



© Bertrant

Untersuchung von Lidardaten für das autonome Fahren mit einem Elektrobus

Elektromobilität und automatisiertes Fahren wachsen zusammen. Bertrant untersucht die Messdaten eines Laserscanners (Lidartechnik), der die Umgebung eines elektrischen Stadtbusses auf seiner rund 4 km langen Route durch Regensburg aufnahm. Auf Basis dieser Daten wurden zwei Verfahren zur Objektdetektion und -klassifikation untersucht. Eine der Herausforderungen war dabei die niedrige vertikale Auflösung im Vergleich zu den häufig üblichen 32 bis 128 Zeilen.

VERSUCHE MIT EINEM ELEKTRO-OMNIBUS

Von Mai bis September 2018 hatte Bertrant die Möglichkeit, den Elektro-Stadtbus Emil aus Regensburg als Forschungsplattform im Live-Betrieb zu nutzen. Ziel war es dabei, die Herausforderungen für ein zukünftiges autonomes Fahren im urbanen Umfeld auf-

zuzeigen und entsprechende Daten zu sammeln. Seit Beginn des Projekts setzte sich ein interdisziplinäres Team intensiv mit den Themen Lokalisierung, Datenübertragung, Lidarbildverarbeitung und Objekterkennung per maschinellem Lernen auseinander.

Der Datenpool nach rund 1800 Fahrten beläuft sich auf 14 TB, die über einen Zeitraum von 120 Tagen auf einer Gesamt-

AUTOREN



Dipl.-Ing. Christian Feller
ist Entwicklungsingenieur für
Informationstechnologie bei
Bertrandt in Regensburg.



Dipl.-Ing. Ulrich Haböck
ist Leiter für Strategie und Innovation
bei Bertrandt in Regensburg.



Stefan Maier, MCSE, MCITP
ist IT-Architekt und Connectivity-
Spezialist bei Bertrandt
in Regensburg.



Dipl.-Ing. Jochen Schwenninger
ist leitender Ingenieur und Funk-
tionsarchitekt für ADAS-Funktionen
und -Systeme bei Bertrandt
in Regensburg.

strecke von 5800 km eingefahren wurden. Bevor der E-Bus des italienischen Unternehmens Rampini Carlo S. p. A. mit der Datenaufnahme startete, wurde zusätzliche Sensorik im Bus verbaut. Dazu zählen im Wesentlichen ein Quanergy-M8-Lidar und ein GNSS-Empfänger mit einer integrierten Inertialsensorik (Inertial Measuring Unit, IMU) von Hexagon. Eine Recheneinheit (Intel-i7-Rechenkern und Nvidia-GPU) übernahm die Sensorsteuerung, die Datenaufnahme und teilweise die Übertragung der Daten in Echtzeit per LTE ins Backend (Microsoft Azure Cloud). Das komplett eigenständige System ohne Verbindung zu den Fahrzeug-Bussystemen oder dem Fahrer wurde über das 24-V-Bordnetz versorgt. Für die Aufzeichnung der Sensordaten und deren Verarbeitung wurde das Software-Framework ROS [1] in der Version „Kinetic Kame“ eingesetzt. **BILD 1** zeigt die Gesamtarchitektur der im Versuchsträger eingebauten Hardware.

AUSWAHL DES SENSORS

Ein Lidarsystem bietet bei einer Sensorfusion für das autonome Fahren eine interessante Mischung aus räumlicher Auflösung sowie Tiefeninformation und ergänzt so die traditionelleren Sensoren Radar und Kamera. Der Lidarsensor mit einer Reichweite von 150 m bei 80 % Reflektivität ermöglicht es, die Umgebung, die der Bus auf seiner rund 4 km langen Route durch die Regensburger Altstadt mehrmals täglich abfährt, wiederholt aufzunehmen. Das Ergebnis ist

ein Graustufenbild (Reflektivität) mit acht Zeilen und bis zu 10.000 Punkten pro Zeile sowie die dazugehörige Tiefeninformation (Abstand).

Montiert wurde der Sensor aufgrund des bereits sehr hohen Busses nicht auf dem Dach, sondern an der Fahrzeugfront in etwa 3 m Höhe. Nutzbar ist deswegen nicht die volle 360°-Sicht, sondern lediglich 180°. Auf Basis dieser Daten wurden zwei Verfahren zur Objektdetektion und -klassifikation untersucht. Eine der Herausforderungen war dabei die niedrige vertikale Auflösung im Vergleich zu den häufig üblichen 32 bis 128 Zeilen, die von den deutlich teureren Referenzsensoren im Automotive-Umfeld geliefert werden.

OBJEKTERKENNUNG ÜBER ZWEI ALGORITHMEN

Die vorherrschende Intention bestand darin, relevante Objekte in der Umgebung des Busses zu detektieren und zu klassifizieren. Dabei wurden die Klassen „Fahrzeug“ und „Fußgänger“ betrachtet. Aufgrund der hohen Variabilität der Daten wurden Ansätze aus dem maschinellen Lernen eingesetzt. Für das Training waren Daten notwendig, die neben den eigentlichen Sensordaten auch das gewünschte Ergebnis der Objekterkennung, sogenannte Labels, enthalten. Diese werden üblicherweise in Form von Bounding Boxes angegeben, die das Objekt vollständig umschließen. Als Quelle für die Trainingsdaten wurde der 3-D-Objekterkennungsdatensatz Kitty genutzt [2], **BILD 2**.

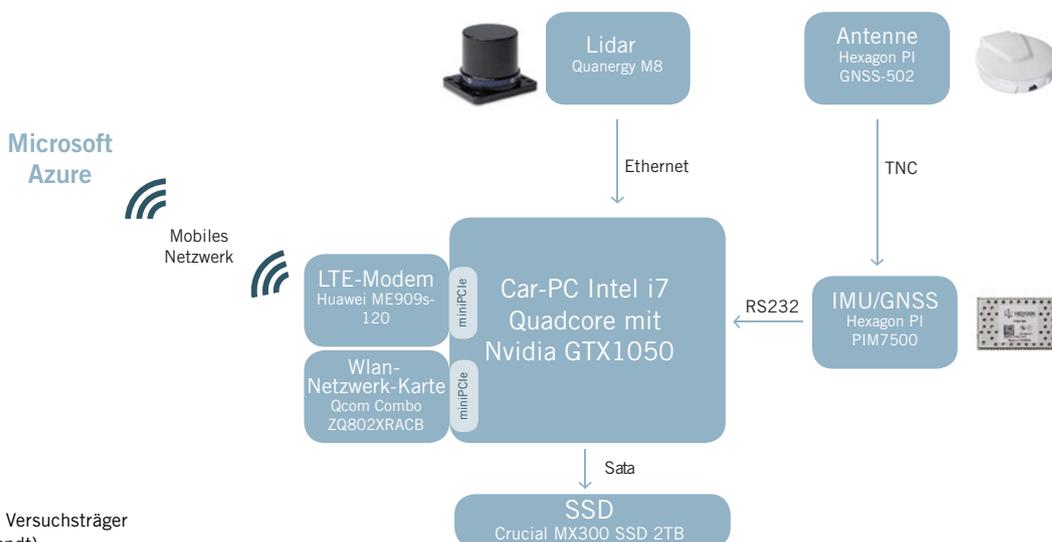


BILD 1 Gesamtarchitektur der im Versuchsträger
eingebauten Hardware (© Bertrandt)

Um die Qualität der Verfahren zu messen, wurden die vom jeweiligen Algorithmus generierten Label mit manuellen Referenzannotationen verglichen. Bei einer ausreichend großen Übereinstimmung der Bounding Boxes wird die Vorhersage als wahr-positiv anerkannt, sonst zählt das Objekt als Fehler. Bewertet werden Verfahren im Umfeld von Kitti normalerweise über die durchschnittliche Präzision (Mean Average Precision, MAP).

Es wurden zwei Algorithmen im direkten Vergleich eingesetzt, um die Objekte zu erkennen: Ein Ansatz orientiert sich an der 2-D-Objektdetektion Yolo [3]. Der zweite Ansatz implementiert und adaptiert für den eingesetzten Lidarsensor die Ideen aus VoxelNet [4].

Der grundlegende Ansatz des Yolo-basierten Algorithmus ist es, aus der 3-D-Punktwolke eine 2-D-Darstellung der Fahrzeugumgebung zu generieren und mit einem etablierten Ansatz aus der Bildverarbeitung in dieser Darstellung die gewünschten Objekte zu detektieren. Implementiert wurde zunächst die Frontansicht. Im Anschluss an die 2-D-Erkennung wurde die fehlende Koordinate aus der zugrunde liegenden Punktwolke bestimmt, um ein echtes 3-D-Objekt angeben zu können. Der Vorteil bei diesem Ansatz ist, dass Zwischenergebnisse wie etwa die 2-D-Projektion oder die 2-D-Objekte gut manuell überprüft werden können und bei der Auswahl geeigneter neuronaler Netze

auf bereits umfangreich evaluierte Architekturen aus der klassischen Bildverarbeitung als neuronale Netze zurückgegriffen werden konnte.

Für den Ansatz auf Basis von VoxelNet wird aus der Punktwolke ein Belegungs-gitter generiert, das den 3-D-Raum um den Sensor in Quader einteilt und für jeden Quader (Größe zum Beispiel $0,4 \times 0,2 \times 0,2$ m) verschiedene Eigenschaften oder Features wie etwa die Anzahl der Punkte im Quader oder ihren typischen Abstand zueinander mithilfe neuronaler Netze bestimmt. Dieses mit den Eigenschaften verbundene Gitter bildet die Eingangsdaten für ein weiteres neuronales Netz (Region Proposal Network), das mit typischen Umrissen und Umgebungen trainiert wird. Am Ausgang dieser Verarbeitung werden dann sowohl die Klasse eines erkannten Objekts als auch seine 3-D Bounding Box ausgegeben.

Für beide Verfahren wurden die Daten für Training und Evaluation aus Kitti an den verwendeten Sensor angepasst. Hierzu zählt unter anderem die Auswahl der acht ähnlichsten Strahlen aus den 64 verfügbaren sowie die Maskierung der dadurch nicht sichtbaren Objekte in der Umgebung und die Korrektur der Sensor-Einbauhöhe, **BILD 3**. Somit ist, aufgrund der reduzierten Informationen, eine Verschlechterung der Erkennungsergebnisse gegenüber einem Verfahren auf Basis der kompletten Daten zu erwarten.

Zum Projektende wurden die beiden implementierten Algorithmen mit den

jeweils besten Ergebnissen im offiziellen Leistungsvergleich für die Detektion von Fahrzeugen in der Draufsicht (Vogelperspektive) verglichen. Das UberTAG-HDNet [5] der Universität Toronto, das zusätzliche Trainingsdaten verwendet, um typische Karten zu lernen und damit die statische Umgebung besser ausblenden kann, liefert bei komplett sichtbaren Fahrzeugen (Kitti Easy) eine MAP von 89 %.

Im Rahmen des Innovationsprojekts bei Bertrand mit nur sechs Monaten Projektlaufzeit wurde beim Ansatz nach VoxelNet mit acht Höhengschichten eine MAP von 59 % für die gleichen Fahrzeuge erreicht. Der alternative Ansatz mit Yolo erzielt immerhin noch 34 % MAP, ohne den Drehwinkel der Objekte in der XY-Ebene zu bestimmen. Auch bei der Aufgabe der Fußgängerdetektion lässt sich dieser Trend beobachten. Die Universität Waterloo (AVOD-FPN [6]) erreichte hier die bisher besten Ergebnisse mit Kitti-Daten (Lidar und Kamera): 59 % der einfachen Daten wurden korrekt klassifiziert. Im Rahmen des vorgestellten Projekts erreichte der Ansatz nach VoxelNet mit den auf acht Höhengschichten reduzierten Daten 41 % MAP, während sich mit Yolo eine Genauigkeit von 19 % erreichen lässt, ohne den Drehwinkel zu schätzen.

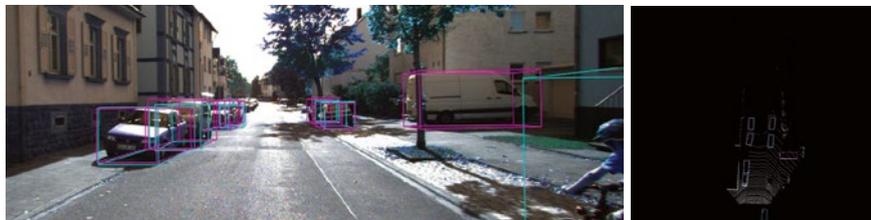
FUSION VON LIDARINFORMATIONEN

Mit den gewonnenen Daten kann die Objekterkennung noch weiter optimiert werden. Neue Ansätze lassen sich evaluieren, ohne dass neue Versuchsfahrten aufgenommen werden müssen. Insbesondere die zeitliche Abfolge der Lidar-Frames kann ausgenutzt werden, um ein genaueres Abbild der Umgebung zu bekommen und potenziell erkannte Objekte zu plausibilisieren oder zu verwerfen. Auch die Fusion der Lidarinformationen mit weiteren Sensordaten wird für zukünftiges automatisiertes Fahren notwendig sein.

LOKALISIERUNG IM URBANEN UMFELD

Ein weiterer Schwerpunkt des Projekts lag auf der Lokalisierung. Eine satellitenbasierte Navigation über GPS oder Galileo bietet ohne Korrekturdaten eine Positionsgenauigkeit von etwa 10 m. Im urbanen Umfeld stellen Häuserschluchten ein Problem dar, da die freie Sicht zum Himmel

Trainingsdaten Kitti



Autos werden zuverlässig erkannt



Fußgänger in der Nähe können erkannt werden

BILD 2 Beispiele für die Lidardaten im Kitti-Leistungsvergleich: Die Referenzlabels sind in Magenta dargestellt, die automatisch erstellten Labels in Türkis (© Bertrand)

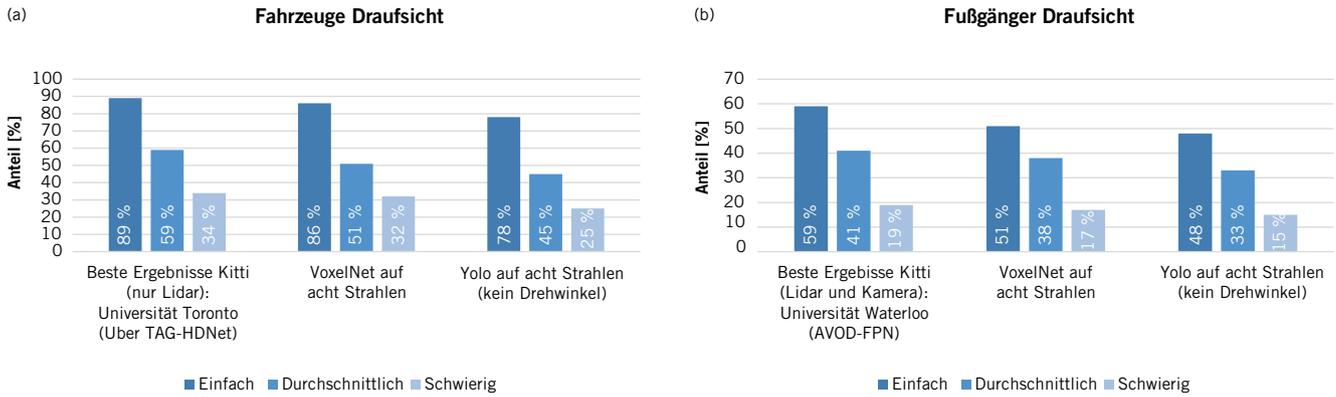


BILD 3 Ergebnisse für die Detektion von Fahrzeugen (a) und Fußgängern (b): Der Schwierigkeitsgrad hängt unter anderem von der Verdeckung und Größe der Objekte ab; die Referenzalgorithmen arbeiten auf allen 64 Strahlen; VoxelNet und Yolo verwenden lediglich acht Strahlen (© Bertrandt)

nicht mehr durchgehend gewährleistet ist und somit die typische Fehlpositionierung ansteigt. Um dem entgegenzuwirken, lässt sich die Eigenbewegung mithilfe einer IMU beobachten und mit der GNSS-Position fusionieren. Leider führen bereits kleine Fehler in den Messwerten zu einer Drift der Position, sodass dieser Ansatz keine langzeitstabile Position liefert.

Im Gegensatz dazu integrieren Simultaneous Localization and Mapping mit der statischen Umgebung eine zweite absolute Informationsquelle für die Position. Bei diesen Verfahren werden anhand von Sensorinformationen sowohl die Position des Fahrzeugs bestimmt als auch eine Karte der Umgebung aufgebaut. **BILD 4**. Die Grundidee dabei

ist, dass geänderte Distanzen zu Wänden oder Ecken auf eine Eigenbewegung hindeuten und damit ein Indiz für eine geänderte Position sind. Unterschiedliche Implementierungen etwa für mobile Roboter sind bereits im Einsatz.

In der Projektlaufzeit wurden mehrere Verfahren evaluiert, die aber bislang noch nicht mit der relativ dynamischen Umgebung einer Innenstadt zurechtkamen. Daher wurde ein Abbild der Umgebung erstellt, indem die Lidareindrücke an die jeweilige GNSS/IMU-Position projiziert und dann über eine komplette Fahrt aggregiert wurden. Über mehrere Fahrten lässt sich diese Darstellung noch weiter verfeinern und präzisieren, sodass am Ende eine genaue Karte der Strecke vorliegt.

HERAUSFORDERUNG DER DATENÜBERTRAGUNG

Ein weiterer zentraler Punkt, der sich während des Projekts deutlich zeigte, war, dass die Netzverbindung in Regensburg grundsätzlich stabil ist, sich aber reproduzierbar an denselben Stellen verschlechtert. In diesen Situationen wurden Latenzen von über 1000 ms beobachtet, gleichzeitig sank die nutzbare Datenrate auf unter 4000 kbit/s, **BILD 5**. Die Gründe dafür lagen unter anderem in überlasteten Funkzellen an hochfrequentierten Stellen und den engen Häuserschluchten in der Regensburger Altstadt.

FAHRZEUGE ALS DATENPRODUZENTEN

Wenn man einen Ausblick zum Thema Datenübertragung geben möchte, wird ein großer Vorteil von (teil-)autonomen Fahrzeugen sein, dass sie umfangreiche Informationen über ihre Sensoreindrücke und geplanten Manöver mit anderen Fahrzeugen in der Umgebung sowie mit Backendsystemen teilen können. Schon heute werden Fahrzeuge immer mehr zu Datenproduzenten.

Um diese Daten für eine Vielzahl an Fahrzeugen zuverlässig übertragen zu können, ist die aktuelle Netzinfrastruktur mit 4G nicht ausreichend leistungsfähig. Daher wird verstärkt auf den kommenden Standard 5G gesetzt, der neben höheren Bandbreiten auch die Möglichkeit für die Funkverbindung von Informationen bieten wird. Neben der Car-to-X-Ver-netzung ist auch die Verteilung von



BILD 4 Automatisch erstellte Darstellung der Fahrzeugumgebung in Echtzeit auf einer Fahrt (daher reduzierte Auflösung); im Hintergrund zum Vergleich ein Satellitenbild (© Bertrandt)



BILD 5 Latenz (oben) und Bandbreite (unten) der Mobilfunkverbindung während einer exemplarischen Fahrt durch Regensburg; beide Werte deuten darauf hin, dass die Kapazitätsgrenze fast erreicht ist (© Bertrandt)

Korrekturdaten für satellitenbasierte Positionslösungen ein Anwendungsfall.

Die gesammelten Rohdaten des Innovationsprojekts wurden bereits während des laufenden Betriebs von internen Projektgruppen zum Kompetenzaufbau im Bereich autonomes Fahren gesammelt und ausgewertet. Im Sinne des Open-Innovation-Ansatzes wird Bertrandt die Erkenntnisse nicht nur mit Kunden teilen, sondern auch mit Kooperationspartnern weiterentwickeln.

LITERATURHINWEISE

- [1] OSR Foundation: Software framework of the Robot Operating System (ROS). Online: <https://www.ros.org/>, aufgerufen am 09.11.2018
- [2] Geiger, A.; Lenz, P.; Urtasun, R.: Are we Ready for Autonomous Driving? The Kitti Vision Benchmark Suite. In: Proceedings of the IEEE Conference on

- Computer Vision and Pattern Recognition (CVPR), Providence (USA), 2012
- [3] Redmon, J.; Diwala, S.; Girshick, R.; Farhadi, A.: You Only Look Once (Yolo): Unified, Real-Time Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas (USA), 2016
- [4] Zhou, Y.; Tuzel, O.: VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City (USA), 2018
- [5] Yang, B.; Liang, M.; Urtasun, R.: HDNet: Exploiting HD Maps for 3D Object Detection. In: 2nd Conference on Robot Learning (CoRL), Zürich (Schweiz), 2018
- [6] Ku, J.; Mozifian, M.; Lee, J.; Harakeh, A.; Waslander, S.: Joint 3D Proposal Generation and Object Detection from View Aggregation. In: International Conference on Intelligent Robots and Systems (Iros), Madrid (Spanien), 2018



READ THE ENGLISH E-MAGAZINE

Test now for 30 days free of charge:
www.atz-worldwide.com