary, the tracking ratio reaches zero for bad lighting conditions, which was expected from the Mobileye specifications [13]. What is more interesting is that in adverse weather conditions we have a poorer tracking stability, i.e. the track is lost more often, the number of tracks created increases and the average track duration decreases. This shows that the track stability of the Mobileye camera clearly benefits from good light conditions present for sunny weather.

In contrast, the distance estimate of the Delphi ESR radar for the Walking Pedestrian scenario is much more precise (errors of about half a meter that lie within our positioning/ground truth precision) and we do not see any trends with increasing distances, nor effects from the adverse environmental conditions (Figure 51, Figure 52). Although the specification sheets of the Delphi ESR Radar report that the sensor only reports tracks, we see many very short tracks in our tests, which look more like single detections rather than filtered tracks. The tracking ratio for long tracks (i.e. tracks longer than 1 s) varies between 20 % and 85 %, with the majority of the tests having a tracking ratio of around 45 %. This means that most of the time the radar cannot track the pedestrian well, or at least not in a continuous manner. Also in terms of the tracking performance, we did not observe any significant effects from the environmental conditions, as expected from the sensors' working principle. Our findings from the real-world tests with the radar are overall in good agreement with the results from the laboratory measurements performed at METAS, where they also measured a high number or spurious tracks and phantom objects. Moreover, our real-world tests showed that there is a strong influence on the radar performance from the buildings and other large objects positioned along the road. In fact, objects with a large radar cross section, such as the T-shaped beams at our laboratory building or the metallic container in our test track, seem to disturb the radar when tracking the pedestrian. This seems to be the cause of the observed track loss of the pedestrian near such objects.

Looking at the results from both sensors, we can conclude that for pedestrian detection and distance estimate, neither the Mobileye camera nor the Delphi ESR radar have an acceptable performance when considered separately, as both sensors have clear and reproducible weaknesses under specific environmental conditions. Specifically, the Mobileye distance estimate is overall very poor and the tracking performance decreases with adverse weather conditions, while the radar, which provides precise distance estimates, shows tracking performance issues when there are buildings and other objects along the road. With our Walking Pedestrian scenario we could therefore show that a single sensor system alone is not reliable enough to work under all conditions and that each sensor has its own specific strengths and weaknesses. Therefore, in order to be safe, an AD system would greatly benefit from the combination of multiple sensor technologies and sensor fusion to reliably detect and track pedestrians.

6.1.4 Mixed Urban Scenario

In order to evaluate the object detection and distance estimation functionalities of an AD-vehicle perform during real world driving conditions, we need some realistic bust still repeatable experiments. The repeatability of the experiments allows us to compare the performance of the sensors during different environmental conditions and to compare the different sensors between each other. Moreover, we want to know the exact positions of the tested obstacles relative to the AD-vehicle so that we can analyze performance changes at different distances, angles or vehicle velocities. With this in mind, we developed the so called "Mixed Urban" scenario, which aims to represent an urban environment. The objects in this scenario are stationary and we know their exact position.

In this section, we will first describe and illustrate the experimental setup of the Mixed Urban scenario. Then we will describe the evaluation procedure for detection and distance estimation with the Mobileye camera and ESR radar. Finally, the results are presented and discussed.

Scenario Description

In this scenario, the AD-vehicle drives with a constant speed of 15 m/s down the front straight of our test tracks. Four mannequins (two female, one male and a child) are installed along the front straight. More details about the mannequins are given in Section 5.4. At the end of the front straight, there is a stationary van (Iveco Daily) (Figure 62 and Figure 63).

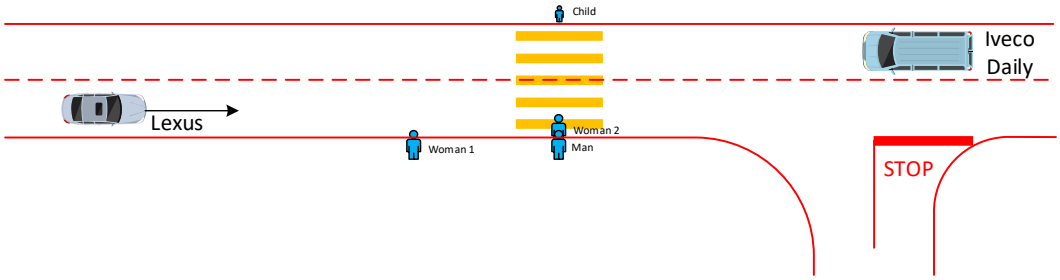

Figure 62: Mixed Urban scenario setup with pedestrian and vehicle.

Figure 63: *Front straight with AD-vehicle and mannequins during a rain test.*

Data processing and analysis

In the Walking Pedestrian scenario we showed how the tracking functionality can be evaluated with the presented tracking ratio. In the Mixed Urban scenario we evaluate the detection capabilities of the black box sensors Mobileye and ESR radar through the analysis of the false negative rates.

False Negative Rate

As mentioned in Section 6.1.1, the Mobileye camera and the ESR radar are black box systems. They provide already processed data in the shape of tracks. Nevertheless, to evaluate these sensors for AD we need to have an understanding of how well the sensors are detecting objects. The performance of the sensors for object detection depends on the AD task and the operational design domain (ODD). In our Mixed Urban scenario, the AD-vehicle is driving with a constant speed and needs to detect possible obstacles on the road and stop before bumping them. Now for this scenario and task, we are fitting our performance evaluation metrics.

In the literature, confusion matrices are often generated to evaluate the object detection algorithms [29]. A confusion matrix allows visualizing the performance of a statistical classification algorithm in the shape of a table or matrix. Thereby, each row contains the instance of the predicted class and the columns contain the instances of the actual or true class. To build such a confusion matrix, an annotated test data set is needed. This test data set is used to evaluate the algorithm performance. In our case, we cannot create such a confusion matrix since we only have processed detections in the shape of tracks from the Mobileye camera and ESR radar sensors, as described in Subsection 6.1.1 therefore, we adapt our sensor evaluation to tracks.

As a result of the design of our experimental setup, we know when the test objects are in the FoV of the evaluated sensors and we know their position at each time step. Thus, we discretize the time when the object is in the FoV of the sensor T_{FoV} into time intervals t_i. and we require that in each time interval the object needs to be detected (or tracked). If the object is not detected, it is a false negative (FN), otherwise it is a true positive (TP). The true positive rate (TPR) is defined as the ratio between the TPs and the sum of the TPs and FNs, as shown in Equation 1. The false negative rate (FNR) can then be computed as 1 minus the true positive rate, as shown in Equation 2 The time interval is a tuning parameter and needs to be reasonable for the sensor type and vehicle velocity.

$$TPR = \frac{TP}{TP + FN}, \tag{1}$$

$$FNR = \frac{FN}{TP + FN} = 1 - TPR. \tag{2}$$

For the sake of evaluating the sensor performance in urban driving, we need to know more about the TPR or FNR. We would like to know what is the FNR at different object distances, for different objects types, for different object sizes or at different environmental conditions. Those dimensions give us a better understanding why certain sensors are better for certain AD tasks. For example, if we consider pedestrian detection, then the sensor performance should be better for close distances then for long distances because close distances are more safety critical. A FNR on some random test data will not allow us to decide if the sensor performance is good enough without this additional spatial information. Moreover, maybe the sensor is detecting perfectly adult pedestrians but it is bad for detecting children. Thus, we evaluated the FNR of the Mobileye camera and Delphi radar with our Mixed Urban scenario for different object distances, object sizes, and environmental conditions.

In the case of FNR dependent on distance, we discretize a certain sensing range $\mathrm{D}_{\max}$ into equally spaced points p with point gap d_p. Then, for each time interval t_i, we calculate the distance to the object and match the distance to the corresponding point. For each matched point p, we add an object appearance n_{pa} to the point $n_{pa} = n_{pa} + 1$ and if there was a detection, also a detection $n_{pd} = n_{pd} + 1$. To avoid sparse data and to increase the number

of appearances per point, we match the detections and appearances of a point around a distance interval d_i.

$$FNR(c) = \frac{n_{cd}}{n_{ca}}. \quad (3)$$

Equation 3 gives us the FNR for all object distances d between $p \cdot d_p - d_i \leq d < p \cdot d_p + d_i$. For the Mobileye camera, we calculated the FNRs for detecting the Iveco Daily vehicle, adult pedestrians, a child pedestrian, all pedestrians (three adults and one children) and all objects (vehicle and pedestrians) for all environmental conditions (explained in Section 5.5). For the ESR radar we only considered the Iveco Daily, since the radar is an active sensor affected by the object material, and therefore the human-like mannequins are not realistic enough for a comprehensive evaluation.

For each environmental condition, we repeat the measurements in the Mixed Urban scenarios at least 5 times. The FNR highly depends on the chosen time interval t_i, point distance d_p, number of measurements and vehicle velocity. Since we are limited by time, manpower and computational recourses, we could not collect enough data to have fine gridded FNRs. Our data is to certain degree sparse and uncertain. Nevertheless, to evaluate the FNR dependence on the distance, we used Gaussian process regression with bounded constraints to fit the FNR curve from our generated data. The advantage of Gaussian process regression for fitting the curve is the obtained uncertainty for the fitted curve. Constrained Gaussian process regression is applied since the rates must be by definition bounded between zero and one. We used the algorithm presented in [30] to generate the FNR curves as a function of the distance with constraint Gaussian process regression.

Distance

Similar as in the Walking Pedestrian scenario, we also evaluate the distance estimation accuracy for different object distances. This is a relevant information, because it tells the AD-vehicle where the test object could be according to the sensor uncertainty once a detection is obtained. Note that, in this scenario, we do not evaluate the performance of the radar in distance measurement. The reason is simply that the true distance to a large 3D-object such as a van is not unique, and therefore not easy definable. Whether the radar detection used in the tracking of a truck results from the closest reflection, some reflection above some threshold, or even from the strongest one, is unknown. Since the evaluation of the distance accuracy is easier and better performed at the METAS-laboratories, we focus on the detection performance.

Data Annotation

The position of the AD-vehicle is determined using the lidar localization algorithm and is given in the map frame coordinates. Since the mannequins and the Iveco Daily are static, their position is known at all time steps.

The Mobileye measurements are annotated manually. For the annotation, we plot the Mobileye tracks with their ID in the map frame and then manually decide to which object the IDs belong (e.g. woman 1, child, Iveco Daily or other not relevant targets on the test track). Figure 64 shows all tracks of the Mobileye camera for the annotation process and the annotated tracks.

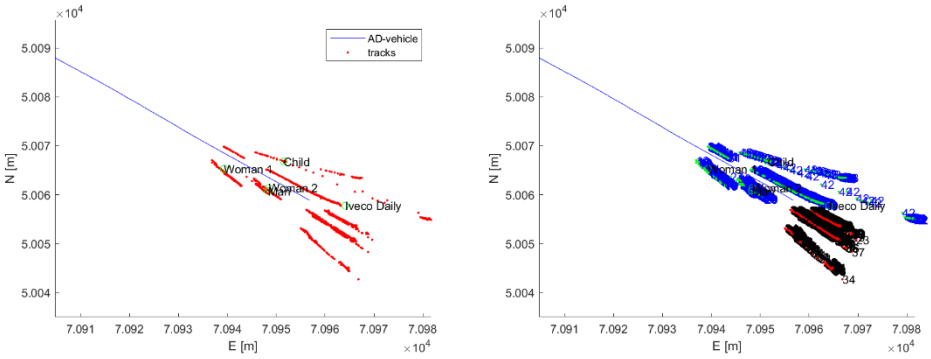

Figure 64: *Left: All Mobileye tracks in the map frame from one run of the Mixed Urban scenario (log 03 from 28.9.2020). The green circles are the true locations of the targets and the blue line is the AD-vehicle trajectory. Right: Manually annotated tracks with target objects in blue and not relevant targets (e.g. parked cars from Empa employees) in black (log 03 from 28.9.2020)*

The radar is more precise in terms of positioning the objects in the map frame and can be annotated with less manual effort. Figure 65 shows all tracks from one test run plotted in vehicle coordinates. The wide and long range FoVs are indicated with green cones. Figure 65 also shows the same tracks but transformed into map-coordinates. The trajectory of the AD-vehicles is visualized with the blue line and the true positions of each test object are marked with green circles. The black bar indicates the wall of the building at the end of the test straight. As can be seen, many track seem to be inside or even behind the building, which is linked to scattering and reflections in the building.

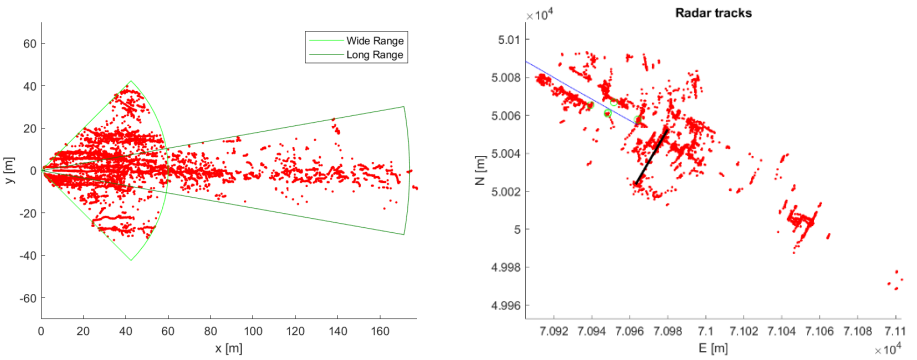

Figure 65: *Left: All radar tracks of one run in vehicle coordinates with the wide and long range FoVs (log 06 from 30.9.2020) Right: AD-vehicle trajectory (blue line), the true object positions (green circle) and radar*

Since mannequins have different radar cross sections than humans, we do not evaluate the performance of the radar to detect the mannequins. Thus, we only consider the detection of the Iveco Daily van for the radar. For this, we filter all tracks based on the Region of Interest around the known positon of the van. Figure 66 shows all the tracks in

the region of interest for the Iveco Daily. Each color represents a different track ID. The right plot illustrates the same tracks plotted against the true distance to the Iveco. It gives a first impression of the overall track quality, the track stability and track errors (ID changes, FNs and double tracks). For the evaluation of the FNR we only care if there is a measurement (track) of the object at a certain time step or not, and therefore, the track stability is not reflected in the FNR.

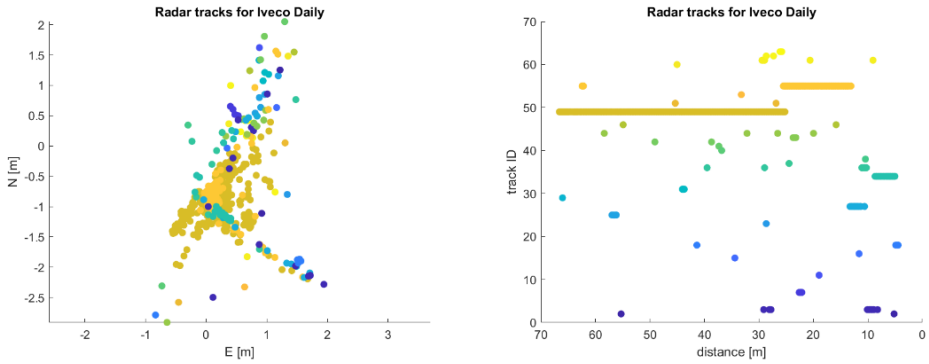

Figure 66: *All radar tracks which belong to the Iveco Daily in maps-coordinates. Different colors indicate different track IDs (log 06 from at 30.9.2020). Right: All track IDs from one test run over time. Same colors indicate same track ID (log 06 from 30.9.2020).*

Results

In this section we present the results from the Mixed Urban scenario. First, we show the results of the Mobileye camera. The FNR and the distance estimate accuracy is presented as a function of the distance for different object types, sizes and environmental conditions. In the second part, the results of the ESR radar are shown. In this case we only consider the FNR curves for the Iveco Daily van for different environmental conditions. The results are discussed at the end of this section.

Mobileye

Figure 67 to Figure 70 show the Mobileye FNR curves for pedestrian detection (adult and child pedestrians) as a function of distance for sunny, rainy, foggy, and snowy weather conditions (see Section 5.5 for the description of the weather conditions), respectively. Contamination is not considered for the Mobileye camera since it is mounted behind the windscreen which can be cleaned continuously. The mean FNR curves are indicated with green lines and the 95 and 5th percentiles with the two black lines, respectively. The grey area shows the uncertainty of all possible FNR curves. The blue dots are the gathered FNRs from the different experiments at different distances which are used to fit the FNR curve with Gaussian process regression. The red triangles are the test data used for the validation of the FNR curve. The test data is generated from an independent experiment in order to avoid dependence between train and test data. Since we have 5 experiments per environmental condition, the test dataset makes 20% of the whole dataset. The lower the green curves are, the better is the sensor performance. The bigger the grey area is, the higher is the uncertainty for the fitted FNR. If we compare the four conditions, the FNR curves show the best sensor performance for sunny (Figure 67) and foggy (Figure 69) conditions, since the curves are always lower (or at least equal) than those of the other conditions. The FNR for rainy conditions (Figure 68) is better than for snow (Figure 70) for distances between 15 and 25 m. For higher distances, the snow curve shows a lower increase of the FNR than the rain curve. All curves have a parabolic shape with a global minima. Only for fog and snow conditions, the curves show a second local minima at around 35 to 40 m. These results are discussed in detail in the next section.

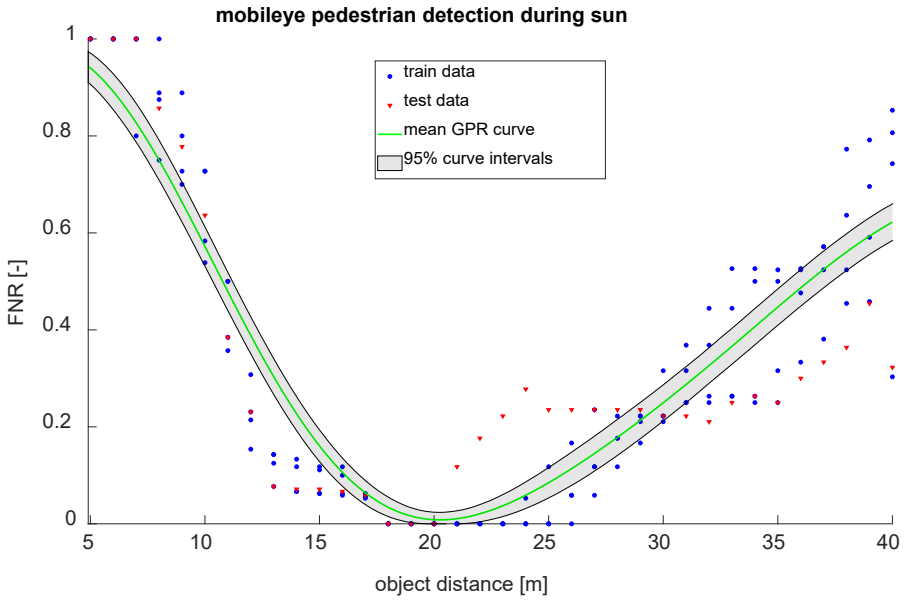

Figure 67: *Mobileye pedestrian detection FNR as a function of the object distance for sunny weather conditions.*

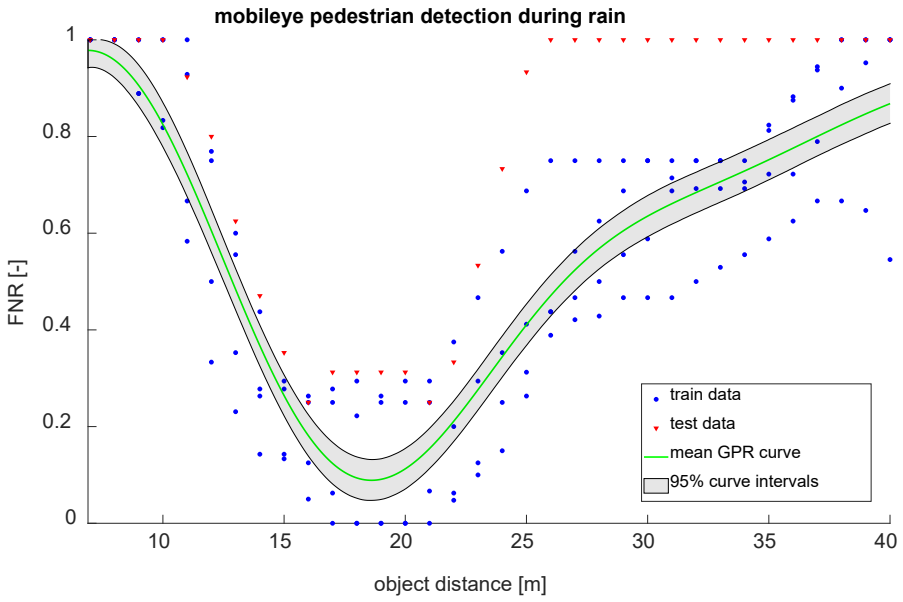

Figure 68: *Mobileye pedestrian detection FNR as a function of the object distance for rainy weather condition.*

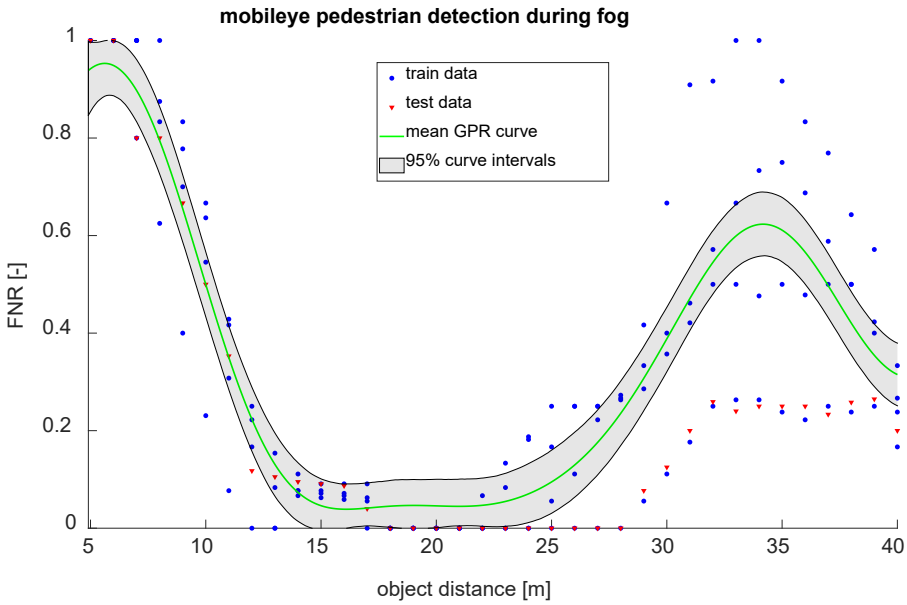

Figure 69: *Mobileye pedestrian detection FNR as a function of the object distance for foggy weather condition.*

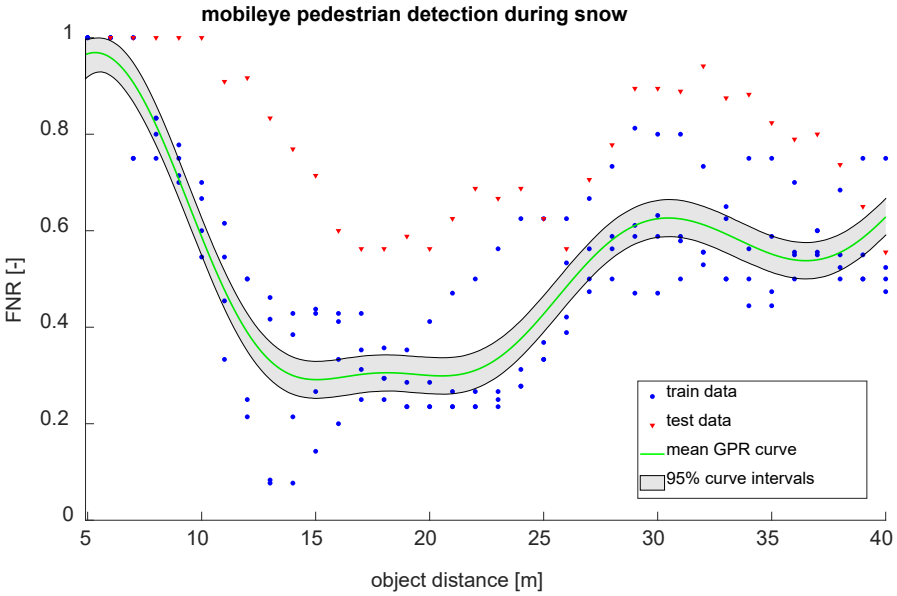

***Figure 70**: Mobileye pedestrian detection FNR as a function of the object distance for snowy weather condition.*

In Figure 71 and Figure 72 we compare the FNR curves of the Mobileye camera for adult and child pedestrian detection during sunny weather conditions. Figure 73 and Figure 74 show this comparison for snowy conditions. Again, the green line indicates the mean FNR function from Gaussian process regression, the grey area (95 and 5th percentiles) shows the fit uncertainty, and the blue dots and red triangles indicate the training and validation data, respectively. These results show that the Mobileye camera performs better for adult pedestrian detection than for child pedestrian detection, since the mean FNR curve of the adult pedestrian is lower for all distances. Moreover, the detection of the child pedestrian is strongly affected by the adverse weather conditions, as for the snowy weather conditions the child mannequin was never detected (Figure 74).

Figure 75 shows the Mobileye camera FNR curve for vehicle detection during different weather conditions. We can see that the FNR curve during sunny weather is the best since it is the lowest one for all distances (constantly 0). During fog condition there are data points which go down from 1 to 0.4 between 10 and 15 m. Whereas the test data and the majority of the training data is at 0 for those distances. After 15 m the data points show similar behavior as during sunny weather. During rainy conditions in the upper right plot, the FNR starts to increase from 0.1 at 30 m all up to 0.6 at 40 m. The test data points start to increase already at 15 m which lead to a high test loss of the fitted FNR (far distance between fitted FNR curve and test data). The snow weather curve shows the highest FNR. It starts already with 0.4 FNR at 7 m and goes all up to 1 at 30 m. If we compare the upper left curve during sunny weather conditions with the curve for pedestrian detection during sunny weather in Figure 67, we clearly see that the Mobileye camera detects vehicles better than pedestrians, as the FNR curve for the vehicle detection is lower than the FNR curve for pedestrian detection for all distances.

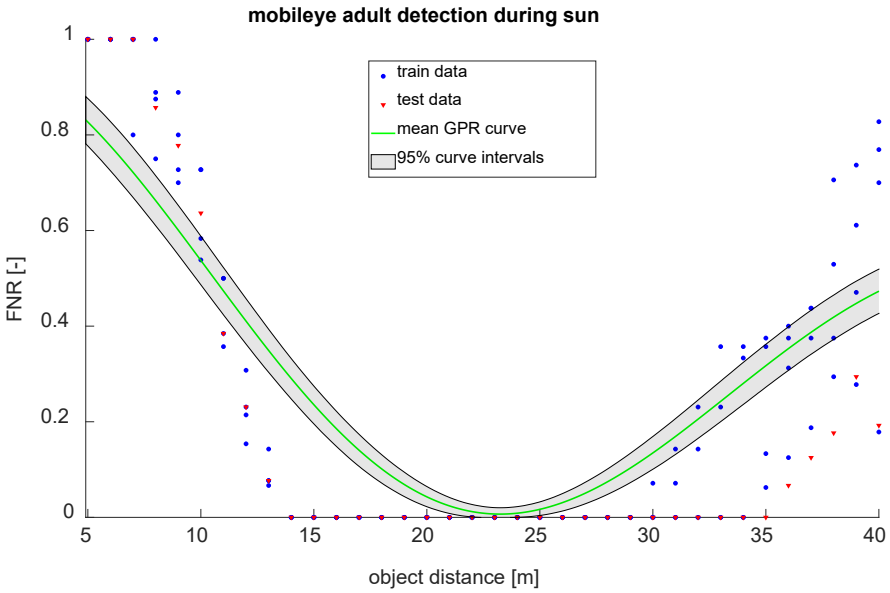

Figure 71: *Mobileye adult pedestrian detection FNR as a function of the object distance for sunny weather condition.*

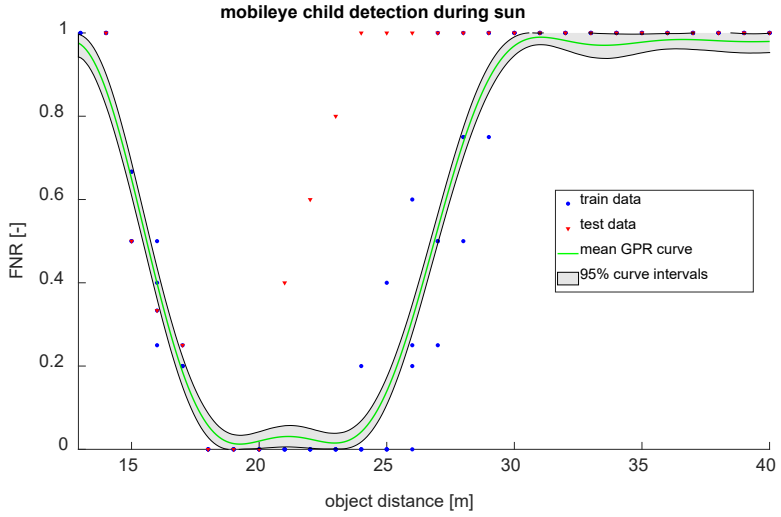

Figure 72: *Mobileye child pedestrian detection FNR as a function of the object distance for sunny weather condition.*

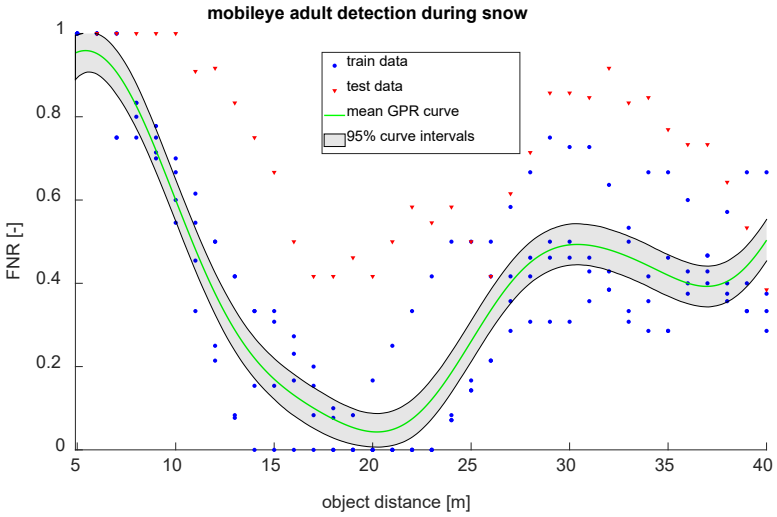

Figure 73: *Mobileye adult pedestrian detection FNR as a function of the object distance for snowy weather condition.*

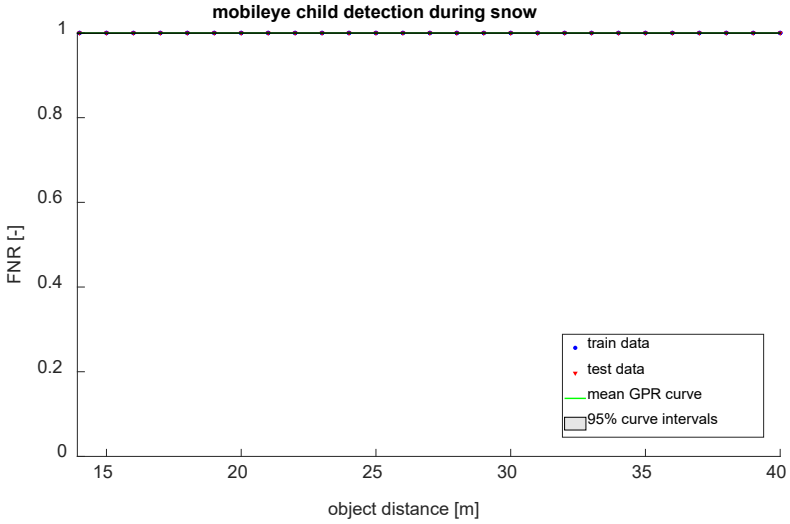

Figure 74: *Mobileye child pedestrian detection FNR as a function of the object distance for snowy weather condition.*

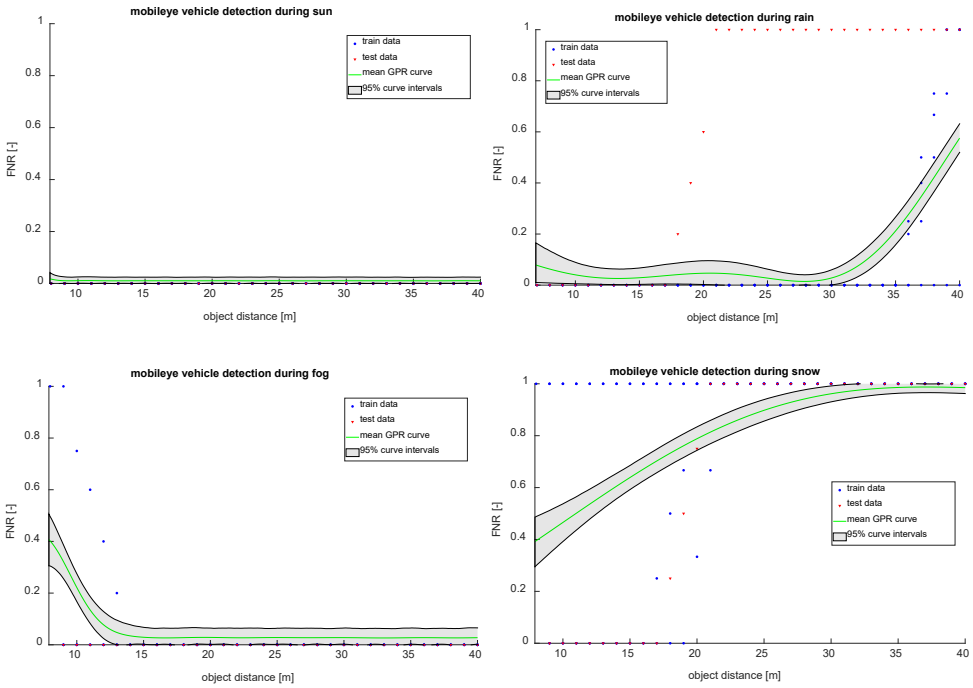

Figure 75: *Mobileye vehicle detection FNR as a function of the object distance for different environmental conditions (upper left: sun, upper right: rain, lower left: fog, lower right: snow).*

In Figure 76 the absolute error distribution for pedestrian distance estimation is presented with box plots for sunny, rainy, foggy and snowy weather conditions. The median is almost constant at 3 m over all distances, but the variability of the absolute error tends to increase with the distance. The absolute error is mostly positive, i.e. the Mobileye camera mostly detects the pedestrian too close.

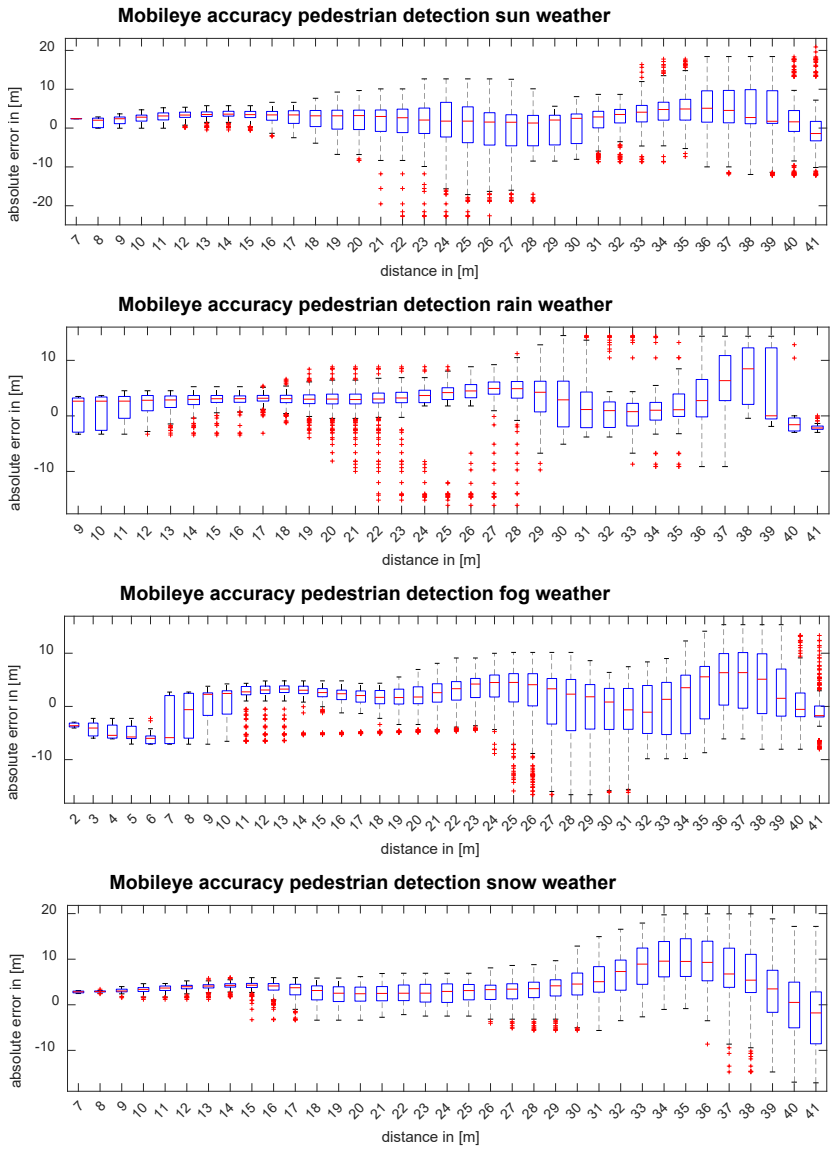

Figure 76: *Mobileye pedestrian distance estimation accuracy as a function of the object distance for different environmental conditions.*

The distance estimate accuracy for children is shown in Figure 77. The median absolute error is between 0 and 1 m for distances up to 19 m. From 20 to 28 m the child pedestrian is detected too far away, as indicated by the large negative absolute errors.

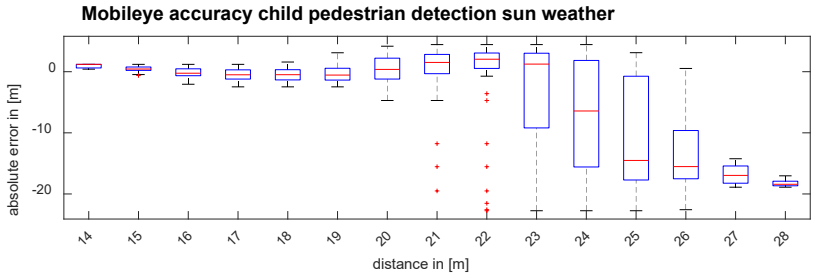

Figure 77: *Mobileye child pedestrian distance estimation accuracy dependent on object distance at sunny weather.*

In Figure 78 the accuracy for estimating the distance of the Iveco Daily at different distances is presented with box plots. We recognize a somehow linear increase of the absolute error for the sunny and foggy conditions. For rainy and snowy conditions, we see a steeper increase of the absolute error, i.e. at 27 m the absolute error is around 8 m, while for sunny and foggy conditions it is only 4 m. Apart from some outliers, the absolute error is mostly positive, indicating that the Iveco Daily is mostly detected too close.

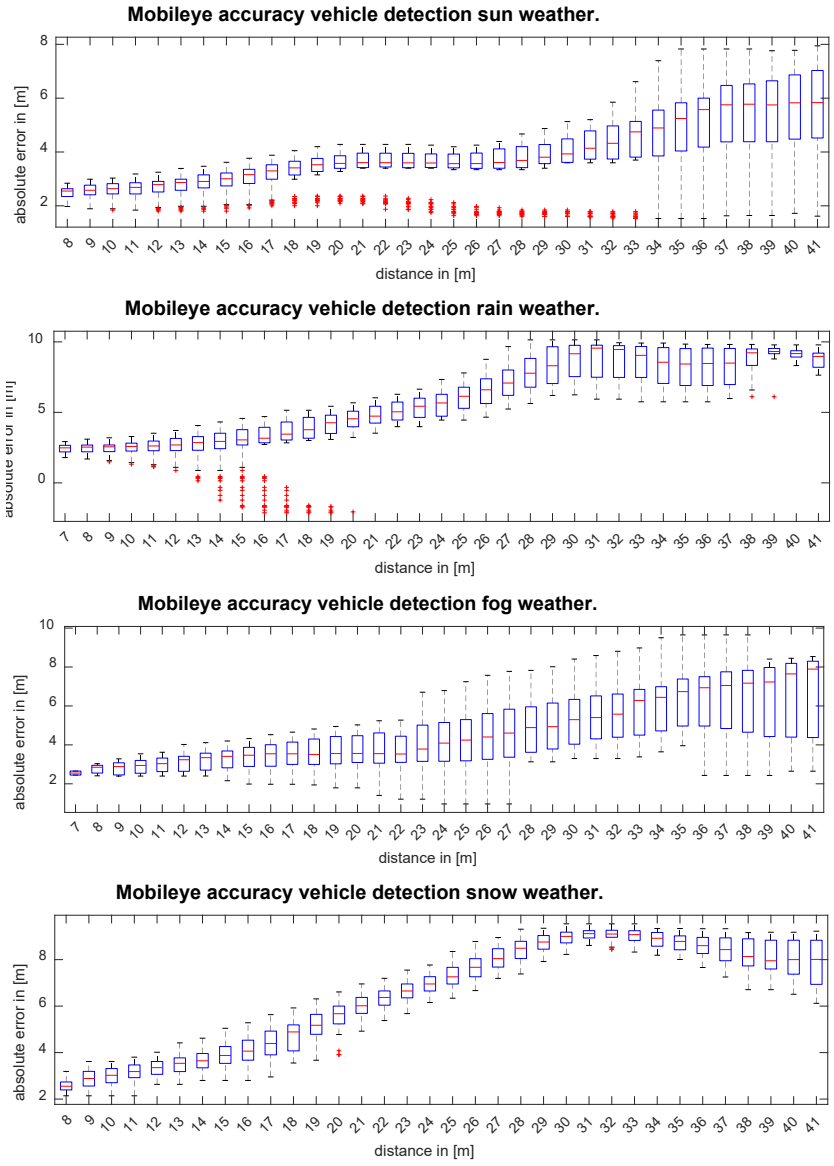

Figure 78: *Mobileye vehicle distance estimation accuracy as a function of the object distance for different environmental conditions.*

ESR Radar

In Figure 79 the FNR curves for vehicle detection of the Delphi radar at different environmental conditions are illustrated (upper left: sun, upper right: rain, lower left: fog, lower right: snow). As for the Mobileye FNR curves, the blue dots are the measured FNRs used for the training of the Gaussian process regression, the red triangles are the test data used for model validation, the green curve is the mean curve from the Gaussian process regression, and the grey area shows the fit uncertainty. We can see that the radar sensor efficiently detects the Iveco Daily at all distances and is weather resistant, since the FNR is close to zero for all distances and weather conditions. In certain cases the FNR curves show a very small increase at very small and large distances (10<d<60) and they become slightly more uncertain for rainy and snowy weather, but the effect is very small.

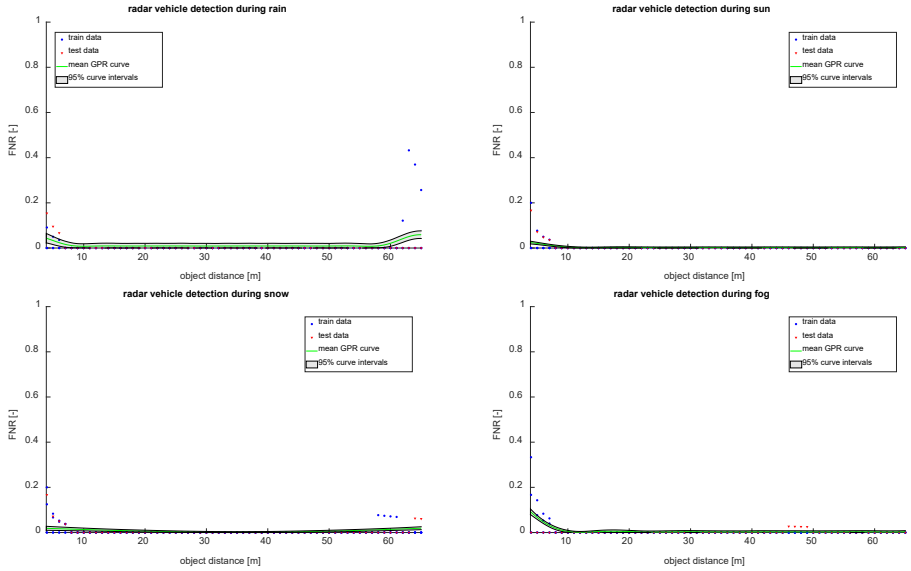

Figure 79: *Delphi radar vehicle detection FNR as a function of object distance for different environmental conditions (upper left: sun, upper right: rain, lower left: fog, lower right: snow).*

Figure 80 shows the FNR curve from the measurements with a contaminated radar. If we compare it with Figure 79 it seems that contamination has a negative influence on the radar performance since the FNR curve of the contaminated sensor strongly increases for distances above 40 m. The train and test data are coming from 5 experiment repetitions and all of them show this effect. We do not know the exact cause of this discovery and it is discussed in the next subsection.

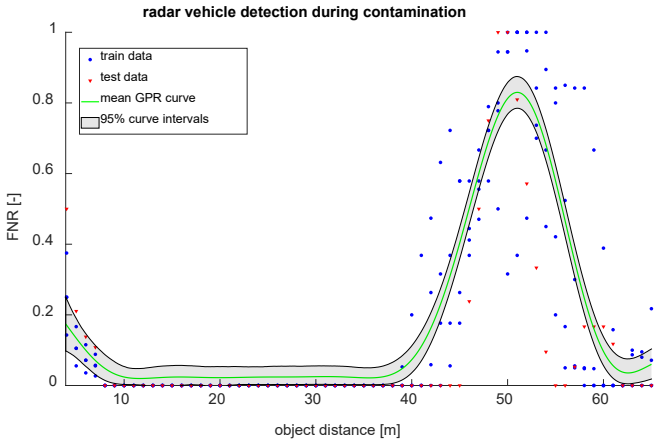

Figure 80: *Delphi radar vehicle pedestrian detection FNR as a function of the object distance for contaminated condition.*

Discussion

Our results show that the performance of the Mobileye camera for object detection strongly depends on the object type/size and the weather conditions.

For pedestrian detection (Figure 67 to Figure 70), we observe the best performance (i.e. lower mean FNR curve for all distances) for sunny and foggy weather conditions. As reported in literature [9], the performance of the Mobileye camera is expected to be the best for sunny conditions, as the passive camera sensor benefits from the good light conditions and the images are not affected by the noise created from water droplets, snowflakes or fog. The similar performance obtained for the foggy conditions as for the sunny conditions was rather unexpected. Also in the Walking Pedestrian scenario reported in Section 6.1.3, the fog did not influence the Mobileye tracking performance, since the tracking ratio was almost 100% for all experiments. The experiments for the Walking Pedestrian and Mixed Urban scenarios during foggy weather were conducted on the same day with a time difference of a couple of minutes, which means the fog density was more or less the same. We assume that the relatively good performance of the sensor for the foggy conditions might be due to the fog during our experiments not being thick enough to have an influence on the camera images and object detection for distances up to 50 m. On the contrary, higher FNRs were obtained for all distances for rainy and snowy conditions, as expected from the additional noise created by the drops/flakes in the camera image, which difficult the object detection.

In terms of the overall performance, the Mobileye shows a rather weak pedestrian detection performance, because the FNR over all weather conditions and distances is rather high. Interestingly, all curves show a parabola shape with a global minimum mostly between 15 and 25 m. The reason for increasing FNRs with the distance is that the camera gradually captures less information from the objects, i.e. the higher the distance, the lower the amount of pixels from the object in the image, which makes it harder to detect it. The lower performance for smaller distances is less intuitive, because closer objects should have more pixels. We assume that the reason for the decreasing performance at smaller distances might be related to the higher relative tangential velocity between the camera and the object. This higher tangential velocity, created by the motion of the vehicle, can produce motion blur in the camera images. Motion blur distorts objects in an image, which makes it harder to detect them [31]. Unfortunately, since we cannot access the raw images produced by the Mobileye camera, we cannot prove the occurrence of the motion blur at smaller distances, but the combination of these two effects could explain the observed parabolic shapes. For the sunny and rainy conditions, the FNR is close to 0 at the local minima between 15 m and 25 m, and increases to 0.8 or even 1 for distances below 15 m or above 25 m. For foggy and snowy conditions, we observed a second local minimum at 40 m and 37 m, respectively. This is against our expectation that for higher distances the object detection performance should continuously decrease. We could not determine yet what the reason for these two minima are. This shows again the difficulty of evaluating the performance of the sensors, as it can be influenced by many parameters. We assume the second local minima could be related to the bumpy road of our test track. In fact, during different experiments we observed that movements in the vertical direction can have an influence on the sensor measurements. Further, the FNR curve is probably over fitted from our data and different models or hyper parameters should be used. Nevertheless, our results show how different unknown effects can influence the sensor performance. We are planning to repeat these experiments in the future to try to explain the additional local minima.

Besides weather effects, we discovered also differences in the object detection performance of the Mobileye camera to detect same object types with different sizes. In Figure 71 to Figure 74 we compare the FNR curves of the Mobileye camera for detecting adult and child pedestrians for sunny and snowy weather conditions. In both weather conditions the performance for adult pedestrian detection is significantly better compared to child pedestrian detection, as the FNR curve for the adult pedestrian detection is lower than that of the child pedestrian detection for all distances. The biggest difference is observed for the snowy weather conditions, for which the child pedestrian was never detected, i.e. the FNR curve equals 1 for all distances. Again, several factors might play an

important role here. From the camera perspective, a child is smaller than an adult, and consequently it appears smaller in the image. This means that there are less pixels or less information about the object. From the algorithm perspective, with learning based methods, the reason could be that the training data set is not balanced enough for the different possible pedestrian size distributions.

In Figure 75 we compare the Mobileye camera FNR curves for vehicle detection during sunny, rainy, foggy and snowy weather conditions. During sunny weather conditions the Mobileye camera shows the best performance since the FNR is around 0 for all distances and the FNR is the lowest one compared to the other conditions. The bright illumination conditions without additional noise from environment such as water and snow are beneficial for vehicle detection. Then during fog conditions the Mobileye camera shows similar performance as during sunny conditions. Only at distances between 7 m and 15 m exist data points greater than 0 which leads to a higher FNR curve. The data points start at 1 and decrease linearly down to 0 at 15 m. We do not know the reason for those data points but the majority of the train data points and all the test data points are always at 0 for all distances. It indicates again the complexity of vehicle detection evaluation but also that more experiments are needed for the evaluation. Then during rainy conditions, the FNR is rather low as during fog and sun but with more uncertainty. At around 30 m the FNR starts to increase all up to 0.6 at 40 m. It indicates that water droplets from the rain have an influence on the vehicle detection and that the influence is higher for larger distances. Further, the illumination conditions are rather dark which makes it harder for a camera to detect objects at far distances (40 m and plus). Interestingly, the test data starts increasing already at 15 m and increases quickly all up to 1 at 21 m and stays 1 for larger distances. The reason for the difference in the test data and train data is not clear to us. In such a case using a random test data would make more sense. Nevertheless, it shows how unreliable the Mobileye camera is, because single experiments can show different behavior as the others. Finally during snow weather, the FNR starts at 0.4 at 7 m and goas all up to 1 at 30 m. The Mobileye camera shows the worst performance for vehicle detection during snow conditions since the FNR curve is the highest for all distances. It seems that snowflakes have a higher impact than raindrops. As for pedestrian detection, the Mobileye camera shows a weather dependent performance for vehicle detection.

Additionally, we also analyzed differences in the Mobileye camera performance to detect different object types, such as pedestrians and vehicles. In Figure 75 it can be seen that the FNR curve for vehicle detection during sunny weather in the upper left plot is much lower than that for pedestrian detection in Figure 67. This indicates that, in sunny weather conditions, the Mobileye camera detects better vehicles than pedestrians. The same was found to be true for foggy and rainy environmental conditions. There could be several explanations for this effect. On the one hand, the Iveco Daily is much bigger than pedestrians and therefore the camera images contain more useful information for its detection. Moreover, bigger objects might be less sensitive to sensor noise or environmental noise originated from the rain or fog. Further, the motion blur of the camera connected to the movement of the AD-vehicle has a higher impact on smaller objects. Finally, vehicles might have easier machine understandable shapes. In the future we plan to make experiments also with smaller vehicles to have a better distribution of possible vehicles. Only during snow it is hard to say if the Mobileye camera detects better vehicles or pedestrians, since the FNR for vehicle detection is lower for distances between 7 m and 15 m and after 15 m it is higher. Generally during snowy conditions the Mobileye camera has rather high FNRs (larger than 0.4) for vehicle and pedestrian detection. The reasons could be the rather dark illumination conditions and noise coming from snowflakes. Since the environment influence during snow is quite random (random snowflakes in the image), it is hard to tell why now the FNRs are similar for pedestrian and vehicle detection during snowy weather.

Finally, we analyzed the accuracy of the distance estimate of the Mobileye camera for pedestrian and vehicle detections. The median absolute error for the pedestrian distance estimation (Figure 76) stays almost constant at around 3 m for all distances and environmental conditions. The pedestrians are most of the time detected too close (positive error) and the spread of the absolute error increases with the distance. However, in the case of the child pedestrian (Figure 77), we observe that the accuracy of the distance

estimation is very bad for distances above 20 m, and that the child pedestrian is mostly detected too far, most probably because it is smaller than the Mobileye pedestrian model. In Figure 78 we show the absolute error of estimating the distance to the Iveco Daily with the Mobileye camera. The absolute error tends to increase linearly for larger distances. For rainy and snowy weather conditions, the absolute error increases faster, e.g. at 20 m distance the median absolute error for sunny weather is 3 m and for snow 6 m. Moreover, the variability of the absolute error significantly increases for larger distances, e.g. for sunny conditions at 7 m the difference between maximum and minimum absolute error is around 0.5 m and for 40 m distance it is around 8 m. The absolute error is mostly positive, which could be linked to the fact that the Iveco Daily is a rather big vehicle (most probably bigger than the vehicles used to fit the Mobileye model) and therefore the distance estimation algorithm estimates the vehicle to be closer. Such data distributions can be used to fit an accuracy model for distance estimation, which provides the uncertainty distribution of the distance estimation.

The ESR radar shows a quite weather independent performance for vehicle detection, as the FNR curves for the four different environmental conditions (sun, fog, rain and snow) are very similar (Figure 79). In particular, the mean FNR curve is almost always zero, i.e. the vehicle was detected at all times. Only for snowy weather conditions we see a slightly higher uncertainty in the FNR curve for longer distances (from 50 to 70 m). This uncertainty comes from data points between 0 and 0.01, i.e. there were some FNs during these measurements. All in all, the radar shows quite a weather resistance for vehicle detection, as expected from the literature [9]. This is because radars are active sensors (no influence from lighting conditions) and their wavelength is big enough to not to be not reflected by snowflakes or rain drops.

Interestingly, we discovered some drops in the radar detection performance when the sensor is contaminated. The FNR curve for vehicle detection with contaminated radar (Figure 80) shows that the FNR is increasing from 0 to 0.9 for distances between 40 and 50 m before decreasing again for distances between 50 and 60 m. The reason for this effect is not clear to us. Probably the dirt, leaves and tire wear have some influence on the radar reflectivity. Moreover, it could be that for distances between 40 and 60 m a leaf fully or partially covered the FoV of the radar exactly in the direction of the vehicle. We plan to repeat these experiments in the future to further investigate the effects from contamination.

Finally, we compared the performance of the Mobileye camera and ESR radar. In particular, the biggest difference can be seen during the hardest environment condition (highest FNR curve), which is snow. The FNR curve of the Mobileye camera in Figure 75 during snow is much higher than the FNR of the ESR radar in Figure 79 during snow. This means that the ESR radar is more reliable and more weather resistant for vehicle detection. The difference in the performance is quite huge, as the radar FNR curve stays very close to zero for all distances, whereas the FNR curve of the Mobileye camera goes from 0.4 and up to 1 at a distance of 40 m. What makes this is even more impressive, is that the radar was evaluated on a much smaller time interval. For the radar the sensor frequency of 20 Hz was used for the time interval because the radar is taking periodic measurements. The Mobileye camera was evaluated on a time interval of 0.3 s, since the measurements are not periodic and 0.3 s seems reasonable given the logged data.

Summing up, with our Mixed Urban scenario we could make discoveries on different levels. First, we could show some sensor specific characteristics. The Mobileye camera showed in general a not reliable pedestrian detection performance and it would therefore not be suitable to support an emergency braking system as in ADAS. The ESR radar has a quite reliable vehicle detection performance for different weather conditions. The drawback of the ESR radar is the very sensitive tracking (high number of track IDs for a single object), which is not included in our analyses of the FNRs. Moreover, the Delphi radar showed some performance drops when it was contaminated. Additional measurements are required to make more precise conclusions and evaluations about both sensors. On the level of general sensor type discoveries, our results demonstrate that the radar sensor is weather resistant, whereas the camera sensor showed performance changes for different weather conditions. Both discoveries were expected from the literature and have been proved with our evaluations. Finally, on the level of AD sensor testing, we showed how it is

not enough to have only a general statistical measure such as the mean average precision on a test data set. More dimensions need to be added to statistical measures such as TPR or FNR. For example we showed how the FNR changes with different distances to the test object. This is a relevant information when we think about sensor fusion, state estimation and planning. Dimensions such as different object sizes have also an influence on the object detection performance. With our comparison between adult and child pedestrian detection we demonstrated how the detection performance changes for two objects with the same shape but different sizes. Weather conditions and contamination are relevant dimensions as well. The Mobileye camera has different performances for different weather conditions. This needs to be considered when an ODD is defined for an AD-vehicle. Many more dimensions could be added, such as the relative orientation to the objects, the dynamics of the objects, or the relative velocity between the AD-vehicle and the obstacle. From the parabolic shape of the FNR curves we deduce that the movement of the AD-vehicle influences the sensor performance. Additional dimensions could be on the sensor level such as different exposure time for the camera sensors. All these additional dimensions affecting the performance of the sensors are relevant for the safety assessment of AD-vehicles, sensor selection, sensor fusion, AD-vehicle co-design or controller design for AD functionalities.

6.1.5 Lidar Targets

Scenario description

To test the influence of weather on the lidar detection performance, we installed the wooden targets painted in different colors on the test track (see Figure 81). In this scenario, we drive towards the lidar targets with the AD-vehicle from a distance of 70 m up to 5 m (Figure 82).

Figure 81: *Black, grey and white lidar targets on the test track.*

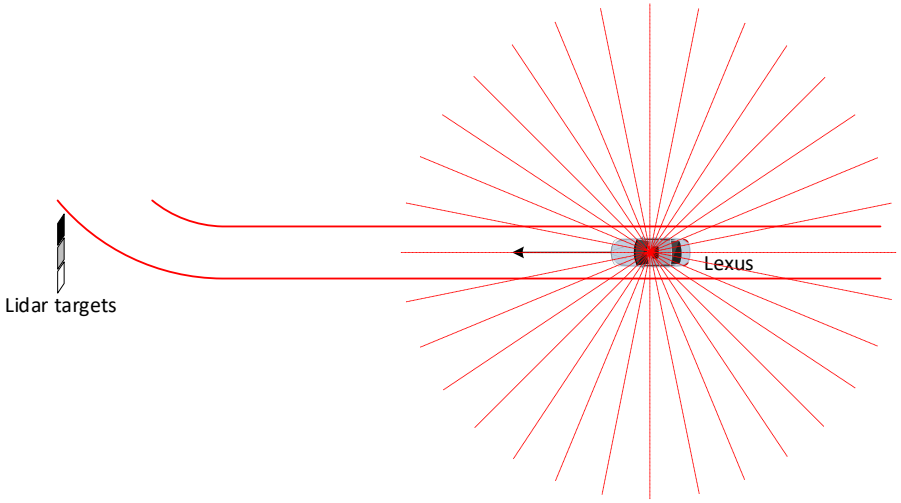

Figure 82: *Lidar Target scenario setup.*

Data Processing and Analysis

In contrast to the Mobileye camera and the Delphi radar with their built in detection algorithms, the lidar provides raw data only. In order to use the same performance metrics, we would therefore have to process the lidar raw data with an appropriate detection algorithm. However, because our goal is to test only the sensor, we defined different metrics. The chosen metrics (Table 4) reflect the information used in lidar detection algorithms and will therefore allow us to analyze and evaluate the potential lidar detection performance based only on the raw data.

Table 4: *Performance metrics for Lidar Target Scenario.*

Metric	Lidar

Number of Points	X
Intensity	X
Reflectivity	X

Resolution and number of points

In order to evaluate the lidar sensor performance with respect to its object detection capabilities, we consider the resolution of the sensor and the number of lidar points reflected by the target object. First, we determine the theoretical number of points (or laser beams) expected to lie on a given target of a fixed size at a certain distance with the given sensor resolution. Then we test the sensor under different environmental conditions using different targets to find out how these parameters affect the real number of points reflected back from the target. The number of points have a direct impact on whether an object can be detected or not.

Since the lidar has an angular resolution, we first calculate the theoretical number of points that lie on each target as a function of the distance. We calculate the number of points by dividing the target width and height by the respective horizontal and vertical resolution at that distance to the target:

$$n_{hor} = \frac{w_{target}}{d \cdot \tan(azi_{res})}$$

$$n_{ver} = \frac{h_{target}}{d \cdot \tan(ele_{res})}$$

where:

Target width:	w_{target}	0.75m
Target height:	h_{target}	2.0 m
Horizontal Resolution @ 600RPM:	azi_{res}	0.2°
Vertical Resolution:	ele_{res}	2°

The results of this calculation for different distances to the target are shown in Table 5.

Table 5: *Average expected number of lidar points on a 0.75 x 2.0 m target at different distances.*

Distance	**70 m**	**60 m**	**50 m**	**40 m**	**30 m**	**20 m**	**10 m**	**5 m**
Average number of lidar channels on target (vertical)	0.82	0.95	1.15	1.43	1.91	2.86	5.73	11.45
Average number of points on target per channel (horizontal)	3.07	3.58	4.30	5.37	7.16	10.74	21.49	42.97
Average number of points on target	2.51	3.41	4.92	7.69	13.67	30.76	123.05	492.22

This simple calculation already shows that with the resolution of our Velodyne VLP-16 lidar the detection yet the classification of cars and pedestrians is almost impossible at anything but short distances (between 5 and 20 m).

In addition to the decreasing spatial resolution with increasing distance, the lidar laser also suffers from beam spreading, i.e. the further away the light travels, the bigger the laser pattern becomes. This means that close to the target edge only part of the laser light is reflected from the target. According to the Velodyne VLP-16 user manual, the laser beams spreads 3.0 mrad in the horizontal direction and 1.5 mrad in the vertical. Therefore, at 50 m the laser pattern already has a size of 14.97 by 8.46 cm.

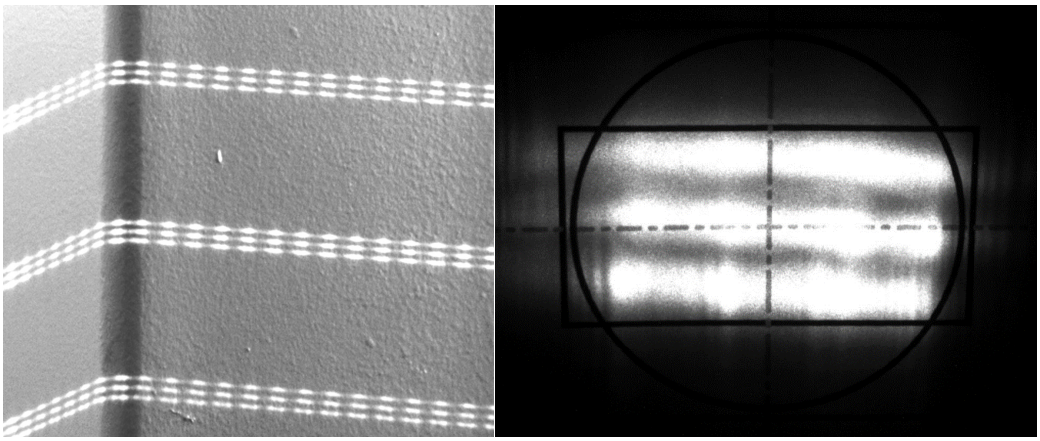

Figure 83: *Left: Laser pattern at short distance (Source: Velodyne VLP-16 user manual* **[10]***). Right: Close up of a single laser spot with the three bands (Source: Velodyne VLP-16 user manual* **[10]***).*

Figure 84 shows four plots with the resulting laser pattern on the lidar target resulting from the combined effect of the angular resolution, beam spreading graphics and distance to the target.

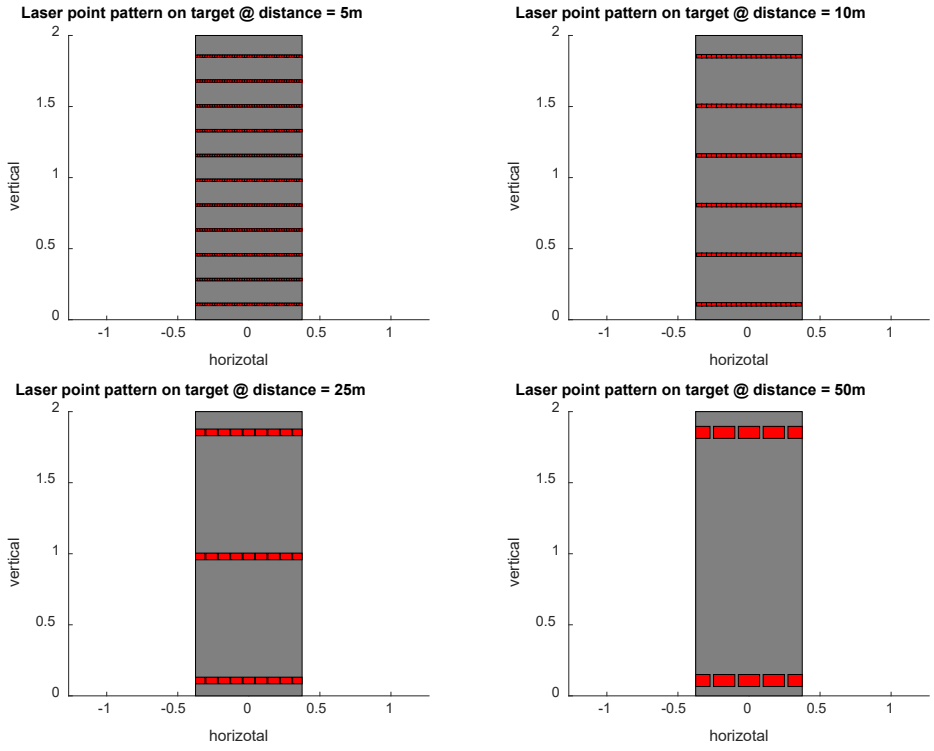

Figure 84: *Laser point pattern on the lidar target at different distances.*

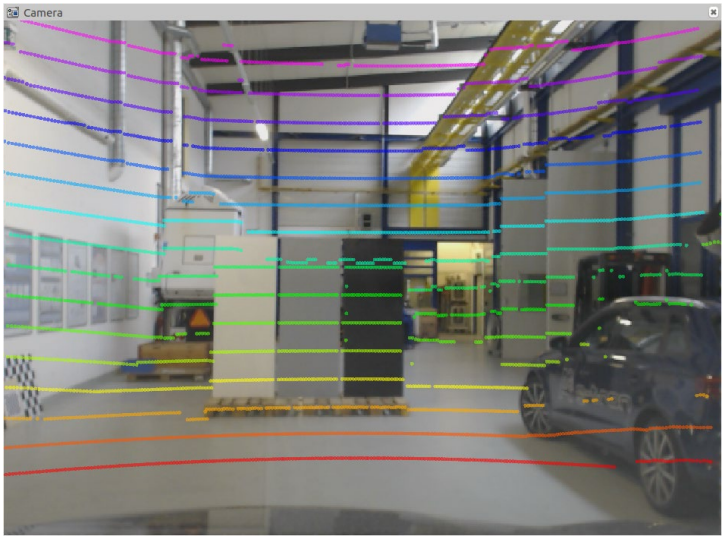

Figure 85: *Image of lidar targets at 10m distance with lidar points overlaid.*

In order to analyze the data from the tests with the lidar targets, we first have to identify all points in each point cloud belonging to the three targets. Figure 86 shows all the lidar points from one test that are attributed to the black, grey and white targets. Note how well the integration of all test data with the NDT-algorithm covers the full target area.

Since we only want to consider points where the complete laser light is reflected from the panel, we further filter the data. Figure 86 shows the same test run data after removing all the points closer than 15 cm to the target edges.

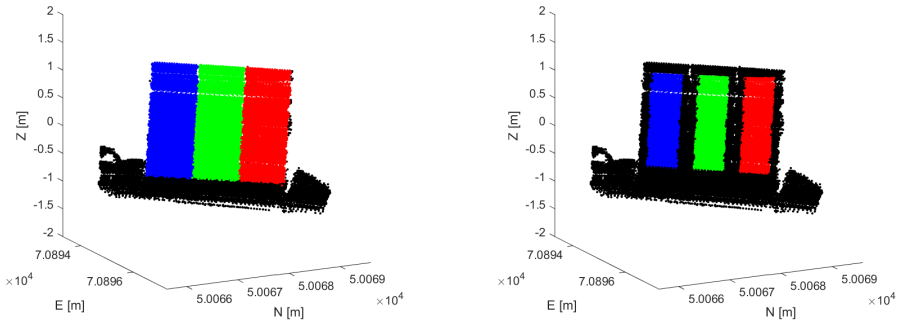

Figure 86: *Left: Lidar data with points separated into white, grey and black targets (blue, green and red). Right: Same test run but with edge points removed (15 cm to each side)*

Reflectivity and Intensity

According to the Velodyne VLP-16 user manual [10], the sensor itself should return the calculated reflectivity of the surface as an eight bit number. Values from 0 to 100 are for normal surfaces, values above 100 indicate a retro reflecting surface such as a number plate, a coated traffic sign or a tail light [10]. However, the ROS-driver for the Velodyne lidar outputs the reflected intensity instead of the reflectivity. Therefore, somewhere in the driver code, the intensity must be calculated from the reflectivity value and the measured distance to the target. Since we know the distance (d) and the intensity (I), the calculation can be inverted to calculate the reflectivity (r) as:

$$r = k \cdot d^2 \cdot I$$

Because the constant k in the formula above is unknown, the reported intensity and reflectivity are normalized, but this is not important for our analyses. In addition, compared

to the classical radar equation, in our formula the distance is only squared instead of to the power of four. The reason for this is that, as long as the target is larger than the beam pattern, we only have to consider the path from the target to the lidar. If the target is larger, the reflecting surface grows with the distance, while in the radar equation the radar cross section of the target remains constant.

The main reason why we transform the intensity back to reflectivity is that the reflectivity is a constant material property and, as such, is independent from the distance to the target and the reflected intensity. Therefore, when driving towards the lidar targets, the reflectivity should stay constant, while the intensity should vary with the squared distance.

Results

In this section, we report some general observations regarding the measurement results from the Lidar Targets scenario, followed by the results obtained from the different environmental conditions. For every condition we ran multiple test runs, and analyzed in total 40 runs (see Table 10 in Appendix I.3 for a list of all Lidar Target test runs in each condition).

Trends in number of points, intensity and reflectivity

Figure 87 shows the number of points lying on each lidar target as a function of the distance to the target. It can be seen, that for the white and grey panels, the increase in the number of points for smaller distances follows very closely the theoretical curve (black solid line). For the black panel at distances below 8 m the number of points seems to remain constant.

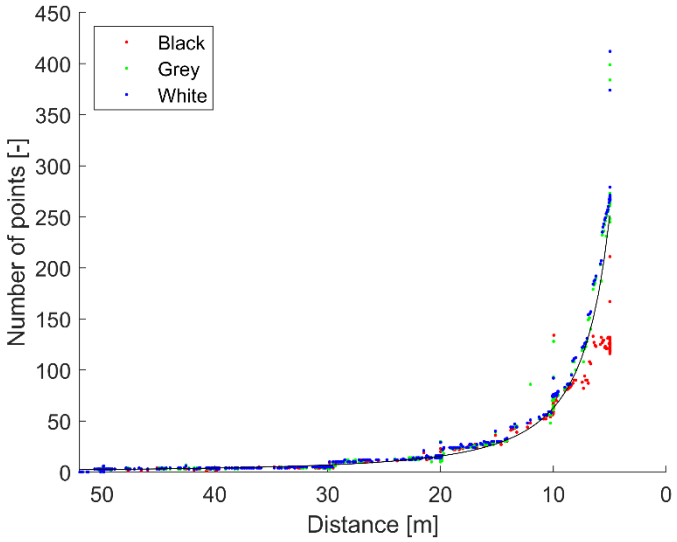

***Figure 87**: Number of points per target and median reflected intensity. The solid black curve in the top plot indicates theoretical expected number of points (log 04 from 28.9.2020).*

The left plot in Figure 88 shows the intensity of the individual points attributed to each panel. The right plot shows the same point distribution after removing all the points close to the panels' edges (see also Figure 86), which clearly results in a reduction in the number of outliers and more clear trends.

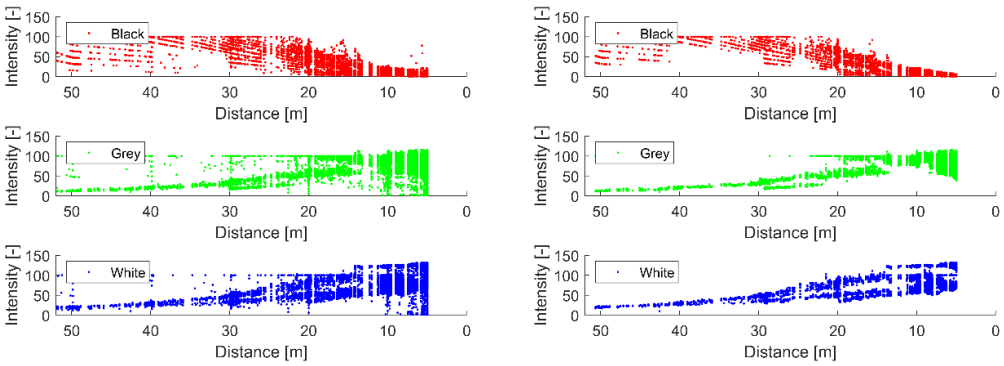

Figure 88: *Left: 0Intensity of each reflected laser beam attributed to each target (log 04 from 28.9.2020). Right: Intensity of each reflected laser beam attributed to each target after the edge points are removed (log 04 from 28.9.2020).*

In Figure 89 the intensity and reflectivity for one representative test run is presented. For distances between 50 and 25 m, the intensity and reflectivity trends for the white and grey panels follow the trends expected from the theory, i.e. the intensity varies with d^2, the reflectivity stays fairly constant, and the white target reflects more light than the grey one. However, for distances below 25 m, the intensity curves become broader and change trend, and the reflectivity decreases with decreasing distances. This is most probably due to a numerical limitation, as according to the Velodyne VLP-16 user manual, the reflectivity value for a non-retro reflecting material is supposed to be limited to less than 100. This limit is clearly visible in the data from the black panel, but for the grey and white panels an important fraction of the points are well above this limit for distances below 25 m. At the time being we do not have an explanation for this observation. Moreover, the intensity curves seem to split into several curves at the short distances (see next paragraph for further discussion). These two effects combined result in erroneous reflectivity trends. The overall behavior for the black panel is completely unexpected for all distances and is discussed in detail later in this section.

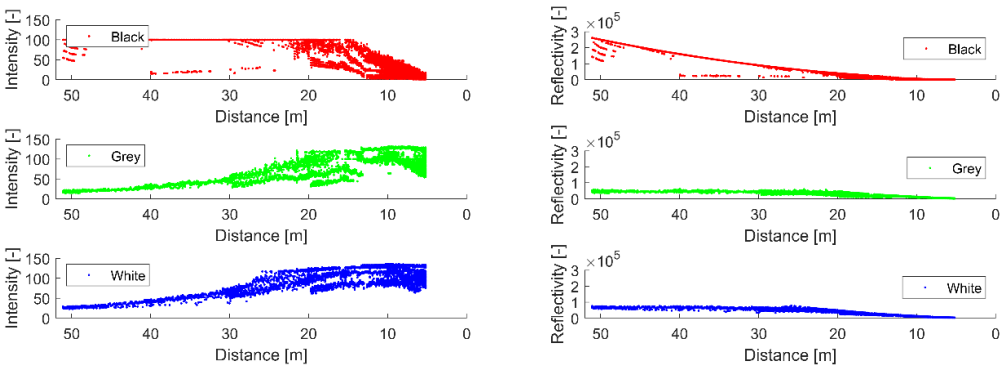

Figure 89: *Left: Reported intensity Right: Back calculated reflectivity (log 02 from 15.10.2020)*

Lidar channel

Figure 90 shows again the intensity distribution of the lidar point clouds as a function of the distance, color-coded with the different lidar channel numbers. For the white and grey panels, we see that up to 28 m we receive only points from the lidar channel 6, which show a clearly visible band over all distances. At 28 m, points from the lidar channel 5 appear on the targets, forming another distinct band with lower intensities than the original one. The same happens again at 20 m, where the lidar channel 4 appears on the targets. At 10 m, we have 6 channels on each target, and we see that for even smaller distances the trends of the single bands start to change. Similar effects are also observed for the black target,

but in this case the intensity tends to decrease for smaller distances (completely unexpected) and the points from each channel are further separated into sub bands, each of them originating from different horizontal locations of the lidar pattern.

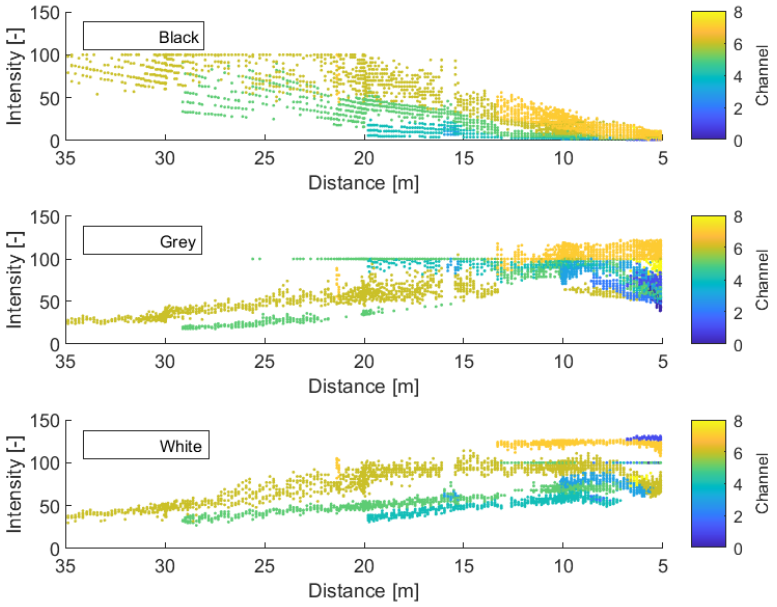

Figure 90: *Visualization of the distribution of intensity values for the different lidar channels (log 02 from 30.9.2020).*

Black colored panel and angle of incidence

In all the test runs and under all environmental conditions the results from the black panel showed consistently unexpected trends. In all cases, at around 7 meters from the wall, the number of points reflecting on the black panel starts to decline and deviates significantly from the expected (theoretical) number of points (see Figure 87).

Our outdoor field test observations with the black colored panel match the observations made by our colleagues at METAS [32] from their laboratory tests. Figure 6.91 show the experimental setup and results from the lidar tests at METAS with two black panels (mat and glossy black, respectively). The test results show that, as soon as the angle of incidence of the laser on the surface starts to deviate from 90°, the lidar cannot measure the black panels correctly. The panel on the right side of the experimental setup (visible in the lower left picture) is not detected at all (see gap in the right side of upper and lower right plot), and from the panel located in front of the lidar only the center area (corresponding to an incidence angle of the laser beam of 90°) is detected (center area of upper and lower right plot).

As can be seen by comparing the upper with the lower right plot, this effect is stronger for glossy black paint than for mat black. Note that both mat and glossy white panels were detected without problems in similar tests [32].

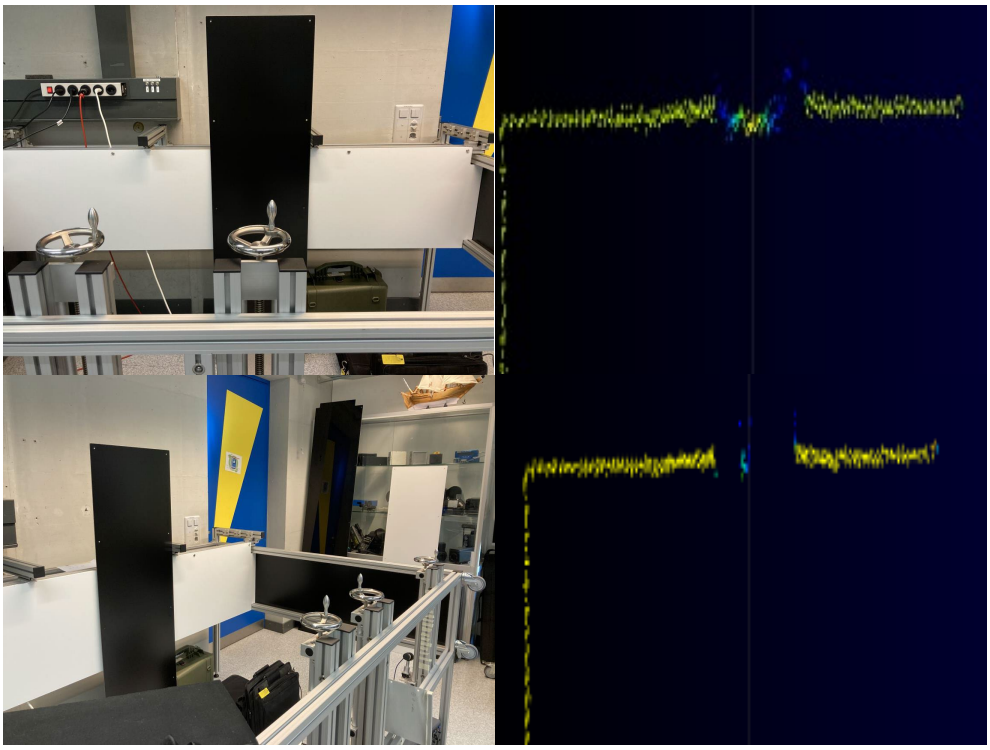

Figure 6.91: *Upper Left: Lab setup with mat black paint. Upper right: Lab results for mate black paint. Lower Left: Lab setup with glossy black paint. Lower Right: Lab results for glossy black paint (Source METAS)* **[32]**.

The observation that the angle of incidence affects the detection of the black panel was verified by rotating our lidar test panels in front of the car in a simple test performed in our workshop (see Figure 92 and Figure 93). The more the panels are turned away from 90°, the more the intensity reflected by the black panel decreases, until eventually the black panel fully disappears.

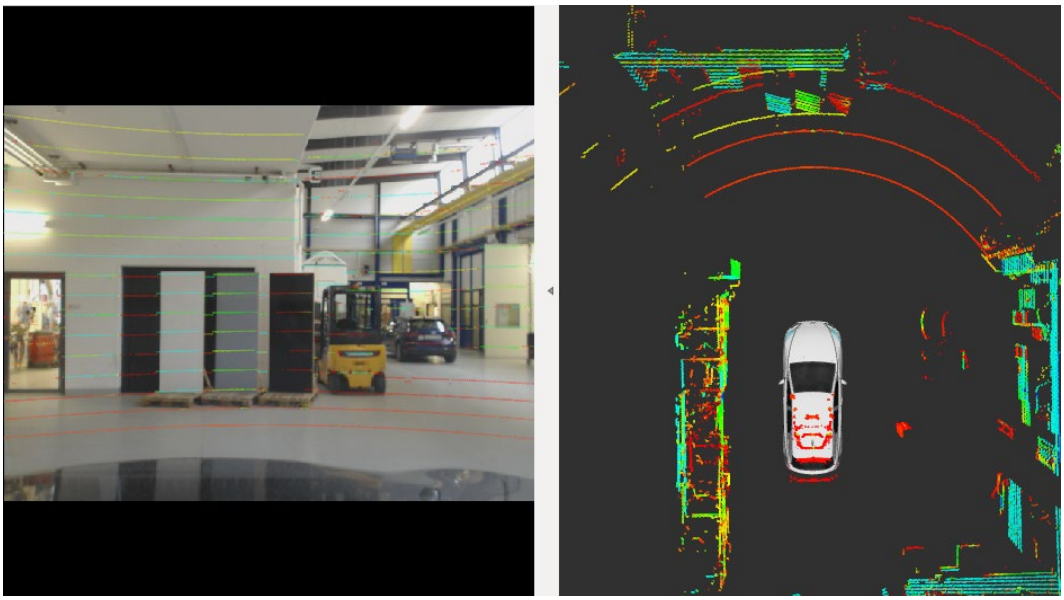

Figure 92: *Slight angled lidar targets: All panels are visible.*

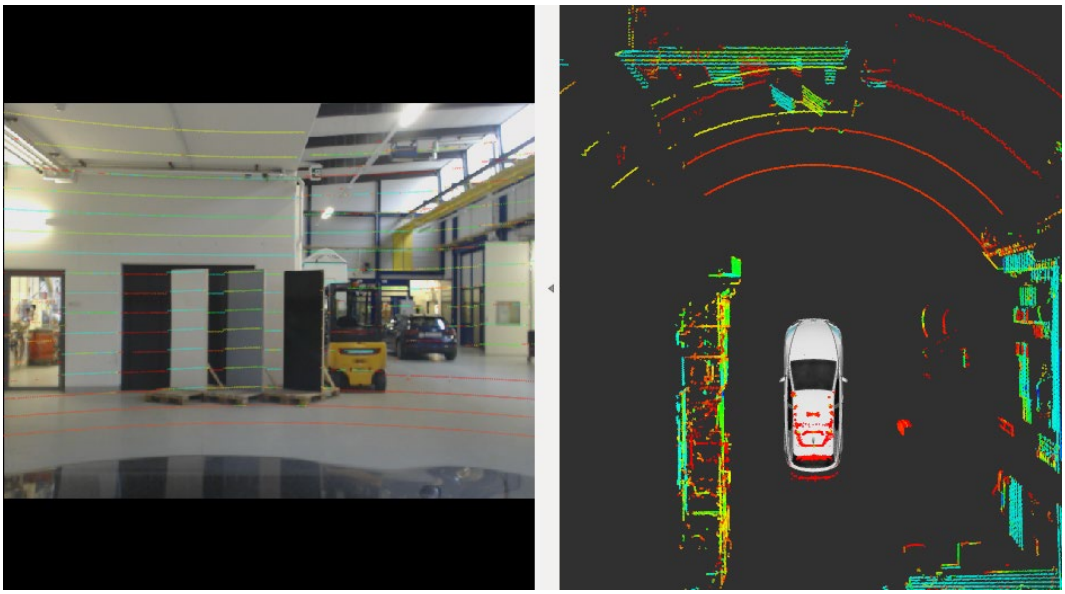

Figure 93: *Invisible black panel at a 45° angle.*

Sunshine

The left plot in Figure 94 shows the number of laser beams (points) reflected on each panel and the corresponding median reflectivity for one test run. The right plot shows the reflectivity from all the lidar points attributed to each panel. We see that, except for the black target, the number of points on each target matches well the expected number from theory (black curve), and the reflectivity follows similar trends as those described above. In addition, for the grey panel we observe some outliers at larger distances and for the black panel we clearly see the above described artifacts of decreasing number of points at short distances and increased reflectivity variability at large distances.

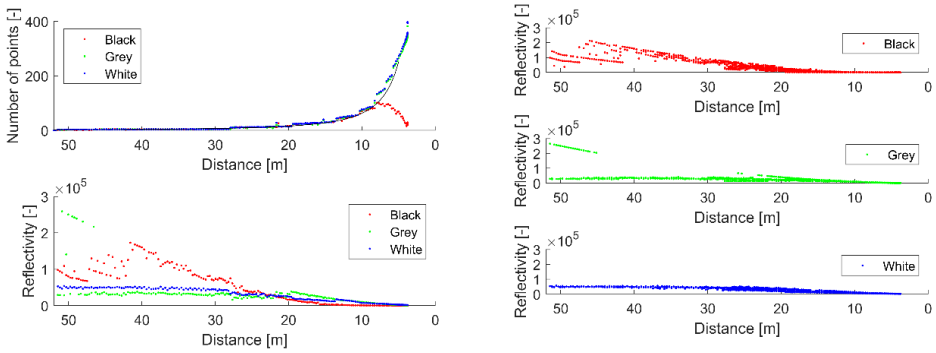

Figure 94: *Left: Number of points and median reflectivity on a sunny day (log 04 from 30.09.2020). Right Reflectivity raw data from the same test run.*

Figure 95 shows a box plot of the reflectivity data for all test runs in sunny conditions. The distance is gridded into 2 m intervals. For the black panel the reflectivity shows significant variability and decreases steadily with decreasing distances. The white and grey panels show little variability in the reflectivity among all test runs, and the mean reflectivity remains fairly constant for distances of up to 30 m and decreases for lower distances. The decrease at smaller distances might be related to the saturation of the intensity values (see also Figure 90). Moreover, we see some outliers in the results from the grey panel at large distances, but at these large distances the number of laser beams that fall completely on the panels is extremely low, and therefore the uncertainty is very high.

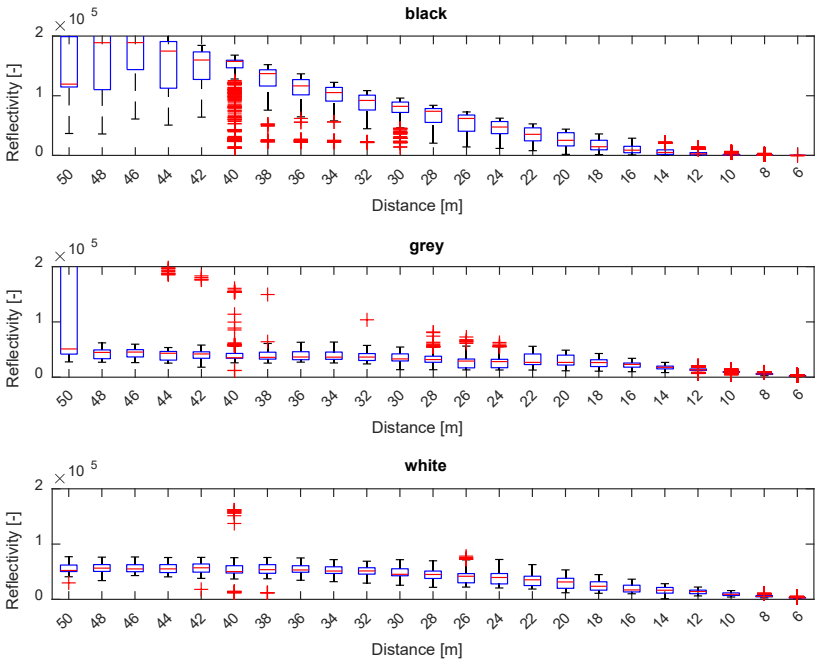

Figure 95: *Boxplot of reflectivity data of all test runs in sunny condition.*

Rain

The left plot in Figure 96 shows the number of laser beams (points) reflected on each panel and the corresponding median reflectivity for one rainy test run. The right plot shows the reflectivity from all the lidar points attributed to each panel from the same test run. Overall, the test results from rainy conditions show the same effects as those from sunny conditions.

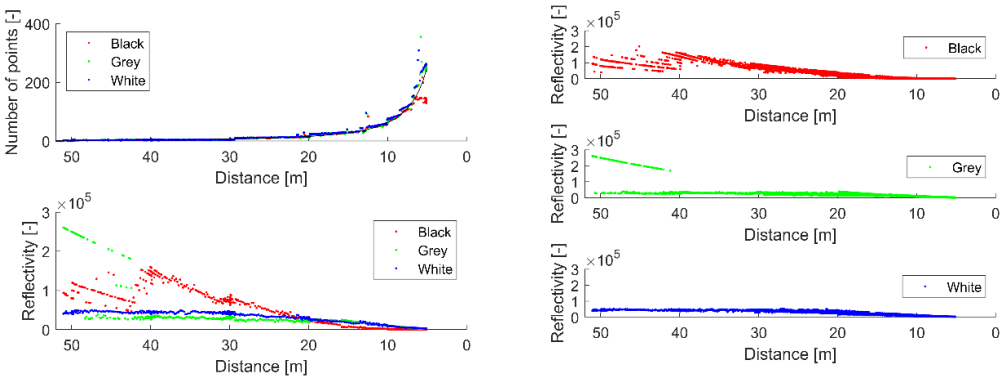

Figure 96: Left: *Number of points and median reflectivity on a rainy day (log 02 from 29.09.2020). Right: Reflectivity raw data from the same rainy test run.*

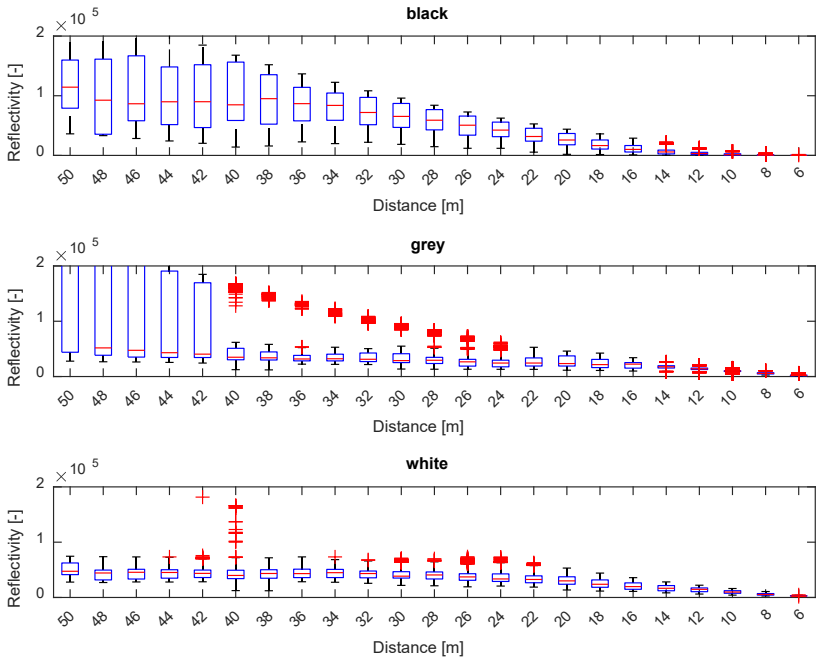

Figure 6.97: *Boxplot of reflectivity data of all test runs in rainy conditions.*

Fog

Figure 98 shows the results from one run on a foggy day. Despite the fog, the number of points on the white and grey targets matches well the expected number from theory, while, as for all other weather conditions, the number of points on the black panel declines for distances below 7 m. With regards to the reflectivity, we observe that for foggy conditions the number of outliers in the white and grey panels significantly increases (high reflectivity bands at large/medium distances that strongly increase with the distance), and that the characteristic sub-bands of the black panel are less visible.

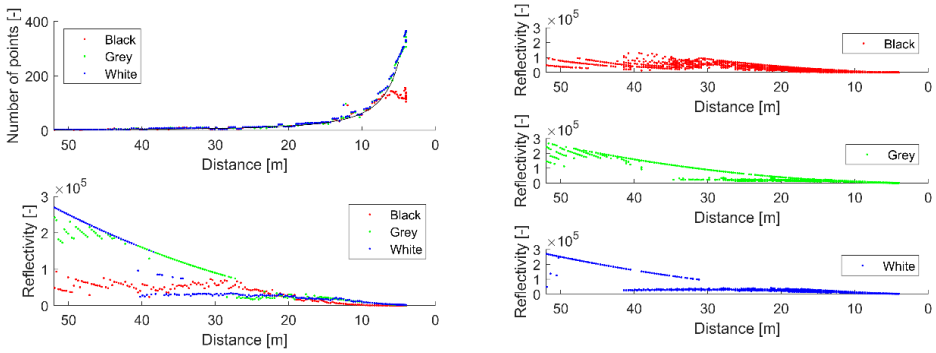

Figure 98: *Number of points and median reflectivity on a foggy morning (log 05 from 18.11.2020). Reflectivity raw data from the same foggy morning test run.*

Figure 99 shows the box plot of all test data from foggy test runs. We can see the increased number of outliers and the overall increased variability among successive test runs. The high reflectivity bands mentioned above result in an increased reflectivity for large distances for the grey panel. When considering the outliers in the box plot, we can also see a similar effect for the white panel, although in this case the median reflectivity stays constant.

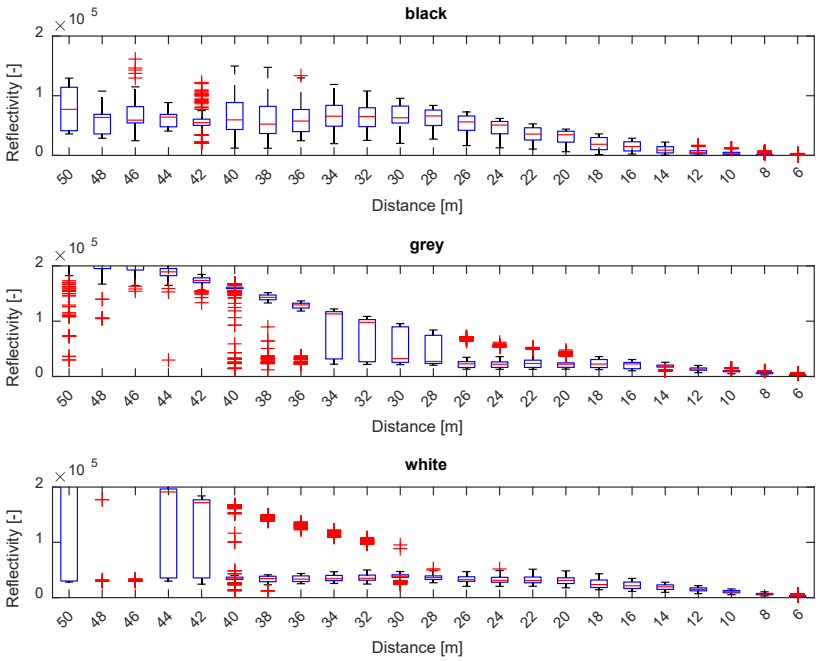

Figure 99: *Box plot for reflectivity data of all test runs with fog.*

Snow

Figure 100 shows the results from one run on a snowy day. As for all weather conditions, the number of points matches the expectations, except again for the black panel at distances below 8 meters. The reflectivity shows similar trends to the foggy conditions, with a high variability between test runs and an increased number of outliers. In snowy conditions and distances larger than 40 m, we observe again the strange behavior of increasing reflectivity for the white and grey target as already seen in foggy condition. This also results in much taller boxes or more outliers in the box plot (Figure 99).

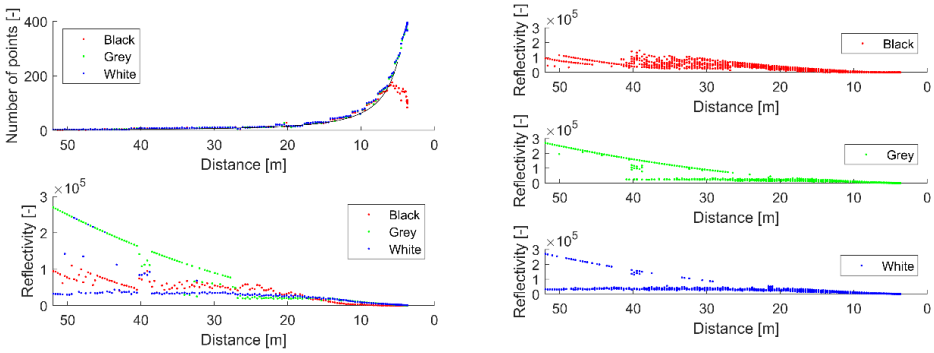

Figure 100: *Left: Number of points and median reflectivity on a snowy morning (log 09 from 01.12.2020). Right: Reflectivity raw data from the same snowy morning test run.*

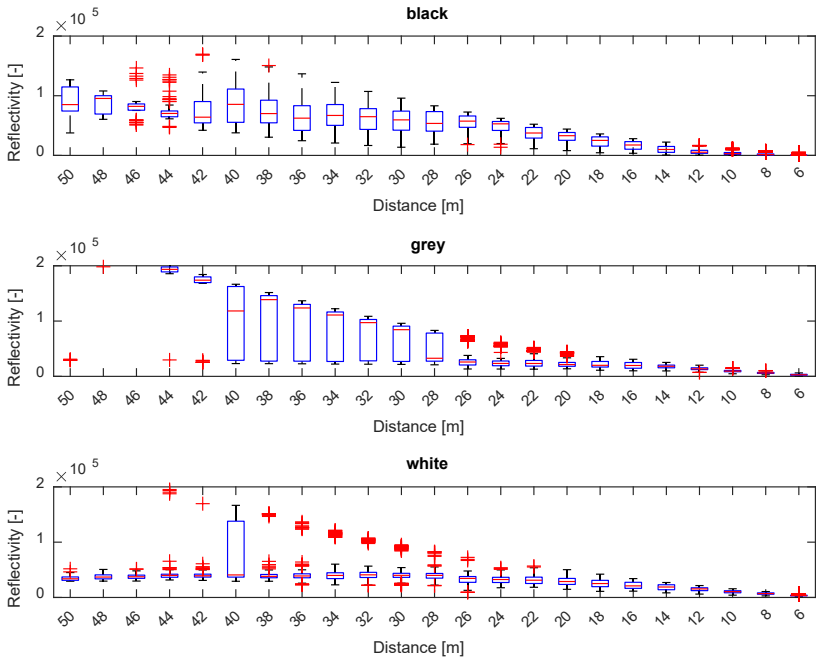

Figure 101: *Box plot for reflectivity data of all test runs with snowfall.*

Contamination

Figure 102 shows the results from one test run with the contaminated sensors. As for all previous conditions, the number of points increase with decreasing distance following the expected trend, except for the black panel, which sharply decreases for distances below 7 m. The trends and variability of the reflectivity (Figure 103 and Figure 103) are comparable to the results obtained for the sunny conditions, with fairly constant reflectivity and small variability among successive test runs for the grey and white panels, and large variability and a decreasing reflectivity for decreasing distances for the black panel.

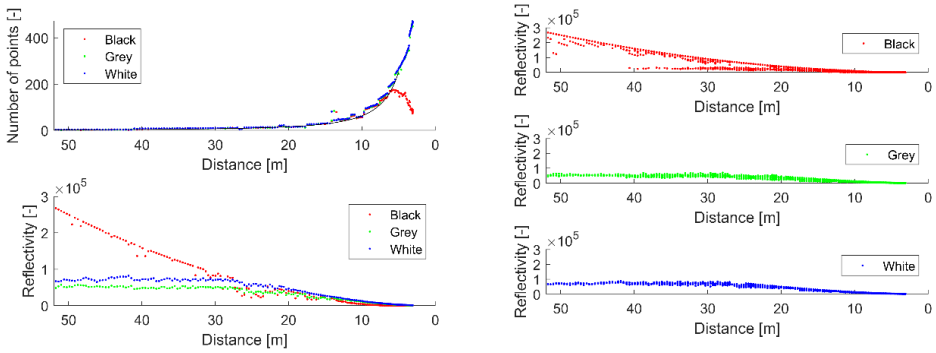

Figure 102: *Left: Number of points and median reflectivity with contaminated sensor (log 03 from 27.11.2020). Right: Reflectivity raw data from the same test run with contaminated sensor.*

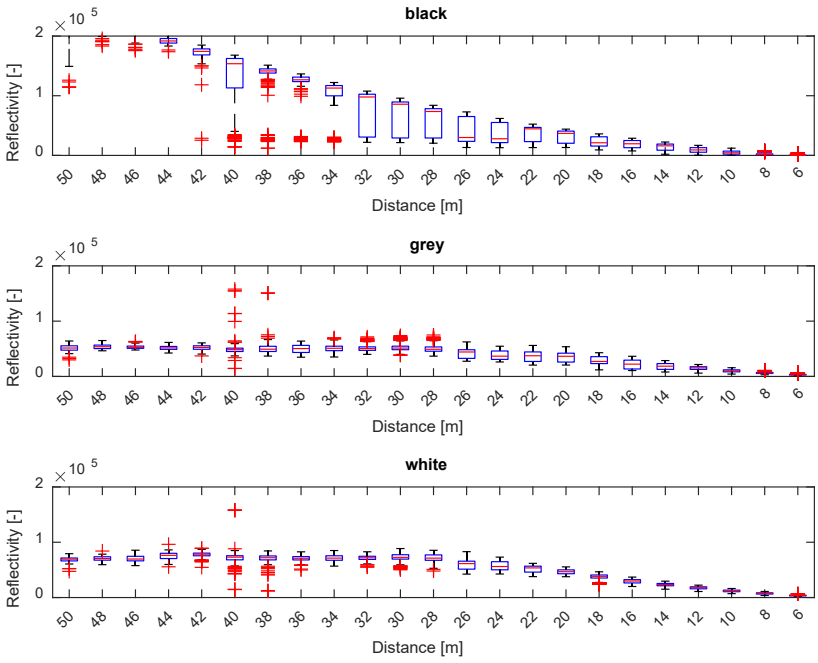

Figure 103: *Box plot for all test runs with contaminated lidar sensors.*

Distribution of number of points

Figure 104 and Figure 105 show the distribution of the number of points on the lidar targets for all tested environmental conditions at 5 and 20 m distance, respectively. At a 5 m distance, we see that the number of points from the black panel is consistently lower than that of the white and grey panels and that, for all environmental conditions, the variability among all test runs is much larger for the black panel. At 20 m distance the variability in the number of points between all the test runs is very low for all environmental conditions and panel colors.

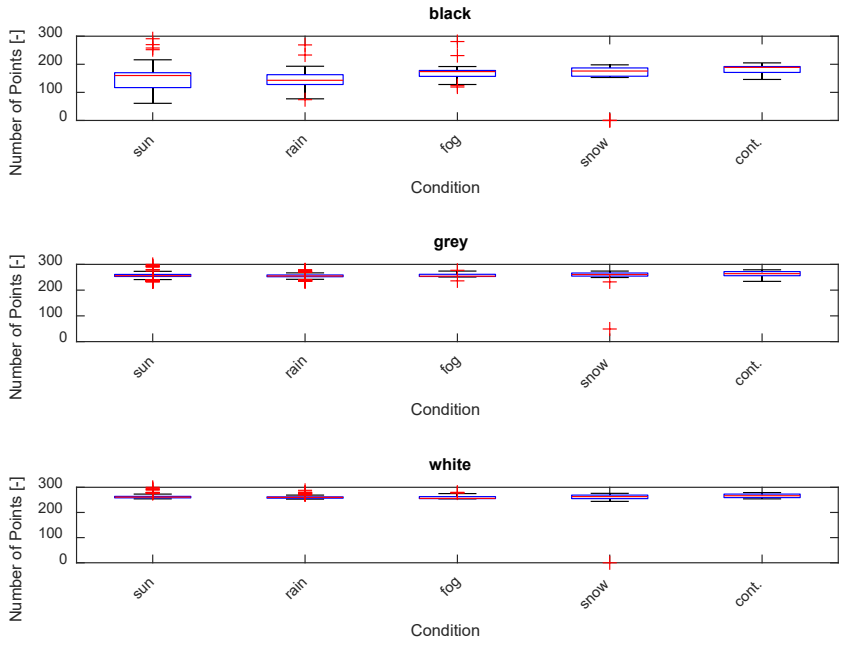

Figure 104: *Box plot of number of points at 5m distance.*

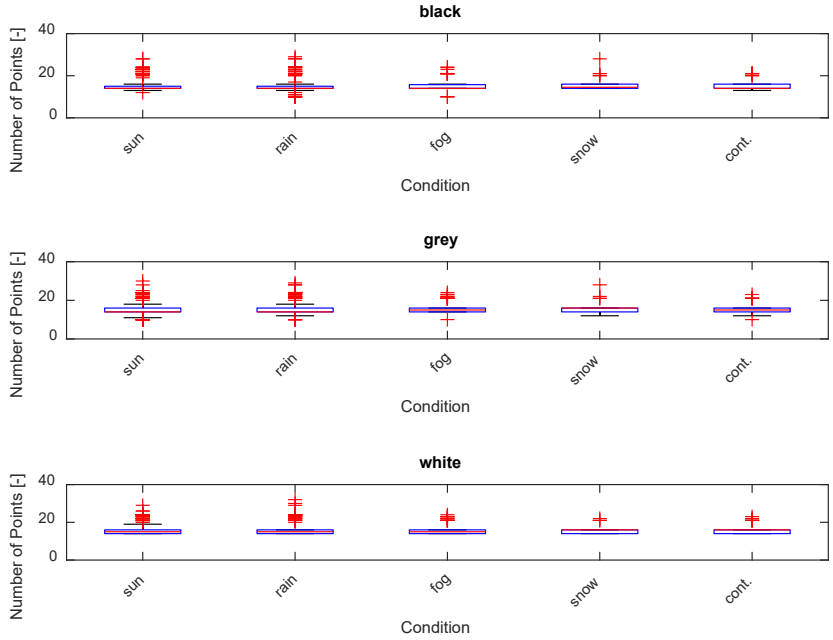

Figure 105: *Box plot of number of points at 20m distance.*

Discussion

Our results from the field tests, the tests in our workshop and the laboratory tests at METAS show that the color of the object and the angle of incidence between the laser and the object's surface have a strong influence on the lidar's performance. Overall, the white and grey surfaces behave as expected from the sensor's functioning principle, as the reflectivity of the white surface is consistently higher than that of the grey one, and the variability among consecutive runs/test conditions is mostly low. However, while the reflectivity is expected to be independent from the distance to the object, we see a continuous decrease in the reflectivity from the white and grey targets for distances below 30 m. This might be linked to the saturation of the intensity measurement at short distances and the beam splitting effect described in the previous chapter. On the contrary, we do not have a comprehensive explanation for the strange behavior of the black surface, for which the reflectivity shows a much higher variability and a completely different trend with the distance than that of the white and grey surfaces. The black panel also results in unexpected results in terms of the number of points, which significantly drop at small distances, instead of increasing exponentially as observed for the white and grey targets and expected from the theory. This is most possibly related to the fact that, at low distances, the target might not be perfectly perpendicular to the laser, which as shown by the tests at METAS and in our workshop, can make the detection of black surfaces even harder. Things get even worse when the surface is glossy instead of matte or semi matte, as seen in the results from METAS [32].

Our results from the Lidar Targets scenario show that overall the reflectivity measurement is noisier and shows an increased variability among consecutive measurements in bad weather conditions such as rain, fog, or snow, than in sunny conditions. In contrast, the contamination of the lidar's surface does not seem to significantly affect the measured reflectivity compared to the sunny weather. On the other hand, the number of points reaching each target does not seem to be affected by the different weather conditions, nor by the sensor contamination.

On the level of general sensor discoveries, we observed that the lidar measurements are not dramatically affected by adverse weather conditions. However, the fact that dark objects do not reflect enough energy at slightly non-perpendicular and at higher distances

even at perpendicular angles poses a major problem for the use of the lidar for object detection purposes.

6.2 Localization

In this chapter we analyze the localization functionality of the lidar sensor (described in Section 3.1.2) and its performance in different environmental conditions.

6.2.1 Lidar localization in urban environment

Scenario description

To evaluate the performance of the lidar localization functionality we simply use the localization results of successive drives around the test-track for different weather and contamination conditions (Figure 106).

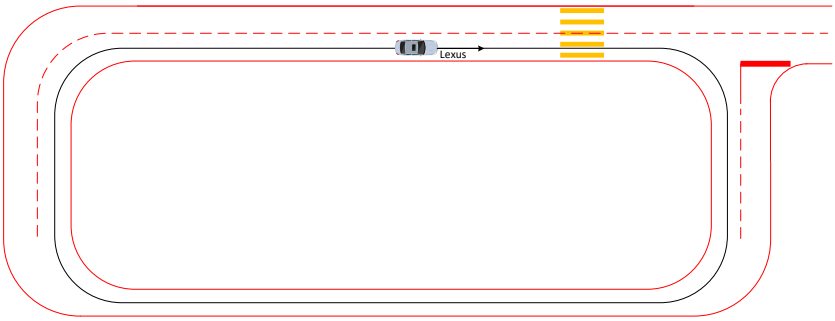

Figure 106: *Lidar Localization scenario setup.*

Data processing and analysis

Ideally, one would use the difference between the true and estimated position as a metric for the localization, such as the root mean square or the maximum distance error. The main challenge here is how to obtain the true position, as the GPS position provided by our xSens GPS receiver (2 m circular error according to its datasheet [33]) is not accurate enough for this. The other issue is that for lidar localization one needs a point cloud map that is accurately enough geo referenced, i.e. with less error than the localization itself.

In order to overcome these two problems, instead of analyzing the localization precision, we use metrics calculated from the localization algorithm to assess how quickly and robustly the localization algorithm converges. Table 6 shows the available metrics from the NDT-algorithm.

Table 6: *NDT-Matching statistics.*

ROS Message	Meaning
ndt_stat.exe_time	Execution time of the algorithm
ndt_stat.iteration	Number of iterations until convergence
ndt_stat.score	Alignment error, i.e. a measure of how well the input point cloud matches the reference point cloud

These metrics are not independent of each other and also depend on the settings of the localization algorithm. Nevertheless, they give a good indication of the quality of the point cloud for the purpose of localization. Since the execution time depends on the computational hardware, we do not use it as a hard metric.

Results

In this section we present the results obtained from the Lidar Localization scenario for different environmental conditions.

Sunshine

Figure 107 shows the position and heading from five laps around our test track on a sunny day. The position and orientation was calculated with the autoware.ai NDT-matching algorithm[4]. Figure 108 shows the NDT-matching statistics from all five laps.

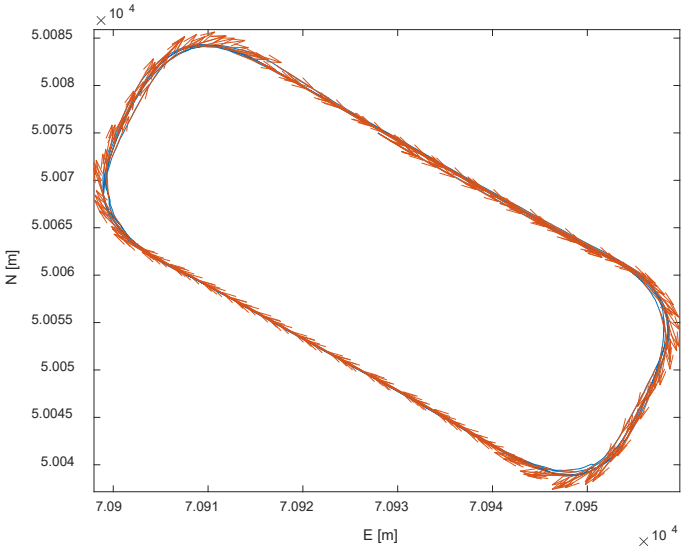

Figure 107: *NDT-Localization track in MGRS Coordinates from five laps around the test track (log 01 from 30.9.2020).*

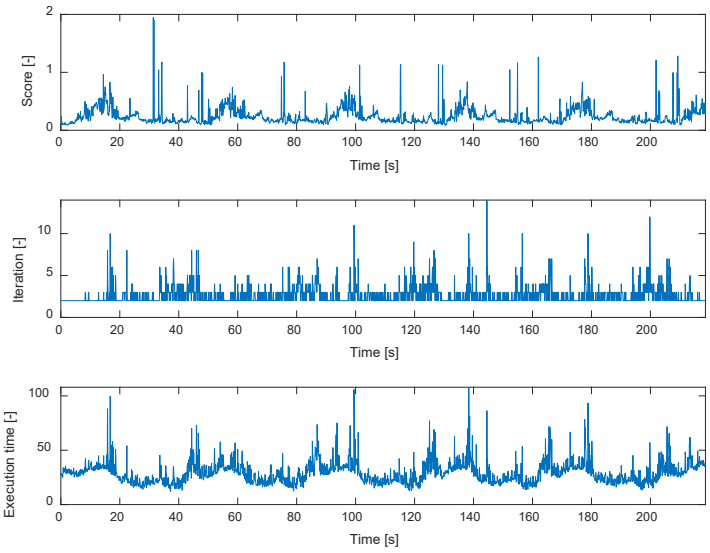

Figure 108: *Statistics from NDT-Localization algorithm during these five laps (log 01 from 30.9.2020).*

[4] https://www.autoware.ai/

To compare the results from successive laps, Figure 109 shows the same data as Figure 108 but separated into laps and plotted against the distance from the lap-start. We observe some repeating patterns depending on the position on the test track, e.g. for every lap there is an increase in the execution time and the number of iterations around the 20 m and 135 m marks.

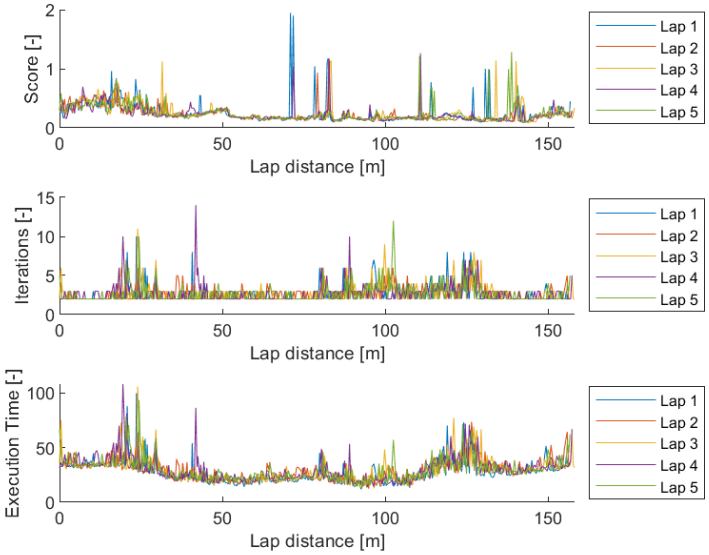

Figure 109: *NDT-Statistics for each lap (log 01 from 30.9.2020)*

For a better visualization, Figure 110 shows the NDT-Matching score plotted against the position on the track in 3D. The plot shows that the score is mostly low on the straights and high in the curves. This pattern is seen in all our test runs, indicating that it's not a weather related phenomena.

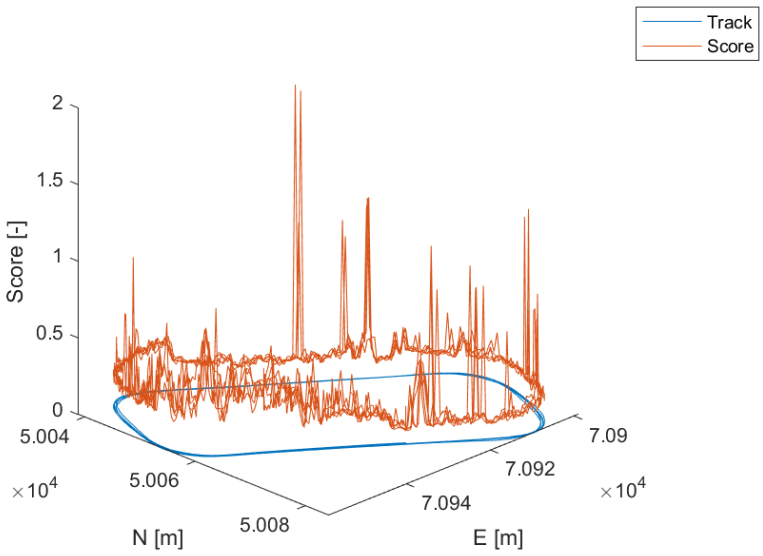

Figure 110: *NDT-Matching score plotted at the corresponding vehicle position on the track (log 01 from 30.9.2020).*

Figure 111 shows the top view of the lidar point cloud at a specific position of our test track from a test run in sunny conditions. The AD-vehicle is situated in the center of the image and the different colors indicate the reflectivity of the materials on which the laser light is reflected**.**The orange points in front of the car correspond to the zebra crossing paint, which contains glass pearls for retro reflecting. The rectangular structures on the left and right from the car are the metallic containers located on the sides of our test track. Due to its position on the vehicle roof, the lidar does not see anything at short distances to the back of the car, as the lower laser beams hit the car's roof. Apart from this, the point cloud for the sunny conditions looks clean, without missing points (except from those to the back) and without visible noise.

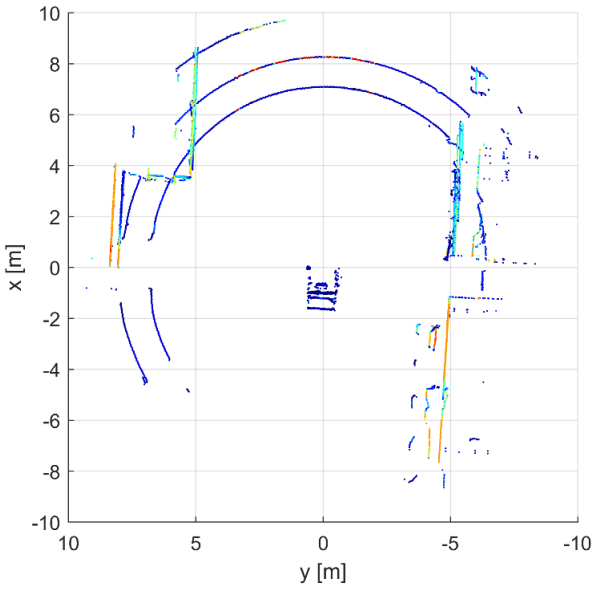

Figure 111: *Point cloud on a sunny day (log 01 from 27.8.2020)*

Rain

Figure 112 shows the lidar point cloud at the same position on the test track from a test run in a rainy day. For rainy and wet conditions the lidar point cloud shows significant fewer points than in sunny and dry conditions. This effect can be seen clearly at the ground in front of the vehicle, where the water layer on the paint of the zebra lines seems to mirror the laser light away, instead of reflecting it back to the lidar. Despite the clear effects on the point cloud, rain does not seem to have an influence on the localization performance (Figure 113).

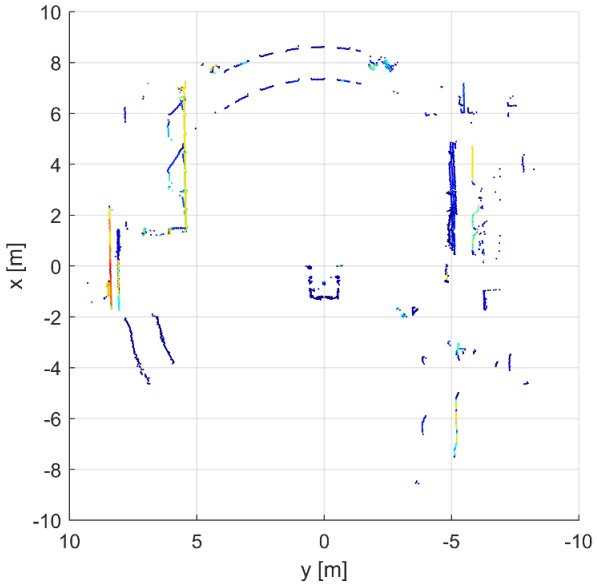

Figure 112: *Point cloud from a rainy and wet test day (log 05 from 29.09.2020).*

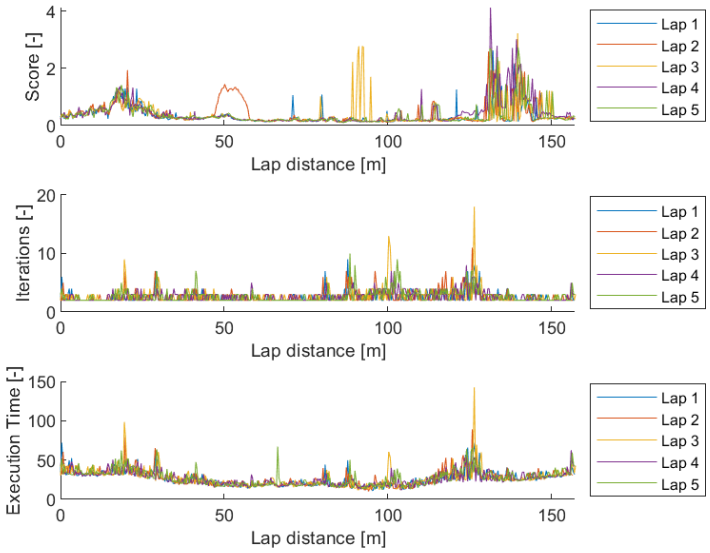

Figure 113: *NDT-Statistics for a rainy and wet test day (log 05 from 29.09.2020).*

Fog

Figure 114 shows the corresponding lidar point cloud from a test run in foggy conditions. The point cloud shows less missing points than the one from the rainy and wet test day (Figure 112), indicating that wet surfaces might have a stronger impact on the point cloud than fog or rain droplets.

As for rain, fog does not seem to have an influence on the localization performance either (Figure 114).

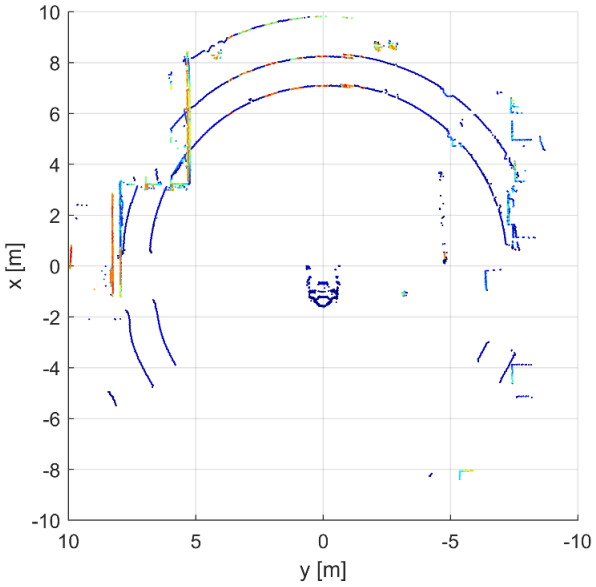

Figure 114: *Point cloud from a foggy morning test day (log 01 from 18.11.2020).*

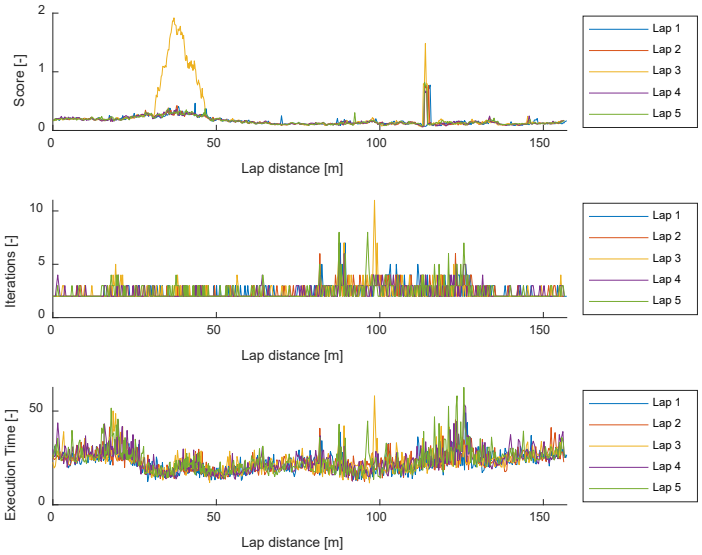

Figure 115: *NDT-Statistics from a foggy morning test day (log 01 from 18.11.2020).*

Snow

Figure 116 and Figure 117 show the lidar point clouds from two test runs (different days) in snowy conditions. The negative effects from snowfall for environmental perception are evident. The point cloud from the first snowfall day (Figure 116) shows a lot of noise (scattered points) and a partially blocked field of view (FoV) to the front of the AD-vehicle. The blocked FoV is due to wet snow adhering to the front surface of the lidar when driving. The point cloud from the second snowfall day (Figure 117) shows no blocking of the field of view and less noise. The snowflakes on that day where smaller and drier, and therefore they did not adhere to the lidar surface.

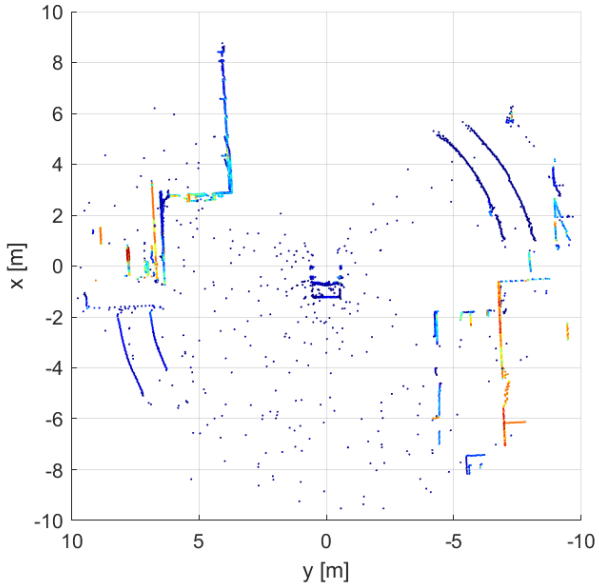

Figure 116: *Partially blocked field of view and noisy point cloud due to adhering snow (log 01 from 27.2.2020).*

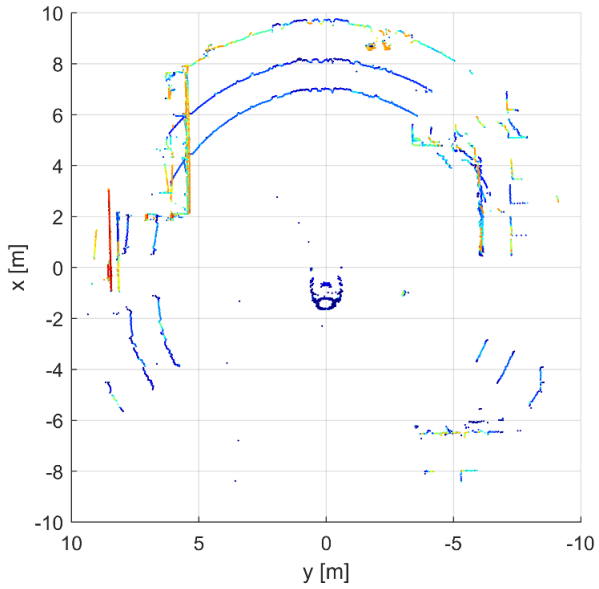

Figure 117: *Point cloud from a second snowfall day (log 01 from 1.12.2020).*

Figure 119 and Figure 118 show the NDT-Statistics from the lidar localization algorithm from the two test days with snowfall. Compared to the results from tests in sunny conditions, we see some increase in the score and, to a smaller extend, in the number of iterations (see Table 7). Surprisingly, even though the lidar point cloud in the heavy snowfall conditions is very noisy and the snow sticking on the lidar surface partially blocks its field of view, the localization algorithm successfully localized the car also for all measurements with snowfall, as indicated by the successful algorithm convergence.

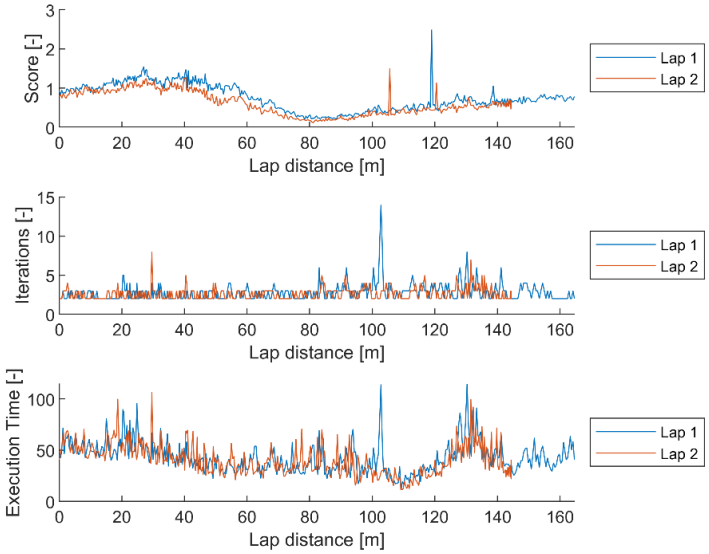

Figure 118: *NDT-Statistics from two laps during heavy snowfall (log 01 from 27.2.2020).*

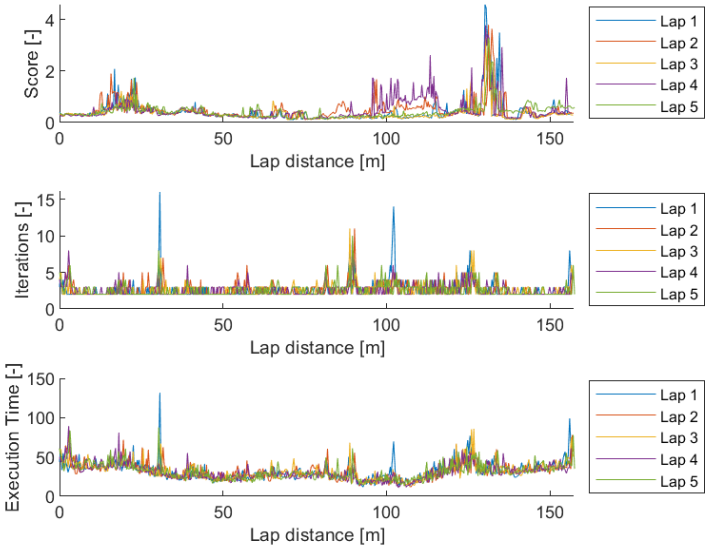

Figure 119: *NDT-Statistics from 5 laps during snowfall (log 01 from 01.12.2020).*

Contamination

Figure 120 shows the lidar point cloud from a test run with the contaminated sensor, and shows that the dirt on the lidar surface does not have a significant influence on the field of view of the sensor, nor does it increase the measurement noise. Figure 121 shows the NDT-parameters from all the test runs with the contaminated sensor. The NDT-statistics (Table 7) show no worsening of the performance of the localization algorithm compared to the sunny scenario. Moreover, the visualization of the point cloud shows no major blocking or increased noise due to the contamination.

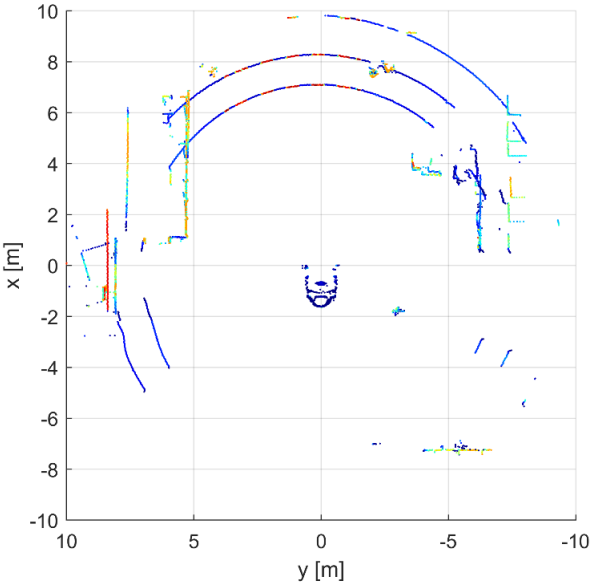

Figure 120: *Point cloud from test run with sensor contamination (log 02 from 27.11.2020).*

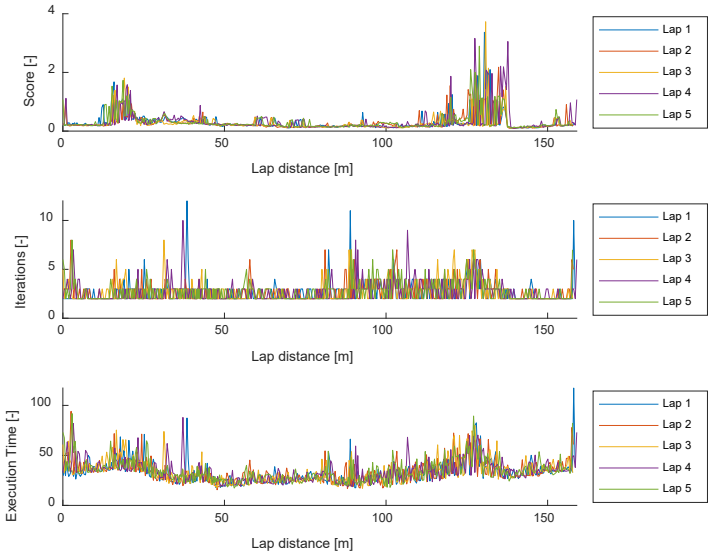

Figure 121: *Statistics from the run with dirty and leave covered lidar (log 02 from 27.11.2020).*

Driving Style

In order to analyze the effects from the driving style on the lidar localization, in Figure 122 and Figure 123 we compare the localization statistics from the steady (average 10 km/h, maximum 13 km/h) and dynamic (average 20 km/h, maximum 36 km/h) test runs from the 27.8.2020. This comparison shows that dynamic driving increases the number of iterations until convergence. The reason for this is that the successive lidar point clouds are more rotated and further apart from each other the more dynamic the vehicle drives. However, in both cases the algorithm can successfully localize the car, i.e. it converges at all time steps. This might not be the case at higher speeds, but the dimensions of the test track limit the maximum speed at which we can perform the tests, so that measurements at other locations would be needed to investigate this further. The convergence speed can be improved by using IMU and other motion data to improve the initial guess of the translation and rotation between the two consecutive point clouds.

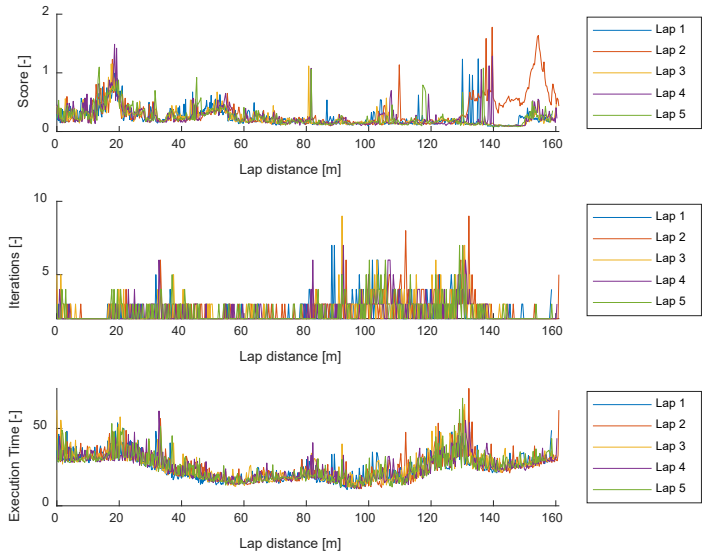

Figure 122: *NDT-Statistics for a gentle, steady driving (log 01 from 27.08.2020).*

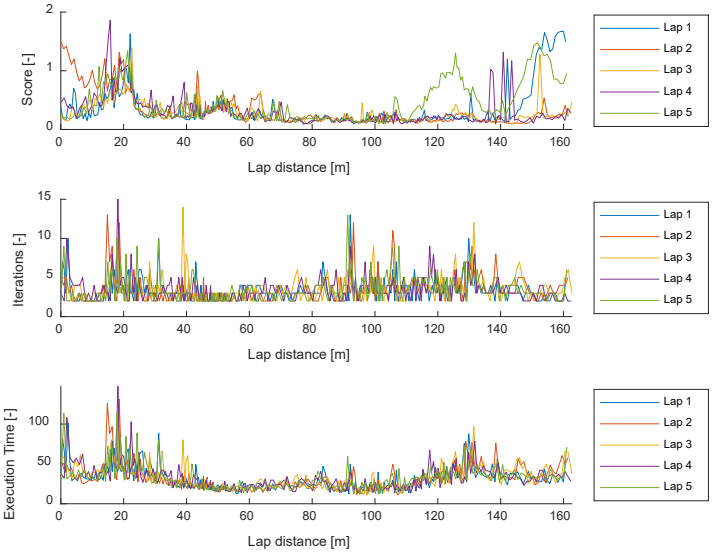

Figure 123: *NDT-Statistics for more dynamic driving (log 02 from 27.08.2020).*

Replay of localization algorithm

In Figure 118 we show the NDT-Stats from the first test runs in snowfall. Looking at the NDT-Stats we can see a spike in the score at around 120 m and a spike in the number of iterations and execution time around 105 m. To further analyze the reason for these spikes, Figure 124 shows the calculated difference between each localization position $\Delta pos = pos(k+1) - pos(k)$ and the derived velocity $v = \frac{\Delta pos}{\Delta t} = \frac{pos(k+1) - pos(k)}{t(k+1) - t(k)}$ in direction north (N) and east (E). The only remaining spike is the one at t=105 s, where even the change in position and the calculated velocity suddenly jump from one time step to the next before going back to the previous values. Except from this single spike, the localization trajectory looks smooth. Even at the region of the spike, the trajectory looks rather smooth as can be seen in the close-up shown in Figure 125. Additionally, Figure 126 shows that the time stamps of the position message are also almost equally spaced and nothing particular happens around the spike time (d = 105 m). Therefore, the reason for the spiking is still unclear.

We guess that the spiking might be due to a timing issue in the replay on the desktop pc. The snowfall runs from February 2020 were the only ones for which we had to run the localization algorithm on replayed data on the pc and we have not seen any other logs with spikes since then.

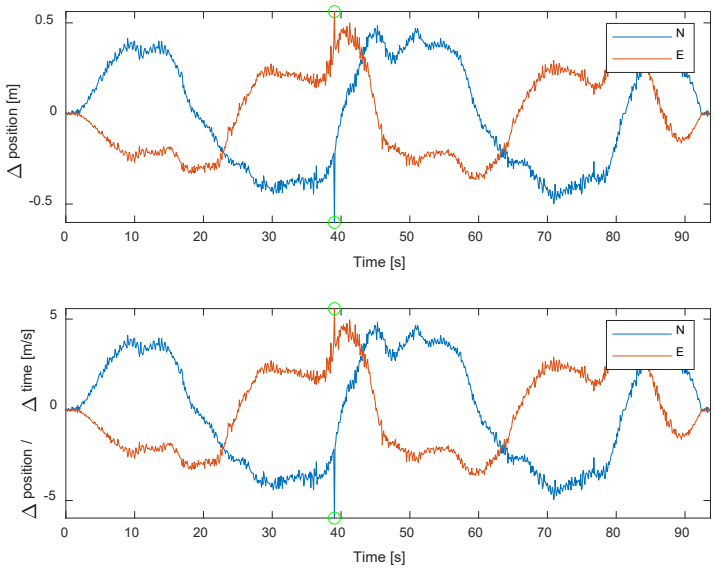

Figure 124: *Delta position and velocity.*

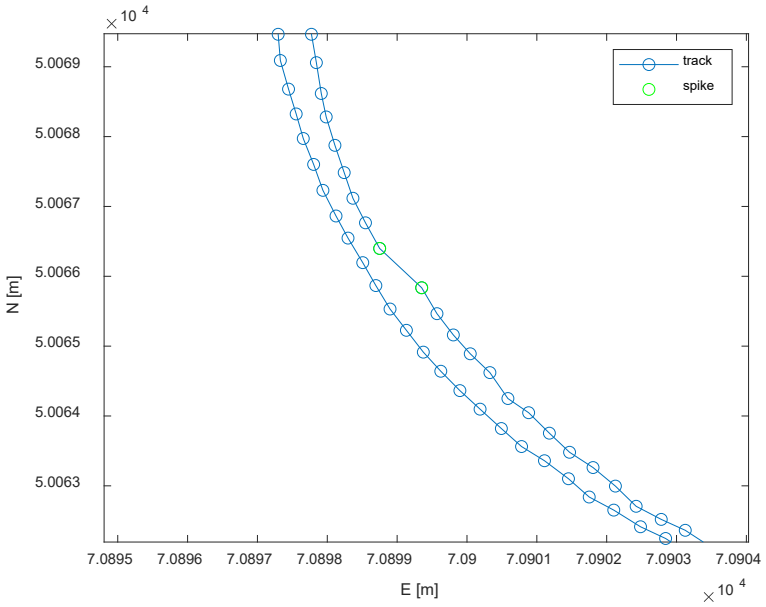

Figure 125: *Close-up of the position at the spike region. Blue circle indicates sampled points; green circle indicate the spike region.*

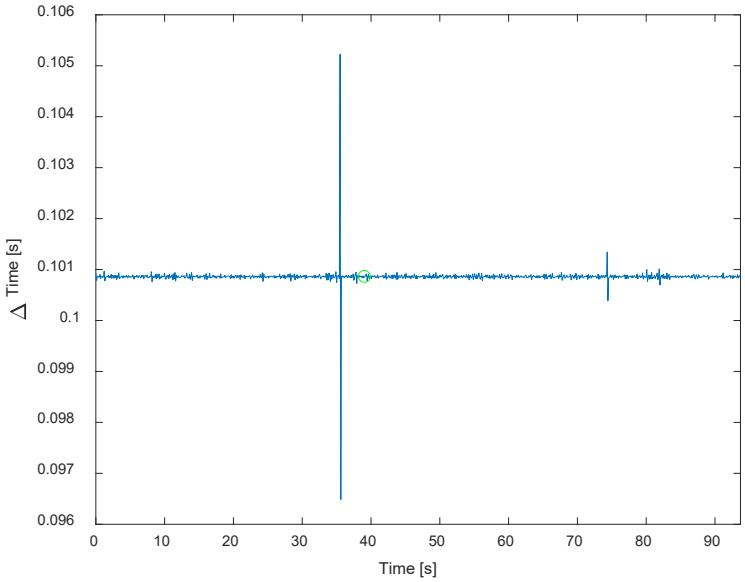

Figure 126: *NDT-Localizer sample time with spike time marked with green circle.*

Statistical Results

Figure 127 and Figure 128 show the distributions of the mean score and number of iterations of the single laps reported in Table 7. The sunny and rainy conditions show very similar distributions for both the score and the number of iterations, with significantly larger variability than other conditions, which is mostly linked to the larger amount of laps (35 and 20 laps, respectively). The tests with the contaminated sensor also result in similar scores and iterations as for sunny conditions, but in this case the variability is much lower (only 5 laps). Unexpectedly, the foggy conditions result in lower scores and iterations, but also in this case the number of laps is very small (5 laps) and we only infer from this that fog does not negatively affect the localization. The only condition that seems to have a negative effect is snow, which results in significantly higher scores. The variability of the score for the snowy condition is the largest among all conditions, despite corresponding to less test

runs (12 laps) than the results for sunny and rainy conditions. The number of iterations in the case of snow remains similar than that of all other conditions.

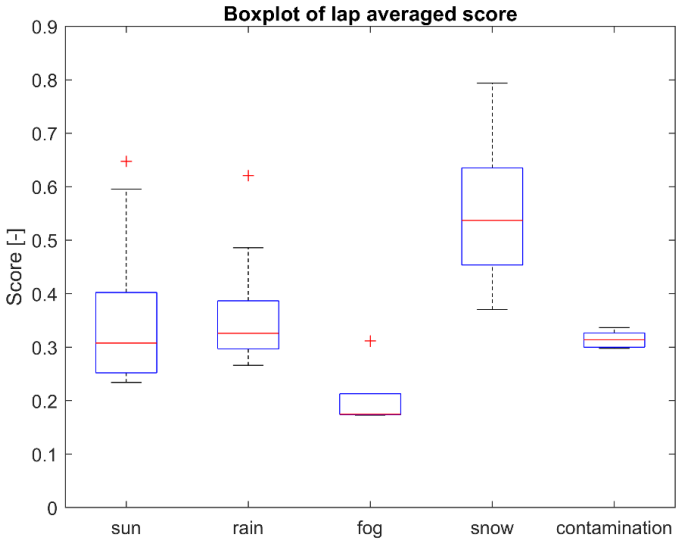

Figure 127: *Boxplot of lap averaged score for all conditions (35 laps for sun, 20 for rain, 5 for fog, 12 for snow and 5 for contamination).*

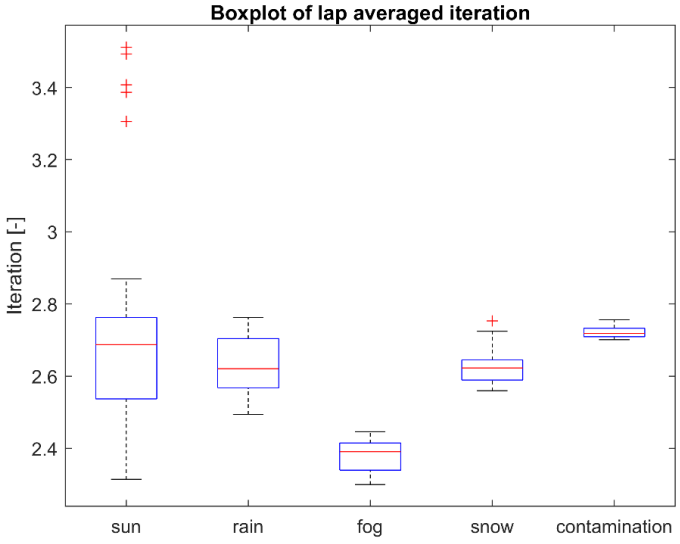

Figure 128: *Boxplot of lap averaged iteration at different conditions (35 laps for sun, 20 for rain, 5 for fog, 12 for snow and 5 for contamination).*

Table 7: Localization test runs for different environmental conditions

Date	Condition	Score				Number of iterations			
		μ	σ	min	max	μ	σ	min	max
7.4.2020	Sunshine, blue sky Steady driving	0.202	0.338	0.018	2.970	2.238	0.603	2	11
27.08.2020	Sunshine, few small clouds Steady driving	0.266	0.178	0.079	1.777	2.402	0.769	2	9
27.08.2020	Sunshine, few small clouds Dynamic driving	0.334	0.265	0.089	1.862	3.421	1.597	2	15
28.08.2020	Sunshine, few clouds Steady driving	0.352	0.282	0.102	2.017	2.554	0.882	2	12
30.09.2020	Sunny afternoon Steady driving	0.241	0.151	0.085	1.945	2.710	1.130	2	14
15.10.2020	Bright morning Steady driving	0.599	0.588	0.119	6.067	2.786	1.248	2	17
18.11.2020	Sunshine, blue sky Shadow light	0.314	0.312	0098	3.836	2.677	1.064	2	17
29.09.2020	Rainy, dim light Steady driving	0.336	0.283	0.105	2.612	2.570	0.909	2	15
29.09.2020	Rainy morning, dim light Steady driving	0.381	0.378	0.103	4.118	2.654	1.075	2	18
16.10.2020	Early morning, light rain 15 min before sunrise	0.386	0.372	0.086	3.207	2.576	0.982	2	12
19.11.2020	Light rain Dim light	0.309	0.350	0.094	4.240	2.77	1.131	2	13
18.11.2020	Foggy morning Dim light	0.203	0.144	0.058	1.920	2.379	0.078	2	11
27.2.2020	Heavy snowfall Steady driving	0.710	0.325	0.118	2.494	2.683	0.931	2	14
1.12.2020	Light Snowfall / Rain Snow covered road	0.449	0.413	0.109	4.589	2.605	1.013	2	16
1.12.2020	Snowfall Snow covered road	0.633	0.392	0.14	3.178	2.636	1.021	2	21
27.11.2020	Contaminated sensors	0.312	0.325	0.095	3.726	2.714	1.075	2	12

Discussion

Our results from the lidar localization tests under different weather conditions and contaminations indicate that, from a pure localization point of view, the effects of adverse weather conditions on lidar are not dramatic. Overall, we do not see much variation within the results and not much worsening of the localization performance in rain, fog or snow compared to sunny. Even when a full quarter of the field of view of the lidar was blocked due to the adhering snow, the localization algorithm never failed to converge. The statistical data in Table 7 and plotted in Figure 127 and Figure 128 support this conclusion, where, except in the case of snowfall, we see no significant increase in the number of iterations and score with worsening conditions. The only clearly visible increase in score is in snowy conditions, but even in this case the number of iterations until convergence did not change.

Our results therefore show that algorithm and sensor cannot always be considered separately. In fact, the localization algorithm had similar performance in all tested cases, even though the pure sensor performance differed significantly, as shown by the point cloud visualizations.

So far, we only investigated the localization in a structured urban environment where there are many structured objects in the FoV. More dimensions could be added and compared, such as driving in locations where the 3D map shows less features, e.g. highways, countryside roads…,. However, GNSS localization performance is less of an issue in these cases, since the line of sight to the satellites is less disturbed. But all in all, these additional considerations will be relevant in case of reliable localization of AD-vehicles in all driving environments.

7 Conclusion

In this project we have shown how AD sensors can be tested with function-scenario-based methods. The function-scenario-based testing provides the advantages from real-world AD testing, i.e. the sensor hardware is tested during real driving and not by mathematical models of the sensors, such as in simulations. Moreover, the sensors take measurements of the real world instead of being spoofed as in vehicle-in-the-loop or hardware-in-the-loop testing systems. Since the scenarios are well defined, the tests are repeatable and the relevant testing specifications such as weather conditions or object distance can be extracted. Moreover, the AD perception functions (e.g. object detection or object tracking) are tested separately, such that weaknesses and failure origins can be better detected. The combination with laboratory experiments allows verifying results from field tests in laboratory environments and vice versa. Besides the testing procedure, we have presented some sensor specific results for the Mobileye camera, Delphi ESR radar and Velodyne VLP-16 lidar.

Among all tested sensors, the Mobileye camera is clearly the most affected by the environmental conditions. In fact, our results clearly show that both the object detection and the object tracking functionalities of the Mobileye camera are negatively affected by adverse weather conditions, such as rain, fog or snow. The results from the Walking Pedestrian scenario show that in well light scenarios the camera has a tracking ratio close to 100 %, whereas in bad light conditions the tracking ratio reaches zero. In addition, adverse weather conditions result in a poorer tracking stability, i.e. the track is lost more often, the number of tracks created increases, and the average track duration decreases. Similar findings emerged from the analyses of the tests in the Urban Mixed scenario with regards to the object detection performance, which showed that the Mobileye camera FNR for vehicle and pedestrian detection significantly increases for snowy and rainy weather conditions. Overall our results demonstrate that the decrease in the light and the noise generated in the camera images from rain drops, snowflakes and fog negatively influence the object detection and tracking functionalities of the Mobileye camera.

The tests in the Walking Pedestrian and the Urban Mixed scenarios were also used to analyze the distance estimation functionality of the Mobileye camera. In both scenarios, the Mobileye camera resulted in very poor distance estimates under all environmental conditions. Overall, the Mobileye camera tends to underestimate the distances to the objects and the absolute error increases linearly with the distance. Only for the child pedestrian the Mobileye camera overestimates the distance at large distances, most probably because instead of a child it sees an adult pedestrian positioned further away.

In the Urban Mixed scenario we also investigated the effect from different object sizes and types. Our results show that the Mobileye camera detects better vehicles than pedestrians, and, among pedestrians, adult pedestrians are detected more easily than child pedestrians. These results are most possibly linked to the different object sizes, as bigger objects are represented by more pixels in the camera images, which facilitates their detection, but also differences in the Mobileye models for different object types might play a role (e.g. vehicles might have more characteristic shape than pedestrians).

Finally, in the Urban Mixed scenario we also looked at changes in the FNR of the Mobileye camera for pedestrian and vehicle detection as a function of the object distances. While the vehicle is properly detected at all tested distances (i.e. FNR close to zero), the detection of pedestrians worsens for larger distances, as the number of pixels of the object on the image decrease. Moreover, the FNR for pedestrian detection also increase at distances below 10 m, which we attribute to motion blur of the camera images due to the high relative tangential velocity between the moving AD car and the static object.

The Walking Pedestrian and the Mixed Urban scenarios were also used to analyze the object detection and tracking performance of the Delphi ESR radar. The tests in the Walking Pedestrian scenario revealed that the radar system has a lower track stability and tracking ratio than the Mobileye camera. Additionally, we observed that the performance of

the ESR radar is heavily influenced by the road surroundings. In fact, in the presence of certain facade geometries, metallic containers and fences, we observed a high number of track losses and gaps in the tracks of the walking pedestrian. On the other hand, when there is a track, the ESR radar provides a very precise distance measurement independent of the weather conditions and the object distance. However, due to the large influence from the surrounding environment, the radar is not reliable enough for pedestrian detection.

From our tests in the Mixed Urban scenario, we only analyzed the detection performance of the radar for vehicles, as the mannequins have a different radar cross-section than the human body, and would therefore not have led to any meaningful results. As in the Walking Pedestrian scenario, the Radar provided a very precise distance measurement, independent of the weather conditions and distance. In the case of the vehicle, also the detection and tracking of the ESR radar were very good, resulting in very low false negative rates for all distances and weather conditions (FNR always below 10 % and very close to 0 % for object distances between 5 and 70 m). Only in the case of the contaminated sensor we observed a drop in the radar detection performance between 40 and 50 m distance. So far, the reason for this effect is not clear to us. Probably the dirt, leaves and tire wear have some influence on the radar reflectivity, but further tests will be needed to confirm this.

One of the major downsides of the ESR radar revealed by our tests is its high sensitivity, which results in ghost detections (false positives) and reflections, as seen both in our real world tests and the laboratory measurements at METAS. For the use in an AD vehicle, this ghost tracks then have to be carefully dealt with and filtered out, leading to higher complexity and potentially increasing the error rate.

Generally, our findings show that, when considered separately, neither the Mobileye camera nor the Delphi ESR radar have an acceptable performance for pedestrian detection, as each sensor shows clear and reproducible weaknesses at specific environmental or test conditions. The Mobileye performance decreases with adverse weather conditions, while the radar shows performance issues when there are buildings and other objects along the road. Therefore an AD system, in order to be safe, would greatly benefit from multiple sensors and sensor fusion to reliably detect and track pedestrian.

We also investigated the performance of the detection and localization functionalities of the Velodyne VLP-16-A lidar sensor. For the evaluation of the detection functionality, we used our Lidar Target scenario. The results of these real world tests showed that the number of points reflected by the test objects. Interestingly, our tests showed big performance issues of the lidar for dark colored surfaces, such as black and dark blue. In particular, these darker colored surfaces are not seen by the lidar if the angle of incidence between the laser beam and the surface is anything but perpendicular. These findings from our real world tests were also confirmed by the results from the measurements at METAS. The fact that dark objects do not reflect energy at slightly non perpendicular angles poses a problem for the use of the lidar for object detection purposes. For the other tested colors (grey and white) we found overall more reasonable results, where the number of points on both targets where always very close to the theoretical number for all distances and weather conditions. For both colors, the measured reflectivity is only slightly affected by adverse weather conditions (at least not as much as the Mobileye camera). For sunny conditions, the reflectivity remains fairly constant over all distances with a decrease at distances below 10 m. The decrease below 10m might be linked to detection saturation and a limited numerical range. We also see an increased number of outliers in the reflectivity values with worse condition and overall slightly more variation.

The results from the Lidar Localization scenario demonstrate that the localization algorithm is quite robust against sensor noise and degradation from adverse environmental conditions. Although certain adverse weather conditions clearly degrade the lidar point clouds, we did not observe any significant worsening of the localization performance for any of the tested conditions, including sun, rain, fog, snow and contamination. In fact, even when a full quarter of the field of view of the lidar was blocked due to the adhering snow, the localization algorithm always converged successfully. Overall, our results show that for pure localization, the performance of the lidar itself is not the limiting factor.

Within this project we have investigated the performance of different AD sensors in real-world driving conditions. The questions that raise next are: how can we use these results for AD validation? How is the final AD performance related to the sensors performance? And what are the effects from sensor fusion and control algorithms? Specifications such as the FNR curves obtained from our tests can be used to create probabilistic sensor models that can then be used in simulations. Thereby, the sensor model would have the same uncertainty as the real hardware and software, which would make the simulation closer to the reality. Once precise and realistic specifications have been determined in real world tests, simulation can be used to validate many more scenarios and thereby speed up the validation process. Moreover, the performance analyses of individual sensor can help to build better sensor fusion models and state estimation models for AD, since they provide important insights into the accuracy and uncertainty of the sensors. Even without direct testing of the whole AD-stack, some general conclusions can also be made directly from the sensor testing results, e.g. our test vehicle would not be able to drive safely and efficiently at night during snowfall using only the currently installed sensors.

7.1 Outlook

The infrastructure and knowledge acquired from this project open up many opportunities and new research topics in the field of AD testing. In order to make the best use of the gained resources, we would like to continue our function-scenario-based sensor testing activities with additional sensors, perception algorithms and scenarios.

First, we want to repeat or extend some tests to further investigate some of the problems in the sensors performance that were discovered in this project. For example, the test with the contaminated radar will be repeated in order to understand the increased FNR observed for the vehicle detection. Also the second local minima in the FNR curves for pedestrian detection with the Mobileye camera for foggy and snowy weather conditions will be further investigated. Moreover, we also plan to test the sensors with thicker fog (or smoke) to see if/when the Mobileye camera performance will drop. Further, we want to repeat the vehicle detection experiments with a smaller vehicle to see differences to the Iveco Daily. The strange behavior of the lidar for short distances and dark surfaces will also be further analyzed, to try to find out if it is related to a sensor defect or if this is common to all lidars. For the localization performance, we want to determine the level of sensor degradation (in the form of noise from the environment) needed for the localization algorithm not to converge.

Moreover, we want to test new cameras and apply different open source perception algorithms. This would allow us to analyze the raw camera images as well as the algorithm working principles. Besides conventional RGB cameras, we plan to test also event-based and infrared cameras. We also want to use open source 3D object detention algorithms to evaluate the Velodyne VLP-16 lidar performance for pedestrian and vehicle detection. In this regards, it would be also interesting to test a lidar with more than 16 channels.

Regarding the environmental conditions, we want to develop methods to quantify the weather and contamination conditions with numbers, in order to, for instance, analyze the changes in the sensing performance during different levels of rainfall or contamination.

Besides the weather and contamination influence, we also want to investigate the aging and wear effects of AD sensors, as well as the interferences between similar sensors mounted on two different AD vehicles.

Our results from the function-scenario-based testing have shown that many parameters can influence the sensing performance: the vehicle velocity, the object distance, the object size, the weather conditions. Therefore, we want to repeat our tests from the Mixed Urban scenario with different vehicle velocities and potentially including also dynamic objects. These additional tests would help us determine all the relevant factors influencing the sensors' performance and make a sensitivity analysis of their relevance for AD perception performance.

The detailed analyses of the performance of the single sensors can then be used to analyze sensor fusion methods and to determine how well different sensor combinations work under different conditions. Finally, by using the sensor models derived from our real world testing in simulations, we can link the sensing performance to AD performance.

Appendix

I Test runs 140
I.1 Walking Pedestrian 140
I.2 Mixed Urban 140
I.3 Lidar Targets 141
I.4 Lidar Localization 142

I Test runs

I.1 Walking Pedestrian

Table 8: *All test runs from the Walking Pedestrian scenario.*

Date	Time	Condition
28.08.2020	15:26	Sunny with clouds
29.09.2020	09:35	Light rain
29.09.2020	09:40	Light rain
29.09.2020	09:44	Light rain
30.09.2020	15:10	Sunny
30.09.2020	15:20	Sunny
15.10.2020	11:10	Sunny with clouds
15.10.2020	11:5	Sunny with clouds
16.10.2020	07:23	Rain
16.10.2020	07:51	Rain
18.11.2020	09:33	Fog
18.11.2020	09:45	Fog
18.11.2020	15:23	Sunny
18.11.2020	15:27	Sunny
19.11.2020	11:41	Light rain
27.11.2020	10:43	Contamination, Sunny
01.12.2020	08:34	Snowfall
01.12.2020	08:40	Snowfall

I.2 Mixed Urban

Table 9: *All test runs from the Mixed Urban scenario.*

Date	Time	Condition
28.08.2020	15:00	Sunny with clouds
28.08.2020	15:07	Sunny with clouds
29.08.2020	09:20	Light rain
29.08.2020	09:31	Light rain
30.09.2020	14:54	Sunny
30.09.2020	15:00	Sunny
30.09.2020	15:02	Sunny
30.09.2020	15:28	Sunny
30.09.2020	15:30	Sunny
15.10.2020	10:55	Sunny with clouds
15.10.2020	10:59	Sunny with clouds
15.10.2020	11:24	Sunny with clouds
16.10.2020	07:06	Rain
16.10.2020	07:09	Rain
16.10.2020	07:11	Rain
16.10.2020	07:13	Rain
16.10.2020	07:33	Rain
16.10.2020	07:37	Rain
18.11.2020	09:08	Fog
18.11.2020	09:09	Fog
18.11.2020	09:13	Fog
18.11.2020	09:27	Fog

18.11.2020	09:28	Fog
18.11.2020	09:29	Fog
18.11.2020	15:43	Sunny
18.11.2020	15:45	Sunny
18.11.2020	15:46	Sunny
18.11.2020	15:47	Sunny
18.11.2020	15:49	Sunny
19.11.2020	11:25	Light rain
19.11.2020	11:26	Light rain
19.11.2020	11:27	Light rain
19.11.2020	11:29	Light rain
19.11.2020	11:30	Light rain
19.11.2020	11:31	Light rain
27.11.2020	10:34	Sunny, Contamination
27.11.2020	10:35	Sunny, Contamination
27.11.2020	10:36	Sunny, Contamination
27.11.2020	10:37	Sunny, Contamination
27.11.2020	10:38	Sunny, Contamination
01.12.2020	08:12	Snowfall
01.12.2020	08:13	Snowfall
01.12.2020	08:14	Snowfall
01.12.2020	08:16	Snowfall
01.12.2020	08:17	Snowfall
01.12.2020	08:47	Snowfall
01.12.2020	08:48	Snowfall
01.12.2020	08:49	Snowfall
01.12.2020	08:50	Snowfall
01.12.2020	08:52	Snowfall

I.3 Lidar Targets

Table 10: *All test runs from the Lidar Target scenario.*

Date	**Time**	**Condition**
28.08.2020	15:14	Sunny with clouds
28.08.2020	15:21	Sunny with clouds
29.08.2020	09:15	Light rain
30.09.2020	14:50	Sunny
30.09:2020	14.57	Sunny
30.09.2020	14:58	Sunny
15.10.2020	10:49	Sunny with clouds
15.10.2020	10:53	Sunny with clouds
15.10.2020	11:01	Sunny with clouds
15.10.2020	11:03	Sunny with clouds
16.10.2020	07:15	Rain
16.10.2020	07:18	Rain
16.10.2020	07:43	Rain
16.10.2020	07:45	Rain
18.11.2020	09:15	Fog
18.11.2020	09:16	Fog
18.11.2020	09:19	Fog
18.11.2020	09:20	Fog
18.11.2020	09:23	Fog

18.11.2020	09:24	Fog
18.11.2020	15:34	Sunny
18.11.2020	15:37	Sunny
18.11.2020	15:38	Sunny
18.11.2020	15:40	Sunny
18.11.2020	14:42	Sunny
19.11.2020	11:33	Light rain
19.11.2020	11:34	Light rain
19.11.2020	11:36	Light rain
19.11.2020	11:37	Light rain
19.11.2020	11:39	Light rain
27.11.2020	10:24	Sunny, Contamination
27.11.2020	10:26	Sunny, Contamination
27.11.2020	10:27	Sunny, Contamination
27.11.2020	10:29	Sunny, Contamination
27.11.2020	10:30	Sunny, Contamination
01.12.2020	08:18	Snowfall
01.12.2020	08:20	Snowfall
01.12.2020	08:23	Snowfall
01.12.2020	08:24	Snowfall
01.12.2020	08:25	Snowfall

I.4 Lidar Localization

Table 11: *All test runs from the Lidar Localization scenario.*

Date	Time	Condition
27.02.2020	14:47	Snow
27.04.2020	14:48	Sunny
27.04.2020	14:59	Sunny
27.8.2020	11:00	Sunny
27.8.2020	11:05	Sunny, dynamic Driving
28.09.2020	14:48	Sunny with clouds
29.09.2020	09:08	Light rain
29.09.2020	09:25	Light rain
30.09.2020	14:42	Sunny
15.10.2020	14:43	Sunny with clouds
16.10.2020	07:28	Rain
18.11.2020	09:00	Fog
18.11.2020	15:50	Sunny
19.11.2020	11:20	Light rain
27.11.2020	10:18	Sunny, Contamination
01.12.2020	08:07	Snowfall
01.12.2020	08:26	Snowfall

Nomenclature

Abbreviation	Definition
AD	Automated Driving
ADAS	Advanced Driver Assistance Systems
AESA	Active Electronically Scanned Array
AV	Automated Vehicle
CCD	Charge-Coupled Device
CMOS	Complementary Metal Oxide Semiconductor
DUT	Device Under Test
FMCW	Frequency Modulated Continuous Wave
FN	False Negative
FNR	False Negative Rate
FoV	Field of View
FP	False Positive
GPS	*Global Positioning System*
IoU	Intersection over Union
Lidar	Light Detection and Ranging
mAP	mean Average Precision
MEMS	Micro-Electro Mechanical System
MGRS	*Military Grid Reference System*
MOTA	*Multiple Object Tracking Accuracy*
NDT	*Normal Distribution Transform*
ODD	Operational Design Domain
RGB	Red Green Blue
ROS	Robot Operating System
RTK	Real Time Kinematic
Rx	Receiver
TN	True Negative
ToF	Time of Flight
TP	True Positive
Tx	Transmitter

Symbol	Definition
μ	Mean
Ae	Effective Antenna Aperture
c	Speed of light
c_x, c_y	Optical Center in x and y
d	Distance
E	East
f_t, f_r	Transmitted and Received Frequency
f_x, f_y	Focal Length in x and y
G	Antenna Gain
I	Intensity
K	Intrinsic camera matrix
N	North
n_{ca}	Number of object appearance in cell c
n_{ca}	Number of detected object in cell c
P_t, P_r	Transmitted and Received Power
r	Reflectivity

s	Skew coefficient
v_r	Radial velocity
σ	Standard Deviation
σ_{RCS}	Radar Cross Section

References

[1] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick and F. Diermeyer, "Survey on Scenario-Based Safety Assessmentof Automated Vehicles," *IEEE Access,* 19 03 2020.

[2] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?," *Transportation Research,* pp. 182-193, December 2016.

[3] S. Shalev-Shwartz, S. Shammah and A. Shashua, "On a Formal Model of Safe and Scalable Self-driving Cars," Mobileye, 2017.

[4] Waymo, [Online]. Available: https://twitter.com/Waymo/status/1214386514873839616. [Accessed 15 06 2020].

[5] C. Wang and H. Winner, "Overcoming Challenges of Validation Automated Driving and Identification of Critical Scenarios," in *IEEE Intelligent Transportation Systems Conference*, Auckland, 2019.

[6] Wikimedia Commons, "commons.wikimedia.org," [Online]. [Accessed 05 12 2019].

[7] Delphi, *Delphi Electronically Scanning RADAR Datasheet.*

[8] AutonomouStuff, "www.autonomoustuff.com," [Online]. Available: https://autonomoustuff.com/product/quantum-storage-solution-kits/. [Accessed 06 12 2019].

[9] F. Rosique, P. J. Navarro, C. Fernández and A. Padilla, "A Systematic Review of Perception System and Simulators for Autonomous Vehicles Research," Sensors, 2019.

[10] Velodyne LiDAR, Inc., *VLP-16 User Manual,* San Jose, CA, USA, 2017.

[11] Teledyne DALSA, "The Evolution of CMOS Imaging Technology," Canada, 2011.

[12] H. Caesar, V. Bakinti, 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," arXiv preprint arXiv:1903.11027, 2019.

[13] AutonomouStuff, *Mobileye Startup Guide,* 2017.

[14] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombar and N. Navab, "Deeper Depth Prediction with Fully Convolutional Residual Networks," in *IEEE International Conference on 3D Vision (3DV)*, 2016.

[15] H. Fu, M. Gong, C. Wang, B. K and D. Tao, "Deep Ordinal Regression Network for Monocular Depth Estimation," in *CVPR*, 2018.

[16] S. Ren, K. Hen, R. Girshick and J. Suan, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2015.

[17] R. Szeliski, Computer Vision, Springer, 2010.

[18] X. Chen, H. Ma, J. Wan, B. Li and T. Xia, "Multi-View 3D Object Detection Network for Autonomous Driving," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[19] Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection," *Computer Vision and Pattern Recognition,* 2018.

[20] C. Qi, W. Liu, C. Wu, H. Su and L. Guibas, "Frustum PointNets for 3D Object Detection from RGB-D Data," *Computer Vision and Pattern Recognition,* 2018.

[21] S. Shi, X. Wang and H. Li, "PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud," *Computer Vision and Pattern Recognition,* 2019.

[22] G. P. Stein, O. Mano and A. Shashua, "Vision-based ACC with a Single Camera: Bounds on Range and," in *IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No.03TH8683)*, Columbus, OH, USA, 2003.

[23] X. H. M. Elhousni, "A Survey on 3D LiDAR Localization for Autonomous Vehicles," in *IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, 2020.

[24] P. Biber and S. W, "The Normal Distributions Transform: A New Approach to Laser Scan Matching," in *Conference: Intelligent Robots and Systems*, 2003.

[25] R. Hartley and A. Zisserman, Multiple view geometry in computer vision, Cambridge: Cambridge University Press, 2003.

[26] D. C. Brown, "Close-Range Camera Calibration," *Photogrammetric Engineering,* pp. 855-866, 1971.

[27] Bernardin, Keni, A. Elbs and R. Stiefelhagen, "Multiple object tracking performance metrics and evaluation in a smart Room environment," in *Proceedings of IEEE International Workshop on Visual Surveillance*, 2006.

[28] Federal Institute of Metrology METAS, "Radar Delphi ESR2.5 24V, Measurement results of test block 4," Bern, 2020.

[29] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote sensing of Environment,* pp. 77-89, 1997.

[30] C. Agrell, "Gaussian processes with linear operator inequality constraints," 2019.

[31] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge and W. Brendel, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," 2019.

[32] Federal Institute of Metrology METAS, «Laser Scanner Measurement Velodyne VLP-16 (Testblock 4),» 2020.

[33] Xsens Technologies B.V., *MTi User Manual,* Enschede, Netherlands, 2018.

Project Completion

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

FORSCHUNG IM STRASSENWESEN DES UVEK Version vom 09.10.2013

Formular Nr. 3: Projektabschluss

erstellt / geändert am: 04.06.2021

Grunddaten

Projekt-Nr.:	ASTRA 2019/004
Projekttitel:	Automated Driving Sensor Testing Vehicle
Enddatum:	30.06.2021

Texte

Zusammenfassung der Projektresultate:

Das Projekt wurde gestartet, um Verfahren zur Sicherheitsvalidierung für automatisierte Fahren zu verstehen und aufzubauen. Sensoren spielen eine entscheidende Rolle im AD-Stack, da sie die relevanten Informationen aus der Umgebung liefern. Das Projekt konzentrierte sich auf die Evaluierung von AD-Sensoren mit Feldtests während des realen Fahrens auf einer Teststrecke und den Vergleich der Ergebnisse mit Daten aus experimentellen Labortests, die am METAS durchgeführt wurden. Zu diesem Zweck wurde ein Lexus RX 450h Fahrzeug mit einem handelsüblichen Satz der wichtigsten AD-Sensoren, d.h. Lidar, Kamera und Radar, ausgestattet. Weiterhin wurde auf dem Gelände des Empa-Campus in Dübendorf eine Teststrecke mit Testobjekten aufgebaut. Für die Bewertung der Leistungsfähigkeit der Sensoren wurde der so genannte funktions-szenariobasierte Testansatz verwendet. Bei diesem Verfahren wird eine bestimmte Wahrnehmungsfunktion im AD, z.B. die Objekterkennung, in einem entworfenen realen Fahrszenario getestet. Dieser Testansatz ermöglicht sowohl die Wiederholbarkeit von Feldtests als auch ein genaues Verständnis des Szenarios. Mit verschiedenen Szenarien haben wir die Objekterkennung, die Distanzschätzung, Tracking und Lokalisierungsleistung der Mobileye-Kamera, des Delphi ESR-Radars und des Velodyne VLP-16-Lidars getestet. Da es sich bei einigen der Sensoren, z. B. bei der Mobileye-Kamera, um Black-Box-Systeme handelt, haben wir neue Bewertungsspezifikationen entwickelt, z. B. die Tracking-Ratio, welche angibt, wie empfindlich bzw. robust die Objekttrackingfunktionalität über die Zeit ist. Das Design unserer Szenarien ermöglichte es uns, präzise Informationen über die Sensorleistung zu erhalten, z. B. die Falsch-Negativ-Rate der Objekterkennung bei unterschiedlichen Objektabständen. Die Feldtests wurden unter verschiedenen Wetterbedingungen und mit verschmutzten Sensoren durchgeführt. Unsere Analysen ergaben, dass die Mobileye-Kamera bei der Objekterkennung eine hohe Wetterabhängigkeit aufweist und grosse Leistungseinbussen bei verschneiten und regnerischen Wetterbedingungen zeigt. Ebenfalls signifikant beeinflusst der Objekttyp, -grösse und -entfernung die Erkennungsleistung der Mobileye Kamera. Fahrzeuge wurden insgesamt leichter erkannt als Fussgänger und grössere (erwachsene) Fussgänger leichter als kleinere (Kinder) Fussgänger. Generell verschlechterte sich die Erkennungsleistung stark bei grossen Entfernungen, aber auch bei sehr geringen Entfernungen, was wir auf Bewegungsunschärfe-Effekte durch das sich bewegende Testfahrzeug zurückführten. Die Entfernungsschätzungen der Mobileye-Kamera erwiesen sich unter allen getesteten Bedingungen als sehr schlecht, mit einer deutlichen Tendenz zur Unterschätzung der wahren Entfernung. Das ESR-Radar zeigte unter allen Umgebungsbedingungen eine gute Erkennungsleistung und lieferte sehr genaue Distanzmessungen. Wir beobachteten jedoch, dass die Objekttrackingfunktionalität recht empfindlich ist, da der Sensor oft Tracks verliert und/oder das gleiche Objekte mehrfach verfolgt. Der Velodyne-Lidar zeigte unter allen getesteten Umgebungsbedingungen eine gute Lokalisierungsleistung. Selbst ein teilweise blockiertes Sichtfeld durch angesammelten Schnee auf der Sensoroberfläche führte nicht zu einem Lokalisierungsfehler. Unsere Feldtests und Laborexperimente zeigten jedoch, dass dunkler gefärbte Oberflächen vom Lidar nicht gesehen werden, wenn der Einfallswinkel zwischen dem Laserstrahl und der Oberfläche nicht senkrecht ist, was ein grosses Problem für die Objekterkennung darstellt.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

Zielerreichung:

Den Projektzielen wurde mit der Literaturrecherche, dem Aufbau des Sensor Test Fahrzeuges und der Test-Infrastruktur inklusive Teststrecke auf dem Empa Campus in Dübendorf sowie der Durchführung und Auswertung von Feldtest zur Prüfung von Sensoren für automatisiertes Fahren unter realen Umweltbedingung Rechnung getragen.
Auch wenn angesichts der Komplexität der Themenstellung, der stetigen Weiterentwicklung der Technologie und der vielen Parameter welche die Leistungsfähigkeit der Sensoren beeinflussen nicht alle Fragen abschliessend beantwortet werden konnten, konnten doch viele interessante Ergebnisse gewonnen und weiter Forschungsbedarf aufgezeigt werden.
Das im Projekt erarbeitete Wissen konnte in mehreren Projektmeetings und Berichten mit den Teilnehmern der Begleitgruppe aus Forschung und Industrie geteilt werden. Daneben konnte ein breites Netzwerk aufgebaut werden um zukünftig weitere Forschung im Bereich des automatisierten Fahrens zu ermöglichen.

Folgerungen und Empfehlungen:

Die Infrastruktur und das erworbene Wissen aus diesem Projekt eröffnen viele Möglichkeiten und neue Forschungsthemen im Bereich der AD-Prüfung. Um die gewonnenen Ressourcen optimal zu nutzen, möchten wir unsere funktions-szenariobasierten Sensortestaktivitäten mit zusätzlichen Sensoren, Wahrnehmungsalgorithmen und Szenarien fortsetzen.

Die Ergebnisse dieses Projekt haben gezeigt, dass szenario-funktionsbasierte Feldtests wertvolle, wiederholbare Ergebnisse zur Performance von Sensoren liefern. Dabei können verschiedene Sensoren im selben Szenario untersucht und miteinander verglichen werden. Generell konnte gezeigt werden, wie die Performance von verschieden Sensor-Typen unterschiedlich durch das Wetter und die Umweltbedingungen beeinflusst werden. Unsere Ergebnisse haben auch gezeigt, dass viele Parameter die Erfassungsleistung beeinflussen können, wie zum Beispiel die Fahrzeuggeschwindigkeit, der Objektabstand, die Objektgrösse, die Wetterbedingungen etc. Zusätzliche Tests würden helfen, alle relevanten Faktoren, die die Leistung der Sensoren beeinflussen, zu bestimmen und eine Sensitivitätsanalyse ihrer Relevanz für die AD-Wahrnehmungsleistung durchzuführen.
Des weiteren zeigte sich, dass idealerweise die Sensoren und die Algorithmen welche die durch die Sensoren gewonnen Daten verarbeiten gemeinsam untersucht werden sollten, um festzustellen, ob die Performance des Sensors für eine Funktion ausreichen ist.

In Bezug auf die Umgebungsbedingungen sollen für zukünftige Projekte Methoden entwickelt werden, um die Wetter- und Verschmutzungsbedingungen mit Zahlen zu quantifizieren, um z.B. die Veränderungen in der Erfassungsleistung bei unterschiedlichen Regenfällen oder Verschmutzungen zu analysieren. Neben dem Wetter- und Verschmutzungseinfluss sollten auch die Alterungs- und Abnutzungseffekte von AD-Sensoren besser untersucht und quantifiziert werden, sowie die Interferenzen zwischen ähnlichen Sensoren, die auf zwei verschiedenen AD-Fahrzeugen montiert sind.

Publikationen:

D. Milojevic, Automated Driving Sensor Testing Vehicle, SCCER Mobility Annual Conference 2019, September 2019, Zürich
D.Milojevic, Sensor Testing and Perception Guaranties in Automated Driving, SCCER Mobility Webinar, März 2020, Zürich
C. Hohl, Evaluation of AD sensors' performance in real-world tests, Referat am SATW-Workshop: Connected and Automated Driving: In-vehicle and Deployment Enablers, November 2020, Zürich

Eine Publikation mit Resultaten aus dem Projekt ist in Vorbereitung

Der Projektleiter/die Projektleiterin:

Name: Hohl Vorname: Christian

Amt, Firma, Institut: Automotive Powertrain Technologies Laboratory, Empa

Unterschrift des Projektleiters/der Projektleiterin:

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

FORSCHUNG IM STRASSENWESEN DES UVEK

Formular Nr. 3: Projektabschluss

Beurteilung der Begleitkommission:

Beurteilung:

Das Projekt hat die Erwartungen hinsichtlich der Resultate grösstenteils erfüllt. Aufgrund der Restriktionen durch die Corona-Pandemie konnte der Zeitplan nicht ganz eingehalten werden. Wie im Antrag vorgesehen, konnten mit diesem Initialprojekt einige Erkenntnisse in Bezug auf die Sensor-Systemleistung (Sensor und Auswertealgorithmus) unter speziellen Witterungsbedingungen und bei verschiedenen Verschmutzungsgraden in realen reproduzierbaren Szenarien gefunden werden. Zudem wurde zusätzlicher Forschungsbedarf adressiert. Mit dem vorliegenden Projekt, welches innerhalb des vorgesehenen Kostenrahmens durchgeführt werden konnte, ist eine tragfähige Basis für weitergehende Untersuchungen gelegt worden. Insbesondere konnten Szenarien definiert und Auswertemethoden entwickelt werden, die auch für kommenden Projekte als Referenz dienen können.

Umsetzung:

Der Einbau und die Inbetriebnahme der Sensorsysteme im Fahrzeug waren erfolgreich. Bei den Versuchsanordnungen wurde der Reproduzierbarkeit grosse Bedeutung beigemessen, womit eine gute Basis für ergänzende Projekte geschaffen wurde. Viele Fragen sind aus der nicht immer ganz klaren Aufteilung von Sensor und deren Algorithmen entstanden. Dieser Black-Box-Situation muss in künftigen Untersuchungen mit entsprechenden Versuchsanordnungen begegnet werden, damit eine differenzierte Aussage möglich wird.

weitergehender Forschungsbedarf:

Im Schlussbericht werden mehrere Fragestellungen herausgestrichen, welche in künftigen Forschungsprojekten zu beantworten wären. Dies trifft einerseits die Distanzschätzungen, der Kamerasysteme, aber auch die Umgebungseinflüsse (Verschmutzung, Dunkelheit, Witterung, Umgebungs-/Hintergrundmerkmale) auf die Qualität der Sensordaten. Zudem könnten neuere Sensorsysteme mit berücksichtigt werden, um die Erkenntnisse auf dem aktuellen Stand der Technik zu halten. Vertiefte Kenntnisse der Sensoren und ihrer Einsatzgrenzen sind hilfreich um künftige Fragestellungen in Zusammenhang der Fahrzeugzulassung sicherer beantworten zu können.

Einfluss auf Normenwerk:

Die Sensortechnik automatisiert fahrender Strassenfahrzeuge unterliegen im Wesentlichen internationalen Vorgaben, wie den Reglementen der UNECE. Mit der vorliegenden Information kann die Vertretung der Schweiz in den entsprechenden GR (Groupe Rapporteur) namentlich GRVA und GRSG ihren Einfluss geltend machen und die Erkenntnisse einbringen.

Der Präsident/die Präsidentin der Begleitkommission:

Name: Gerster Vorname: Bernhard

Amt, Firma, Institut: ehem. Geschäftsführer DTC AG und Abteilungsleiter Automobiltechnik der BFH-TI

Unterschrift des Präsidenten/der Präsidentin der Begleitkommission: