



KLASSIFIKATION HYPERSPÉKTRALER DATEN
ZUR BEFAHRBARKEITSANALYSE

von

Christian Winkens

Genehmigte Dissertation zur Verleihung
des akademischen Grades eines

Doktor der Naturwissenschaften (Dr. rer. nat.)

Fachbereich 4: Informatik
Universität Koblenz-Landau

Vorsitzender der Promotionskommission **Prof. Dr. Patrick Delfmann**

Vorsitzender des Promotionsausschusses **Prof. Dr. Ralf Lämmel**

Erster Berichterstatter und Betreuer **Prof. Dr.-Ing. Dietrich Paulus**

Zweiter Berichterstatter **Prof. Dr.-Ing. Bernhard Hill**

Datum der wissenschaftlichen Aussprache **11.03.2021**

ZUSAMMENFASSUNG

Der Wettbewerb um die besten Technologien zur Realisierung des autonomen Fahrens ist weltweit in vollem Gange. Trotz großer Anstrengungen ist jedoch die autonome Navigation in strukturierter und vor allem unstrukturierter Umgebung bisher nicht gelöst. Ein entscheidender Baustein in diesem Themenkomplex ist die Umgebungswahrnehmung und Analyse durch passende Sensorik und entsprechende Sensordatenauswertung. Insbesondere bildgebende Verfahren im Bereich des für den Menschen sichtbaren Spektrums finden sowohl in der Praxis als auch in der Forschung breite Anwendung. Dadurch wird jedoch nur ein Bruchteil des elektromagnetischen Spektrums genutzt und folglich ein großer Teil der verfügbaren Informationen zur Umgebungswahrnehmung ignoriert. Um das vorhandene Spektrum besser zu nutzen, werden in anderen Forschungsbereichen schon seit Jahrzehnten sog. spektrale Sensoren eingesetzt, welche das elektromagnetische Spektrum wesentlich feiner und in einem größeren Bereich im Vergleich zu klassischen Farbkameras analysieren. Jedoch können diese Systeme aufgrund technischer Limitationen nur statische Szenen aufnehmen. Neueste Entwicklungen der Sensortechnik ermöglichen nun dank der sog. Snapshot-Mosaik-Filter-Technik die spektrale Abtastung dynamischer Szenen. In dieser Dissertation wird der Einsatz und die Eignung der Snapshot-Mosaik-Technik zur Umgebungswahrnehmung und Szenenanalyse im Bereich der autonomen Navigation in strukturierten und unstrukturierten Umgebungen untersucht. Dazu wird erforscht, ob die aufgenommenen spektralen Daten einen Vorteil gegenüber klassischen RGB- bzw. Grauwertdaten hinsichtlich der semantischen Szenenanalyse und Klassifikation bieten. Zunächst wird eine geeignete Vorverarbeitung entwickelt, welche aus den Rohdaten der Sensorik spektrale Werte berechnet. Anschließend wird der Aufbau von neuartigen Datensätzen mit spektralen Daten erläutert. Diese Datensätze dienen als Basis zur Evaluation von verschiedenen Klassifikatoren aus dem Bereich des klassischen maschinellen Lernens. Darauf aufbauend werden Methoden und Architekturen aus dem Bereich des Deep-Learnings vorgestellt. Anhand ausgewählter Architekturen wird untersucht, ob diese auch mit spektralen Daten trainiert werden können. Weiterhin wird die Verwendung von Deep-Learning-Methoden zur Datenkompression thematisiert. In einem nächsten Schritt werden die komprimierten Daten genutzt, um damit Netzarchitekturen zu trainieren, welche bisher nur mit RGB-Daten kompatibel sind. Abschließend wird analysiert, ob die hochdimensionalen spektralen Daten bei der Szenenanalyse Vorteile gegenüber RGB-Daten bieten.

ABSTRACT

The competition for the best technologies to achieve autonomous driving is in full swing worldwide. The idea of autonomously acting systems fascinates mankind. Despite great efforts, autonomous navigation in structured and especially unstructured environments has not yet been solved. A crucial component in this field is the perception and analysis of the environment, which is accomplished by suitable sensor technology and corresponding sensor data analysis. Imaging sensors, which imitate the functionality of the human eye, are widely used for this purpose. However, only a fraction of the electromagnetic spectrum emitted by the sun or other light sources is used. Consequently, a large part of the available information on environmental perception is neglected. In other research areas, so-called spectral sensors have been used for decades. They are used to analyse the electromagnetic spectrum in a much finer and wider range. However, due to technical limitations these systems can only capture static scenes. The latest developments in sensor technology now enable spectral measurements of dynamic scenes thanks to the so-called Snapshot-Mosaic filter technique. In this dissertation the application and suitability of the Snapshot-Mosaic technique for environmental perception and scene analysis in the field of autonomous navigation in structured and unstructured environments is investigated. For this purpose it is explored whether the captured spectral data offer an advantage over classical RGB or greyscale data with respect to semantic scene analysis and classification. Thus, different aspects of the evaluation and analysis of these data are examined in consecutive thematic blocks. A suitable pre-processing system is presented first, which computes spectral values from the raw data of the sensor system. Afterwards the assembly of novel datasets with spectral data is explained. These datasets serve as a basis for the evaluation of different classifiers from the field of classical machine learning, which will be discussed in the following. Based on this, methods from the field of deep learning are presented, which have established themselves as state-of-the-art in the field of semantic scene analysis of classical color image data. By applying selected architectures it will be examined whether they can be trained with spectral data. Furthermore, the use of deep learning methods for data compression is discussed. In a next step, the compressed data will be used to train network architectures that are currently only tailored for RGB data. Finally, it will be evaluated whether the high-dimensional spectral data offer advantages over RGB data in scene analysis.

EIGENE PUBLIKATIONEN

- [1] Christian Winkens, Volkmar Kobelt und Dietrich Paulus. Robust features for snapshot hyperspectral terrain-classification. In Michael Felsberg, Anders Heyden und Norbert Krüger (Editoren), *Computer Analysis of Images and Patterns (CAIP)*, Seiten 16–27, Cham, 2017. Springer International Publishing. ISBN: 978-3-319-64689-3.
- [2] Christian Winkens und Dietrich Paulus. Context aware hyperspectral scene analysis. *Electronic Imaging*, 2018(14):346–1–346–7, 2018. DOI: 10.2352/ISSN.2470-1173.2018.17.AVM-346. ISSN: 2470-1173.
- [3] Christian Winkens, Florian Sattler, Veronika Adams und Dietrich Paulus. Hyko: A spectral dataset for scene understanding. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Seiten 254–261, Oct 2017. DOI: 10.1109/ICCVW.2017.39. ISSN: 2473-9944.
- [4] Christian Winkens, Florian Sattler, Veronika Adams und Dietrich Paulus. Vorverarbeitung hyperspektraler bilddaten von snapshot-mosaik-kameras. In *23. Workshop Farbbildverarbeitung, 05.-06. Oktober 2017*, Seiten 47–58. Fogra Forschungsgesellschaft Druck e.V. München, 2017. ISBN: 978-3-00-057941-7.
- [5] Christian Winkens, Florian Sattler und Dietrich Paulus. Hyperspectral terrain classification for ground vehicles. In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISIGRAPP (VISIGRAPP 2017)*, Seiten 417–424, Porto, Portugal, 2017. INSTICC, SciTePress. ISBN: 978-989-758-226-4.
- [6] Christian Winkens, Florian Sattler und Dietrich Paulus. Deep dimension reduction for spatial-spectral road scene classification. *Electronic Imaging*, 2019(15):49–1–49–9, 2019. DOI: doi:10.2352/ISSN.2470-1173.2019.15.AVM-049. ISSN: 2470-1173.

Die eigenen Veröffentlichungen sind durch einen numerischen Referenzschlüssel ([4]) kenntlich gemacht, wohingegen referenzierte Publikationen einen alphanumerischen Schlüssel ([scho7a]) aufweisen. Die Schlüssel von Internetquellen sind durch ein vorangestelltes @ ([@2]) gekennzeichnet.

DANKSAGUNGEN

Die Grundlagen dieser Dissertation entstanden während meiner Zeit als wissenschaftlicher Mitarbeiter in der Arbeitsgruppe Aktives Sehen bei Herrn Prof. Dr.-Ing. Paulus. Ihm danke ich für das entgegengebrachte Vertrauen, das diese Arbeit erst ermöglichte und die vielen Gespräche und die daraus entstandenen konstruktiven Anmerkungen und Anregungen, welche die Arbeit maßgeblich beeinflusst und vorangetrieben haben. Weiterhin gilt mein Dank der ganzen Arbeitsgruppe für die vielen konstruktiven Diskussionen.

Besonders möchte ich hier Christian Fuchs, Florian Sattler und Frank Neuhaus für die vielen konstruktiven Diskussionen und Anregungen in all den Jahren danken. Weiterhin danke ich auch meinen wissenschaftlichen Hilfskräften für deren fleißige Arbeit besonders beim Annotieren der Daten.

Des Weiteren gilt mein ausdrücklicher Dank der WTD 41 sowie dem BAAINBw U 6.2 für die umfangreiche Unterstützung der Arbeiten im Rahmen der Dissertation. Auch möchte ich meinen Eltern dafür danken, dass sie mir eine gute Ausbildung ermöglicht haben. Weiterhin bedanke ich mich für das Korrekturlesen bei Klaus, Christian, Nick und Irina.

Mein Dank für ihre Unterstützung und Hilfe gilt zudem meiner ganzen Familie und auch allen Freunden. Ein ganz besonderer Dank geht an meine Frau und meinen Sohn, ohne deren Unterstützung und Geduld es nicht möglich gewesen wäre, die Dissertation fertigzustellen.

INHALTSVERZEICHNIS

1	EINLEITUNG	1
1.1	Eigener Beitrag	10
2	STAND DER TECHNIK	13
2.1	Einführung	13
2.2	Spektrale Bildgebung	14
2.3	Fernerkundung	15
2.3.1	Geschichte der spektralen Bildgebung	17
2.3.2	Klassifikationsverfahren	19
2.3.3	Fazit	22
3	GRUNDLAGEN	23
3.1	Einführung	23
3.2	Bildsegmentierung	23
3.2.1	Sensoren	24
3.2.2	Effekte der Bildgebung	27
3.3	Verwendete Sensorik	28
3.3.1	Kameras	28
3.3.2	Auslöser-Board	31
3.3.3	Sensor-Plattform	32
3.4	Bilderzeugung	33
3.5	Strahlung und Wahrnehmung	34
3.5.1	Elektromagnetisches Spektrum	34
3.5.2	Licht und Farbe	36
3.5.3	CIE-Normvalenzsystem	40
3.5.4	Spektral nach RGB	43
3.6	Spektrale Bildgebung	46
3.6.1	Einführung	46
3.6.2	Snapshot-Mosaik-Technik	47
3.6.3	CMOS-Implementierung	50
3.7	Bilddefinition	55
3.8	Evaluationsmetriken	58
4	VORVERARBEITUNG	63
4.1	Einführung	63
4.2	Stand der Technik	63
4.3	Datenvorverarbeitung	65
4.4	RGB-Generierung	75
4.5	Fazit	76
5	SPEKTRALE DATENSÄTZE	79
5.1	Einführung	79
5.2	Stand der Technik	80

5.2.1	Hyperspektrale Luftaufnahmen	80
5.2.2	Hyperspektrale Daten	81
5.2.3	RGB-Datensätze	84
5.2.4	Zusammenfassung	86
5.3	Eigener Datensatz	86
5.4	Datenaufnahme	88
5.4.1	Vorverarbeitung	89
5.4.2	Datenstruktur	89
5.5	Datenextraktion	89
5.5.1	Beispiele	91
5.5.2	Annotation	92
5.5.3	Datenanalyse	94
5.6	Fazit	95
6	OPTISCHE INDIZES	99
6.1	Einführung	99
6.2	Stand der Technik	100
6.3	Evaluation	102
6.4	Fazit	104
7	ÜBERWACHTE KLASSIFIKATION	109
7.1	Einführung	109
7.2	Stand der Technik	109
7.3	Per-Pixel Klassifikation	111
7.3.1	Stützvektormaschinen	112
7.3.2	Random Forest	114
7.4	Training	117
7.5	Evaluation	118
7.5.1	NIR semantisch	121
7.5.2	VIS semantisch	122
7.5.3	NIR offRoad	123
7.5.4	VIS offRoad	124
7.5.5	Dimensionsreduktion	125
7.6	Fazit	125
8	MERKMALSBASIERTE KLASSIFIKATION	129
8.1	Einführung	129
8.2	Stand der Technik	129
8.3	Extraktion von Merkmalen	131
8.3.1	Superpixel	131
8.3.2	Merkmalsextraktion	131
8.4	Evaluation	134
8.5	Fazit	138
9	KONTEXTUELLE KLASSIFIKATION	139
9.1	Einführung	139
9.2	Stand der Technik	139

9.3	Szenenanalyse	142
9.3.1	Bedingtes Zufallsfeld	142
9.3.2	Fully Connected CRF	145
9.3.3	Eigener Ansatz	147
9.4	Evaluation	147
9.4.1	<i>VIS</i> semantisch	148
9.4.2	<i>NIR</i> semantisch	150
9.4.3	<i>VIS</i> offRoad	150
9.4.4	<i>NIR</i> offRoad	152
9.5	Fazit	153
10	NEURONALE NETZE	157
10.1	Einführung	157
10.2	Stand der Technik	158
10.2.1	Faltungsnetze (CNN)	158
10.3	Semantische Segmentierung	162
10.3.1	Vollständig gefaltete Netze	163
10.3.2	Architektur-Übersicht	165
10.4	Evaluation von Netzarchitekturen	167
10.5	Fazit	171
11	DATENKOMPRESSION	173
11.1	Einleitung	173
11.2	Stand der Technik	174
11.3	Autoencoder zur Dimensionsreduktion	175
11.3.1	Autoencoder-Design	177
11.4	Evaluation	178
11.4.1	Fazit	180
11.5	Integration in Semantische Segmentierung	182
11.6	Zusammenfassung	187
12	FAZIT	189
12.1	Zusammenfassung und Bewertung	189
A	STATISTIK DER DATENSÄTZE	193
	ABBILDUNGSVERZEICHNIS	195
	TABELLENVERZEICHNIS	199
	MATHEMATISCHE SYMBOLE	201
	LITERATURVERZEICHNIS	207
	INTERNETQUELLEN	239

EINLEITUNG

In Zukunft werden autonome Systeme und Agenten eine immer größere Rolle spielen. Sie werden in immer mehr Bereichen des Alltags eingesetzt und werden auch die Mobilität der Zukunft verändern. Der Ursprung der autonomen mobilen Systeme liegt in den 80ern des vorigen Jahrhunderts. Im Jahre 1986 startete auf Initiative von Daimler Benz ein europäisches Großprojekt mit dem Namen *Prometheus* [Wil88], an dem sich 13 Autohersteller und insgesamt 19 Länder beteiligten. Ziel der Initiative war es, neue Technologien zu entwickeln, um die Effizienz und Sicherheit der Mobilität zu erhöhen. Fast parallel dazu startete die Carnegie Mellon Universität [THKS88] im Jahr 1988 ein Projekt mit dem Titel *Navlab*, um Algorithmen und Verfahren zur Umgebungswahrnehmung und Navigation zu entwickeln. Sie erreichten 1995 einen Durchbruch, in dem sie nahezu autonom von Pittsburgh nach San Diego fuhren.

Die Arbeiten und Fortschritte dieser Pionierzeit wurden von Bertozzi et al. im Jahr 2000 zusammengefasst [BBFoo]. Die Autoren kommen zu dem Schluss, dass die nötige Rechenleistung inzwischen verfügbar ist, aber Schwierigkeiten wie Reflexionen, nasse Fahrbahn, direkte Sonneneinstrahlung, Tunnel und Schatten noch große Probleme bei der Analyse bereiten. In den darauf folgenden Jahren wurden weitere Fortschritte im Bereich des autonomen Fahrens erzielt.

Trotz dieser Fortschritte ist jedoch die völlig autonome Navigation in beliebig komplexen Umgebungen bisher nicht vollständig gelöst. Auch wird der Terminus *autonome Navigation* in Publikationen ambivalent verwendet. Es muss zunächst klar unterschieden werden zwischen *automatisiertem* und *autonomen* Fahren.

Beim *automatisierten* Fahren werden verschiedene Aufgaben während der Fahrt vom Fahrzeug übernommen, beim *autonomen* Fahren absolviert das Fahrzeug eine bestimmte Aufgabe oder Strecke ohne einen Fahreingriff. Um die Stufen von der Automatisierung bis hin zur Autonomie klarer voneinander abzugrenzen, wurde im Jahr 2014 die Norm *SAE J3016* [C⁺14] publiziert, welche 6 Fähigkeitsstufen einführt nach denen der Grad der Autonomie bestimmt wird. Diese Stufen sind in Tabelle 1 dargestellt.

Aufgrund des großen Potenzials vor allem für die Fahrzeugindustrie wurde in den letzten Jahren viele Ressourcen in die Erforschung zur Lösung der Probleme des autonomen Fahrens investiert. Allein Volkswagen investierte kürzlich mehr als 2,6 Milliarden Euro in diesen Bereich [3]. Bosch will bis 2022 mehr als 4 Milliarden Euro investieren [6] und unter anderem 3 000 KI-Experten einstellen. Entsprechend

Stufe	Name	Beschreibung
0	Selbstfahrer	Nur rudimentäre Assistenz wie ESP und ABS
1	Assistenz	Rudimentäre Assistenz wie Abstandsregelung
2	Teilautomatisierung	Assistenzfunktionen wie automatisches Einparken oder Spurhalten
3	bedingte Automatisierung	Autonomes fahren in definierten Szenarien wie z. B. auf der Autobahn
4	Hochautomatisierung	Automatische und dauerhafte Führung des Fahrzeugs. Der Fahrer wird bei Bedarf zum Eingreifen aufgefordert.
5	Vollautomatisierung	Vollständig autonomes Fahren. Keine Bedienung durch den Fahrer mehr erforderlich.

Tabelle 1: Stufen der Autonomie nach [C⁺14]

stand die autonome Navigation von Fahrzeugen in wohldefinierten und strukturierten Umgebungen, wie dem Straßenverkehr oder Innenräumen, in den vergangenen Jahren im Fokus der wissenschaftlichen Forschung im Bereich der Robotik. Öffentlichkeitswirksam haben verschiedene Wettbewerbe wie beispielsweise die *DARPA Grand Challenge 2004* [7] und die *DARPA Urban Challenge 2007* [8] dieser Forschung zur Seite gestanden.

Und trotz dieser immensen Anstrengungen sind nicht alle von Bertozzi erwähnten Probleme gelöst. Das Handelsblatt titelte 2019 „Das vollkommen autonome Fahren wird vorerst nicht kommen“ [5]. Es zeigt sich, dass die Probleme komplexer und die Lösungen teurer sind als zunächst angenommen. Daher lässt das vollautonome Fahren auf definierten Straßen noch auf sich warten [4].

Die Methoden und Algorithmen aus dem Stand der Technik haben noch viele Schwächen, sodass eine umfassende autonome Navigation noch nicht möglich ist. Aktuelle Serienfahrzeuge mit Autonomiefunktionen liegen daher etwa auf dem Niveau von Stufe 3 (Tabelle 1).

Die Gründe dafür sind vielfältig, so erfordern autonome Systeme, welche in komplexen dynamischen Umgebungen agieren, entsprechende Methoden und Verfahren welche in der Lage sind, sich auch an unvorhersehbare Situationen anzupassen und korrekte Entscheidungen zu treffen.

Korrekte Entscheidungen beruhen im Kern auf einer präzisen Wahrnehmung und Analyse der Umgebung, welche durch entsprechende Sensorik und passende Sensordatenauswertung realisiert wird.

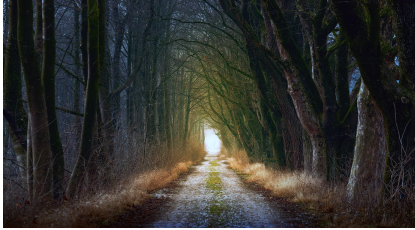
Weiterhin ist zu berücksichtigen, dass der Fokus der Automobilindustrie auf urbanen Umgebungen mit strukturiertem und wohlde-



(a) Quelle: [@18]



(b) Quelle: [@19]



(c) Quelle: [@20]



(d) Quelle: [@21]

Abbildung 1: Beispiele für unstrukturierte Umgebungen

finiertem Umfeld liegt. Neben dem autonomen Fahren in urbanen Umgebungen etabliert sich aber auch ein immer höherer Bedarf an autonom agierenden Zugmaschinen und Fahrzeugen abseits der Straße, in unwegsamem und unstrukturiertem Gelände. Dieser Bereich hat bisher nur einen Bruchteil der Aufmerksamkeit erfahren. Aufgrund der Komplexität und Vielseitigkeit der unstrukturierten Umgebung stellen sich den Forschern neue Aufgaben, welche bisher auch noch nicht gänzlich gelöst werden konnten. Denn in der Regel sind dort keine befestigten oder asphaltierten Fahrbahnen mit Begrenzungen und Markierungen vorzufinden, welche Vereinfachungen wie z. B. die „Flat World Assumption“ ermöglichen. Diese Vereinfachung nimmt an, dass die Welt nicht gekrümmt sondern flach sei, was eine starke Vereinfachung der Algorithmik ermöglicht. In unstrukturierten Umgebungen muss jedoch der jeweilige Untergrund gezielt bezüglich Geometrie und Zusammensetzung analysiert und auf befahrbare Flächen untersucht werden. Auch muss über die eventuelle Durchfahrbarkeit von vorhandener Vegetation entschieden werden. Beispiele für unstrukturierte Umgebungen sind in Abbildung 1 dargestellt. Dort sind Wege in unterschiedlicher Ausprägung in diversen Umgebungen zu sehen, welche auch unterschiedliche Beleuchtungssituationen zeigen. An diesen Beispielen ist gut der große Facettenreichtum unstrukturierter Umgebungen zu erkennen.

Die erfolgreiche autonome Navigation unbemannter Fahrzeuge in solchen komplexen Umgebungen ist unter anderem von einem effektiven Pfadplanungsalgorithmus abhängig. Dieser nutzt die Daten einer sog. semantischen Szenenanalyse, um für den Untergrund Befahrbarkeitswerte zu berechnen und eine Heuristik zu generieren, welche die

Planung einer Trajektorie vom Start zum Ziel ermöglicht.

Die Qualität der Heuristik beeinflusst somit, unabhängig vom Pfadplanungsalgorithmus, das Ergebnis und die Laufzeit der Trajektorienfindung und damit auch die Qualität der autonomen Navigation. Entsprechend kommt der Umgebungswahrnehmung und im Speziellen der Szenenanalyse eine Schlüsselrolle zu.

Die semantische Szenenanalyse selbst ist ein komplexes Thema aus dem Bereich des Bildverstehens, bei der ein System aus Sensorinformationen ableiten soll, welche Elemente und Objekte in einer Szene vorhanden sind. Grundsätzlich ist die semantische Segmentierung ein wichtiger Bereich des maschinellen Sehens, welcher den Weg zum Szenenverständnis ebnet. Immer mehr Anwendungsbereiche des Alltags profitieren davon, dass sie Wissen aus Bilddaten ableiten. Auf das autonome Fahren bezogen ist es relevant festzustellen, ob befahrbare Flächen vorhanden sind und wo sich eventuell Hindernisse wie Bäume, Mauern oder Gebäude befinden.

Die Bewertung der Befahrbarkeit, insbesondere in unstrukturierter Umgebung, ist eine komplexe Funktion, die sich aus Umgebungseigenschaften und Fahrzeugparametern zusammensetzt und die sich abhängig vom Terrain auf eine unterschiedliche Anzahl von sog. Befahrbarkeitsklassen abbildet. Nur wenn diese Zuordnung zu jeder Zeit korrekt umgesetzt wird, ist es möglich zu bewerten, ob, wo und wie eine optimale hindernisfreie Überquerung des Terrains möglich ist.

In der Regel werden bildgebende Sensoren wie Kameras oder Laserscanner zur Umgebungswahrnehmung eingesetzt. Bei der Verwendung von Laserscannern wird meist nur aufgrund der geometrischen Beschaffenheit der Umgebung bzw. des Höhenunterschieds einzelner Segmente auf die Befahrbarkeit geschlossen. Dies führt aber beispielsweise im Falle von Hindernissen, welche nicht durch ihre Geometrie bestimmt sind, wie z. B. Sand potentiell zu einer folgenreichen Fehleinschätzung. Es ist folglich notwendig die Zusammensetzung des Terrains näher zu bestimmen. Doch die etablierte Sensorik und die zugehörigen Analyseverfahren reduzieren die Problematik stark, so dass die Komplexität unstrukturierter Umgebung sich nicht adäquat abbilden und erfassen lässt.

Um zu einer besseren Repräsentation der Umgebung zu gelangen, muss z. B. der Einsatz alternativer bildgebender Sensorik geprüft werden. Neu entwickelte Kamerasysteme erlauben eine präzisere Wahrnehmung der Umgebung und somit potentiell die Durchführung einer umfassenderen semantischen Szenenanalyse.

Diese Sensoren sind dem Themenkomplex der multispektralen Bildgebung zuzuordnen, welche die digitale Bildgebung mit der Spektroskopie kombiniert. Dies ist ein stark wachsender Bereich, welcher seit einigen Jahren immer größere Aufmerksamkeit erlangt. Dabei messen Spektrometer sehr feingranular die Intensität von Strahlung

in einem Teilbereich des elektromagnetischen Spektrums in Abhängigkeit von der Wellenlänge. RGB-Kameras messen diese Strahlung ebenso, allerdings wird die Strahlung auf drei Kanäle reduziert was zu einer Unterabtastung führt, wodurch viele Informationen verloren gehen. Bezogen auf die Umgebungswahrnehmung analysieren spektrale Bildsensoren entsprechend spektrale Strahldichten an jedem Punkt in einer Szene. Dabei ist die spektrale Strahldichte eines Objekts die reflektierte Lichtintensität als Funktion der Wellenlänge. Diese sogenannte spektrale Verteilung eines Reflexionsgrades lässt dann spezielle Rückschlüsse auf die materielle Zusammensetzung an der Oberfläche eines Objektes zu. Dies ist ein großer Vorteil gegenüber der etablierten Sensorik, welche die menschliche Wahrnehmung zum Vorbild hat. Denn obwohl die menschliche Wahrnehmung die Form von verschiedenen Objekten sehr gut differenzieren kann, erkennt es Attribute wie die spektralen Reflexionseigenschaften nicht annähernd so genau. Das menschliche Sinnesorgan nimmt nur bestimmte schmale Bereiche der Strahlung, die sogenannte sichtbare Strahlung wahr. Und diese wird auch nicht wellenlängensensitiv wahrgenommen, sondern integral über bestimmte Wellenlängenbereiche. Weiterhin ist Farbe keine Eigenschaft des Objektes, sondern ein Sinneseindruck, welcher vom Gehirn des Menschen interpretiert wird. Dieser Sinneseindruck wird durch das Zusammenwirken von einer Beleuchtung mit Licht und den Reflexionseigenschaften eines beleuchteten Objektes ausgelöst. So ist Farbe aufgrund des Effekts der Metamerie [FANFo6] keine eindeutig diskriminative Eigenschaft um verschiedene Materialien zu unterscheiden, da derselbe Farbeindruck aus unterschiedlichen spektralen Leistungsverteilungen konstruiert werden kann. Ein Beispiel hier kommt aus der Automobilindustrie, wo die Karosserie in beliebiger Farbe markiert werden kann, ohne das dazu die Zusammensetzung des Materials geändert wird bzw. sich die Objektklasse (Auto) ändert. Folglich lassen sich sowohl das menschliche Gehirn als auch die Sensoren, welche nach diesem Vorbild geschaffen wurden, täuschen.

Die spektrale Bildgebung hingegen nimmt die Spektralinformationen in einem definierten Bereich des elektromagnetischen Spektrums mit einem schmalen Wellenlängenbereich wie z. B. 10 nm auf. Die folgenden Definitionen spezifizieren die relevante Terminologie im Rahmen dieser Arbeit.

Definition 1: Spektrale Bildgebung

Die spektrale Bildgebung wird nach Garini et al. [GYM06] auch als bildgebende Spektroskopie bezeichnet. Sie kombiniert konventionelle Bildgebungs- und Spektroskopiemethoden, um sowohl räumliche als auch spektrale Informationen eines Objekts zu erhalten. Denn die normale Bildgebung bestimmt die Intensität an jedem Pixel eines Bildes. Bei der Spektroskopie hingegen wird ein Spektrum an einem einzelnen Punkt bestimmt. Bei der spektralen Bildgebung wird nun ein Spektrum für jeden Pixel eines Bildes bestimmt. Diese Technologie wurde von Goetz et al. [GVSR85] für die Fernerkundung definiert. Gemäß der Definitionen lässt sich die spektrale Bildgebung in multispektrale, hyperspektrale und ultraspektrale Bildgebung unterteilen. Der Unterschied liegt jeweils in der Menge, Breite und dem Abstand der jeweiligen Abtastpunkte im Spektrum.

Definition 2: Hyperspektral

Die hyperspektrale Bildgebung sammelt Informationen aus dem elektromagnetischen Spektrum. Hyperspektrale Daten bestehen im Allgemeinen aus vielen Spektralbändern mit geringem Abstand und enger Bandbreite (5 – 10 nm) [HK13, Jen15]. Somit kommt es zu einer Überabtastung und so ist der Unterschied im Spektrum für das menschliche Auge nicht mehr unterscheidbar. Dies ist auch so im technischen Report (CIE 223:2017)[CIE17] der CIE von 2017 beschrieben.

Definition 3: Multispektral

Multispektrale Daten werden von Sensoren erzeugt, welche die reflektierte Energie in mehreren definierten Abschnitten (Bänder) des elektromagnetischen Spektrums messen. Die Sensoren verfügen in der Regel über 3 bis 10 verschiedene Spektralbänder in jedem Pixel der von ihnen erzeugten Daten mit großer Bandbreite (70 – 400 nm) [HK13, Jen15]. Durch die größere Bandbreite und die geringe Anzahl an Bändern im Vergleich zur hyperspektralen Bildgebung kommt es zu einer Unterabtastung. Dies macht eine Interpolation der Daten notwendig, um wieder zu einem kontinuierlichem Spektrum zu gelangen.

Dadurch können unterschiedliche Materialien anhand ihrer unterschiedlich reflektierten spektralen Leistungsverteilung des Lichtes unterschieden werden. Dies funktioniert dann auch bei Materialien, welche beim menschlichen Betrachter einen identischen Farbeindruck erzeugen. Im Prinzip lassen sich so viele Objekte anhand ihres Reflexionsverhaltens identifizieren.

Die hyperspektrale Sensorik wurde anfänglich in den 70ern und 80ern primär für die Fernerkundung und Astronomie entwickelt [GVSR85]. Eines der Ziele der Fernerkundung ist dabei die automatisierte Extraktion von Oberflächen, Objekten und Informationen aus den erfassten Daten. Der Bedarf an detaillierten Informationen von Objekten wie Gebäuden und Straßen nimmt aufgrund ihrer Anwendungen in der Navigation von autonomen Fahrzeugen rasant zu. So werden die extrahierten Informationen genutzt, um automatisch präzise Straßenkarten zu erzeugen und sie gleichzeitig auch zu aktualisieren. Inzwischen werden auch immer neue Anwendungsgebiete erschlossen. Entsprechende Systeme finden mittlerweile ihre Anwendung z. B. in der Agrarindustrie [CBJNo4], der Lebensmittelkontrolle [GOC⁺07] oder anderen Bereichen der Qualitätskontrolle von Produkten.

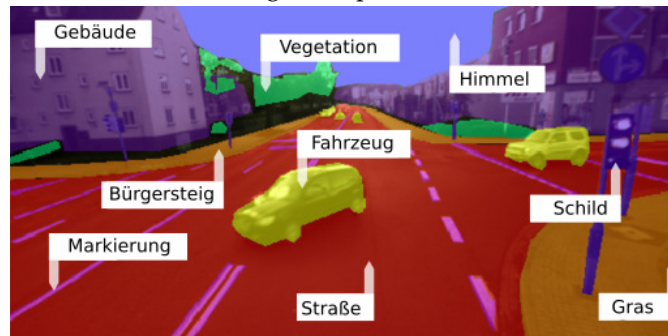
Grundsätzlich lassen sich die Einsatzszenarien dabei in Erdoberflächenbeobachtung und Umgebungen mit definierter Beleuchtung unterteilen.

Die primär verwendeten hyperspektralen Sensoren nutzen das sog. Line-Scanning-Prinzip oder ähnliche Techniken, um Daten aufzunehmen. Dabei sind sie auf die lineare Bewegung des Sensors angewiesen, um eine komplette Szene erfassen zu können. Weiterhin sind diese Sensoren auf die Aufnahme von statischen Szenen beschränkt und eignen sich primär zur Erdbeobachtung oder für eine Verwendung unter Laborbedingungen.

Unerforscht dagegen ist speziell der Einsatz von spektraler Sensorik auf mobilen Systemen aufgrund der zuvor erwähnten Einschränkungen bspw. mit dynamischen Umgebungen. Es ergeben sich aufgrund der rapide voranschreitenden Technik sowohl neue Chancen als auch Herausforderungen bei der Analyse dieser Daten. Folglich stellt die Analyse spektraler Bildinformationen zur Umgebungswahrnehmung und Szenenanalyse einen wichtigen Forschungsbereich dar und findet aufgrund neuer Bildgebungstechniken neue Anwendungsgebiete. Spektrale Informationen erlauben dabei systembedingt einen detaillierteren Einblick in die Zusammensetzung und Beschaffenheit von Materialien, Pflanzen und Bodenbelägen als normale Kameras. Insbesondere im Bereich der Interpretation dieser Daten, z. B. zur Herleitung von Befahrbarkeitsinformationen oder der semantischen Analyse lässt sich daher potentiell wertvoller Zusatznutzen generieren. Ein Beispiel wie ein Ergebnis einer solchen Analyse aussehen kann, ist in Abbildung 2 gegeben.



(a) RGB-Bild erzeugt aus spektralen Informationen



(b) RGB-Bild mit semantischen Klassen

Abbildung 2: Beispiel für eine semantische Klassifikation multispektraler Bilddaten

So ist im Bereich der spektralen Sensorik die sog. Snapshot-Mosaik-Filter-Technik eine vielversprechende neue Entwicklung. Sie ist eine spezielle Implementierung eines Linienscanners, welcher das parallele Abtasten mehrerer räumlicher Linien pro Spektralband ermöglicht und somit derartige Sensorik in dynamischen Umgebungen einsetzbar macht.

Es existieren verschiedene Methoden zur Implementierung der Snapshot-Mosaik-Filter-Technik. Eine dieser Implementierungen nutzt für jedes Pixel des Sensors einen eigenen Spektralfilter, welcher durch eine spezielle Transmissionskurve gekennzeichnet ist. Der Sensor enthält so unterschiedliche Bandpassfilter, welche spektrale Wellenlängen nach dem Fabry-Pérot-Interferenzprinzip isolieren. Diese Methode wird von den in dieser Arbeit verwendeten Kameras umgesetzt. Entgegen der Deklaration des Herstellers sind die Kameras gemäß der Definitionen 2 und 3 eher der multispektralen Sensorik zuzuordnen. Denn die Einschränkungen des genutzten Prinzips begrenzen aktuell die mögliche Anzahl der Kanäle zur Abtastung des Spektrums. Die Technik entwickelt sich zwar rasant, allerdings ist aktuell noch keine hyperspektrale bildgebende Sensorik unter Verwendung der Snapshot-Mosaik-Technik absehbar. Die in dieser Arbeit untersuchten Methoden zur spektralen Datenverarbeitung können aber

theoretisch auch für hyperspektrale Daten genutzt werden, da sie überwiegend auch aus diesem Bereich stammen.



(a) AVIRIS Satellitendaten, mit 3 Bändern als RGB kodiert [23]



(b) Rohbild einer Snapshot-Mosaik-Kamera



(c) Eine suburbane Szene aufgenommen mit einer RGB-Kamera



(d) Eine suburbane Szene aufgenommen mit einer Spektralkamera

Abbildung 3: Beispieldaten unterschiedlicher Sensortechnologien

Ein Vergleich der aufgenommenen Daten mit denen von Line-Scannern ist in Abbildung 3 dargestellt. Aufgrund der Neuartigkeit der Technologie existieren wenige Veröffentlichungen, welche die in dieser Dissertation zu verwendende Hardware und ihre möglichen Anwendungen thematisieren.

Ziele der Dissertation

In dieser Dissertation soll der Einsatz und die Eignung der Snapshot-Mosaik-basierten Sensorik zur Umgebungswahrnehmung und Szenenanalyse im Bereich der autonomen Navigation in strukturierten (urban) und unstrukturierten (suburban) Umgebungen untersucht werden. Diese Sensorik liefert spektral feiner aufgelöste Bildinformationen als sie z. B. mit RGB-basierten Kameras möglich sind. Es soll erforscht werden, ob diese feiner aufgelösten spektralen Daten einen Vorteil gegenüber klassischen RGB- bzw. Grauwertdaten hinsichtlich der semantischen Szenenanalyse und Klassifikation bieten. Dies wird am Beispiel der Nutzung einer Kamera mit 16 Spektralkanälen im sichtbaren und einer Kamera mit 25 Spektralkanälen im Infrarotbereich durchgeführt.

Dazu wird zunächst eine geeignete Vorverarbeitung genutzt, welche aus den rohen Pixelinformationen spektrale Werte berechnet, mit denen dann geeignete Klassifikatoren getestet und trainiert werden. Ba-

sierend auf den vorliegenden Daten sollen diese eine semantische Bildanalyse vornehmen und einen Vergleich sowie Evaluation der unterschiedlichen Bildgebungen und Daten erlauben.

1.1 EIGENER BEITRAG

Die in dieser Dissertation erläuterten Arbeiten wurden bereits auf internationalen Konferenzen präsentiert und publiziert. Eine Literaturliste der eigenen Veröffentlichungen ist gleich zu Beginn der Arbeit zu finden. Der weitere Aufbau der Arbeit gliedert sich wie folgt. In Kapitel 2 wird ein Überblick über den Stand der Technik zur semantischen Segmentierung und spektraler Datenverarbeitung gegeben. Darauf folgt in Kapitel 3 ein Abschnitt zu den theoretischen Grundlagen dieser Arbeit und der Sensortechnologie. Die eigenen Arbeiten und Publikationen werden von Kapitel 4 bis Kapitel 11 beschrieben und gliedern sich auf, wie nachfolgend beschrieben. Kapitel 12 beinhaltet noch mal eine Zusammenfassung der Arbeiten und eine abschließende Bewertung.

Der eigene Beitrag:

- Veröffentlichung annotierter Datensätze mit spektralen Daten von strukturierter (urban) und unstrukturierter (suburban) Umgebung
- Evaluation der spektralen Daten mit optischen Indizes zur Separation von Vegetation und befahrbaren Bereichen in strukturierter Umgebung
- Evaluation der spektralen Daten mit Verfahren des maschinellen Lernens
- Evaluation der spektralen Daten mit Netzarchitekturen aus dem Bereich des Deep-Learning
- Entwicklung eines Autoencoders zur Kompression von spektralen Daten

verteilt auf die folgenden Kapitel:

Kapitel 4: Vorverarbeitung

Die durch spektrale Sensorik aufgenommenen Daten stellen zunächst nur ein Rohsignal dar, welches durch einen Intensitätswert repräsentiert wird. Um die Daten effektiv zu nutzen und weiter zu verarbeiten, muss der Prozess der Bildentstehung bei dieser Art Sensorik analysiert und verstanden werden. In diesem Abschnitt wird eine Vorverarbeitung erläutert und modelliert, welche die Rohdaten der Kameras in Form eines Grauwertbildes in spektralen Werte transformiert. Die entwickelte Vorverarbeitung wurde für das gegebene Szenario, die

Aufnahme von Szenen in strukturierter und unstrukturierter Umgebung mit dynamischer Beleuchtung entwickelt.

Kapitel 5: Datensätze

Da die in dieser Arbeit verwendete Sensorik neuartig ist, sind bisher keine öffentlich zugänglichen Datensätze vorhanden. Dieser Abschnitt beschreibt die Erstellung, den Aufbau und die Zusammensetzung der im Rahmen dieser Arbeit aufgebauten neuen Datensätze, welche bis dato so nicht verfügbar sind. Zum Aufbau eines Datenbestandes an spektralen Sensordaten wurden Messdaten unterschiedlicher Terrainoberflächen sowie Hindernistypen und Materialien aufgezeichnet. Dazu wurde entsprechende Sensorik synchronisiert und auf verschiedenen Fahrzeugen montiert. Die aufgezeichneten Daten wurden manuell annotiert, um eine Grundwahrheit für spätere Evaluationen zu schaffen.

Kapitel 6: Optische Indizes

Sind spektrale Daten verfügbar, so ist es möglich, mittels optischer Indizes verschiedene Materialien zu unterscheiden. Es gibt eine ganze Reihe von optischen Indizes, welche den Reflexionseffekt von Chlorophyll im Nahinfrarotbereich nutzen, um Aussagen über den Zustand der Navigation zu machen. In diesem Abschnitt wird untersucht, ob verschiedene etablierte Indizes auch mit den verwendeten Sensoren kombiniert werden können, um eine rudimentäre Szenenanalyse zu ermöglichen.

Kapitel 7: Klassifikation

In diesem Abschnitt werden geeignete Verfahren vorgestellt und untersucht, um aufgenommene Szenen aus den in Kapitel 5 beschriebenen veröffentlichten Datensätzen semantisch zu klassifizieren. Hier werden vor allem etablierte Verfahren aus den Bereichen des maschinellen Lernens untersucht. Ziel ist dabei die Herleitung definierter semantischer Klassen für verschiedene Regionen im Bild.

Kapitel 8: Erweiterte Klassifikation von Merkmalen auf spektralen Daten

Dieser Abschnitt beschäftigt sich mit der sinnvollen Kombination von spektralen und räumlichen Informationen, um Nachteile der per-Pixel-Klassifikation zu kompensieren. Dazu wird untersucht, wie die zuvor genutzten Klassifikatoren entsprechend mit speziellen Merkmalen kombiniert werden können, um eine bessere und stabilere Klassifikation zu ermöglichen.

Kapitel 9: Kontextuelle Klassifikation

Bisher wurde in den vorherigen Kapiteln eine klassische per-Pixel-Klassifikation verfolgt. Diese ist jedoch anfällig für sogenanntes *Klassifikationsrauschen*. Dabei werden Bereiche einer Szene, welche einer einzigen Oberfläche jedoch nicht durchgehend als eine solche klassifiziert. Daher ist es von großer Bedeutung bei der Klassifikation der einzelnen Pixel auch deren Umgebung mit zu betrachten um so flächige Klassifikationsergebnisse zu erzeugen. Dazu werden in diesem Abschnitt die zuvor untersuchten Klassifikatoren mit einem Graphenmodell kombiniert. Dieses Graphenmodell erlaubt die Integration von Kontextinformationen in den Klassifikationsprozess, durch den Abhängigkeiten zwischen den einzelnen Pixeln im Hyperwürfel modelliert werden können.

Kapitel 10: Klassifikation mittels Neuronaler Netze

In diesem Abschnitt werden verschiedene Netzarchitekturen aus dem Bereich des Deep Learning untersucht und mit spektralen Daten trainiert. Es wird untersucht, ob sich das Potential der neuronalen Netze auch auf spektrale Daten übertragen lässt.

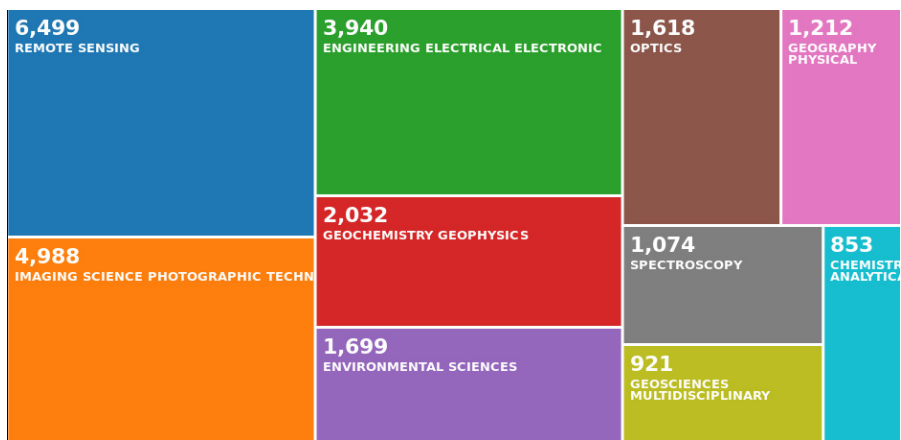
Kapitel 11: Datenkompression

Da die hier verwendete Sensorik potentiell sehr große Datenmengen bereitstellt, welche mitunter gar nicht in Echtzeit verarbeitet werden können, wird in diesem Abschnitt untersucht, inwiefern die Sensordaten unter Erhaltung der relevanten Informationen reduziert werden können. Dazu werden spezielle Methoden aus dem Bereich des Deep-Learnings nutzbar gemacht und auf die spektralen Daten angewendet.

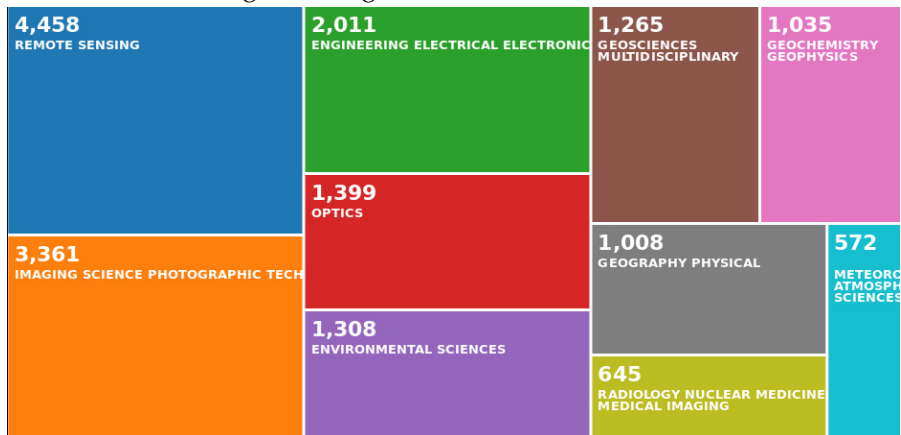
STAND DER TECHNIK

2.1 EINFÜHRUNG

In diesem Abschnitt wird eine grundsätzliche Einführung in die Thematik der Arbeit gegeben. Dazu werden die geschichtlichen Hintergründe sowie der aktuelle Stand der Technik beleuchtet. Generell ist das Gebiet der spektralen Bildgebung sehr groß und hat viele Anwendungsbereiche, welche schon Jahrzehnte zurückreichen. Dazu gehö-



(a) Veröffentlichungen mit dem Thema *Hyperspektral* in den letzten 30 Jahren nach Anwendungsbereich geclustert



(b) Veröffentlichungen mit dem Thema *Multispektral* in den letzten 30 Jahren nach Anwendungsbereich geclustert

Abbildung 4: Anzahl der Veröffentlichungen zu den Themen *Hyperspektral* und *Multispektral* in den letzten 30 Jahren nach Anwendungsbereich geclustert. Quelle: [17]

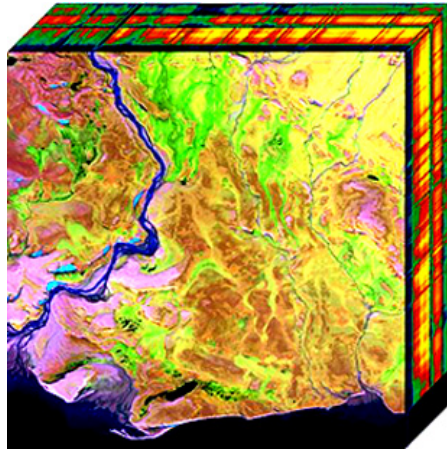


Abbildung 5: Beispiel eines Hyperwürfels einer Satellitenaufnahme [Com18]

ren z. B. die Bereiche der Prüftechnik, Farbmeterik, Medizin, Industrie und Fernerkundung. Alle Bereiche zu betrachten würde den Rahmen dieser Arbeit übersteigen. Daher beschäftigen sich die folgenden Abschnitte vornehmlich mit den Entwicklungen der spektralen Bildgebung im Bereich der Fernerkundung. Die Fernerkundung ist eines der wichtigsten Anwendungsgebiete der Spektraltechnik und in diesem Bereich wurden in den letzten 30 Jahren ca. 33% der Paper zu diesen Themen veröffentlicht wie Abbildung 4 zeigt. Weiterhin ist eines der Ziele der Fernerkundung die Klassifikation von Oberflächen, was sich auch mit dem Thema dieser Arbeit deckt. So wird zunächst ein Überblick über die Sensormodalitäten und den geschichtlichen Verlauf dargestellt. Anschließend wird eine Übersicht über Verfahren zur Klassifikation gegeben, welche diese Daten nutzen.

2.2 SPEKTRALE BILDGEBUNG

Die Sonne bildet das Zentrum unseres Sonnensystems und ist mit ihrer Strahlung die Grundlage für alles Leben auf unserem Planeten. Entsprechend stellt die ausgesandte solare Strahlung die primäre Lichtquelle für das wichtigste Sinnesorgan des Menschen dar, das Auge. Das von der Sonne ausgesandte Licht umfasst ein breites Spektrum, welches von der Umgebung reflektiert wird. Ein Teil dieses Spektrums wird von der Netzhaut in elektrische Reize umgewandelt. So ist es dem Menschen möglich, seine Umgebung wahrzunehmen und mit dieser zu interagieren. Bei der Entwicklung von bildgebender Sensorik wurde dieses Prinzip kopiert bzw. adaptiert und zum Beispiel in Form von Sensoren mit drei spektralen Kanälen (RGB) ähnlich dem Aufbau des menschlichen Auges adaptiert.

Wird nun speziell der Bereich der spektralen Bildgebung betrachtet, wird das reflektierte solare Spektrum in einzelnen Wellenlängenbereichen durch definierte Aufnahmekanäle/Bänder (engl. *bands*) gemessen und insgesamt als Hyperwürfel wie in Abbildung 5 dargestellt. Qin et al. [QCK⁺13] und Vagni [Vago7] präsentieren dazu einen Überblick über die in diesem Bereich eingesetzten Technologien:

Abhängig von der Sensortechnik und der Optik des Sensors variiert das messbare Spektrum sowie die Anzahl der Kanäle und deren Bandbreite. Sensoren, die viele definierte Kanäle mit einer hohen spektralen Auflösung (geringe Bandbreite und kleiner Abstand zwischen den Bändern) aufnehmen, werden auch als *bildgebende Spektrometer* bezeichnet. Darauf aufbauend ist die spektrale Auflösung definiert als die Anzahl und Größe der Wellenlängenintervalle/Bänder im elektromagnetischen Spektrum, für die ein Sensor empfindlich ist. Ein erheblicher Faktor bei der Datenaufnahme ist das Übersprechen (engl. *crossstalk*). Dies ist ein Effekt, bei dem zusätzlich zum eigentlich empfindlichen Pixel noch weitere umliegende Pixel mit aktiviert werden. Aufgrund technischer Grenzen bei der Realisierung von spektralen Filtern in einer Kamera sind die spektralen Empfindlichkeiten in Kameras nicht scharf begrenzt. Vielmehr sind die Sensoren über einen Bereich im Spektrum mit der maximalen Empfindlichkeit in der Mitte der Bandbreite eines Bandes empfindlich. Dieser Zusammenhang ist in Abbildung 6 dargestellt, die Verteilung ähnelt einer Gaußkurve [LWTG14]. Die volle Breite wird durch die Halbwertsbreite (engl. *Full Width at Half Maximum*) (FWHM) definiert und bezeichnet die Breite des Spektralkanals. Beim Sensordesign werden die entsprechenden Wellenlängenempfindlichkeiten oft so gewählt, dass der Kontrast zwischen gesuchtem Material und dem Hintergrund maximal wird. Die sich aus der Bildgebung ergebenden spektralen Messungen ermöglichen es, Rückschlüsse auf die Beschaffenheit und Zusammensetzung der Oberfläche zu ziehen [SB03]. Die Messungen sind abhängig von Reflexionseigenschaften der Oberfläche. Die Zusammensetzung des Oberflächenmaterials lässt sich durch eine charakteristische Spektralkurve, auch als *Spektralsignatur* bezeichnet, angemessen darstellen.

2.3 FERNERKUNDUNG

Die Fernerkundung (engl. *Remote Sensing*) ist ein Sammelbegriff für Verfahren und Sensoren, welche die von der Erde reflektierte Strahlung messen, um ein Abbild der Erdoberfläche zu generieren. Aus diesem Abbild können dann Rückschlüsse auf die Beschaffenheit und Zusammensetzung gezogen werden, wie Richards und Jia [Rico6] sowie Camps-Valls et al. [CVTGC⁺11] umfassend darlegen.

Die dazu verwendete Sensorik ist häufig an einer Luft- oder Satellitenplattform montiert. Hier ergeben sich aufgrund der Höhenunterschie-

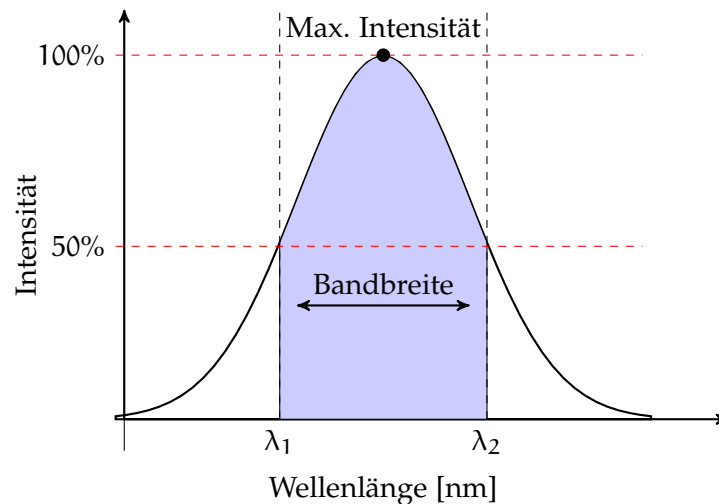


Abbildung 6: Schematische Halbwertsbreite (FWHM) am Beispiel einer Gauss-Glocke

de zwischen den Systemen auch unterschiedliche Bildeigenschaften, welche in einer Datenvorverarbeitung betrachtet werden müssen. Die in der Fernerkundung verwendeten Wellenlängen liegen dabei oft außerhalb des Bereichs des menschlichen Sehvermögens. Spektrale Informationen können in der Fernerkundung grundsätzlich im ultravioletten, sichtbaren und Infrarotbereich aufgenommen werden, da die Sonne als Lichtquelle dient. Zweck der Erfassung dieser Sensorarten ist die Analyse der Zusammensetzung von Oberflächenmaterialien der Erde. Auf der Anwendungsebene wird im Bereich der Auswertung der Fernerkundungsdaten eine deterministische Beziehung bzw. ein Modell zwischen der Zusammensetzung der reflektierten bzw. emittierten Strahlung und den chemischen, biologischen und physikalischen Eigenschaften der Oberfläche hergestellt.

Werden die Projekte zur Fernerkundung betrachtet, so lassen sich grundsätzlich zwei Klassen unterscheiden. Zum einen die Satelliten, welche in einer geostationären Erdoberfläche platziert sind und im Allgemeinen für Wetter- und Klimamessungen genutzt werden. Und zum anderen die Satelliten, welche die Erdoberfläche in einem niedrigeren Radius umkreisen und somit zur Erdoberflächenanalyse und Ozeanographie verwendet werden. Hier befinden sich die Satelliten auf einer sonnensynchronen Umlaufbahn, sodass die Daten von der Erdoberfläche immer zur selben Uhrzeit aufgenommen werden. Die verschiedenen Systeme liefern unterschiedliche räumliche Auflösungen, die zwischen 100 m und 1000 m pro Pixel variieren. Die räumliche Auflösung ist neben der spektralen Auflösung ein wichtiger Aspekt. Während die Auflösung von 1000 m pro Pixel für die Wetterbeobachtung ausreichend ist, werden für die Erkennung einzelner Pflanzen und ihres Zustandes wesentlich höhere Auflösungen benö-

tigt. Sensoren mit einer so hohen spektralen und räumlichen Auflösung sind nicht nur teuer, sie produzieren auch große Mengen an Daten, welche ausgewertet werden müssen. Daher wird zu diesem Zweck oft eine Auflösung gewählt, welche in einem Pixel z. B. nicht einzelne Teile eines Strauches zeigt, sondern vielmehr mehrere nebeneinander gelegene Sträucher, sodass deren Zusammensetzung gemessen wird. Die Unterscheidung zwischen einzelnen Klassen wird aufgrund der Annahme getroffen, dass verschiedene Oberflächen spektrale Verteilungen erzeugen, welche für sie innerhalb eines Datensatzes einzigartig sind.

2.3.1 Geschichte der spektralen Bildgebung

Die ersten Konzepte zur spektralen Bildgebung stammen aus den 1960ern. Sie basieren auf dem Erzeugen einer spektralen Signatur, um verschiedene Materialien und Oberflächen in einer Szene zu erkennen und zu unterscheiden. Dazu wurden Experimente [Lan02] mit Zeilensensoren, die bis zu 20 Spektralbänder im sichtbaren und infraroten Bereich bereitstellen, durchgeführt. Ein tabellarischer Überblick über die relevanten Sensorsysteme der Fernerkundung ist in Tabelle 2 gegeben. Eine weitere graphische Visualisierung der Sensorsysteme und ihrer Kanäle ist in Abbildung 8 dargestellt. Einer der

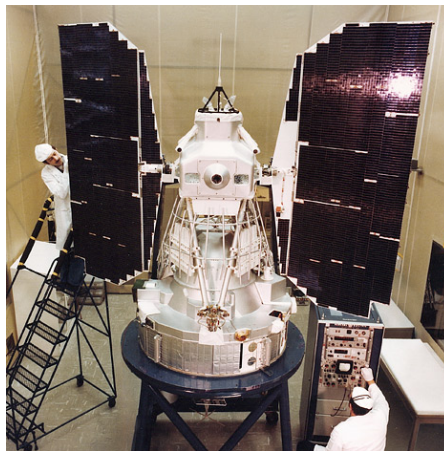


Abbildung 7: Landsat 3 Satellit [Com17]

ersten multispektralen Sensoren eines Satelliten (M-7) des Environmental Research Institute of Michigan (ERIM) wurde 1963, vorgestellt [Lan05]. Der erste Sensor für den Weltraum hatte vier Kanäle mit der Bezeichnung *MSS* (engl. *Multispectral Scanner*) und wurde im Jahr 1972 in den Landsat 1-Satelliten [9] integriert und in den Orbit geschossen. In Abbildung 7 ist der nahezu baugleiche Landsat 3 dargestellt. Dieses Sensorsystemexperiment erwies sich als erfolgreich, aber die geringe Anzahl an Kanälen schränkte die Unterscheidbarkeit

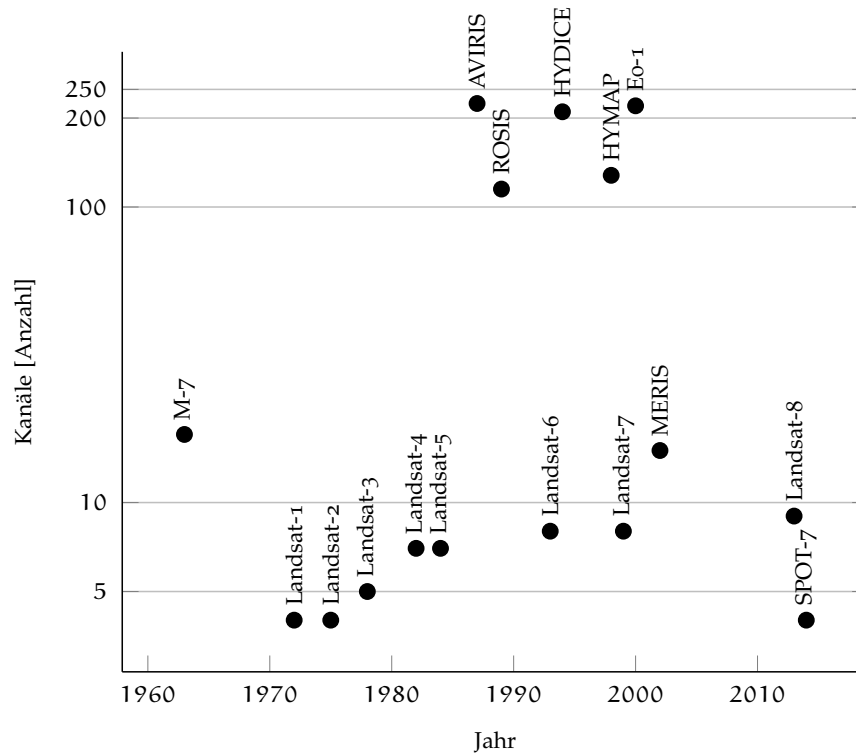


Abbildung 8: Visualisierung der Sensorsysteme und Projekte nach Jahreszahlen und Kanälen

der einzelnen Materialien ein. Daher wurde eine verbesserte Version mit sieben Kanälen und der Bezeichnung *Thematic Mapper* entwickelt. Dieses System wurde 1982 gestartet und danach immer weiter entwickelt. Die Kanäle des Landsat TM wurden ausgewählt, um die dominanten Faktoren, welche die Blattreflexion von Pflanzen steuern, wie Blattpigmentierung, Blatt- und Vordachstruktur und Feuchtigkeitsgehalt, maximal zu nutzen.

Im Jahre 1987, wurde der erste luftgestützte hyperspektrale Sensor mit der Bezeichnung *AVIRIS* [TBEG17], in Betrieb genommen. *AVIRIS* war das erste bildgebende Spektrometer, welches das Spektrum der Sonne in eng aneinandergrenzenden Spektralkanälen abdeckte. Einige Jahre später im Jahr 2000 wurde mit Hyperion [PBS⁺03] der erste weltraumgestützte hyperspektrale Sensor im Orbit ausgesetzt. Dieses System verfügt über 220 Kanäle und eine Auflösung von 30 m pro Pixel. Die NASA-Erdbeobachtungsmission EO-1 (Earth Observing-1) enthält neben dem Hyperion Sensor ein multispektrales Instrument (Advanced Land Imager), das eine Verbesserung gegenüber dem Landsat-7 darstellt.

Die Mehrheit der Hyperspektralsensoren arbeitet im VNIR/SWIR-Spektrum und nutzt so die Sonnenstrahlung, um Materialien anhand der Reflexionsspektren zu identifizieren. Zugrunde liegt die Annahme, dass das spektrale Reflexionsverhalten eines Materials einzigartig

Name	Jahr	Bänder	Spektrum (nm)	Sensor	Träger	Referenz
M-7	1963	17	660-12500	M-7	Flugzeug	[Lan05]
Landsat-1	1972	4	500-1100	Multispectral Scanner	Satellit	[@9]
Landsat-2	1975	4	500-1100	Multispectral Scanner	Satellit	[@10]
Landsat-3	1978	5	500-1100	Multispectral Scanner	Satellit	[@11]
Landsat-4	1982	7	450-2350	Thematic Mapper	Satellit	[@12]
Landsat-5	1984	7	450-2350	Thematic Mapper	Satellit	[@13]
Landsat-6	1993	8	450-2350	Enhanced Thematic Mapper	Satellit	[@14]
Landsat-7	1999	8	450-2350	Enhanced Thematic Mapper Plus	Satellit	[@15]
Landsat-8	2013	9	450-2200	Operational Land Imager	Satellit	[@16]
SPOT-7	2014	4	450-890	NAOMI	Satellit	[BK16]
AVIRIS	1987	224	400-2500	AVIRIS	Flugzeug	[TBEG17]
HYDICE	1994	210	400-2500	HYDICE	Flugzeug	[RBZ ⁺ 93]
Eo-1	2000	220	400-2500	Hyperion	Satellit	[PBS ⁺ 03]
HYMAP	1998	128	440-2500	HYMAP	Flugzeug	[CJS ⁺ 98]
ROSIS	1989	115	400-900	ROSIS	Flugzeug	[vdPD93]
MERIS	2002	15	390-1040	Envisat	Satellit	[RBB99]

Tabelle 2: Übersicht der spektralen Sensoren zur Erdbeobachtung

ist. Der Begriff der *spektralen Signatur* suggeriert dabei eine eindeutige Übereinstimmung zwischen Material und spektraler Reflexion. Die Anzahl der zu unterscheidenden Materialien bestimmt beim Sensor-design auch die Anzahl und Lage der Spektren, die aufgenommen werden müssen. Sind die zu identifizierenden Materialien vorab bekannt, so können Spezialsensoren konzipiert werden, welche die für die Materialien relevanten Spektren messen können.

2.3.2 Klassifikationsverfahren

Parallel zu der Entwicklung der Sensorik wurden entsprechende Verfahren zur Analyse der erzeugten Daten entwickelt.

Die gemessenen hyperspektralen Daten werden als ein geordneter Vektor von Realzahlen repräsentiert. Die Länge des Vektors entspricht der Anzahl der Spektralbänder. Zu beachten ist dabei, dass Bänder, die spektral nahe beieinander liegen, hoch korreliert sind, wie Kumar et al. [KGC01] darlegen. Ein Standardverfahren zur Dekorrelation der hyperspektralen Daten ist die Hauptachsentransformation, (engl. *Principal Component Analysis*) (PCA) vorgestellt von Richards und Jia [XR99]. Einen analogen Ansatz zeigt das (engl. *Minimum Noise Fraction (MNF)*) Verfahren von Green et al. [GBSC88], welches auch zur Reduktion der hochdimensionalen Hyperspektraldaten genutzt werden kann. Demnach wird durch die zweifache Anwendung der Hauptachsentransformation das Rauschen der Daten berücksichtigt. Weiterhin wurden zur Klassifikation der hyperspektralen Daten verschiedene weitere Klassifikationsverfahren entwickelt, welche sich in zwei Klassen aufteilen und vom Ansatz grundlegend unterscheiden. Relativ simple Verfahren zur Klassifikation vergleichen eine spektrale Referenz mit gemessenen Daten und versuchen so das entsprechende Material zuzuordnen. Aufgrund von entsprechenden Entscheidungs-

regeln oder Schwellwerten wird dem Pixel in einem finalen Schritt dann eine Klasse zugeordnet. Ein Beispiel für ein solches Verfahren ist der von Kruse et al. publizierte *Spectral Angle Mapper (SAM)* [KLB⁺93], welcher ein zu klassifizierendes Spektrum mit einer Datenbank von gegebenen Referenzspektren vergleicht. Daraus wird ein Winkel im spektralen Raum abgeleitet. Je kleiner der berechnete Winkel, desto größer ist die Übereinstimmung zwischen Material und Referenz. Dazu ist jedoch eine entsprechende Referenzdatenbank notwendig. Ein bekanntes Beispiel für eine solche Referenzspektraldatenbank ist die Bibliothek des Geological Survey [CSK⁺93] mit über 500 Spektren von unterschiedlichen Materialien. Ein weiteres Beispiel für solch eine Datenbank stammt von Roberts et al. [RGC⁺98], welche Feld- und Labormessungen von Pflanzen und Dächern beinhaltet. Zu beachten ist hierbei, dass der Einfallswinkel der Strahlung und der Aufnahmewinkel des Sensors signifikanten Einfluss auf das gemessene Spektrum haben [PWS⁺01]. Weiterhin existieren Verfahren, die jedes zu klassifizierende Spektrum als eine Mischung der Spektren von verschiedenen Materialien betrachten und entsprechend eine Entmischung der Materialien versuchen. Dem zugrunde liegt die Beobachtung, dass die Pixel selten nur das Reflexionsspektrum eines Materials aufnehmen.

Innerhalb der Abdeckung eines Pixels gehen die aufgenommenen Oberflächen oft ohne scharfe Grenzen ineinander über. Das Ergebnis der Entmischung ist im Gegensatz zum vorherigen Verfahren keine Klassenzugehörigkeit einer Klasse, sondern eher ein Materialanteil einer vorgegebenen spektralen Referenz, was auch als (engl. *Soft Classifier*) bezeichnet wird. Ein bekanntes Verfahren dieser Kategorie ist das *Linear Spectral Unmixing* [KLB⁺93], welche das Spektrum eines Pixels als Resultat einer linearen Kombination von verschiedenen Materialreflexionen betrachtet. Ein weiteres Verfahren ist das *Matched Filtering* von Harsanyi und Chang [HC94], welches auf gängigen Verfahren der Signalverarbeitung beruht und durch Filteroperationen eine Zielsignatur von einem gemessenen Signal trennt. Eine Weiterentwicklung dieses Verfahrens ist das *Mixture Tuned Matched Filtering (MTMF)* von Boardman [Boa98], welches weitere Bedingungen einführt.

Des Weiteren kann die *Spectral Mixture Analysis (SMA)* verwendet werden, um Informationen aus Pixeln zu extrahieren [LW07]. Dazu wird jedes gemessene Spektrum als eine lineare Kombination aus einer Reihe von Einzelspektren verschiedener Materialien angesehen. Weiterhin können auch geometrische Merkmale oder eine bekannte Form der zu detektierenden Objekte bzw. Oberflächen zur Klassifikation genutzt werden, wie Segl et al. [SRHK03] demonstrieren. Die meisten Verfahren zur Klassifikation der spektralen Daten basieren auf dem Prinzip einer per-Pixel Verarbeitung der Szene, wie Blaschke und Strobl konstatieren [Bla01]. Dabei kommen häufig auch nicht-

parametrische Nächster-Nachbar-Klassifikatoren zum Zuge. Der einfache Nächste-Nachbar-Klassifikator berechnet den euklidischen Abstand vom zu klassifizierenden Pixel zum nächsten Trainingsdatenpixel im n -dimensionalen Merkmalsraum und ordnet ihn dieser Klasse zu [Scho7b, MGB⁺11]. Der Klassifikator des nächstgelegenen Nachbarn macht dabei keine Annahmen über die Verteilung der Daten im Merkmalsraum. Sie können aber sehr gute Ergebnisse erzielen, wenn die Trainingsdaten gut im Merkmalsraum verteilt sind. Diese Klassifikatoren basieren dabei in erster Linie auf der Konstruktion von Grenzen im Merkmalsraum. Dazu werden multispektrale Distanzrechnungen basierend auf der Trainingsklasse durchgeführt.

Die Maximum-Likelihood-Methode ist ein weiteres Verfahren, welches auf dem Gesetz der Wahrscheinlichkeit beruht. Es ordnet jedem Pixel eine Klasse zu, basierend auf der Annahme, dass die Verteilung der Werte am wahrscheinlichsten einer bestimmten Referenzklasse angehören [LY07, DBVG09]. Dazu wird für jede Klasse einer vordefinierten Menge von Klassen die Wahrscheinlichkeit berechnet, mit welcher das jeweilige Pixel ihr zugehörig ist. Das Pixel wird dann der Klasse mit der höchsten Wahrscheinlichkeit zugeordnet. Die Maximum-Likelihood-Methode ist immer noch eine der am weitesten verbreiteten überwachten Klassifikationsalgorithmen [Cam11, MGB⁺11].

Sind keine Trainingsdaten verfügbar, ist die unüberwachte Klassifikation, auch allgemein als Clustering bezeichnet, eine weitere Methode zur Klassifikation von Sensordaten. Bei diesen Algorithmen sind keine Trainingsdaten nötig, da versucht wird, in den gegebenen Daten Muster zu finden und die Daten anhand dieser Muster zu trennen bzw. zu gruppieren. Huang et al. [Hua02] demonstrieren, wie diese Verfahren zur Extraktion von Oberflächeninformationen genutzt werden können. Dazu werden numerische Operationen durchgeführt, die nach natürlichen Gruppierungen der Messungen im spektralen Merkmalsraum suchen. Dieser Clustering-Prozess führt dann zu einer Klassifikation, welche aus einer definierten Menge von spektralen Klassen besteht, wie Lo et al. [LY07] erläutern. Da keine Trainingsdaten notwendig sind, werden diese Verfahren verbreitet eingesetzt, sodass viele verschiedene Clustering-Algorithmen entwickelt [Dudo1, Scho7b] wurden. So ist die *Self-Organizing Data Analysis Technique* (ISODATA) ein weiterer weit verbreiteter Clustering-Algorithmus. ISODATA repräsentiert dazu eine Reihe von heuristischen Regeln, die in einen iterativen Prozess integriert wurden, wie Pascher et al. und Rich et al. [PK10, RFRB10] aufzeigen. Eine Zusammenfassung der Eigenschaften und eine Implementation des Algorithmus wird von Memarsadeghi et al. [MMNM07] geliefert.

2.3.3 *Fazit*

Entsprechend [Rico6, CVTGC⁺11] ist das Feld der spektralen Datenaufnahme und Analyse sehr groß. Ein Bereich in diesem Feld ist die Fernerkundung. Haupteinsatzgebiet der Daten hier sind die Generierung von Kartendaten zur Erdbeobachtung und Analyse der Erdoberfläche für verschiedene Zwecke. Neben der Fernerkundung ist die Analyse von Stoffen wie Lebensmitteln unter kontrollierten Bedingungen ein weiterer großer Anwendungsbereich. Viele Algorithmen zur Klassifikation und Analyse spektraler Daten entstammen diesen Bereichen. Die in dieser Dissertation untersuchte Sensorik ist relativ neu und unbekannt. Auch ist das Einsatzszenario zur Umgebungswahrnehmung bei der autonomen Navigation bisher soweit bekannt, nicht mit derlei Sensorik untersucht worden.

3.1 EINFÜHRUNG

Ein zentrales Forschungsthema dieser Arbeit bilden spektrale Daten. Um diese Daten mit üblichen Bilddaten vergleichen zu können, wird hier zunächst die Technik der Bildgebung näher untersucht und beschrieben. Weiterhin werden wichtige Begriffe erläutert und spezifiziert. Dazu wird im Abschnitt 3.3 die in dieser Arbeit verwendete Sensorik erläutert. Anschließend wird in Abschnitt 3.5 eine kurze Einführung in die Wahrnehmung sowie Erläuterungen zum *CIE*-Normalbeobachter gegeben. Darauf aufbauend wird ein Verfahren zur Konvertierung von spektralen Daten zu Farb- und Grauwertbildern erläutert. Dies ist notwendig, um einen späteren Vergleich der Daten zu ermöglichen. Im Abschnitt 3.6 werden die Grundlagen der spektralen Bildgebung der Snapshot-Mosaik-Technik erläutert. Darauf folgen in Abschnitt 3.7 formale Definitionen zur verwendeten Terminologie.

3.2 BILDSEGMENTIERUNG

*Angelehnt an
[FP02, HB20]*

Definition 4: Szenenanalyse

Bei der Szenenanalyse oder semantischen Segmentierung ist es nach [FP02] Ziel, dem System ein Verständnis der Szene zu vermitteln. Semantische Segmentierung ist hier eine klassische Aufgabe des Rechnersehens, bei der den verschiedenen Teilen eines visuellen Inputs semantisch interpretierbare Klassen zugeordnet werden. „Semantisch interpretierbar“ bezeichnen dabei Klassen, die für den Menschen eine gewisse reale Bedeutung haben.

Die Bildsegmentierung ist eine Aufgabe der Bildverarbeitung, bei der bestimmte Bereiche eines Bildes in Klassen unterteilt und entsprechend annotiert werden. Genauer gesagt, ist das Ziel der semantischen Bildsegmentierung oder auch Szenenanalyse, jedes Pixel eines Bildes mit einer entsprechenden Klasse zu kennzeichnen. Dies ist beispielhaft in Abbildung 9 dargestellt. Wichtig ist noch zu beachten, dass nicht zwischen Instanzen derselben Klasse unterschieden wird. Relevant ist nur die Klasse jedes einzelnen Pixels. Wenn

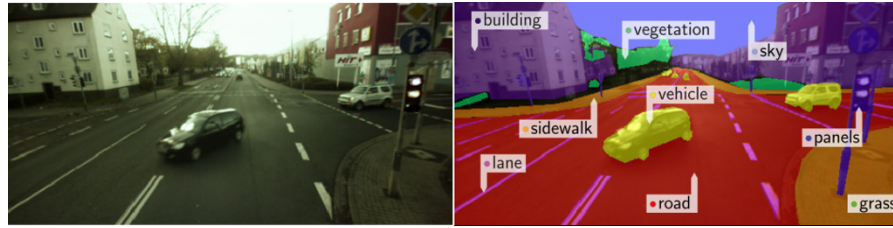


Abbildung 9: Beispiel zur semantischen Segmentierung. Links dargestellt ist ein aus spektralen Daten generiertes RGB-Bild. Rechts daneben ist die zugehörige semantische Analyse dargestellt.

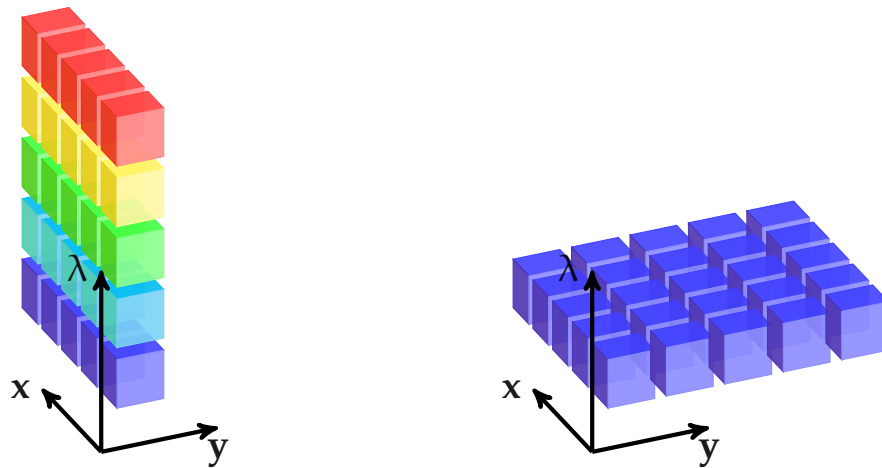
zwei Objekte derselben Klasse sich in der Szene befinden, unterscheidet die Segmentierung nicht von Natur aus zwischen den Objekten. Dafür existiert eine weitere Klasse von Methoden, die sogenannten *Instanzsegmentierungsmethoden*, die zwischen einzelnen Objekten derselben Klasse unterscheiden, wie von Hafiz et al. [HB20] erläutert. Die semantische Segmentierung ermöglicht ein viel detaillierteres Verständnis von Bildern als eine reine Bildklassifizierung oder Objekterkennung. Dieses detaillierte Verständnis ist in einer Vielzahl von Bereichen entscheidend, darunter autonomes Fahren und Robotik. Zum Lösen dieser Aufgabe existieren grundsätzlich mehrere Methoden aus den Bereichen des überwachten und unüberwachten Lernens. Während unüberwachte Methoden wie Clustering zur Segmentierung verwendet werden können, sind durch die Ergebnisse nicht unbedingt semantisch. Diese Methoden segmentieren nicht Klassen, auf denen sie trainiert wurden, sondern finden Regionengrenzen innerhalb der Daten.

3.2.1 Sensoren

Angelehnt an
[SB03]

Grundsätzlich existieren in der etablierten spektralen Bildgebung mehrere Varianten von Messprinzipien. Bei einer Variante werden die verschiedenen Spektren durch eine zeitliche Abfolge von Bildaufnahmen mit verschiedenen Wellenlängen in einem Hyperwürfel übereinander gestapelt. Dazu wird ein durchstimmbarer Filter genutzt, welcher für jede Aufnahme auf eine andere Wellenlänge eingestellt wird oder ein Filterrad mit einer Vielzahl von Einzelfiltern für jeweils einen Spektralzug. Bei der anderen werden unter Verwendung von sog. Zeilensensoren (engl. *Line-Scanning*) eindimensionale Linienbilder erzeugt, welche direkt alle spezifizierten Wellenlängen enthalten und durch Bewegung des Sensors zu einem räumlichen Bild zusammengesetzt.

Zeilensensoren sind die in der Fernerkundung, der spektralen Bildaufnahme von Bildern mit großer Farbexaktheit in der Bildre-



(a) Räumliche Abtastung mit einem Zeilensensor

(b) Spektrale Abtastung mit durchstimmbaren Filtern

Abbildung 10: Überblick über die Techniken zur spektralen Bildgebung visualisiert als Ausschnitte eines Hyperwürfel für ein Frame bzw. eine Aufnahme. Bei der räumlichen Abtastung liefert jede Aufnahme das gesamte vom Sensor messbare Spektrum von einem schmalen Streifen der Szene. Bei der spektralen Abtastung liefert jede Aufnahme ein Abbild der Szene jedoch nur mit einer definierten Wellenlänge.

produktion, der Materialprüfung in der Industrie und der Medizin am häufigsten eingesetzten Sensortypen. Sie besitzen entweder ein Prisma oder ein Gitter als spaltendes Element. Dadurch wird die eintreffende Strahlung in einzelne Wellenlängenbereiche aufgetrennt und dann durch die Sensoren registriert. Kombiniert mit der Flugbewegung eines Sensorträgers entsteht eine fortlaufende zeilenweise Aufnahme z. B. der Erdoberfläche oder eines zu prüfenden Materials in einer industriellen Produktion. Dies ist eine Form der zeitsequenzierten Bildgebung, welche als dreidimensionaler Datenwürfel dargestellt wird. Dieses Sensorprinzip erzeugt durch eine geringe mögliche Kanalbreite Sensordaten mit einer hohen spektralen Auflösung. Daher werden diese Sensorsysteme auch der Kategorie der hyperspektralen Sensoren zugeordnet. Sie grenzen sich damit von multispektralen Sensoren ab, welche neben der wesentlich höheren Anzahl von Kanälen auch eine geringe spektrale Breite des jeweiligen Kanals aufweisen.

Zur Unterscheidung der jeweiligen Sensormodularität existieren vier Abtastoperationen. Diese Operationen erzeugen jeweils auf unterschiedliche Art und Weise einen Hyperwürfel. Eine schematische Übersicht dieser Techniken ist in Abbildung 10 und Abbildung 11 dargestellt.

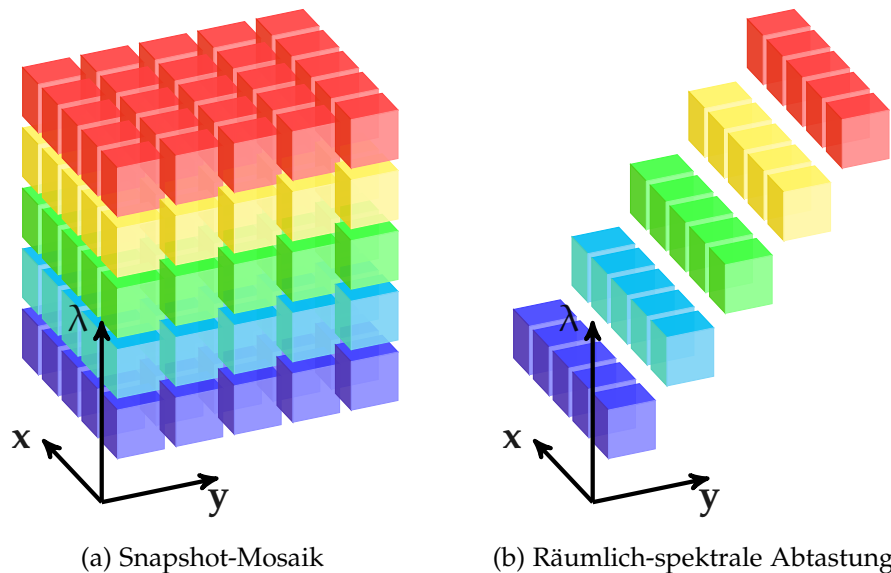


Abbildung 11: Überblick über die Techniken zur hyperspektralen Bildgebung visualisiert als Ausschnitte eines Hyperwürfel für ein Frame bzw. eine Aufnahme. Bei der Snapshot-Mosaik-Technik wird mit einer Aufnahme sowohl die gesamte Szene als auch alle vom Sensor messbaren Wellenlängen erfasst. Bei der räumlich-spektralen Technik liefert jede Aufnahme eine spektral kodierte räumliche Karte der Szene.

RÄUMLICHE ABTASTUNG Bei der räumlichen Abtastung wird mit jeder Aufnahme ein spezifiziertes Spektrum komplett gemessen, wie in Abbildung 10a dargestellt. Dazu wird in der Regel ein spektraler Zeilensensor genutzt. Bei einem spektralen Zeilensensor wird meistens das einfallende Licht senkrecht zur physikalischen Anordnung der Bildzeile des Objektes spektral aufgespalten und dann mit einer zweidimensionalen Anordnung von elektrischen Sensorelementen, einer Bildmatrix mit $m \times n$ Elementen, aufgenommen. Wenn m Bildpunkte einer Bildzeile des Objektes aufgenommen werden sollen, dann kann mit den n Zeilen der Bildmatrix für jedes der m Pixel eine spektrale Auflösung von n spektralen Kanälen erreicht werden. Vorausgesetzt ist, dass das Element zur spektralen Aufspaltung des Lichtes (z. B. Gitter oder Prisma) eine ausreichende Trennung der Spektralanteile in n unterscheidbare Anteile ermöglicht. Die Zeilensensoren werden in der Regel in der Fernerkundung auf Satelliten eingesetzt. Alternativ werden sie eingesetzt, um Materialien, z. B. in der Lebensmittelindustrie zu untersuchen, die auf Förderbändern transportiert werden.

SPEKTRALE ABTASTUNG Bei der spektralen Abtastung wird ein angenähert monochromatisches Bild integriert über einen begrenzten,

schmalen Wellenlängenbereich aufgenommen. Der Wellenlängenbereich wird durch die Bandbreite des verwendeten Bandpassfilters bestimmt. Dies ermöglicht Aufnahmen einer Szene mit einer begrenzten Bandbreite der Abtastwerte, wie in Abbildung 10b dargestellt. Diese Systeme nutzen üblicherweise einstellbare oder feste Bandpassfilter zur Erzeugung der monochromatischen Aufnahme. Um nun ein volles Spektrum einer Szene aufzunehmen, wird der Filter immer neu konfiguriert oder ausgetauscht. Während des Messvorgangs darf der Sensor nicht bewegt werden bzw. die Szene muss statisch sein, da sonst Schliereffekte auftreten, was die Erstellung eines vollständigen Hyperwürfels verhindert.

SNAPSHOT-MOSAİK Beim Snapshot-Mosaik-Prinzip enthält eine Aufnahme direkt sowohl die spektralen wie auch die räumlichen Informationen, wie in Abbildung 11a dargestellt. Es ist sozusagen eine Kombination aus beiden vorherigen Sensortechniken. Sie liefert in einer Aufnahme den vollen Hyperwürfel. So stellt eine Aufnahme im Prinzip eine perspektivische Projektion der Umgebung auf einen Hyperwürfel dar. Der Vorteil dieses Prinzips ist offensichtlich: Aufgrund der kurzen Aufnahmezeit für ein volles spezifiziertes Spektrum können auch dynamische Szenen erfasst werden. Der Nachteil ist die oft komplexe Vorverarbeitung der Daten, um aus der Aufnahme einen Hyperwürfel zu berechnen und die vergleichsweise niedrigere Anzahl an messbaren Wellenlängen. Ein weiterer Nachteil ist, dass aufgrund des Makropixelmusters die Pixel, welche für dieselbe Wellenlänge empfindlich sind, auf dem Sensor nicht nebeneinander angeordnet sind. Damit nehmen die jeweiligen Pixel auch Informationen aus der Szene auf, die je nach Entfernung weit auseinander liegen. Dies führt insbesondere bei einer Kante im Bild, welche zwischen Pixeln mit verschiedener geometrischer Lage links und rechts der Kante angeordnet sind, zu Schwierigkeiten der Bestimmung des zugeordneten Spektrums links und rechts der Kante. Dies ist ein grundsätzliches Problem dieser Sensortechnologie, welches bisher nicht gelöst wurde. Ein Überblick über die verschiedenen Implementierungen dieser Sensortechnologie wird von Hagen et al. [HK13] gegeben.

RÄUMLICH-SPEKTRALE ABTASTUNG Bei der räumlich spektralen Abtastung wird mit einer Aufnahme eine Reihe von diagonalen Schichten des Hyperwürfels gemessen, wie in Abbildung 11b dargestellt. Jede Aufnahme ist eine wellenlängenkodierte räumliche Karte der Szene.

3.2.2 Effekte der Bildgebung

Es gibt einige Faktoren, welche bei der Konzeption eines spektralen Sensors berücksichtigt werden müssen. Dazu gehören die räumliche

*Angelehnt an
[SB03]*

und spektrale Auflösung eines Sensors, sowie die atmosphärischen Effekte wie Absorption und Streuung, welche bei der Fernerkundung eine wichtige Rolle spielen. Weitere Faktoren bzw. Effekte der Bildaufnahme sind die spektrale Variabilität, Umwelteffekte wie Blickwinkel, Sekundärbeleuchtung und Schattenbildung. Die Erdatmosphäre moduliert das Sonnenspektrum, bevor es den Boden erreicht. Ändern sich dazu der Einstrahlwinkel des Lichts und der Blickwinkel des Sensors, werden dadurch die gemessenen Werte beeinflusst. Des Weiteren gelangen potentiell Anteile des Sonnenspektrums ohne Reflexion an einer Oberfläche in den Sensor. Auch der Sonnenwinkel und die Oberflächenausrichtung beeinflussen die Menge des Lichts, welches aufgenommen wird. Es wurden in Experimenten mit Feld- und Labordaten Schwankungen in den Reflexionsspektren von vielen Materialien beobachtet, wie Shaw und Burke [SB03] zeigen.

Es können viele Effekte für diese Schwankungen verantwortlich sein. Dazu gehören unkompenzierte Fehler in der Sensorik sowie Atmosphären- und Umwelteffekte, Oberflächenverunreinigungen und Materialschwankungen. Auch zu berücksichtigen sind hier Nachbarschaftseffekte, bei denen Reflexionen von nahe gelegenen Oberflächen die Beleuchtung der gemessenen Oberfläche verändern. Ein Beispiel sind die Veränderungen in einem Laubwald im Frühjahr, Sommer, Herbst und Winter, an denen auch der Umfang der spektralen Variabilität verständlich gemacht werden kann.

3.3 VERWENDETE SENSORIK

3.3.1 Kameras

Zentraler Gegenstand dieser Arbeit sind zwei Kameras der Firma Ximea. Als Teil der xiQ Kameraserie zeichnen sich beide Kameras durch eine kompakte Bauweise (26,4 mm × 26,4 mm × 31 mm) und ein geringes Gewicht (31 g) aus. Betrieben werden die Kameras über eine USB-3.0-Schnittstelle. Durch die vom Hersteller zur Verfügung gestellte *xiAPI*-Schnittstelle ist es möglich, die Rohdaten der Kamera softwareseitig zu empfangen und Veränderungen an den Betriebsparametern wie z. B. der Belichtungszeit vorzunehmen. In beiden Kameramodellen werden 2/3 Zoll CMOSIS CMV2000 Sensoren der Firma Interuniversity Microelectronics Centre (IMEC) verwendet, welche Rohbilder mit einer Auflösung von 2048 px × 1088 px aufnehmen und dabei eine theoretische Bildwiederholrate von bis zu 170 fps erreichen können. Die Bezeichnung und Spezifikationen der beiden Kameras sind in der Tabelle 3 aufgelistet.

Um spektrale Messungen zu ermöglichen, wurde durch den Hersteller auf dem CMOS-Sensor eine pixelgenaue Filtermatrix aufgebracht, sodass jeder Pixel individuell sensitiv auf einen bestimmten Wellenlängenbereich reagiert.

Modell	MQ022HG-IM-SM ₄ X ₄ -VIS	MQ022HG-IM-SM ₅ X ₅ -NIR
Spektralbereich	400 – 675 nm	675 – 975 nm
Bänder	16	25
Bildrate	bis zu 170 fps	bis zu 170 fps
Auflösung (nativ)	2048 × 1088 Pixel	2048 × 1088 Pixel
Auflösung	512 × 272 Pixel	409 × 217 Pixel
Sensor	CMOSIS CMV2000	CMOSIS CMV2000
Abmessungen	26,4 × 26,4 × 31 mm	26,4 × 26,4 × 31 mm
Kürzel	VIS	NIR

Tabelle 3: Spezifikation der in dieser Arbeit verwendeten bildgebenden Sensorik



Abbildung 12: Darstellung der beiden Kameramodelle der Firma Ximea mit unterschiedlichen Objektiven

Die verwendete Filtermatrix besteht hierbei aus verteilten Bragg-Spiegeln unterschiedlicher Höhe, um einen lichtbrechenden Effekt nach dem Fabry-Pérot-Interferometer Prinzip [Fab99] zu erzeugen. So wird das durch den CMOS-Sensor aufgenommene Signal spektral gefiltert. Abbildung 13 und Abbildung 14 zeigen den Aufbau und die Verteilung mehrerer Wellenlängenfilter mit einer Halbwertsbreite zwischen 5 – 20 nm auf dem jeweiligen Sensor. Dargestellt ist eine quadratische Einheit von Wellenlängenfiltern, hier als Makropixel bezeichnet, welche sich über der gesamten Sensorfläche wiederholt. Unter der Annahme, dass die Pixel innerhalb eines Makropixels nahezu den gleichen Ort abbilden, ergeben die Messwerte dieser Pixelgruppe eine spektrale Messung für einen definierten Ort im Bild. Je nach Größe und Wellenlängensensitivität der Makropixel ändert sich das Spektrum, das durch den Sensor abgedeckt werden kann. Da der Annahme der gleichen Ortsauflösung physikalisch Grenzen gesetzt sind, lässt sich aber die spektrale Auflösung nicht beliebig präzise wählen. Auch sind dieser Filtertechnik physikalische Grenzen gesetzt.

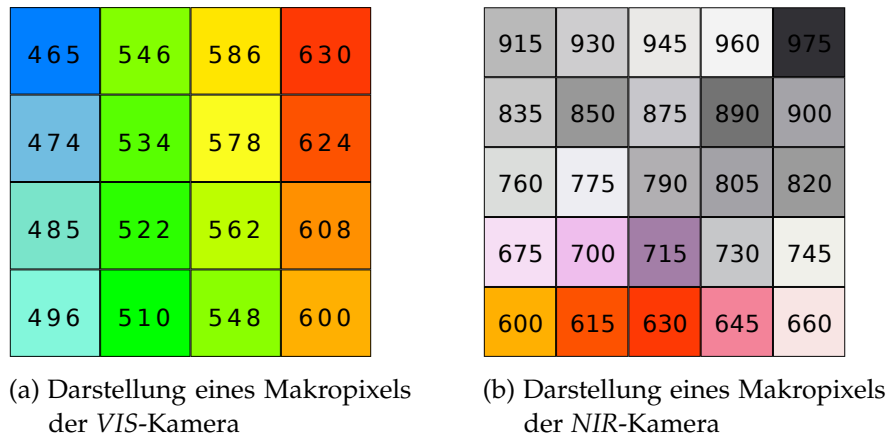


Abbildung 13: Schema je eines Makropixels der VIS- und NIR-Kamera mit entsprechenden Wellenlängenempfindlichkeiten in nm

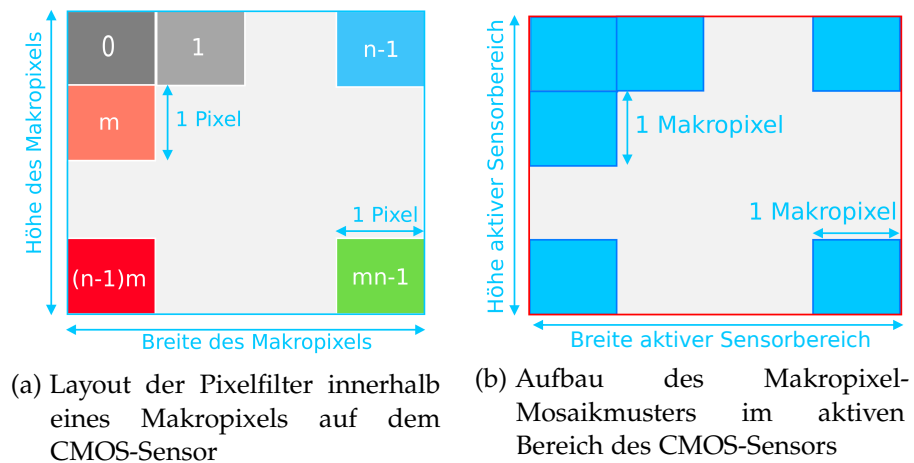


Abbildung 14: Schema von Layout und Aufbau der Makropixel und Filter auf dem Sensorchip der Kameras von Ximea

Ximea stellt daher zwei Kameras mit unterschiedlichen spektralen Empfindlichkeiten zur Verfügung. Die erste Kamera vom Modelltyp MQ022HG-IM-SM4X4-VIS (VIS) ist mit einem 4×4 Makropixelmuster - also 16 Messwerten- ausgestattet und bildet das Spektrum des sichtbaren Lichts im Bereich von etwa 400 – 675 nm ab. Die zweite Kamera vom Typ MQ022HG-IM-SM5X5-NIR (NIR) ist für den Bereich von 675 – 975 nm ausgelegt und deckt somit sowohl das visuelle Rotpektrum als auch Teile des Nahinfrarotbereichs ab. Hier kommt ein 5×5 Filtermuster zum Einsatz, so dass eine Spektralmessung also 25 Messwerte umfasst. Kombiniert man beide Kameras, lässt sich also theoretisch der Wellenlängenbereich von 400 – 975 nm durch insgesamt 41 Messwerte abbilden.



Abbildung 15: Fahrzeug mit Kameras (rot hervorgehoben) und zwei 3D-Laserscannern (blau hervorgehoben)

OBJEKTIVE Aufgrund der Winkelabhängigkeit des Fabry-Pérot-Filters ist nicht jede Linse zur Aufnahme spektraler Daten geeignet. Die Linse transformiert den Einfallswinkel des Lichts auf den Sensor. Für eine optimale Leistung der Kameras sollten die Strahlen orthogonal zum Sensor einfallen. Dies ist jedoch nicht bei allen Objektiven der Fall. Idealerweise wird daher eine beidseitig telezentrische Linse verwendet, bei der sowohl auf der Objektseite als auch auf der Sensorseite die Hauptstrahlen parallel zur optischen Achse verlaufen. Weiterhin sollte das Objektiv auch geeignet sein, Wellenlängen sowohl im sichtbaren als auch im nahen Infrarotbereich zu erfassen. Das bedeutet, dass die Linse eine geeignete Entspiegelung (ARC) für Licht im gesamten Bereich von 400 – 1000 nm aufweisen muss. In dieser Arbeit wird eine Linse mit einer festen Brennweite von Kowa mit der Bezeichnung LM5JC10M¹ verwendet. Nach Angaben des Handbuchs des Objektivs wurde hier die chromatische Aberration in jedem Abstand erheblich reduziert und eine hohe Transmission vom sichtbaren zum NIR-Bereich des Spektrums kann gewährleistet werden. Die Linse verfügt über einen großen Öffnungswinkel, was es erlaubt, einen großen Bereich der Szene vor einem Fahrzeug aufzunehmen.

3.3.2 Auslöser-Board

Damit die verwendeten Kameras synchron Bilder aufnehmen können, müssen sie möglichst präzise synchronisiert werden. Dazu wurde ein Arduino-Nano-Board [1] verwendet, welches in Abbildung 16 abge-

¹ <https://lenses.kowa-usa.com/10mp-jc10m-series/397-lm5jc10m.html>

bildet ist. Das Board lässt sich mit einfachen Mitteln programmieren und bietet eine kostengünstige Möglichkeit Hardware entsprechend anzusteuern. Zur Synchronisierung der Kameras wurde ein spezielles Programm für den Arduino genutzt und modifiziert, welches in konfigurierbaren zeitlichen Abständen ein Triggersignal sendet und somit eine Kameraaufnahme auslöst. Die Kameras wurden entsprechend den Handbüchern und Dokumentationen mit passenden Pins und Trigger-Kabeln am Arduino Board angeschlossen und erlauben so eine zeitlich synchronisierte Aufnahme von Sensordaten. Mit dem Rechner wird das Board dann über eine USB-Schnittstelle verbunden.

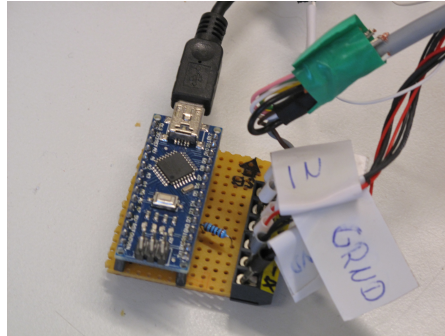
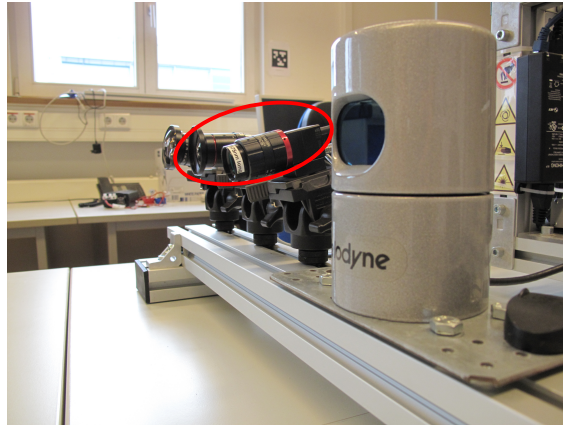


Abbildung 16: Als Auslöser-Board programmierter Arduino-Nano mit entsprechender Verkabelung zur Steuerung der angeschlossenen Kameras

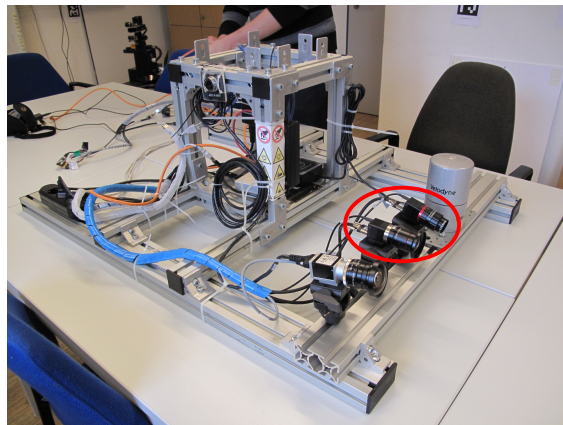
3.3.3 Sensor-Plattform

Zur Montage der Sensorik und Aufnahme von Daten mit der verwendeten Sensorik wurde ein Alukonstrukt verwendet, welches auf die Dachträger eines normalen Pkws montiert werden kann. Bei der Konstruktion des Aufbaus wurde auf Flexibilität geachtet, um später ohne großen Aufwand weitere Sensoren auf dem Dachträger montieren zu können. Der Aufbau ist in Abbildung 17 dargestellt. Darauf montiert wurden unter anderem zwei spektrale Kameras, eine RGB-Kamera und der 3D-Laserscanner.

Alternativ ist die Sensorik auch auf einem LKW montiert und in Benutzung wie in Abbildung 50a dargestellt. Der LKW wird im Rahmen einer F&T-Studie des Bundesamts für Ausrüstung, Informationstechnik und Nutzung der Bundeswehr (BAAINBw) entwickelt. Hier sollen verschiedene Autonomiealgorithmen auf einem Fahrzeug erprobt werden, wobei insbesondere Algorithmen zur Umgebungswahrnehmung hier eine wichtige Rolle spielen. Die in dieser Arbeit verwendeten Daten wurden mit den zuvor genannten Sensorplattformen aufgenommen.



(a) Seitenansicht des Sensoraufbaus



(b) Sensorikaufbau

Abbildung 17: Sensor-Plattform mit montierter Sensorik zur Anbringung auf einem PKW. Die Spektralkameras sind rot umrandet.

3.4 BILDERZEUGUNG

Nach Bernd Jähne [Jä12] und dem EMVA 1288 Standard [A⁺10] lässt sich jedes digitale Signal eines Sensors als Funktion der Photonenzahl ausdrücken, welche über einen definierte Belichtungszeit auf das Sensorelement auftreffen:

*Angelehnt an [Jä12]
und [A⁺10]*

$$I = K(\eta \cdot N_p + I_d) \quad (1)$$

mit folgenden Komponenten:

- I als der gemessene Signalwert
- K als Kameraverstärkung
- η als Quanteneffizienz
- N_p als mittlere Photonenzahl akkumuliert über die Belichtungszeit

- I_d Dunkelstrom in Anzahl der Ladungsträger

Die Energie der elektromagnetischen Strahlung ist in den Photonen quantisiert und hängt direkt von der jeweiligen Wellenlänge λ ab. Bei der Aufnahme trifft eine Anzahl von Photonen N_p auf das Sensorelement und wird entsprechend der Quanteneffizienz η absorbiert und aufgrund des photoelektrischen Effekts in Elektronen überführt. Hinzu kommt der Dunkelstrom I_d welcher auch vom Sensor generiert wird, wenn er kein Signal aufnimmt. Die Kameraverstärkung K generiert dann schlussendlich aus der empfangenen Strahlung einen digitalen Signalwert I . Die Werte der einzelnen Sensorelemente bilden dann zusammengenommen ein digitales Abbild der aufgenommenen Szene.

3.5 STRAHLUNG UND WAHRNEHMUNG

3.5.1 Elektromagnetisches Spektrum

Angelehnt an [SB03,
MSA⁺78, ISO89]

Die Intensität von Licht als elektromagnetischer Welle wird durch die Bestrahlungsstärke (engl. *irradiance*) E beschrieben. Sie ist definiert als die Lichtenergie, welche pro Zeiteinheit auf eine definierte Oberfläche trifft und wird als W/m^2 beschrieben. Trifft das Licht auf eine Oberfläche, wird es abhängig vom Material in entsprechenden Anteilen absorbiert, reflektiert und unter Umständen auch transmittiert. Der von der Oberfläche reflektierte Strahlungsfluss wird von einem Sensor gemessen. Wird nun der gemessene Strahlungsfluss $\phi_r(\lambda)$ ins Verhältnis zu dem auf der Oberfläche einfallenden Strahlungsfluss $\phi_i(\lambda)$ gesetzt, resultiert daraus ein Proportionalitätsfaktor $\rho(\lambda)$, welcher von der Wellenlänge λ abhängig ist und als Reflexionsgrad bezeichnet wird. Dies gilt allerdings nur unter der Annahme, dass die einfallende Strahlung isotrop ist [ISO89]. Zur allgemeinen Beschreibung des Reflexionsgrades wird eine Funktion benötigt, die für alle möglichen Einfallrichtungen das reflektierte Licht für alle möglichen Ausfallrichtungen angibt. Dafür wurde die bidirektionale Reflexionsverteilungsfunktion (BRDF) definiert. Für alle auf einer Oberfläche auftreffenden Lichtstrahlen mit gegebenem Eintrittswinkel liefert die Funktion den Quotienten aus Bestrahlungsstärke $E_i(\lambda)$ und $L_r(\lambda)$ für jeden austretenden Lichtstrahl. Abgesehen von Laborumgebungen mit definierter Beleuchtungssituation lässt sich allerdings die winkelabhängige Bestrahlungsstärke nicht präzise bestimmen, daher wird eine Alternative zur BRDF verwendet. Die Alternative wird als Reflexionsfaktor bezeichnet und ist durch die Standardisierung der reflektierten Strahldichte definiert. Dazu wird eine Platte verwendet, die als perfekt diffus und vollständig reflektierend spezifiziert ist und unter denselben Bestrahlungsbedingungen und in derselben Geometrie wie die Messoberfläche betrachtet wird.

Nach Shaw et al. [SB03] kann bei gegebenem Beleuchtungsspektrum

das Reflexionsspektrum oder das Reflexionsvermögen des Materials grundsätzlich aus der spektralen Strahldichte, unter Vernachlässigung der Fluoreszenz, berechnet werden. Das Ermitteln von Reflexionsinformationen aus gemessenen Strahldichten, welche das wellenlängenabhängige (spektrale) Reflexionsvermögen einer Oberfläche widerspiegeln, ist die Grundlage der hyperspektralen Datenverarbeitung. Der Reflexionsgrad ist unabhängig von der Beleuchtung, was es ermöglicht, Materialien und Oberflächen in einer Szene zu klassifizieren. Dazu ist es nötig, möglichst präzise die Zusammensetzung der ursprünglich emittierten Strahlung zu kennen. Doch dies ist außerhalb von Laborumgebungen ein schwieriges Unterfangen. Bei der solaren Beleuchtung ist die spektrale Zusammensetzung des in die Atmosphäre einfallenden Lichts gut messbar. Allerdings wird diese Zusammensetzung vor dem Auftreffen auf die Erdoberfläche durch diverse Umwelteinflüsse wie die sich ständig verändernde und ortsabhängige Zusammensetzung der Atmosphäre beeinflusst [CVTGC⁺11].

Definition 5: Reflexionsgrad

Der spektrale Reflexionsgrad ist entsprechend dem ISO-Standard 9288 [ISO89] der Quotient aus von einer Oberfläche reflektiertem und einfallendem Strahlungsfluss ϕ :

$$\rho(\lambda) = \frac{\phi_i(\lambda)}{\phi_r(\lambda)} \quad (2)$$

Dies gilt unter der Annahme, dass der auftreffende Strahlungsfluss isotrop ist. Grundsätzlich ist der Reflexionsgrad eine einheitslose Zahl zwischen 0 und 1, die den Anteil des von einer Oberfläche reflektierten Lichts charakterisiert. Die Abhängigkeit des Reflexionsgrads von der Wellenlänge (λ) wird als Reflexionsspektrum oder spektrale Reflexionskurve bezeichnet [SB03].

Definition 6: Bidirektionale Reflexionsverteilungsfunktion

Die bidirektionale Reflexionsverteilungsfunktion definiert sich nach [MSA⁺78] durch das Verhältnis zwischen der in den Halbraum reflektierten Strahldichte $L_r(\lambda)$ und auf der betrachteten Oberfläche aus dem Halbraum einfallenden Bestrahlungsstärke $E_i(\lambda)$. Daraus folgt:

$$f_r(\varphi_r, \theta_r, \varphi_i, \theta_i) = \frac{dL_r(\varphi_r, \theta_r)}{dE_i(\varphi_i, \theta_i)} = \frac{dL_r(\varphi_r, \theta_r)}{L_i(\varphi_i, \theta_i) \cdot \cos \varphi_i d\omega_i} \quad (3)$$

Die Strahlungsquelle strahlt aus einer bestimmten Position auf die Oberfläche, welche durch einen Horizontalwinkel φ und einen Vertikalwinkel θ von einer Flächennormalen angegeben wird. Eine natürliche Oberfläche reflektiert die Strahlung nicht perfekt diffus, das heißt die gemessene Strahldichte ist auch abhängig vom Winkel des Sensors relativ zur Oberfläche. Die BRDF ist folglich eine Funktion in Abhängigkeit von der Geometrie der einfallenden und reflektierten Strahlung.

Definition 7: Reflexionsfaktor

Der Reflexionsfaktor definiert sich nach [MSA⁺78] als das Verhältnis zwischen der von einem Objekt in eine bestimmte Richtung reflektierten Strahldichte $L_r(\lambda)$ und der von einem perfekt diffus reflektierenden weißen Lambert-Reflektor bei gleicher Beleuchtung reflektierten Strahldichte $L_t(\lambda)$.

$$R(\lambda, \varphi_r, \theta_r, \varphi_t, \theta_t) = \frac{L_r(\lambda, \varphi_r, \theta_r)}{L_t(\lambda, \varphi_t, \theta_t)} \quad (4)$$

3.5.2 *Licht und Farbe*

Angelehnt an
[SBo2, RKA]08]

Die folgenden Informationen basieren auf den Arbeiten von Sharma und Bala [SBo2] sowie von Rheinard et al. [RKA]08]. Die relevanten radiometrischen und photometrischen Größen sind zur Übersicht in Tabelle 4 dargestellt.

Der physikalische Reiz für Farbe im menschlichen Auge ist elektromagnetische Strahlung im sichtbaren Bereich des elektromagnetischen Spektrums, welches allgemein als *Licht* bezeichnet wird. Der sichtbare Bereich des elektromagnetischen Spektrums wird in Luft oder Vakuum typischerweise durch den Wellenlängenbereich zwischen

Radiometrie	Symbol	Einheit	Beschreibung	Photometrie	Symbol	Einheit
Strahlungsfluss	Φ	W	Strahlungsenergie pro Zeit	Lichtstrom	Φ_v	lm
Strahlungsenergie	Q	J	Energie einer Anzahl von Photonen	Lichtmenge	Q_v	lm · s
Strahlstärke	I	W/sr	Strahlung pro Raumwinkel	Lichtstärke	I_v	cd
Bestrahlungsstärke	E	W/m ²	Strahlungsfluss pro Empfängerfläche	Beleuchtungsstärke	E_v	lx
Spektrale Bestrahlungsstärke	\mathcal{E}	W/m ² · λ	Strahlungsfluss pro Empfängerfläche pro Wellenlänge			
Strahldichte	L	W/m ² · sr	Strahlungsfluss pro Raumwinkel pro Senderfläche	Leuchtdichte	L_v	cd/m ²

Tabelle 4: Überblick über die relevanten radiometrischen Größen und deren photometrisches Gegenstück.

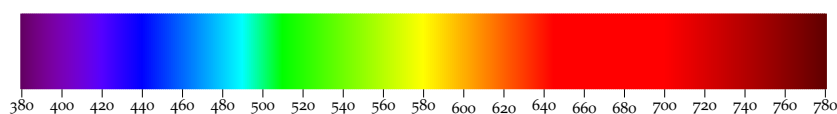


Abbildung 18: Darstellung des sichtbaren Lichts als Teil des elektromagnetischen Spektrums mit Wellenlängenangaben

$\lambda_{\min} = 380\text{nm}$ und $\lambda_{\max} = 780\text{nm}$ definiert, wie in Abbildung 18 dargestellt.

Definition 8: Elektromagnetisches Spektrum

Das elektromagnetische Spektrum definiert sich durch den Frequenzbereich der elektromagnetischen Strahlung und den zugehörigen Wellenlängen. Das *Licht* ist ein Teil des elektromagnetischen Spektrums, welcher vom menschlichen Auge wahrgenommen werden kann und den Bereich von $\lambda_{\min} = 380\text{nm}$ und $\lambda_{\max} = 780\text{nm}$ abdeckt [HP11].

Licht stimuliert Rezeptoren der Netzhaut im Auge eines Menschen, was letztendlich zu der Wahrnehmung von Farbe führt. Das heutige Verständnis von Licht und Farbe geht zurück auf Sir Isaac Newton. Newtons Experimente mit Sonnenlicht und einem Prisma führten zur Erkenntnis, dass Licht in ein Spektrum aus elementaren monochromen Komponenten zerlegt werden kann. Dementsprechend lässt sich Licht physikalisch durch seine spektrale Zusammensetzung definieren. Die Sonne, als unsere primäre Lichtquelle, emittiert Licht über viele Wellenlängen. Die spektrale Bestrahlungsstärke (engl. *Spectral Power Distribution*) (SPD) ist in Abbildung 19 dargestellt. Objekte in der Umgebung absorbieren und reflektieren einige dieser Wellenlängen unterschiedlich stark. Dies ist primär abhängig von der strukturellen Zusammensetzung des Objekts oder der Oberfläche.

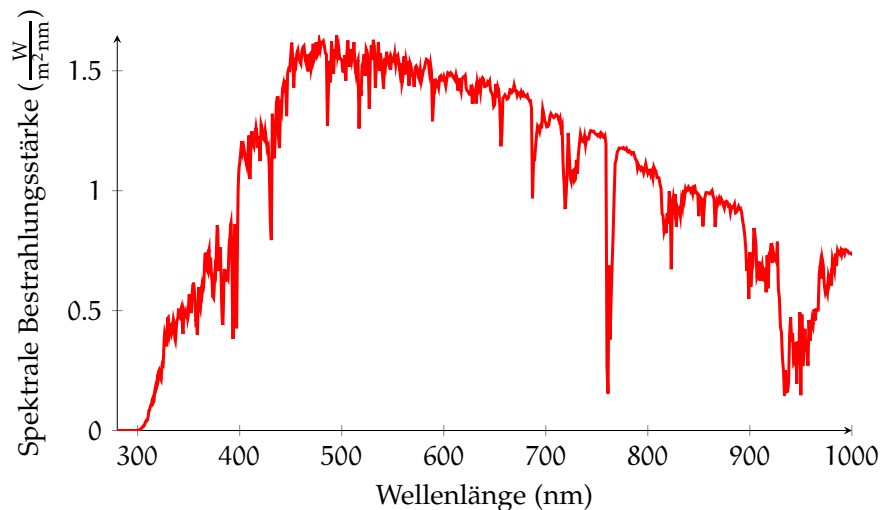


Abbildung 19: Ausschnitt des globalen Standardspektrums [GME02] des Sonnenlichts beim Auftreffen auf die Erdoberfläche nach ISO 9845 – 1. Die Integration ergibt den Wert der Solarkonstante von etwa $1,4\text{KW}/\text{m}^2$.

Der Mensch charakterisiert und beschreibt die Objekte in seiner Umgebung oft anhand ihrer Farbe. Der Himmel wird bspw. als blau, ein Apfel als rot und Gras als grün bezeichnet. In Wirklichkeit wird die Farbe jedoch durch den Beobachter bestimmt, das bedeutet im Umkehrschluss, dass die genaue Zuordnung einer Farbe zu einem Objekt nicht eindeutig ist und somit subjektiv. Der Farbeindruck im visuellen System des Menschen wird primär definiert durch die Reflexionseigenschaften eines Objektes sowie die spektrale Verteilung des Lichtes der Lichtquelle, welche ein Objekt beleuchtet. Weiterhin kann es zu individuellen Unterschieden bei der Farbwahrnehmung verschiedener Menschen kommen.

Definition 9: Hellempfindlichkeitskurve

Die dimensionslose Hellempfindlichkeitskurve $V(\lambda)$ definiert die normalisierte durchschnittliche spektrale Empfindlichkeit der menschlichen visuellen Wahrnehmung von verschiedenen Wellenlängen im Tageslicht [WS00] wie in Abbildung 20 dargestellt.

Im Inneren des Auges gibt es drei Arten von Zapfen-Photorezeptoren zur Farbwahrnehmung welche als *Long* (L), *Medium* (M) und *Short* (S) bezeichnet werden. Sie sind individuell empfindlich gegenüber unterschiedlichen, sich jedoch überlappenden Lichtwellenlängen. Daher sind sie mit der Farbe verbunden, für die sie am empfindlichsten sind, L = rot, M = grün, S = blau, wie in Abbildung 21

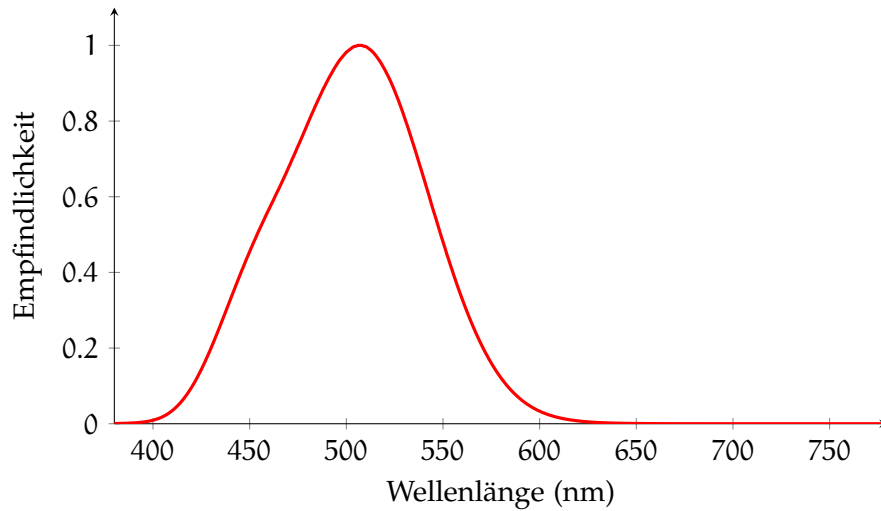


Abbildung 20: Relative spektrale Hellempfindlichkeitskurve $V(\lambda)$ des menschlichen Auges bei Tageslicht [WSoo].

dargestellt. Das Auge sieht daher eine komplexe Spektralverteilung

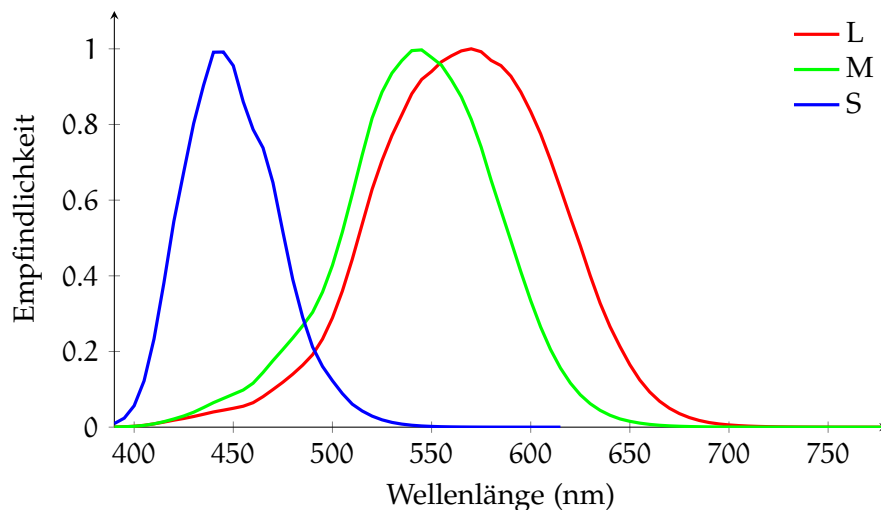
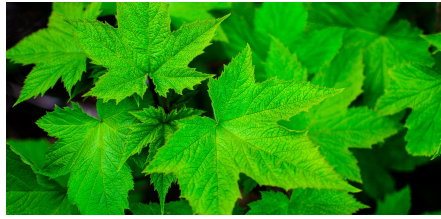


Abbildung 21: Dargestellt sind auf 1 normierte Empfindlichkeitsspektren der menschlichen Zapfen-Photorezeptoren [SSoo]

und reduziert sie auf drei Zahlenwerte, die jeweils angeben, wie stark die drei Rezeptoren stimuliert wurden. Dieser trichromatische Prozess entspricht einer Abtastung des Lichtspektrums mit nur drei Bändern, welche über alle wahrgenommenen Spektralanteile integrieren. Folglich kann aus der Wahrnehmung der Rezeptoren die ursprüngliche Spektralverteilung nicht rekonstruiert werden. So können unterschiedliche spektrale Verteilungen die Rezeptoren auf genau die gleiche Weise stimulieren. Ein Beispiel ist ein Blatt eines Baums und ein grünes Auto. Farblich sehen sie gleich aus, aber physikalisch haben sie unterschiedliche Reflexionseigenschaften. Der



(a) Grüne Blätter eines Baums. Quelle: [18]



(b) Grün lackierter Sportwagen. Quelle: [19]



(c) Ein Gepard in der Steppe. Quelle: [20]



(d) Ein Soldat im Tarnanzug. Quelle: [21]

Abbildung 22: Verschiedene Beispiele, welche die Möglichkeiten der Metamerie illustrieren

Mathematiker Germann Graßmann befasste sich mit der Theorie der Farbmischung und Wahrnehmung. Aus seiner Arbeit gingen vier sog. Graßmannschen Gesetze [Gra53] hervor. Diese Gesetze befassen sich mit der auf dem trichromatischen Prozess basierenden Farbwahrnehmung. Dieses Gesetz geht auf Newtons Erkenntnis zurück, nach der jede Farbe als eine Mischung von Spektralfarben zusammen einem bestimmten oder Helligkeitsgrad beschrieben werden kann. Dies beschriebenen Effekte werden von der Natur und dem Menschen auf vielfältige Weise genutzt, wie die Beispiele in Abbildung 22 illustrieren. Aufgrund dieses Effekts ist es möglich, jede Farbe, die das Auge wahrnimmt, aus vielen verschiedenen Spektralverteilungen zu erzeugen. Dank dieser Effekte kann eine Farbe generiert werden, die wie spektrales Gelb auf einem LCD-Display aussieht, welches aber keinen gelben Spektralanteil hat. Die Steuerung des Displays kombiniert rot und grün erscheinendes Licht, um einen perzeptiv äquivalenten Farbeindruck zu vermitteln.

3.5.3 CIE-Normvalenzsystem

Angelehnt an
[scho7a]

Bevor die grundlegende Arbeit an Farbmetriken begann, veröffentlichte die (engl. *Commission Internationale de l'Eclairage*) (CIE) 1924 eine Funktion, welche die Empfindlichkeit des menschlichen Auges gegen-

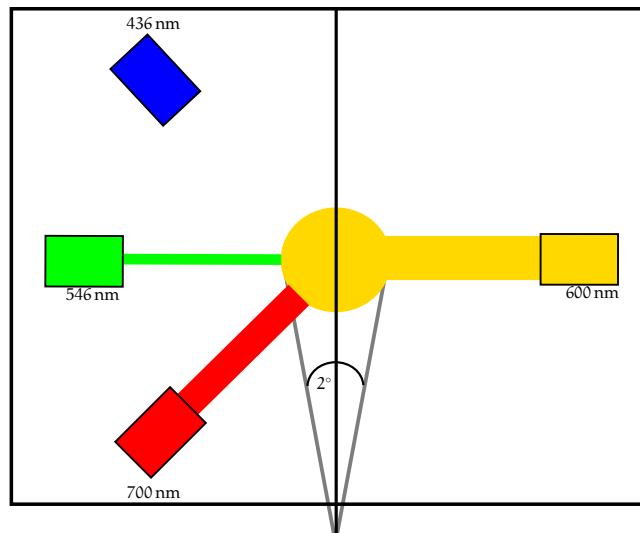


Abbildung 23: Schematische Darstellung der Experimente von Wright und Guild. Rechts zu sehen die monochromatische Lichtquelle und im linken Bereich die drei Primärlichtquellen deren Strahldichte variiert werden muss um die Lichtfarbe der monochromatischen Lichtquelle zu reproduzieren.

über Licht bei verschiedenen Wellenlängen im Tageslicht beschreibt [WS00], so wie in Abbildung 20 dargestellt. Die CIE ist eine Organisation, welche sich mit der Standardisierung von Farbmetriken und Terminologie beschäftigt.

Diese zeigt auf, dass zwei Lichtquellen mit Frequenzen von 500 nm und 400 nm, die die gleiche Strahldichte haben, als unterschiedlich hell wahrgenommen werden können. In den 1920er Jahren führten zwei Farbwissenschaftler, W. D. Wright und J. Guild, Farbsehversuche durch, welche auch auf die Graßmannschen Gesetze zurückgehen. Wright führte sein Experiment an 10 Probanden durch, Gilde wählte 7 Probanden. Ihre Ergebnisse stimmten so gut überein, dass sie von CIE kombiniert wurden, um RGB-Spektralwertfunktionen zu erstellen. Die Funktionen der CIE aus den Jahren 1931 und 1964 [SB10, WS00] definieren einen standardmäßigen farbmtrischen Beobachter, indem sie Spektralwertfunktionen (engl. *Color Matching Functions*) (CMF) bereitstellen, welche in Abbildung 24 dargestellt sind. Die einzelnen Kurven entsprechen monochromatischen Primärlichtquellen mit Wellenlängen von 700 nm für rotes Licht \mathcal{R} , 546,1 nm für blaues Licht \mathcal{B} und 435,8 nm für grünes Licht \mathcal{G} .

Der Versuchsaufbau von Wright und Guild aus dem Jahr 1931 ist in Abbildung 23 dargestellt. Sie forderten einen Probanden auf, die Strahldichte $\bar{r}, \bar{g}, \bar{b}$ der Primärlichtquellen so lange anzupassen, bis sie eine Farbe erzeugen, die von einem monochromatischen Referenzlicht nicht mehr zu unterscheiden ist. Dies wird mit einem Referenz-

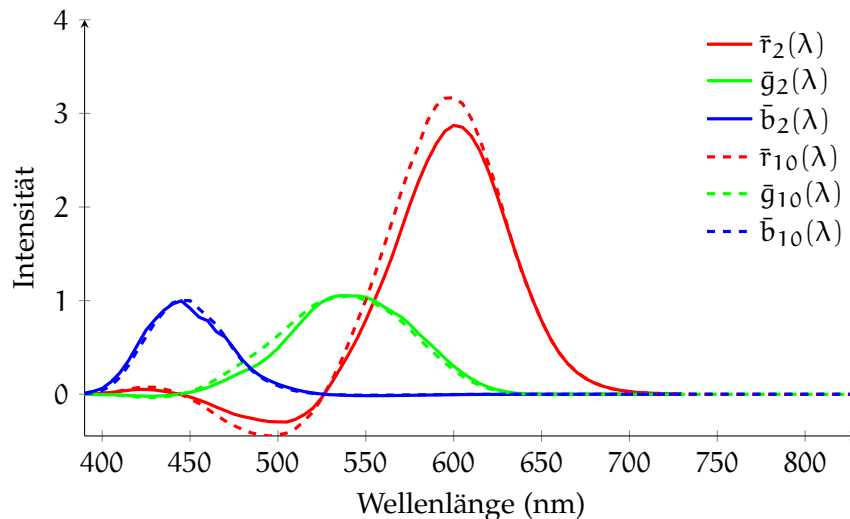


Abbildung 24: Die Grafik zeigt die eingestellten Strahldichten $\bar{r}, \bar{g}, \bar{b}$ der Primärlichtquellen $\mathcal{R}, \mathcal{B}, \mathcal{G}$ für den 1931 2° und den 1964 10° Normbeobachter [SB₁₀, WS₀₀].

licht für jede sichtbare Wellenlänge wiederholt. Aus den so erhaltenen Werten kann eine Funktion $F(\lambda)$ aufgestellt werden, die für jede Wellenlänge λ eine Intensität $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$ liefert, welche denselben Farbeindruck wie die monochromatische Lichtquelle erzeugt.

$$F(\lambda) = \bar{r}(\lambda) \cdot \mathcal{R} + \bar{g}(\lambda) \cdot \mathcal{G} + \bar{b}(\lambda) \cdot \mathcal{B} \quad (5)$$

Auffällig bei der Betrachtung der Funktionen ist, dass einige Wellenlängen negative Werte erfordern, um eine Übereinstimmung zu erreichen, wie z. B. 520 nm. In diesem Fall wurden einige der Primärlichtquellen auf der gegenüberliegenden Seite des Bildschirms mit dem Referenzlicht so lange gemischt, bis eine Farbanpassung vorgenommen werden konnte. Bei den Experimenten stellte sich heraus, dass die Primärfarben nicht jede Spektralfarbe erzeugen können. Mit dem negativen Lichttrick konnten die Forscher aber eine Farbübereinstimmung in Spektralbereichen quantifizieren, die so nicht erreicht werden konnten. Der CIE 1931 2° -Normbeobachter ist nur für Farben gültig, die in einem Sichtfeld von 2° betrachtet werden. Dadurch wird das Licht auf eine Stelle auf der Rückseite des Auges gerichtet, die Fovea genannt wird. Dies ist ein Punkt mit hoher Dichte der farbeempfindlichen Zapfen, der eine maximale Farbunterscheidung bietet. Später wurden die Experimente bei 10° wiederholt und 1964 als der CIE 1964 10° -Normbeobachter veröffentlicht [WS₀₀]. Mit dem 10° -Normbeobachter wird die Betrachtung großer Farbflächen wie z. B. Hauswände besser beschrieben, für die Betrachtung von nahen Objekten funktioniert der 2° -Normbeobachter besser. Diese Normbeobachter-Funktionen wurden in den *ISO/CIE 11664-1* Standard überführt. Dieser Standard spezifiziert Farbanpassungsfunktionen zur Verwendung in der Farbmeterik.

Soweit bekannt, sind die zuvor beschriebenen Normbeobachter-Funktionen in der praktischen Farbmeterik weitgehend ungenutzt. Ihnen vorgezogen werden die XYZ-Farbanpassungsfunktionen, welche eine lineare Transformation der 1931 RGB-Farbanpassungsfunktionen sind. Die XYZ-Farbanpassungsfunktionen basieren auf dem XYZ-Farbraum und verfügen über einige mathematisch günstige Eigenschaften. Die Idee hinter den XYZ-Funktionen ist, dass eine der drei Funktionen so transformiert werden kann, dass sie sehr gut mit der Helligkeitsfunktion $V(\lambda)$ von 1924 übereinstimmt. So kann die Helligkeit einer Farbe vollständig aus der Betrachtung eines der Primärwerte dieser Farbe bestimmt werden. Außerdem ist die manuelle Berechnung, die bei der Verwendung der Normbeobachter-Funktionen vor dem Computer anfang, schwierig. Hier bestand der Wunsch, Funktionen zu haben, die keine negativen Werte enthalten. Um dies zu erreichen, wurden bei der Erstellung des XYZ-Farbraums Primärfarben verwendet, welche keinen echten Farben entsprechen. Die neuen Achsen x, y und z , wurden so ausgewählt, dass bei der Transformation der RGB-Daten die y -Kurve der $V(\lambda)$ -Kurve entspricht und alle Werte positiv sind. Für gewöhnlich wird der Farbwert Y_2 als Leuchtdichte bezeichnet und korreliert mit der wahrgenommenen Helligkeit des Strahlungsspektrums. Die Leuchtdichte wird in Einheiten von Candela pro Quadratmeter (cd/m^2) beschrieben. Daraus ergeben sich neue Kurven wie in Abbildung 25 dargestellt.

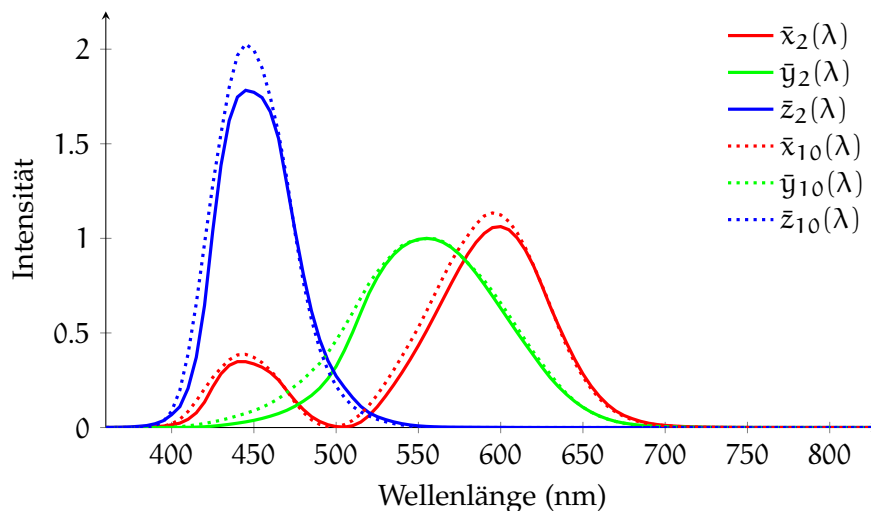


Abbildung 25: Die Grafik zeigt die eingestellten Strahldichten $\bar{x}, \bar{y}, \bar{z}$ der virtuellen Primärlichtquellen x, y, z für den 2° und den 10° Normbeobachter [CIE32, WS00].

3.5.4 Spektral nach RGB

In der Regel wird ein beobachtetes Objekt von einer äußeren Lichtquelle beleuchtet, wie z. B. Tageslicht im Freien oder Licht von einer

Angelehnt an
[cho14]

Lampe im Innenbereich. In solchen Situationen ist die spektrale Leistungsverteilung des ins Auge eintretenden Lichts das Produkt aus der spektrale Bestrahlungsstärke der Lichtquelle und der spektralen Reflexion des Objekts. Die spektrale Bestrahlungsstärke der Lichtquelle wird durch $\mathfrak{S}(\lambda)$ definiert und der spektrale Reflexionsgrad des Objekts als $R(\lambda)$ formuliert. Dann resultiert die spektrale Bestrahlungsstärke des von einem Objekt reflektierten Lichts aus dem Produkt $\mathfrak{S}(\lambda) \cdot R(\lambda)$. Dieses Produkt misst die Kamera und wandelt es in Intensitätswerte I um, welches aber noch vorverarbeitet werden muss, um daraus einen spektralen Wert bzw. das gemessene Spektrum χ zu erhalten.

Es ist anzumerken, dass diese mathematische Beziehung auf einem idealisierten Modell der Wechselwirkung zwischen Beleuchtung und Objekt basiert. Geometrie- und Oberflächeneffekte werden dabei nicht berücksichtigt. Die von einer Kamera gemessenen Werte pro Pixel setzen sich so aus der spektralen Bestrahlungsstärke der Lichtquelle und den Reflexionseigenschaften des beobachteten Objekts zusammen. So lässt sich Folgendes definieren:

Definition 10: Intensitätswert

Jedes Pixel eines digitalen Bildes wird durch einen Intensitätswert I_k repräsentiert, welcher proportional ist zu einem gewichteten Integral über einem Teilbereich des elektromagnetischen Spektrums:

$$I_k = \int \rho(\lambda) \cdot \mathfrak{S}(\lambda) \cdot \mathfrak{E}_k(\lambda) \quad (6)$$

mit $\rho(\lambda)$ als den Reflexionsgrad des Objekts bei λ , $\mathfrak{S}(\lambda)$ als die spektrale Bestrahlungsstärke der Lichtquelle bei λ und $\mathfrak{E}_k(\lambda)$ als die spektrale Empfindlichkeit des k-ten Bandes der Kamera.

Der Intensitäts- bzw. Signalwert stellt zunächst nur das Rohsignal dar. Durch eine geeignete Vorverarbeitung (vgl. Kapitel 4) muss der Einfluss der spektralen Empfindlichkeit der Kamera für jeden Kanal entfernt werden.

Definition 11: Gemessenes Spektrum

Das gemessene Spektrum ist, im Rahmen dieser Dissertation, definiert als die reflektierte Spektralverteilung von einem Punkt in der Szene quantisiert über die Kanäle bzw. Bänder \mathcal{N}_λ der jeweiligen spektralen Kamera.

$$\chi : \mathcal{L}_\lambda \rightarrow [0,1]^{\mathcal{N}_\lambda} \quad (7)$$

Die reflektierte Spektralverteilung, im Rahmen dieser Dissertation auch als gemessenes Spektrum χ bezeichnet, kann nun wie folgt in den XYZ-Farbraum überführt werden [SBo2, cho14]:

$$x = k \sum_{i=0}^{\mathcal{N}-1} \bar{x}(\lambda_i) \underbrace{\mathfrak{S}(\lambda_i)R(\lambda_i)}_{\chi} \quad (8)$$

$$y = k \sum_{i=0}^{\mathcal{N}-1} \bar{y}(\lambda_i) \underbrace{\mathfrak{S}(\lambda_i)R(\lambda_i)}_{\chi} \quad (9)$$

$$z = k \sum_{i=0}^{\mathcal{N}-1} \bar{z}(\lambda_i) \underbrace{\mathfrak{S}(\lambda_i)R(\lambda_i)}_{\chi} \quad (10)$$

- $R(\lambda)$ den Reflexionsfaktor des Objekts bei λ im Bereich $[0 - 1]$.
- $\mathfrak{S}(\lambda)$ die spektrale Bestrahlungsstärke der Lichtquelle bei λ in $\text{W}/\text{m}^2 \cdot \text{nm}$.
- $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ sind die Spektralwertfunktionen des XYZ-Farbraums.
- k ist ein Normalisierungsfaktor, welcher wie folgt definiert ist:

$$k = \frac{100}{\sum_{i=0}^{\mathcal{N}-1} \bar{y}(\lambda_i) \mathfrak{S}(\lambda_i)} \quad (11)$$

Die korrespondierenden Werte im XYZ-Farbraum ergeben sich also durch eine Multiplikation der vom Sensor gemessenen spektralen Energieverteilung mit den passenden Spektralwertfunktionen aus dem gewählten CIE-Standard. Zu beachten ist noch, dass für Werte außerhalb des von der Kamera gemessenen Wellenlängenbereichs keine Informationen vorliegen. Hier wird das Spektrum als Null angenommen und somit gehen diese Werte nicht in die Berechnung mit ein. Weiterhin setzt sich die reflektierte spektrale Bestrahlungsstärke nur aus vom Sensor gemessenen diskreten Werten zusammen. Dies gilt auch für die ermittelten Spektralwertfunktionen. Daher werden die Funktionen in der Praxis auch als Matrix \mathfrak{B} zusammen gefasst.

$$\mathfrak{B} = (\bar{x}, \bar{y}, \bar{z})^T \quad (12)$$

Womit sich die Transformation der Werte in den XYZ-Farbraum wie folgt vollziehen lässt:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathfrak{B} \cdot \mathfrak{S}(\lambda) \mathbf{R}(\lambda) \quad (13)$$

Somit sind die gemessenen Werte nun im XYZ-Farbraum definiert. Um nun in den RGB-Farbraum zu gelangen ist eine weitere Transformation notwendig:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (14)$$

Die dazu nötige Transformation beinhaltet eine 3×3 Matrix \mathbf{M} dessen Werte abhängig vom jeweiligen Endgerät sind und durch Messen bestimmt werden müssen. Liegen diese Werte nicht vor, kann auf den sRGB-Standard [C⁺99] zurückgegriffen werden. Durch invertieren der 3×3 -Matrix können so einfach die RGB-Farbwerte berechnet werden.

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \mathbf{M}^{-1} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (15)$$

3.6 SPEKTRALE BILDGEBUNG

3.6.1 Einführung

In diesem Abschnitt wird näher auf die Grundlagen der spektralen Bildgebung eingegangen. Die spektrale Bildgebung ist eine Kombination aus Spektroskopie und Fotografie und ermöglicht es, detaillierte Spektralinformationen von jedem Pixel in einer Szene zu erhalten. Die hyperspektrale Bildgebung ist mit der multispektralen Bildgebung eng verbunden. Die multispektrale Bildgebung bezieht sich auf relativ breite Bänder, die das Spektrum vom sichtbaren bis zum langwelligen Infrarot abdecken. Im Gegensatz dazu handelt es sich bei hyperspektralen Daten um die Aufnahme von schmalen Spektralbander aus einem kontinuierlichen Spektrum. Zur genaueren Abgrenzung der einzelnen Begriffe folgen entsprechende Begriffsdefinitionen.

Definition 12: Spektrales Band

Ein spektrales Band, auch als Frequenzband bezeichnet, definiert nach [IEE] einen begrenzten Teilbereich des elektromagnetischen Spektrums. Sensoren sind jeweils für bestimmte Teilbereiche zur Detektion entlang des elektromagnetischen Spektrums konfiguriert.

3.6.2 *Snapshot-Mosaik-Technik*

Der Kern einer Spektralkamera ist die Spektraleinheit, eine optische Komponente, welche die Wellenlängentrennung implementiert. Die zuvor in Kapitel 3.3.1 beschriebenen Sensoren basieren auf der Snapshot-Mosaik-Technik und setzen dementsprechend auf einen speziellen Sensor zur Bildgebung. Das Messprinzip dieser Sensoren basiert auf dem Fabry-Pérot-Interferometer [Fab99, PF99], welches den Effekt der Vielstrahlinterferenz nutzt und vom Hersteller des Sensors in den letzten Jahren mehrfach publiziert wurde [TLSH12, GTL14, LGG⁺14, ATG⁺16]. Die folgenden Abschnitte basieren auf Informationen aus diesen Publikationen.

*Angelehnt an
[Fab99, PF99,
TLSH12, GTL14,
LGG⁺14, ATG⁺16]*

Definition 13: Fabry-Pérot-Interferometer

Ein Fabry-Pérot-Interferometer ist ein optischer Resonator, der aus zwei teilreflektierenden Spiegeln zusammengesetzt ist [Fab99], welche als optische Filter dienen. So kann aus einem emittierten Spektrum ein Teilbereich herausgefiltert werden. Die Parametrisierung des Abstands zwischen den Spiegeln ermöglicht es, den Wellenlängenbereich der gefilterten Strahlung zu variieren.

Die Vielstrahlinterferenz wird definiert durch die Überlagerung mehrerer kohärenter Wellen [GK07], welche in ihrer Phase um den gleichen Betrag gegeneinander verschoben sind. Diese Wellen besitzen die gleiche Frequenz und werden durch Vielfachreflexion zwischen zwei Oberflächen erzeugt, wie z. B. bei einem Fabry-Pérot-Interferometer oder an einer Lummer-Gehrcke-Platte. Überlagern sich Wellen lokal, so kann dies zur Erhöhung oder Schwächung der Intensität führen. Überlagern sich zwei kohärente Wellenzüge gleicher Frequenz, so tritt Interferenz auf. Es sei λ weiterhin definiert als die Wellenlänge und es gelte $n \in \mathbb{N}$. Sind nun ihre Phasen gleich oder um

$$\Delta_s = n \cdot \lambda \quad (16)$$

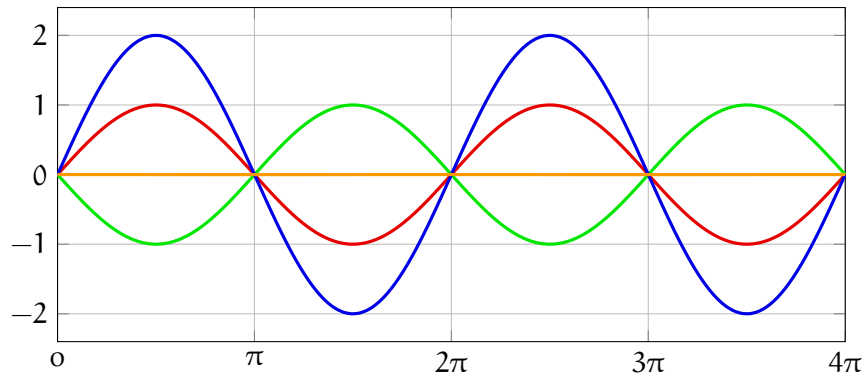


Abbildung 26: Visualisierung von Sinuswellen zur exemplarischen Darstellung von Interferenzen. Eine exemplarische Sinuswelle ist rot, eine zweite kohärente um π verschobene Sinuswelle ist grün dargestellt. Die Überlagerung dieser beiden Wellen ist in orange dargestellt. Dies demonstriert eine destruktive Interferenz. Eine Überlagerung der roten Welle mit sich selbst definiert eine konstruktive Interferenz, welche in blau dargestellt ist.

verschieden, so addieren sich die Amplituden, was eine konstruktive Interferenz definiert. Hier verstärken sich die kohärenten Wellen maximal. Unterscheiden sich die Phasen um

$$\Delta_s = (n + 0,5) \cdot \lambda \quad (17)$$

so subtrahieren sich die Amplituden, was eine destruktive Interferenz definiert. In diesem Fall werden die Wellen maximal abgeschwächt. Ein Beispiel für solche Interferenzen ist in Abbildung 26 gegeben. Bei der blau dargestellten konstruktiven Interferenz addieren sich die Wellen und erreichen eine doppelt so große Amplitude wie die Ausgangswelle, welche in rot dargestellt ist. Eine Überlagerung der grünen und roten Welle führt zu der in orange eingezeichneten destruktiven Interferenz, bei der die Amplitude auf null reduziert wird. Ein Fabry-Pérot-Interferometer nutzt nun diesen Effekt, um eintreffendes Licht zu spalten. Dazu setzt es sich aus zwei parallelen, teildurchlässigen Spiegeln zusammen. Ihre Parallelstellung und ihr Abstand d werden dabei mechanisch oder Piezo-elektrisch im Größenbereich der Lichtwellenlänge verändert, sind aber in der Regel während der Messung fixiert.

Licht ist eine Form von elektromagnetischer Strahlung. Trifft diese elektromagnetische Strahlung auf Materie, so wird ein Teil der Strahlung reflektiert und die restliche Strahlung dringt in das Medium ein. Dort wird die Strahlung diversen Prozessen unterworfen und geschwächt. Zunächst wird ein Teil vom Medium absorbiert, wobei ein Bruchteil der Energie z. B. in Wärme umgewandelt wird. Ein weiterer Teil ist dem Streueffekt (Richtungsänderung) unterworfen. Beides wird als *Extinktion* bezeichnet. Nicht absorbierte Strahlung durch-

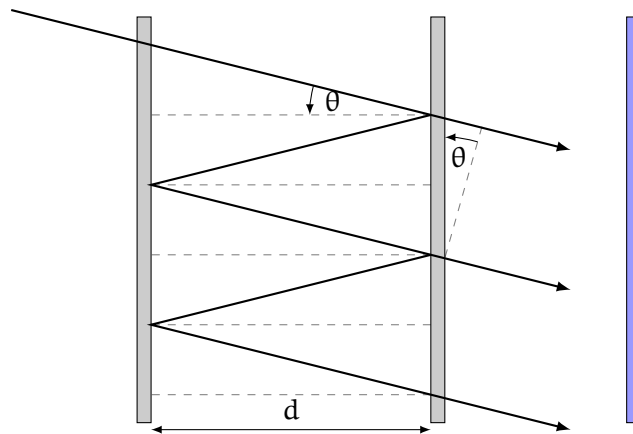


Abbildung 27: In dieser Grafik wird der physikalische Effekt der Vielstrahlinterferenz, welcher dem Fabry-Pérot-Interferometer zugrunde liegt, dargestellt. Zwei parallel angeordnete Spiegelplatten (grau) mit einem Sensorelement (blau) dahinter.

dringt das Medium und verlässt es wieder. Dies ist insbesondere bei transparenten Medien der Fall. Wird nun ein Lichtstrahl auf die teildurchlässigen Spiegelplatte eines Fabry-Pérot-Interferometers gerichtet, durchdringt er diese und wird an der gegenüberliegenden Platte reflektiert. Somit wird ein Großteil des eintreffenden Lichtes zurückgeworfen. Es wird angenommen, dass der Reflexionsgrad der Spiegel sehr hoch ist und dass der Anteil, der von der Platte absorbiert wird, derart klein ist, dass er für die folgenden Betrachtungen keine Rolle spielt. Folglich wird der Lichtstrahl weiter zwischen den Spiegelplatten reflektiert, sodass eine Überlagerung der Lichtwellen entsteht, was zu einer konstruktiven Interferenz führt. Abbildung 27 zeigt den schematischen Aufbau dieser zwei Spiegelplatten mit einem Sensorelement, für das gilt

$$2 \cdot n \cdot d \cdot \cos(\theta) = m \cdot \lambda, \quad (18)$$

wobei m die harmonische Ordnung, d den Abstand der beiden Spiegelplatten zueinander, θ den Winkel des Lichteinfalls und n den Brechungsindex definiert [ATG⁺ 16]. Die weiteren Erläuterungen basieren nun auf dieser Darstellung.

Gegeben sei ein orthogonal einfallender Lichtstrahl, welcher durch eine beidseitig telezentrische Linse auf die Spiegelplatten geworfen wird, wodurch sich das aus verschiedenen Wellenlängen bestehende Licht überlagert. Telezentrische Linsen sorgen dafür, dass die Hauptstrahlen im Objektraum alle parallel zur optischen Achse verlaufen. Die einzelnen Wellenlängen lassen sich aufgrund des Superpositionsprinzips [GK07] getrennt betrachten, da die Überlagerung der Wellen diese nicht verändert. Da die beiden Platten einen geringen Abstand zueinander aufweisen und die Wellen mit Lichtgeschwindigkeit auf-

treffen, werden sie in großer Zahl reflektiert und zurückgeworfen. Dabei kommt es sowohl zur konstruktiven wie auch destruktiven Interferenz. So unterliegen die Teilwellen der konstruktiven Interferenz, für welche der Abstand der Platten dem ganzzahligen Vielfachen ihrer Wellenlänge entspricht. Diese Teilwellen steigern ihre Amplitude immer weiter, da die sich überlagernden Teilwellen lediglich phasenverschoben sind, während Teilwellen mit anderer Wellenlänge der destruktiven Interferenz unterliegen und abgeschwächt werden (vgl. Formel (18)). Zu beachten ist, dass nicht nur eine Wellenlänge der konstruktiven Interferenz unterliegt, sondern auch alle harmonischen Vielfachen entsprechend ihre Amplituden steigern.

Da die Spiegelplatten teildurchlässig sind, wird ein Teil des auftreffenden Lichts nach außen abgegeben. Aufgrund der erhöhten Amplituden der verstärkten Wellen, kommt es so zu einer entsprechenden Transmission der definierten Wellenlänge. Hinter der zweiten Spiegelplatte ist nun ein Sensorelement positioniert, welche das eintreffende Licht registriert und die Interferenz darstellt. Trifft das Licht nicht orthogonal auf, so ist die Interferenz zusätzlich abhängig vom Winkel des eintreffenden Lichts.

3.6.3 CMOS-Implementierung

Angelehnt an
[GTL14]

Auf dem CMOS-Sensor der Kameras sind entsprechend pixelgenaue Filter aufgebracht [GTL14], welche den zuvor beschriebenen Aufbau mit zwei Spiegelplatten realisieren. Die Dicke der Filter wird entsprechend pro Pixel variiert, um unterschiedliche Wellenlängenempfindlichkeiten zu erreichen. Diese Technik erzeugt jedoch Artefakte, welche für harmonische Vielfache und gegebenenfalls für ähnliche Wellenlängen gelten, was die Bildgebung entsprechend beeinflusst. Diese Effekte müssen bei der entsprechenden Vorverarbeitung berücksichtigt und korrigiert werden, vgl. Kapitel 4. Die Messung durch einen idealen Filter hat eine schmale Spitze, die um jede für diesen Filter spezifizierte harmonische Wellenlänge zentriert ist und ohne weitere Antwort im Spektralbereich außerhalb dieser Spitze.

Die zentrale Wellenlänge des Peaks λ_0 befindet sich im Zentrum der Halbwertsbreite FWHM. Die Bandbreite eines Peaks ist entsprechend definiert als das dreifache der FWHM. Daher ist der Bereich, welcher von einem Filter abgedeckt wird, definiert als $[\lambda_0 - 1,5 \cdot \text{FWHM}; \lambda_0 + 1,5 \cdot \text{FWHM}]$.

Die Transmissionskurve (engl. *response curve*) eines Filters ist die Kombination aus der Quanteneffizienz (QE) des Sensors und den Transmissionseigenschaften des Filters, wobei die maximale Antwort eines Filters durch die QE des Sensors limitiert ist. Die QE der in den Kameras verbauten CMOS-Sensoren ist in Abbildung 28 dargestellt. Aufgrund physikalischer Eigenschaften der Filter sind allerdings Störfaktoren in der Transmissionskurve eines Filters enthalten. Diese Ef-

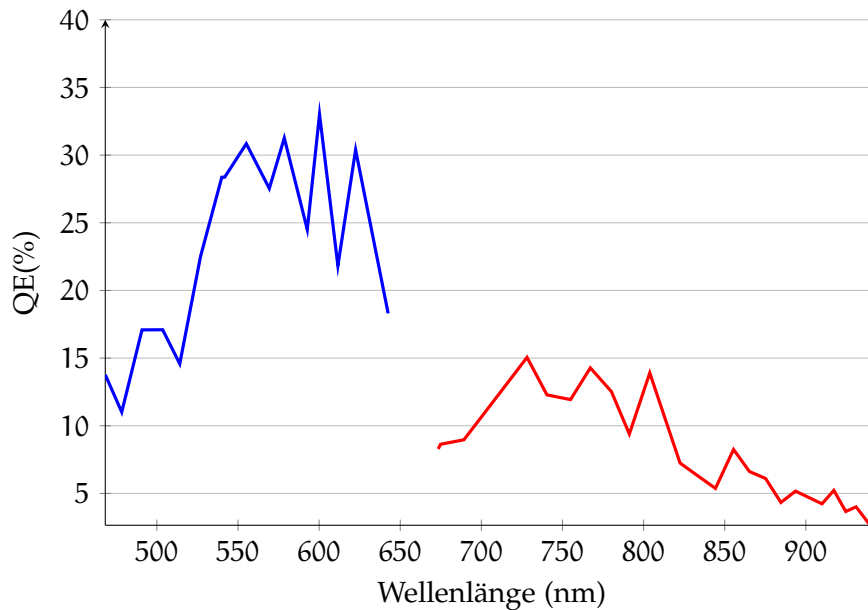


Abbildung 28: Quanteneffizienz der in den Kameras verbauten CMOS-Sensoren basierend auf den Daten des Herstellers. Die Kennlinie des VIS-Sensors ist blau und die des NIR-Sensors ist rot.

fekte lassen sich als spektrale Verschiebung (engl. *Spectral Shift*), Leck-Effekt (engl. *Spectral Leaking*) und Übersprechen (engl. *Crosstalk*) zusammenfassen.

IMEC nutzt hier zur Implementierung der Fabry-Pérot-Filter auf dem Sensor verteilte Bragg-Spiegel. Diese bestehen aus dielektrischen, alternierenden dünnen Schichten mit jeweils niedrigem und hohem Brechungsindex. Um das maximale Reflexionsvermögen für eine Wellenlänge zu erreichen, müssen dabei alle Schichten eine optische Dicke von genau einem Viertel der Wellenlänge aufweisen. Beim Fabry-Pérot-Interferometer und der dazugehörigen Formel wird von einer Phasenverschiebung des Lichts um 180 Grad bei jeder Reflexion ausgegangen. Dies gilt aber nur bedingt für Bragg-Spiegel, da es an den Grenzflächen von niedrigem zu hohem Brechungsindex zu einer Phasenverschiebung des Lichts von einer halben Wellenlänge $\frac{\lambda}{2}$ kommt [She95], was die spektrale Verschiebung begründet. Folglich ist die Wellenlänge der harmonischen zweiten Ordnung leicht verschoben. Das Verhalten der Filter für Wellenlängen außerhalb des spezifizierten Bereichs ist materialabhängig und nicht definiert. Tritt nun Licht mit Wellenlängen außerhalb dieses Bereichs in den Filter ein, führt dies zu dem spektralen Leckeffekt (engl. *Spectral Leaking*). Der Effekt des Übersprechens (engl. *Crosstalks*) tritt auf, wenn das Signal an einem Pixel das Signal eines anderen Pixels beeinflusst. Dies spielt eine wichtige Rolle an der Grenze zwischen den einzelnen Filtern. Elektronen, die auf einen Filter treffen, beeinflussen auch die Reaktion eines

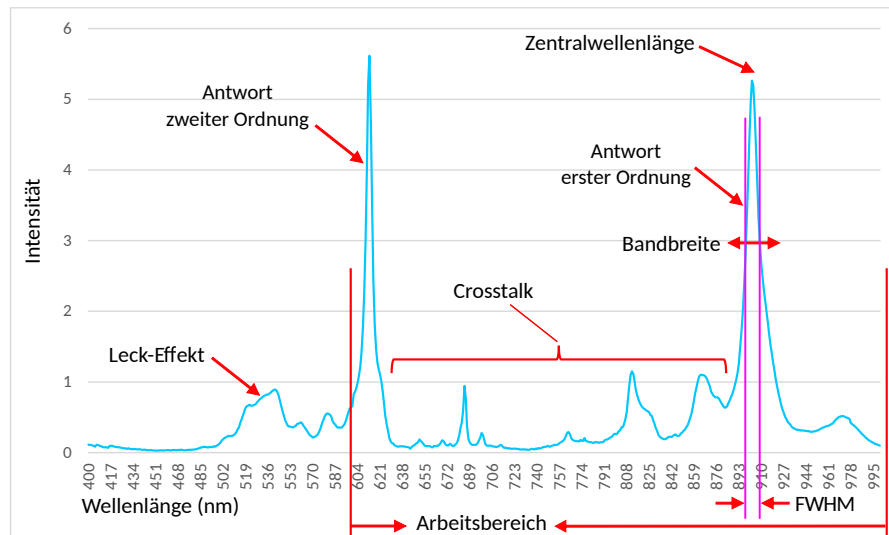
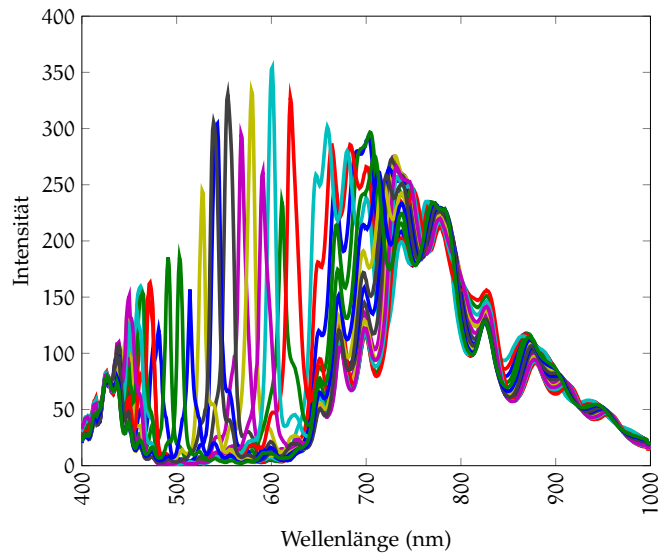


Abbildung 29: Beispielhafte Transmissionskurve eines spektralen Filters. Die Daten wurden aus einer Kalibrierdatei extrahiert. Sie basieren auf den vom Hersteller ermittelten Werten eines 5×5 Mosaiksensors für den Nahinfrarotbereich.

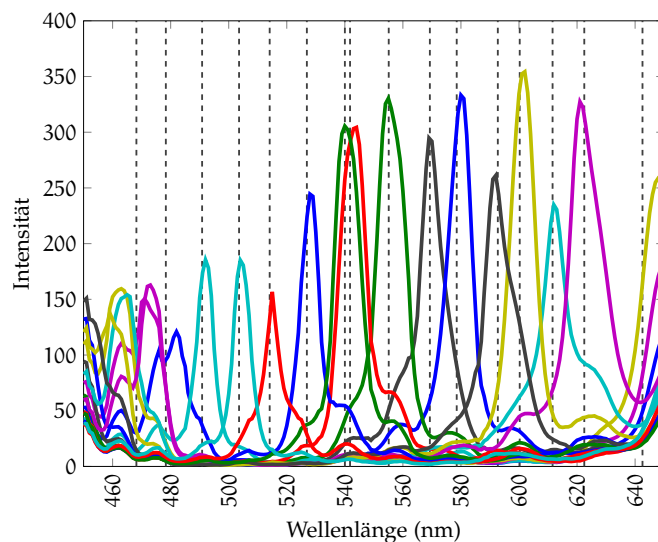
benachbarten Filters. Dies führt zu unerwünschten Reaktionen außerhalb der Resonanzspitzen der einzelnen Filter. Es hat auch einen stärkeren Einfluss auf Sensoren mit einer hohen räumlichen Variation der Filter.

Die soeben genannten Effekte sind in Abbildung 29 illustriert. Dargestellt ist die Transmissionskurve eines Filters mit einer Spezifikation für eine Wellenlänge von 913 nm.

Zu sehen ist eine Antwort zweiter Ordnung bei 603 nm, sowie Wellenlängen außerhalb des aktiven Sensorbereichs, die in den Filter eindringen. Weiterhin ist ein Übersprechen der im Makropixel umliegenden Pixel mit anderen Wellenlängenempfindlichkeiten zu erkennen. Diese sichtbaren Effekte müssen im Rahmen einer Vorverarbeitung entfernt oder kompensiert werden, um effektiv mit den Daten arbeiten zu können. Da die Wellenlängensensitivität der Filtermatrix produktionsbedingten Schwankungen unterworfen ist, ist es erforderlich, jeden Sensorchip individuell nach der Produktion zu kalibrieren. Eine entsprechende Kalibrierung wird vom Hersteller des CMOS-Sensors IMEC durchgeführt [ATG⁺16]. Dazu wird der Sensor mit einer durchstimmbaren Lichtquelle mit monochromatischem Licht im Spektralbereich von 400 – 1000 nm mit Intervallschritten von 1 nm bestrahlt und die entsprechenden Sensorantworten für jede der 16, beziehungsweise 25, Messstellen in einem Makropixel aufgenommen. Das Ergebnis der Kalibrierung ist ein Sensormodell in Form einer Antwortmatrix. Jede Spalte der Antwortmatrix enthält den Beitrag jeder Wellenlänge von 400 – 1000 nm zu einer Gesamtreaktion ei-



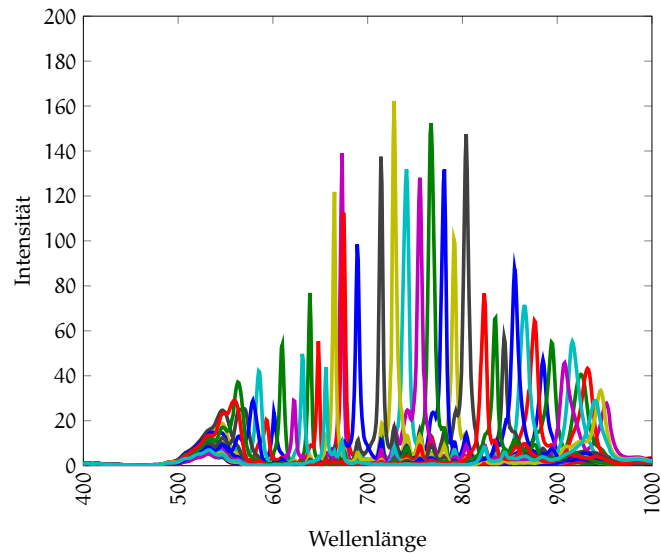
(a) Sensorantwort im Bereich 400-1000 nm



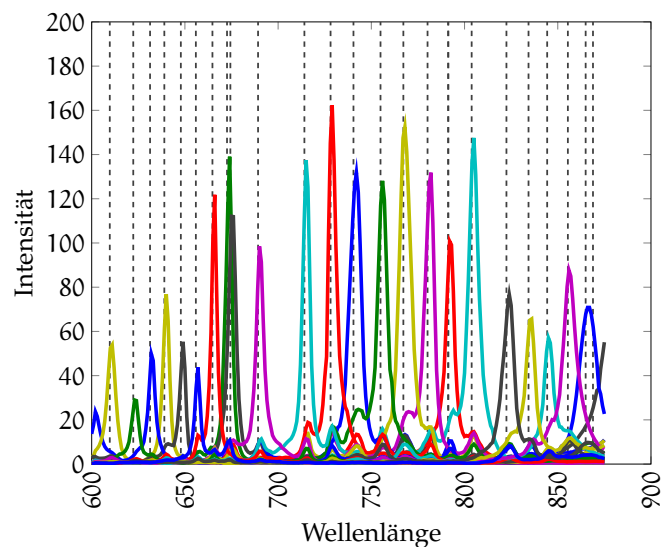
(b) Sensorantwort im Bereich 450-650 nm

Abbildung 30: Das Verhalten der 16 Spektralfilter der Ximea VIS-Kamera aufgetragen über den Wellenlängenbereich von (a) 400-1000 nm und (b) 450-650 nm. Die Referenzlinien in Grafik b zeigen die Wellenlängen, denen die Messung mit der jeweiligen spektralen Empfindlichkeit zugeordnet wird. Die Grafik basiert auf den Kalibrierdaten welche vom Hersteller zur Verfügung gestellt werden.

nes Pixels. Dieses Spektrum aus 601 Proben ergibt die Antwortzusammensetzung so wieder, wie sie eigentlich vom Sensor gemessen wurde. Entsprechend ist die Anzahl der Zeilen der Antwortmatrix gleich



(a) Sensor Antwort im Bereich 400-1000 nm



(b) Sensor Antwort im Bereich 675-975 nm

Abbildung 31: Das Verhalten der 25 Spektralfilter der Ximea NIR-Kamera aufgetragen über den Wellenlängenbereich von (a) 400-1000 nm und (b) 675-975 nm. Die Referenzlinien in b zeigen die Wellenlängen denen die Messung mit der jeweiligen spektralen Empfindlichkeit zugeordnet wird.

der Anzahl der Pixel in einem Makropixel. Die gemessenen Antwortmatrizen werden vom Kamerahersteller Ximea der Kamera beigelegt bzw. in ihr gespeichert. Abbildung 30 zeigt die spektralen Transmissionskurven der 16 Spektralfilter bzw. Messstellen, die in der VIS-Kamera zum Einsatz kommen. In Abbildung 30a und Abbildung 31a

zeigt sich das Verhalten des Filtermusters über den gesamten Messbereich. Die Transmissionskurven jedes Sensors, die während des Kalibrierprozesses gemessen werden, sind in einer Kalibrierdatei gespeichert. Jeder Sensor hat seine eigene individuelle Kalibrierdatei. Die Verwendung der Kalibrierdatei eines anderen Sensors zur Interpretation von erfassten Daten führt unweigerlich zu falschen Spektralinformationen und macht die Daten unbrauchbar. In Abbildung 30a ist ersichtlich, dass sich alle Spektralfilter im Bereich von 400-450 nm und von 700 nm aufwärts ähnlich verhalten und lediglich im Bereich von 450-650 nm sind dediziert unterschiedliche Transmissionskurven zu erkennen. Dies ist der Messbereich, für den die VIS-Kamera ausgelegt ist, welcher in Abbildung 30b noch einmal vergrößert dargestellt wird. Analog dazu ist die NIR-Kamera für den Bereich von 675-975 nm ausgelegt, was auch in Abbildung 31b deutlich wird. Den Abbildungen ist zu entnehmen, dass Licht außerhalb des definierten Messbereiches signifikante Störungen des gemessenen Signals erzeugt. Daher wird vom Sensorhersteller IMEC empfohlen, dass in die Kamera einfallende Licht durch einen zusätzlichen optischen Filter zu führen, sodass nur Licht aus dem zulässigen Messbereich aufgenommen wird. Hierzu werden Lang- und Kurzpassfilter mit sehr steilen Flanken zu einem Bandpassfilter kombiniert, der das gewünschte Verhalten zeigt. Für die Kameras von Ximea ist ein entsprechender Filter direkt in der Kamera verbaut. Im Gegensatz zu den Wellenlängen außerhalb des aktiven Bereichs können einige unerwünschte Effekte nicht durch die Verwendung der Sperrfilter entfernt werden. Diese unerwünschten Effekte tragen zur Messung bei und sind in den aufgenommenen Daten enthalten. Ihr Beitrag zur Messung kann durch Anwendung einer Spektralkorrektur im Anschluss an die Aufnahme unterdrückt werden, vgl. Kapitel 4.

3.7 BILDDEFINITION

Im Folgenden erfolgt eine formale Definition verschiedener terminologischer Begriffe zur Verwendung in dieser Arbeit. In der Literatur [HS⁺73, Pri15a] wird das Bild f als eine Funktion aus dem Ortsbereich des Bildes Loc in einen Wertebereich Val definiert. Dies lässt sich ohne weitere Einschränkungen wie folgt darstellen:

Angelehnt an
[HS⁺73, Pri15a]

$$f \subseteq Loc \rightarrow Val \quad (19)$$

So wird jedem Ortswert $l \in Loc$ ein Wert $v \in Val$ mit $(l,v) \in f$ durch $v = f(l)$ zugeordnet.

Definition 14: Pixel

Ein elementares Bildelement auch als Pixel \mathbf{p} von \mathbf{f} bezeichnet wird gemäß [Pri15a] wie folgt definiert:

$$\mathbf{p} = (\mathbf{l}, \mathbf{v}) \in \text{Loc} \rightarrow \text{Val} \quad (20)$$

Somit ist ein Pixel \mathbf{p} ein Paar bestehend aus einem Ort \mathbf{l} und einem Wert \mathbf{v} welcher einen Signalwert I darstellt.

Die Mengen Loc und Val sind kartesische Produkte spezifischer Mengen:

$$\text{Loc} = \mathcal{L}_1 \times \cdots \times \mathcal{L}_m, \text{Val} = V_1 \times \cdots \times V_n \quad (21)$$

Mit m als Dimension des Bildes und n die Zahl der Kanäle des Bildes. So gilt entsprechend $\text{Loc} = (\mathcal{L}_x, \mathcal{L}_y)$ wobei $\mathcal{L}_x, \mathcal{L}_y$ die räumliche Dimension in X- und Y-Richtung beschreiben. Weiter lässt sich definieren, dass $N_x = |\mathcal{L}_x|$ als Anzahl der Bildpixel in X-Richtung und $N_y = |\mathcal{L}_y|$ als Anzahl der Pixel in Y-Richtung gilt. Der Wertebereich richtet sich nach dem Bildtyp. Ein Kamerasensor misst bis zu einer maximal möglichen Intensität und daher wird die Sensorantwort im Wertebereich $[0,1]$ definiert. Es ergeben sich als Konsequenz Bilddefinitionen für verschiedene Bildtypen.

Definition 15: Bild

Ein Bild lässt sich für verschiedene Bildtypen wie folgt definieren:

$$\mathbf{f}^G : \mathcal{L}_x \times \mathcal{L}_y \rightarrow [0,1] \quad (22)$$

$$\mathbf{f}^{\text{RGB}} : \mathcal{L}_x \times \mathcal{L}_y \rightarrow [0,1]^3 \quad (23)$$

$$\mathbf{f}^H : \mathcal{L}_x \times \mathcal{L}_y \rightarrow [0,1]^{N_\lambda} \quad (24)$$

Dabei definiert \mathbf{f}^G ein Grauwertbild, \mathbf{f}^{RGB} ein RGB-Bild oder Farbbild und \mathbf{f}^H ein spektrales Bild (Hyperwürfel) mit N_λ Wellenlängen. Dazu definiert N_λ eine spektrale Achse \mathcal{L}_λ analog zu N_x und N_y als geometrische Achsen.

Aus der Bilddefinition ergibt sich ein dreidimensionaler Körper, der hier als Hyperwürfel bezeichnet wird. Diese Struktur besitzt zwei geometrische und eine spektralen Achse wie in Abbildung 32 schematisch dargestellt.

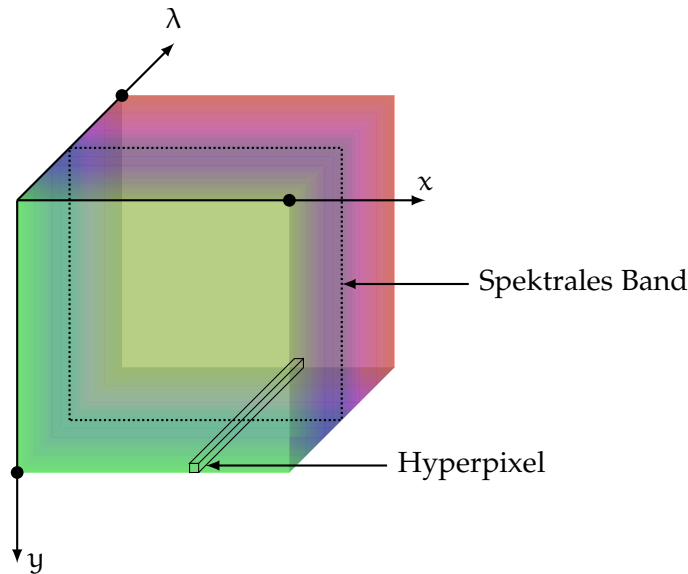


Abbildung 32: Die Grafik zeigt eine geometrische Interpretation der spektralen Daten als Hyperwürfel. Eingezeichnet ist ein Hyperpixel durch einen dünnen Rahmen und ein spektrales Band mithilfe einer gepunkteten Kontur. Das Pixel erstreckt sich entlang der spektralen Achse. Das Band entspricht einem Schnitt durch den Würfel, welcher parallel zur Ebene, die durch die X- und Y-Achse aufgespannt wird, verläuft.

Um einen Vergleich der oben definierten Bildmodalitäten zu ermöglichen wird nun eine geschlossene Bilddefinition gewählt, in der alle Bildmodalitäten als Würfel mit zwei geometrischen und einer spektralen Achse modelliert sind.

$$f: \mathcal{L}_x \times \mathcal{L}_y \times \mathcal{L}_\lambda \rightarrow [0; 1] \quad (25)$$

Hier wird der Ortsbereich eines Bildes um eine spektrale Achse \mathcal{L}_λ erweitert, womit auch die Gleichung 19 weiterhin Gültigkeit besitzt. Entsprechend gilt $\text{Loc} = \mathcal{L}_x \times \mathcal{L}_y \times \mathcal{L}_\lambda$. Daraus folgt $\mathcal{N}_\lambda = 1$ für ein Grauwertbild und $\mathcal{N}_\lambda = 3$ für ein Farbbild. Weiterhin wird ein spektrales Band in diesem Kontext dadurch beschrieben, dass alle Werte auf den geometrischen Achsen für einen definierten Wert der spektralen Achse ausgewählt werden.

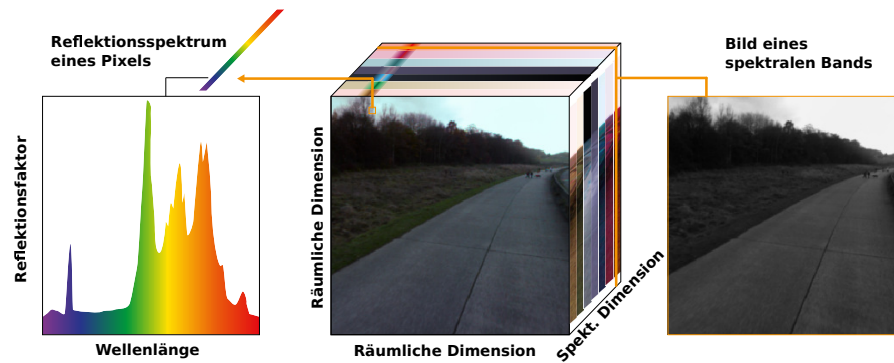


Abbildung 33: Schematische Darstellung eines Hyperwürfels

Definition 16: Spektrales Band im Hyperwürfel

Ein spektrales Band entspricht einem Schnitt durch einen Hyperwürfel parallel zu den geometrischen Achsen:

$$\mathbf{B} : \mathcal{L}_\lambda \rightarrow [0,1]^{N_x \times N_y} \quad (26)$$

Dadurch ergibt sich ein Bild, bei dem jedes Bildpixel die spektralen Werte entsprechend einer Wellenlängenempfindlichkeit darstellt.

Definition 17: Hyperpixel

Ein Hyperpixel wird analog zu einem Pixel definiert:

$$\mathbf{p}^H = (\mathbf{l}, \chi) \in \text{Loc} \times \text{Val} \quad (27)$$

So ist ein Hyperpixel \mathbf{p}^H ein Paar bestehend aus einem Ort \mathbf{l} und einem an diesem Ort gemessenen Spektrum χ .

Eine grafische Darstellung dieser Definitionen ist durch Abbildung 33 gegeben.

3.8 EVALUATIONSMETRIKEN

Um die in den sich anschließenden Kapiteln durchgeführten Evaluation durchführen zu können, müssen entsprechende Metriken eingeführt werden. Dazu wird zunächst ein binäres Klassifikationsproblem angenommen, wie in Abbildung 34 dargestellt. Komplexere Klassifikationsprobleme mit mehreren Klassen lassen sich auf dieses Basisproblem reduzieren. Bei dem hier gezeigten Beispiel muss ein Klassifikator entscheiden, ob die Eingabedaten in Form von Merkmals-

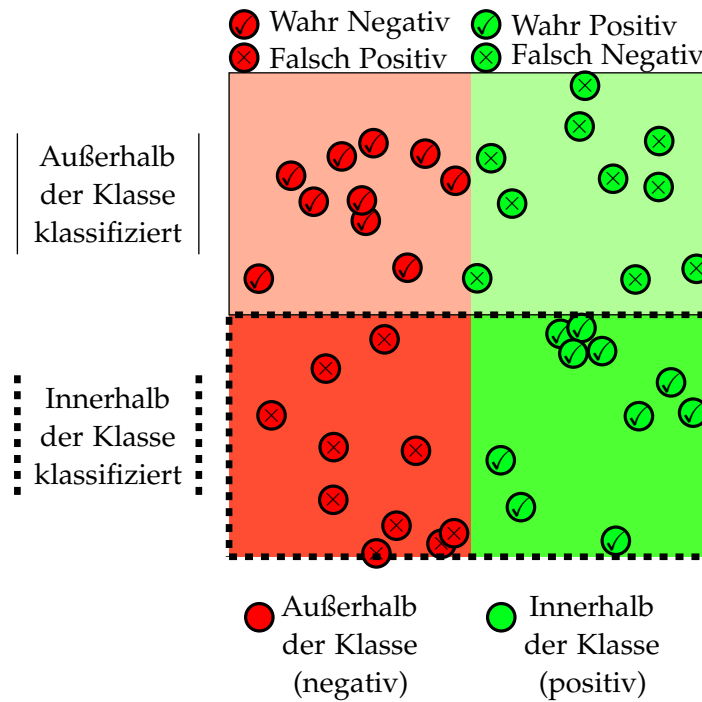


Abbildung 34: Schematische Darstellung der binären Klassifikation von Daten mittels eines Klassifikators. Der grüne Sektor ist Bestandteil der Klasse, während der rote Bereich außerhalb der Klasse liegt. Der Klassifikator entscheidet, welche Daten innerhalb und welche außerhalb der Klasse liegen. Alle Elemente innerhalb des gestrichelten Bereichs wurden vom Klassifikator als der Klasse zugehörig definiert. Korrekt klassifizierte Daten sind mit ✓ gekennzeichnet und falsch klassifizierte Daten mit ×.

vektoren zu einer Klasse gehören oder nicht. Dazu wird der Merkmalsraum entsprechend unterteilt, sodass ein Teilbereich der Klasse zugeordnet wird und der andere nicht. Im hier vorliegenden Beispiel wird zwei gerade Linien durch den quadratischen Merkmalsraum gezogen. Dadurch entstehen grundsätzlich vier Bewertungsfälle, die zu betrachten sind. Hat der Klassifikator Daten der Klasse zugeordnet, die auch der Klasse zugehörig sind wird von *richtig positiv* (engl. *true positive*) gesprochen. Wurden umgekehrt Daten nicht der Klasse zugeordnet, welche auch tatsächlich nicht dazu gehören, wird dies als *richtig negativ* (engl. *true negative*) bezeichnet.

Im negativen Fall, wenn Daten der Klasse zugewiesen wurden, welche nicht dazu gehören, wird dies als *falsch positiv* (engl. *false positive*) definiert. Wurden Daten nicht der Klasse zugewiesen, obwohl sie zur Klasse gehören, wird dies als *falsch negativ* (engl. *false negative*) bezeichnet.

Angelehnt an
[DGo6]

Unter Verwendung dieser vier Definitionen lassen sich verschiedene Evaluationsmetriken definieren. Dazu gehören die Sensitivität (engl. *recall*) und die Genauigkeit (engl. *precision*), welche im Folgenden beschrieben werden.

SENSITIVITÄT Die Sensitivität (engl. *recall*) [DGo6] zeigt die Fähigkeit eines Modells alle relevanten Elemente innerhalb eines Datensatzes zu finden die zur Klasse gehören. Sie definiert sich durch die Menge der richtig positiven \mathbb{E}_{TP} Elemente geteilt durch die richtig positiven plus die Menge der falsch negativen \mathbb{E}_{FN} Elemente.

$$\mathcal{K}_{\text{R}} = \frac{\mathbb{E}_{\text{TP}}}{\mathbb{E}_{\text{TP}} + \mathbb{E}_{\text{FN}}} \quad (28)$$

GENAUIGKEIT Genauigkeit (engl. *Precision*) [DGo6] ist definiert als die Menge der richtig positiven \mathbb{E}_{TP} Elemente dividiert durch die Menge der richtig positiven plus die Menge der falsch positiven Elemente \mathbb{E}_{FP} .

$$\mathcal{K}_{\text{P}} = \frac{\mathbb{E}_{\text{TP}}}{\mathbb{E}_{\text{TP}} + \mathbb{E}_{\text{FP}}} \quad (29)$$

KORREKTKLASSIFIKATIONSRATE Die Korrektklassifikationsrate (engl. *Accuracy*) ist die Anzahl der richtigen Vorhersagen, geteilt durch die Gesamtzahl aller Vorhersagen:

$$\mathcal{K}_{\text{A}} = \frac{\mathbb{E}_{\text{TP}} + \mathbb{E}_{\text{TN}}}{\mathbb{E}_{\text{TP}} + \mathbb{E}_{\text{TN}} + \mathbb{E}_{\text{FP}} + \mathbb{E}_{\text{FN}}} \quad (30)$$

Die Sensitivität drückt die Fähigkeit des Klassifikators aus, alle der Klasse zugehörigen Elemente in einem Datensatz zu identifizieren. Die Genauigkeit definiert wie viele der Elemente, welche vom Klassifikator der Klasse zugeordnet wurden, tatsächlich auch zu der Klasse gehören. Wie so oft bei der Optimierung von Parametern gibt es auch hier einen Kompromiss bei den Kennzahlen. Wird die Sensitivität erhöht, sinkt oft die Genauigkeit.

DER JACCARD-KOEFFIZIENT Bei der semantischen Segmentierung besteht die Aufgabe darin, die Klasse jedes Pixels in einem Bild vorherzusagen. Jedoch ist hier nicht unbedingt eindeutig, was in diesem Fall als richtig positiv bezeichnet werden kann bzw. wie die Ergebnisse der Segmentierung zu bewerten sind. Dazu wird in diesem Fall die (engl. *Intersection over Union*) (IoU) Metrik bestimmt, welche auch als Jaccard-Koeffizient [Jaco2] bezeichnet wird und in Abbildung 35 schematisch dargestellt ist. Diese Metrik ist im Wesentlichen eine Methode zur Quantifizierung der prozentualen Überlappung zwischen der Annotationsmaske, der Grundwahrheit und der Vorhersagemaske des Klassifikators. Dazu misst die IoU-Metrik die Anzahl der Pixel, die zwischen Grundwahrheits- und Vorhersagemaske gemeinsam sind, geteilt durch die Gesamtzahl der Pixel, welche

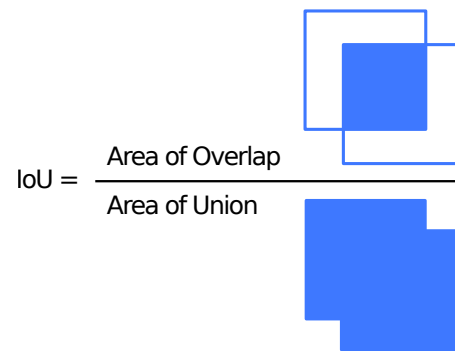


Abbildung 35: Schema von Intersection over Union. Die Menge der Pixel die Grundwahrheit und Maske gemeinsam haben, wird geteilt durch die Gesamtzahl der Pixel beider Masken.

über beide Masken verteilt sind, wie auch in Abbildung 35 dargestellt:

$$\text{IoU} = \frac{\text{Grundwahrheit} \cap \text{Vorhersage}}{\text{Grundwahrheit} \cup \text{Vorhersage}} \quad (31)$$

Da es sich bei der semantischen Segmentierung in der Regel um ein Mehrklassenproblem handelt, wird die Metrik für jede Klasse separat berechnet. Anschließend wird über alle Klassen gemittelt, um einen globalen, mittleren IoU-Wert der semantischen Segmentierung zu erhalten.

VORVERARBEITUNG

4.1 EINFÜHRUNG

Die Daten der verwendeten Sensoren sind zunächst als Rohdaten zu verstehen. Sie stellen zunächst lediglich einen digitalen Signalwert dar. Aufgrund der spezifischen Mosaikstruktur dieser Sensoren ist eine spezielle Vorverarbeitung erforderlich, um aus den Rohdaten einen Hyperwürfel mit spektralen Werten zu erhalten. In den Daten jedes Pixels ist trotz der Fabry-Pérot-Interferometer-Filtertechnik eine Komposition von verschiedenen Wellenlängen enthalten, denn die spezifizierten Wellenlängenempfindlichkeiten werden nicht exakt gemessen. Es wird immer ein Bereich des Spektrums bei einer einzigen Wellenlänge gemessen, da die Messung einer einzelnen Wellenlänge auch physikalisch unmöglich [Pri15b] ist. Diese Mischung von Wellenlängen führt zu Spektralinformationen, die mit wachsender Halbwertsbreite des Messbereiches stärker verfälscht werden. Die Kameras selbst nehmen diesbezüglich hardwareseitig keine Korrektur der Daten vor. Entsprechende Korrekturwerte sind zwar in jeder Kamera gespeichert, werden aber nicht von der Kamera selbst verwendet. Um die Daten zu bereinigen und die unerwünschte Informationen aus den Antworten des Filters zu entfernen existieren verschiedene Möglichkeiten. Zunächst lässt sich durch geeignete optische Filter schon bei der Aufnahme eine Konditionierung im Hinblick auf unerwünschte Wellenlängen erreichen, indem Wellenlängen außerhalb des spezifizierten Bereichs geblockt werden. Weiterhin müssen die restlichen Fehlerquellen über eine geeignete Vorverarbeitung kompensiert werden. Eine entsprechende Vorverarbeitung, welche aus dem Rohbild und den enthaltenen Rohdaten einen Hyperwürfel erzeugt, wird im folgenden Abschnitt erläutert. Das Verfahren basiert auf Informationen und Anweisungen des Kameraherstellers bezüglich der Sensoreigenschaften, den hinterlegten Korrekturkurven und deren Anwendung. Es wurde im Hinblick auf das gegebene Szenario mit dynamischer Beleuchtung entwickelt.

4.2 STAND DER TECHNIK

Es existieren verschiedene Techniken, um spektrale Messungen zu ermöglichen, wie in Kapitel 2.2 beschrieben. Die verschiedenen Technologien für spektrale Messverfahren sind grundsätzlich seit vielen Jahren etabliert. Weiterhin wurden in den letzten Jahren neue Technologien entwickelt, welche Interferenzfilter nutzen, um multispek-

trale Filterarrays auf Sensorebene zu implementieren, wie von Lapray et al. [LWTG14] publiziert. Diese Filterarrays orientieren sich am Prinzip des Bayer-Sensors und erweitern dieses auf mehr als drei Kanäle. Dementsprechend sind einzelne Pixel innerhalb des Filterarrays für unterschiedliche Wellenlängen empfindlich, was das Erfassen von verschiedenen Wellenlängen mit einer Belichtung ermöglicht. Diese Methode wird auch als Snapshot-Mosaik-Technik bezeichnet.

Ein Überblick über etablierte Snapshot-Sensoren wurde 2013 von Hagen et al. [HK13] publiziert. Auf dieser Technologie aufbauend stellten Lambrecht et al. [LGG⁺14] und Geelen et al. [GTL14] im Jahr 2014 ein neuartiges Sensorkonzept vor, bei dem die spektrale Einheit monolithisch auf einem Standard-CMOS-Sensor integriert ist. Lambrecht et al. erläutern auch das eine Vorverarbeitung der Rohdaten bei den hier verwendeten Sensoren notwendig ist, um spektrale Werte zu erhalten. Sie nutzen dafür Aufnahmen einer Kachel mit Referenzweiß. Dies ist, wie sie beschreiben, aber nicht für Anwendungen mit weitem Sichtfeld praktikabel. Dementsprechend sind andere Verfahren notwendig. Zuletzt hat Tsagkatakis et al. [TJGT16] mehrere Vorverarbeitungsmethoden für die Rekonstruktion von Spektraldaten vorgeschlagen, die mit den hier verwendeten Kameras aufgenommen wurden. Sie liefern einen Ansatz zur Abschätzung und Rekonstruktion fehlender spektraler Messungen, indem sie es als ein Problem der Matrixvervollständigung mit niedrigem Rang formulieren. In einem weiteren Ansatz [TT16] wird die Selbstähnlichkeiten über Skalen in Bildern ausgenutzt um eine Schätzung des Hyperwürfels mit voller Auflösung zu ermöglichen. Während Degraux et al. [DCJ⁺15] das *Demosaiicing* als *3-D-Inpainting*-Problem formuliert, um dieses zu lösen und die Auflösung des Datenvolumens zu erhöhen. Allerdings werden auch hier nur Laborszenarien angenommen. Eine Übersicht der Kalibrierverfahren für andere spektrale Messsysteme ist bei Tansock et al. [YK15] und [JDS⁺16] zu finden. Weiterhin zeigen Aasen et al. [AHLZT18] einen Überblick über die verschiedenen Sensortechniken und welche sich zur Verwendung auf Drohnen eignen und verweisen auch auf Verfahren zur spektralen Korrektur und Kalibrierung der aufgenommenen Daten. So stellt laut Aasen die Normalisierung der dynamischen Beleuchtungsänderungen in der Drohnenfernerkundung nach wie vor ein Problem dar. Dieses Problem trifft auch auf das in dieser Arbeit betrachtete Szenario zu, da die Sensorik zur Umgebungswahrnehmung beim autonomen Fahren eingesetzt werden soll. Eine weitere Möglichkeit der Rekonstruktion von spektralen Informationen ist die Wiener Inverse wie von Hubel et al. [HSF94], Nishidate et al. [NMNA13] und Yoshida [YNI⁺15] vorgestellt. Julie Klein hat in ihrer Dissertation [KHM16] verschiedene Rekonstruktionsalgorithmen untersucht. Laut ihren Untersuchungen liefert die Wiener Inverse die besten Ergebnisse bei der Rekonstruktion von spektralen Werten. In einem Patent von Biosensing Systems wurden

die Wiener Inverse und die Rekonstruktionsmethode des Chipherstellers miteinander verglichen, welche auf der Bestimmung einer Antwortmatrix (engl. *Response Matrix*) mittels eines Monochromators basiert. Laut den Angaben im Patent ist der Rekonstruktionsfehler bei der Verwendung der Wiener Inverse geringfügig besser als gegenüber der vom Hersteller genutzten Methode.

4.3 DATENVORVERARBEITUNG

Die spezielle Funktionsweise der in dieser Arbeit verwendeten Kameras wurde in Kapitel 3.6.2 beschrieben. In dem Abschnitt werden auch Artefakte beschrieben, welche den Umgebungsbedingungen entspringen und von der Messtechnik an sich erzeugt werden und die gemessenen Daten verfälschen. Diese Artefakte treten bei den gesuchten Wellenlängen und harmonischen Vielfachen auf und sind in den aufgenommenen Daten enthalten. Dies beeinflusst die anschließende Verarbeitung der Daten, da diese Effekte kompensiert werden müssen, um bereinigte Werte über die reflektierte Strahlung zu erhalten. Das hier beschriebene Verfahren nutzt Daten und Informationen des Chipherstellers. Um den Sensor bzw. den Sensorchip zu kalibrieren, wird die Antwort (engl. *Response Matrix*) jedes Chips nach der Herstellung bestimmt. Unter Verwendung eines Monochromators wird die Sensorantwort für diskrete Wellenlängen in 1 nm Schritten in einem Bereich von 400-1000 nm gemessen und bestimmt. Das Ergebnis dieses Vorgangs ist eine Antwortmatrix, bei der jede Zeile den Beitrag der jeweiligen Wellenlänge zur Gesamtantwort des Filters enthält. Die Zeilen der Antwortmatrix entsprechen der Anzahl der unterschiedlichen Filter in einem Makropixel, und die Anzahl der Spalten entspricht der Anzahl der bei der Kalibrierung durchgeführten Messpunkte im Spektrum. Diese Antwortmatrix, im Folgenden auch als Korrekturkurven bezeichnet, kann dann genutzt werden, um aus den gemessenen Signalwerten spektrale Werte zu bestimmen. Es lässt sich folgendes definieren:

$$I_i = \eta \sum_{\lambda=400}^{1000} \mathbf{R}_{i\lambda} \cdot \mathcal{S}(\lambda) \quad (32)$$

mit

- I_i als der gemessene Signalwert
- η als Quanteneffizienz
- \mathbf{R} als Antwortmatrix
- $\mathcal{S}(\lambda)$ als die spektrale Bestrahlungsstärke der Lichtquelle bei λ

Da die Lichtquelle (Monochromator) und der jeweilige Signalwert bekannt sind, lässt sich die Antwortmatrix entsprechend bestimmen

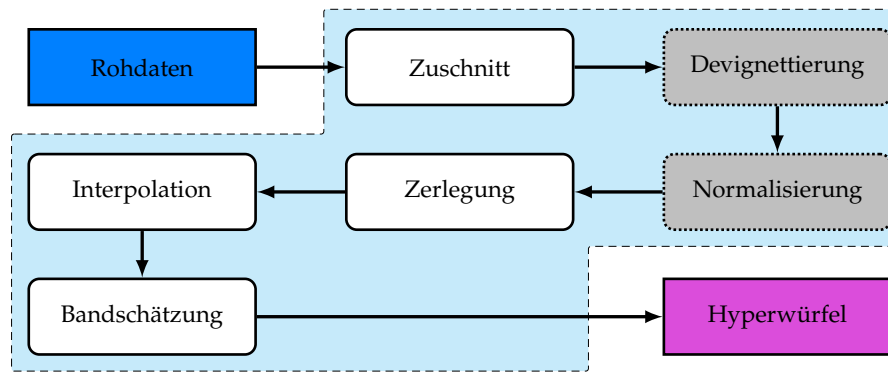


Abbildung 36: Die Abbildung zeigt den Ablauf der Vorverarbeitung (cyan) von den Rohdaten der Kamera (blau) bis zum Hyperwürfel (violett). Feste Komponenten sind weiß und optionale Stufen sind grau dargestellt.

und nach der Invertierung zur Ermittlung von spektralen Werten nutzen.

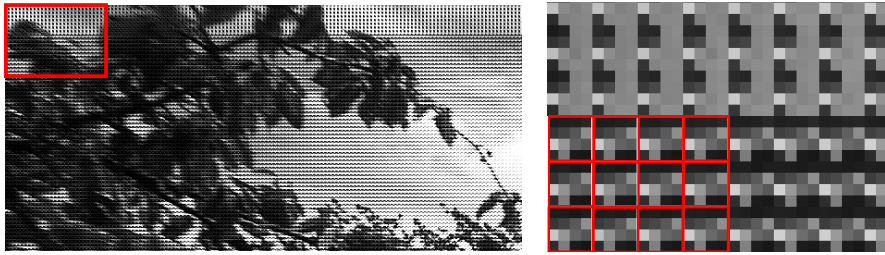
Wie Abbildung 36 zeigt, werden in dieser Arbeit mehrere Schritte unternommen, um die Daten des Rohbilds zu verarbeiten. Ein solches Rohbild f^G , das von der Kamera erzeugt wird, ist definiert als

$$f^G : \mathcal{L}_x \times \mathcal{L}_y \times \mathcal{L}_\lambda \rightarrow [0,255] \quad (33)$$

wobei \mathcal{L}_x und \mathcal{L}_y die räumliche Domäne des Bildes definieren und \mathcal{L}_λ mit $\mathcal{N}_\lambda = 1$ definiert ist. Solch ein Bild wird in Abbildung 37 gezeigt. Die Pixelresonanzwerte im Rohbild müssen zu einem Hyperwürfel umgewandelt werden, um eine weitere Analyse und Nutzbarkeit herzustellen. Dabei ist die Konvertierung der rohen Pixelwerte zu spektralen Werten von zentraler Bedeutung. Zusätzlich relevant sind die Entfernung von Artefakten, welche durch die verlustbehaftete Messtechnik und Umgebungseinflüsse entstanden sind. Dies geschieht in einem Vorverarbeitungsprozess, deren einzelne Schritte im Folgenden näher erläutert werden.

Zuschnitt

Zuerst muss das Rohbild zugeschnitten werden, da die auf die Pixel aufgebrachte Beschichtung nicht ganz bis zu den Rändern reicht wie in Abbildung 37 dargestellt. Diese Randbereiche enthalten nicht die korrekten Makropixel und nehmen damit fehlerhafte Daten auf. Um in der anschließenden Analyse nur mit korrekten Daten zu arbeiten wird das Bild im Ortsbereich nach Gleichung (19) entsprechend der Herstellerangaben beschnitten: $Loc' \subset Loc$



(a) Rohbild der VIS-Kamera mit erkennbarem Makropixelmuster und erkennbar fehlenden Filtern am oberen Bildrand (b) Vergrößerung des linken Bildausschnitts. Makropixel sind rot markiert

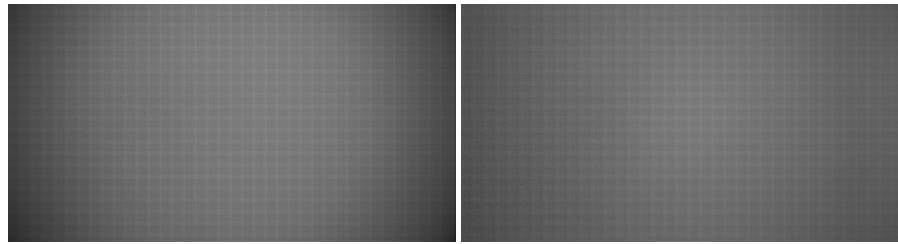
Abbildung 37: Das Makropixelmuster ist nicht vollständig auf der Sensorfläche aufgebracht. In der oberen Bildhälfte ist die Beschichtung nicht vollständig. Eine Auswahl gültiger Makropixel ist im rechten Bild rot hervorgehoben.

Devignettierung

In Abhängigkeit von der verwendeten Optik (Objektiv) ist im Bild eine Vignettierung enthalten. Abbildung 38a zeigt ein Beispielbild, das vor einem monotonen weißen Hintergrund aufgenommen wurde. Die Optik der Kamera reduziert hier die Bestrahlungsstärke des einfallenden Lichts am Rand des Bildes. Wie von Ray et al. [Ray02] beschrieben, führen in der Regel zwei Faktoren zur Bildung einer Vignette im Bild:

- Die Reduzierung des einfallenden Lichtstroms wird durch den Winkel des einfallenden Lichtes auf die Optik der Kamera verursacht.
- Die mechanische Vignettierung ist der andere Faktor, welcher durch Okklusion des einfallenden Lichts, durch Objekte oder Linsenschattierungen definiert ist. Hierzu gehört auch die Verdunkelung durch das Objektiv oder das Objektivgehäuse, was zu harten Kanten führen kann.

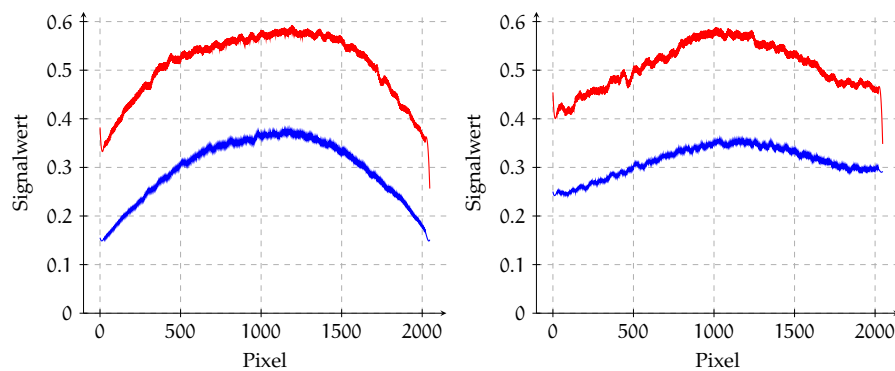
Gerade bei den hier eingesetzten Kameras spielt die mechanische Vignettierung jedoch keine Rolle. Die Kameras sind frei hängend montiert und verfügen über keine montierte Gegenlichtblende, sodass nur der Einfluss der Optik wichtig ist. Die natürliche Vignettierung reduziert das auf den Sensor einfallende Licht um einen bestimmten Faktor ähnlich der Wirkung eines Objektivs. Dabei ist der Faktor abhängig von der Position des Pixels auf dem Sensor. Um dies zu korrigieren muss üblicherweise ein entsprechender inverser Faktor je Pixel gefunden werden, welcher dann mit dem Pixelwert multipliziert wird. Lopez et al. haben [LFOM15] haben eine Methode vorgeschlagen, bei der die Entropie der Intensitäten des Eingabebildes minimiert wird, so wird für jedes Bild eine neue Korrektur ermittelt. Dies



(a) Vignette 1

(b) Vignette 2

Abbildung 38: Eine weiße diffuse Fläche wurde zweimal hintereinander mit identischer Kameraposition aufgenommen. Zwischen Abbildung 38a und Abbildung 38b wurde der Lichteinfall auf den Sensor und die Belichtungszeit der Kamera variiert.



(a) Vignette 1

(b) Vignette 2

Abbildung 39: Dargestellt sind Plots je zweier Zeilen der Bilder aus Abbildung 38a und Abbildung 38b. Anhand der Plots ist gut der Verlauf der Vignettierung zu sehen und dass er sehr stark vom Lichteinfall und der Position des Lichtes abhängig ist.

ist jedoch relativ komplex und eher eine Nachbearbeitung von Bildern und daher nicht für den Dauerbetrieb ausgelegt. Goldman et al. [Gol10] stellen ein anderes Verfahren vor, bei dem eine radialsymmetrische Funktion mit wenigen Parametern bestimmt wird. Dies bietet dann die Möglichkeit, für jedes Pixel einen Faktor zu finden, der über das Eingangsbild definiert wird und damit ein einziges Modell für eine Kamera und ihr Objektiv definiert.

In dieser Arbeit wird auch jeweils ein Faktor bestimmt, indem zunächst ein monoton helles Graubild aufgenommen wird. Dies kann z. B. durch die Aufnahme einer Weißreferenzplatte erreicht werden, wie sie in Abbildung 40 dargestellt ist. Eine solche Platte ist nahezu



Abbildung 40: Platte zum Weißabgleich von Spherooptics

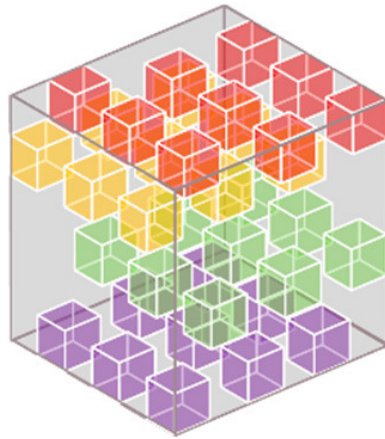
ideal diffus für alle untersuchten Lichtwellenlängen und reflektiert das Licht gleichmäßig über den gesamten Wellenlängenbereich. Allerdings ist es bei der Verwendung einer lambertschen Oberfläche, wie beispielsweise einer radiometrischen Referenzplatte schwierig, das gesamte Ziel homogen auszuleuchten. Auch ist es wichtig, dass die Beleuchtung so geregelt wird, dass kein Pixel vollständig weiß und damit gesättigt ist und dass keine sehr schwach beleuchteten Pixel existieren, um ein gleichmäßig ausgeleuchtetes Bild zu erhalten. Der Effekt der Vignettierung ist in der Bildmitte am wenigsten signifikant, sodass dort die Intensitäten des Makropixels gemittelt werden. Dieser Signalwert dient dann als Referenz für die Berechnung der inversen Faktoren. Da jede Bildaufnahme für sich genommen Schwankungen der Lichtsituation und des Rauschens unterworfen ist, werden die Werte einer größeren Anzahl geeigneter Bilder gemittelt.

Die vollständige oder teilweise Entfernung der Vignettierung ist als ein optionaler Schritt in der Vorverarbeitung enthalten. Beobachtungen haben gezeigt, dass die Vignette aufgrund von Lichtveränderungen stark variiert. Daher ist es möglich, dass eine Korrektur solcher Bilder zu starken Artefakten führen kann und damit der Verarbeitung mehr schadet als nutzt. Denn Änderungen der Beleuchtung gegenüber der Lichtsituation während der Kalibrierung sind, im vorliegenden Szenario, allein aufgrund der veränderlichen relativen Position der Sonne und sich verändernder Wettersituation im Freien nicht zu vermeiden. Abbildung 38b zeigt ein zweites Bild des *NIR*, in dem die Beleuchtung leicht variiert wurde. Hier ist klar zu sehen, dass die Vignette eine andere Form und Verteilung aufweist. Dies wird auch von Abbildung 39 bestätigt, in dem jeweils die Signalwerte zweier Zeilen der Aufnahmen dargestellt sind. Daher ist eine Vignettenkorrektur eher für Szenarien mit statischer Beleuchtung, wie z. B. bei der Nahrungsmittelinspektion auf einem Förderband, geeignet. Da die Kameras auf einem autonomen Fahrzeug zum Einsatz kommen sollen, bzw. dieses Szenario hier untersucht wird, unterliegen die aufgenommenen Daten ständigen Änderungen der Beleuchtungssituation.

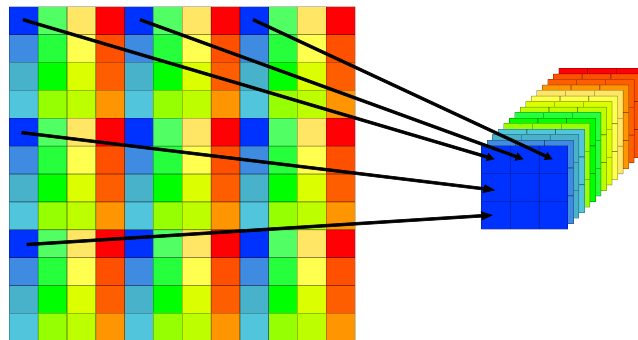
Daher ist eine Vignettenkorrektur in diesem Szenario eher als nachteilig anzusehen.

Normalisierung

Ein Rohbild enthält eine akkumulierte, digitalisierte und diskrete Bestrahlungsstärke in jedem Pixel. Dieser Wert besitzt jedoch keine physikalische Einheit, da die auf der Sensoroberfläche gemessene Lichtmenge von mehreren Faktoren wie z. B. Belichtungszeit, Quanteneffizienz des Sensors und Blendeneinstellung abhängt, wie auch bereits in Gleichung (1) in Kapitel 3.4 beschrieben. Der Einfluss dieser teilweise nicht messbaren Faktoren lässt sich nicht vollständig umkehren. Um dennoch eine physikalische Aussage treffen zu können, kann analog zur Vignettierungskorrektur ein Referenzweiß aufgenommen und daraus ein Makropixel als Referenzwert extrahiert werden. Wurde zuvor eine Devignettierung durchgeführt, muss auch das Weißspektrumbild entsprechend korrigiert werden. Die so aufgenommenen Werte enthalten in diesem Fall in jedem Pixel eine definierte Beleuchtung in Bezug auf ein bekanntes Lichtspektrum und eine bekannte Oberfläche. Teilt man die Werte der einzelnen Makropixel durch diese Weißspektrum-Makropixel, so erhält man einen relativen Beleuchtungswert bezogen auf die gemessene und damit bekannte Beleuchtung. Dieser Schritt ist wie die Devignettierung optional, da für jedes Makropixel die gleiche lineare Skalierung durchgeführt wird. Wie sich eine solche Skalierung jedoch auf nachfolgende Klassifizierungsalgorithmen auswirkt, ist nicht unmittelbar ersichtlich. Es ist denkbar, dass sich die Konditionierung wie in [Har97] ändert, jedoch bleiben lineare Beziehungen in den Daten linear und nicht lineare nichtlinear. Weiterhin ist zu beachten, dass dieses Vorgehen sich nur für Einsatzszenarien unter Laborbedingungen eignet, da die Beleuchtung als unveränderlich angenommen wird. Dies trifft allerdings nicht zu, sobald die Kameras auf bewegten Plattformen im Außenbereich eingesetzt werden, da hier die Beleuchtung als dynamisch anzunehmen ist. Zusätzlich ist das Verfahren auch nicht anwendbar bei Anwendungen, wie z. B. Aufnahmen mit großem Öffnungswinkel der Linse. Zu beachten ist auch, dass der effektiv nutzbare Öffnungswinkel durch die Wirkungsweise des Fabry-Pérot-Interferometer begrenzt wird. Dieses Problem wird auch von Aasen et al. [AHLZT18] beim Einsatz von Hyperspektralkameras auf Drohnen und von Hagen et al. [HK13] beschrieben. Zur Kompensation dieses Problems müsste ein zweiter identischer spektraler Sensor auf dem Fahrzeug montiert werden, welcher kontinuierlich die Beleuchtung der Umgebung aufnimmt, dies schlagen auch Aasen et al. [AHLZT18] in ihrer Publikation vor. Allerdings war dieses Vorgehen im vorliegenden Szenario nicht möglich, da jeweils nur ein Kamerasystem eingesetzt werden konnte. Daher ist dieser Schritt zwar in der entwickelten Soft-



- (a) Darstellung des Hyperwürfels vor der Interpolation. Aufgrund des Makropixelmusters sind die einzelnen Pixel für eine definierte Wellenlänge im Hyperwürfel räumlich verteilt angeordnet. Bildquelle: [AHLZT18]



- (b) Zerlegung der jeweiligen Makropixel zu einem Hyperwürfel. Die einzelnen Pixel mit unterschiedlichen Wellenlängenempfindlichkeiten werden Teilbildern zugeordnet, die als Bänder mit definierten Wellenlängen in einem Hyperwürfel definiert sind.

Abbildung 41: Schematische Darstellung des Hyperwürfels vor der Interpolation und nach der Zerlegung.

ware modelliert, wird aber nicht aktiviert, wenn Daten zur Umgebungswahrnehmung beim Autonomen Fahren analysiert werden.

Zerlegung

In diesem Schritt wird das Rohbild und die enthaltenen Makropixel in einen Hyperwürfel transformiert. Dazu müssen die vorhandenen Makropixel in ihre einzelnen Pixel zerlegt werden. Jeder Makropixel enthält 16 bzw. 25 Pixel mit unterschiedlicher Filterhöhe und entsprechend definierten Wellenlängenempfindlichkeiten. Eine Kamera, de-

ren Sensor pro Makropixel \mathcal{N}_λ Pixel enthält, wird in \mathcal{N}_λ Teilbilder der Größe

$$\frac{\mathcal{N}_x}{\mathcal{M}_x} \times \frac{\mathcal{N}_y}{\mathcal{M}_y} \quad (34)$$

zerlegt, wobei gilt $\mathcal{M}_x = \mathcal{M}_y = \sqrt{\mathcal{N}_\lambda}$ da alle Makropixel quadratisch sind. Die aus dem Makropixel extrahierten Pixel werden in Teilbilder entsprechend der Makropixelposition einsortiert. Jede Pixelposition dieser neuen Bilder entspricht dann einem Makropixel des Originalbildes. Ein solches Teilbild enthält dann den Messwert jedes Makropixels für einen definierten Wellenlängenbereich. Schließlich werden die Teilbilder nach Wellenlänge sortiert und in einem Würfel angeordnet. Jedes Teilbild im Hyperwürfel kann dann auch als ein *Band* \mathbf{B} einer definierten Wellenlänge bezeichnet werden. Das Vorgehen bei der Zerlegung wird in Abbildung 41b schematisch dargestellt. Die Auflösungen der Teilbilder bzw. Bänder entsprechen bei der VIS-Kamera einer Bildgröße von 512×256 und bei der NIR-Kamera von 409×216 .

Interpolation

Im vorherigen Schritt wurde ein Hyperwürfel erzeugt, der jedoch noch systematische Fehler enthält. Das Hyperpixel \mathbf{p}^H an der Position (i,j) definiert, aufgrund der Makropixelstruktur des Sensors, für jede Wellenlänge eine andere geometrische Position. Dies ist in Abbildung 41a schematisch dargestellt. Bei der Zerlegung wurden die Pixel einer Wellenlänge lediglich aus den einzelnen Makropixeln extrahiert und in einem Teilbild nebeneinander geschrieben. Dadurch sind geometrische Differenzen zwischen den Teilbildern verschiedener Wellenlängen entstanden. Denn ein Pixel, welches im Bild neben einem anderen Pixel positioniert war, ist im Hyperwürfel nun hinter oder vor dem Nachbarpixel angeordnet. Pixel, die im Teilbild nebeneinander liegen, hatten vorher einen Abstand, welcher der Seitenlänge \mathcal{M} des Makropixels entspricht. Diese geometrischen Verschiebungen können vor allem bei Materialübergängen zu Problemen führen, wie in Abbildung 42 beispielhaft dargestellt. In dieser Abbildung ist auch zu sehen, dass das mit 16 markierte Pixel einen engen Nachbarn im benachbarten Makropixel mit einem Abstand von 1 hat. Diese geometrischen Fehler können durch die geometrische Interpolation der einzelnen Bänder ausgeglichen werden. Zuerst wird die Auflösung aller Bänder um den Faktor $\sqrt{\mathcal{N}_\lambda} = \mathcal{M}$ mittels einer Interpolation erhöht. Die vergrößerten Bänder werden dann so gegeneinander verschoben, dass die eigentlichen Positionen der Pixel in den Bändern übereinander liegen. Dazu werden die Bänder entsprechend ihrer Position im Makropixel verschoben, so liegen alle Originalpixel an ihrer ursprünglichen Position. Dadurch bilden sich an den Rändern aber Bereiche, die nicht mehr von allen Bändern abgedeckt werden können. Da nicht alle Wellenlängen verfügbar sind, werden die unvollständigen

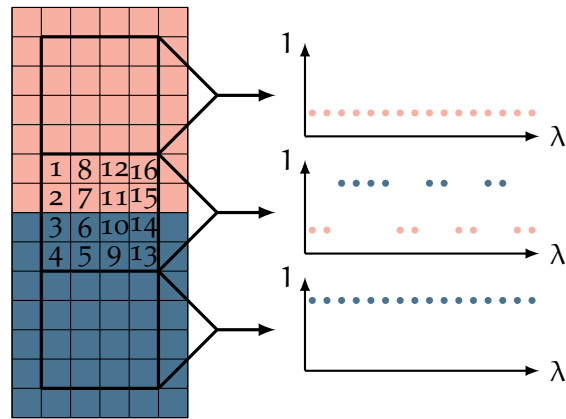


Abbildung 42: Die Grafik zeigt die Abtastung einer Kante zwischen zwei Materialien (orange und blau) durch Makropixel. Auf der linken Seite befindet sich ein Ausschnitt des Bildes, auf der rechten Seite das Messsignal aus den Makropixeln. λ bezeichnet die Wellenlängenempfindlichkeit. Die Pixelintensität wird auf die vertikale Achse ohne Einheiten angewendet. Es zeigt sich, dass das mittlere Spektrum unbeständig ist.

gen Hyperpixel an allen vier Rändern jeweils abgeschnitten und die Bänder so geometrisch verkleinert. Das Vorgehen ist in Abbildung 43 schematisch dargestellt. Die neue Auflösung des Hyperwürfels ist entsprechend kleiner. Für jedes Band der *VIS*-Kamera ergibt sich eine Auflösung von 510×254 Pixeln und jedes Band der *NIR*-Kamera hat eine Auflösung von 407×214 Pixeln.

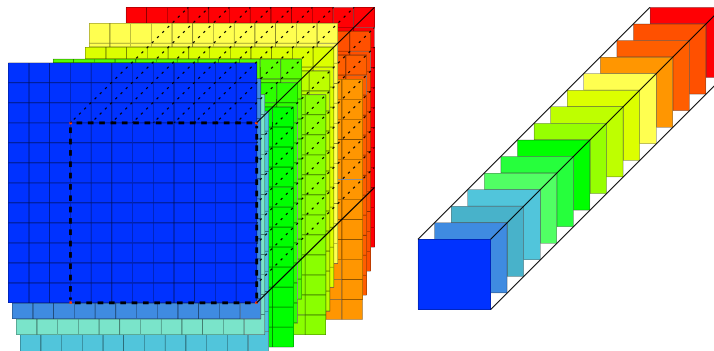


Abbildung 43: Schema der geometrischen Korrektur des Hyperwürfels durch Verschieben der Bänder und Abschneiden der unvollständigen Hyperpixel

Bandschätzung

Da nun die Bänder des Würfels ihren geometrischen Positionen entsprechen, müssen die Pixelwerte in spektrale Werte umgewandelt werden. Der gemessene digitale Signalwert eines Pixels wird durch drei wesentliche Faktoren beeinflusst, welche in Kapitel 3.6.2 beschrieben werden:

1. Licht einer definierten Wellenlänge wird im Filter durch Interferenz verstärkt und strahlt dann auf die Sensoroberfläche. Es wird aber nicht nur genau eine Wellenlänge übertragen, sondern ein Wellenlängenbereich, da die Bandbreite des Sensorelements etwa bei von 15 nm liegt.
2. Durch benachbarte Pixel kommt es zu einem Übersprechen (engl. *Crosstalk*), welches bei der Belichtung entsteht. Dabei kann ein Pixel *überstrahlt* werden und umliegende Pixel mit belichten.
3. Ein Pixel reagiert nicht nur auf eine definierte Wellenlänge, sondern auch auf harmonische Vielfache dieser. Dies kann auch als Antwort zweiter Ordnung (engl. *Second Order Response*) bezeichnet werden. Dieser Effekt wird durch die in der Kamera verbauten Bandpassfilter begrenzt. Wären diese nicht vorhanden, würde eine große Anzahl anderer Frequenzen den Pixel zusätzlich belichten und die Messung massiv beeinträchtigen. Die Bandpassfilter verhindern aber nicht alle Interferenzen.

Um aus den Pixelwerten, welche momentan nur ein digitales Signal definieren, Spektralwerte zu erhalten, wird eine Korrektur aus allen gemessenen Signalwerten berechnet, die für eine geometrische Position zur Verfügung stehen. Essentiell dafür sind Korrekturkurven \mathcal{C} , welche vom Hersteller des Sensors experimentell für jede Kamera einzeln ermittelt wurden.

Unter Verwendung dieser Korrekturkurven werden aus den vorhandenen Pixelwerten neue virtuelle Spektralbänder wie folgt berechnet. Jeder Wert an einer (i,j) -Position des neuen virtuellen Bandes mit der Empfindlichkeit λ wird aus dem Skalarprodukt des Punktspektrums χ des entsprechenden Hyperpixels \mathbf{p}^H mit der korrespondierenden Korrekturkurve \mathcal{C}_λ für das jeweilige Band berechnet:

$$\mathbf{B}_\lambda(i,j) = \chi(i,j) \circ \mathcal{C}_\lambda \quad (35)$$

Die Korrekturkurven bilden somit Gewichte für die einzelnen Messungen mit unterschiedlichen Empfindlichkeiten. Die Wellenlänge, welche dem Band entspricht, erhält ein hohes Gewicht. Messungen,

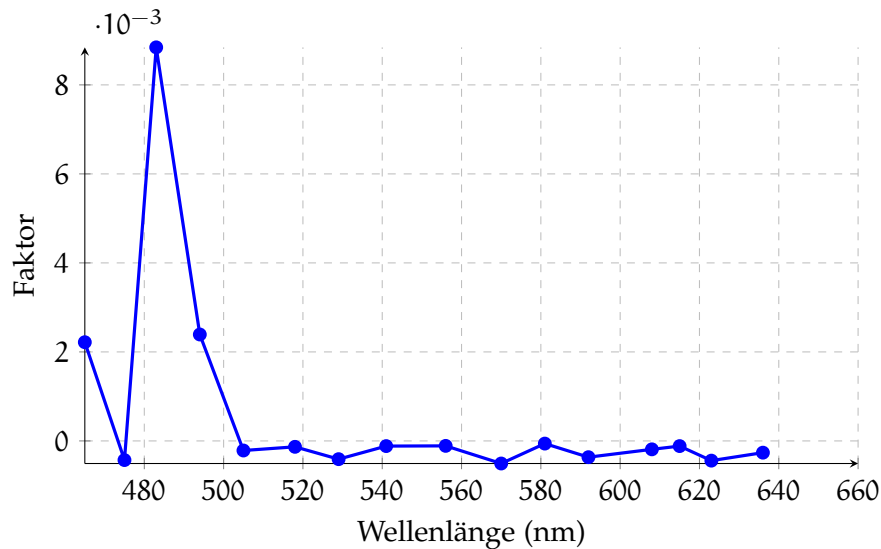


Abbildung 44: Korrekturkurve für 483 nm

die zu einem geometrischen Nachbarn gehören, erhalten einen kleinen negativen Faktor, um dem Übersprechen benachbarter Pixel entgegenzuwirken. Dies ist in Abbildung 44 anhand einer Korrekturkurve zu sehen. Darüber hinaus erhalten harmonische Vielfache ein signifikantes negatives Gewicht, um ihren Einfluss zu reduzieren. Die Korrekturkurven werden also entsprechend auf alle Pixel des Rohbildes angewandt woraus sich ein korrigierter Hyperwürfel mit spektralen Werten ergibt. Um einen spektralen Reflexionsgrad gemäß der Definition in Kapitel 3.5 zu erhalten, müsste der Wert ins Verhältnis zur in der Szene vorherrschenden Beleuchtung gesetzt werden. Dies ist aber aufgrund des in dieser Arbeit vorliegenden Szenarios, wie bereits erläutert, nicht möglich, aber auch nicht notwendig. Denn die Werte sollen nicht mit Referenzdatenbanken verglichen werden, sondern genutzt werden um Klassifikatoren zu trainieren.

4.4 RGB-GENERIERUNG

Entsprechend Kapitel 3.5.4 wird untersucht ob aus den spektralen Daten der Kameras unter Verwendung der Vorverarbeitung RGB-Bilder erzeugt werden können. Dazu wird eine Transformation $T_{H \rightarrow RGB}$ durchgeführt, welche die 16 Bänder der VIS-Kamera auf 3 Bänder reduziert. Da die NIR-Kamera Daten aus dem nicht sichtbaren Bereich liefert, werden nur Daten der VIS-Kamera genutzt.

$$T_{H \rightarrow RGB} : (\mathcal{L}_x \times \mathcal{L}_y \times \mathcal{L}_\lambda) \rightarrow (\mathcal{L}_x \times \mathcal{L}_y \times \{0,1,2\}) \quad (36)$$

Basierend auf dem zuvor erwähnten Kapitel kann ein Spektrum des sichtbaren Lichts, unter Verwendung der Spektralwertfunktionen, welche aus den Grundspektralwertkurven abgeleitet sind, in

den XYZ-Farbraum überführt werden. Dazu müssen aber zunächst die vorverarbeiteten Daten der Kamera in Form von spektralen Bändern interpoliert werden, da die Kameras natürlich kein vollständiges Spektrum messen, sondern nur punktuell das Spektrum abtasten. Nach der Interpolation werden die Werte mittels der Spektralwertfunktionen des CIE 1931 2°-Normbeobachters (8), (9) und (10) in den XYZ-Farbraum überführt. Da kein gemessenes kontinuierliches Spektrum vorliegt und auch die Spektralwertfunktionen nur diskret vorliegen, werden die Funktionen als Matrix $\mathfrak{B} = (\bar{x}, \bar{y}, \bar{z})^T$ aufgefasst und die XYZ-Werte aus dem von der Kamera gemessenen Spektrum χ des Hyperwürfels durch eine Matrixmultiplikation wie folgt berechnet:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathfrak{B} \cdot \chi \quad (37)$$

Die überführten Werte definieren einen Farbeindruck den ein CIE-Normbeobachter wahrnimmt. Um nun zu RGB-Informationen zu gelangen, müssen die Werte aus dem XYZ-Farbraum in RGB-Farbraum überführt werden. Die dazu nötige Transformation beinhaltet eine invertierte 3×3 Matrix \mathbf{M} , für die auf den sRGB-Standard [C+99] zurückgegriffen werden kann.

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \mathbf{M}^{-1} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (38)$$

So definieren dann R,G,B die gesuchten Werte im RGB-Farbraum. Um die Werte aus dem vorverarbeiteten Hyperwürfel zu erhalten, muss diese Transformation für jeden Hyperpixel durchgeführt werden. Einige Ergebnisse der beschriebenen Transformationen sind in Abbildung 45 dargestellt. Die Beispiele zeigen, dass die Transformation von spektralen Daten zu RGB-Daten grundsätzlich plausible Ergebnisse produziert. Dies ist insbesondere an den Verkehrsschildern in den einzelnen Bildern zu sehen und am Grün der Wiese sowie der Farbe der Straße. Weiterhin ist aber ein leichter Blaustich erkennbar. Dieser wird wahrscheinlich dadurch ausgelöst, dass die Kamera im blauen Bereich des Spektrums nur einen Messpunkt bei 465 nm besitzt. Der Mensch kann allerdings schon ab 380 nm Licht wahrnehmen, so kommt es bei der Kamera zu einer Unterabtastung in diesem Bereich, welcher wahrscheinlich den Blaustich erklärt.

4.5 FAZIT

Die vorgeschlagene spektrale Vorverarbeitung rekonstruiert aus den Rohdaten der Sensoren mit Fabry-Pérot-Interferometer-Filtertechnik



Abbildung 45: Beispiele für die Transformation der spektralen Daten zu RGB-Daten basierend auf den vorgestellten Datensätzen.

spektrale Informationen. Die Vorverarbeitung erzeugt dazu in mehreren Schritten aus einem Rohbild einen Hyperwürfel, welcher zur spektralen Analyse genutzt werden kann. Das Verfahren beruht auf den Anweisungen des Chipherstellers und wurde unter Berücksichtigung der Anforderungen an das hier vorliegende Szenario und die Aufgabenstellung konzipiert. Basierend auf den Informationen aus dem Patent von Biosensing Systems ist dieses Verfahren nur geringfügig schlechter als die Verwendung der Wiener Inverse. Da die Sensoren nicht unter kontrollierten Laborbedingungen, sondern in dynamischen Szenen eingesetzt werden, lassen sich einige Aspekte der klassischen Vorverarbeitung wie die Weißspektrumskorrektur und Devignettierung nicht praxistauglich umsetzen. Basierend auf den rekonstruierten spektralen Daten kann dann im weiteren Verlauf ein Klassifikator trainiert werden, wie es in Kapitel 7 beschrieben wird. Dieser erlaubt es, verschiedene Oberflächen in der Szene zu klassifizieren.

5.1 EINFÜHRUNG

Verfahren zur Umgebungsanalyse zielen darauf ab, die einzelnen Bestandteile einer ganzen Szene sowie deren Zusammenhänge auf einer lokalen Pixel- und Instanzebene zu identifizieren. Trotz erheblicher Fortschritte bleibt das visuelle Szenenverständnis eine Herausforderung, insbesondere wenn man die menschliche Kognition als Referenz nimmt. Der Fortschritt der neuronalen Netze hat einen großen Einfluss auf den aktuellen Stand der Technik im Bereich des maschinellen Lernens und des Rechnersehens. Viele leistungsstarke Methoden in einer Vielzahl von Anwendungen basieren heute auf tiefen neuronalen Netzwerken. Ein wesentlicher Faktor für den Erfolg ist die Verfügbarkeit umfangreicher, öffentlich zugänglicher Datensätze. Infolgedessen werden erhebliche Forschungsanstrengungen in neue und umfangreiche annotierte Datensätze gesteckt.

Während in den letzten Jahren viele große Datensätze öffentlich verfügbar wurden, bei denen RGB- und Laserdaten kombiniert sind, gab es keine bekannten neuen Datensätze im Bereich der spektralen Bildgebung. Weiterhin ist bis dato kein Datensatz bekannt, bei dem spektrale Daten von dynamischen, strukturierten und unstrukturierten Umgebungen enthalten sind. Wird die Literatur zur hyperspektralen Klassifikation von terrestrischen Hyperspektral-Daten betrachtet, gibt es nur wenige Forschungsarbeiten, bei denen Daten nicht von einer Erdumlaufbahn oder einem Flugzeug aufgenommen wurden. Weiterhin sind öffentliche Datensätze, die mit spektraler Sensorik erfasst wurden, welche auf landgestützten Fahrzeugen montiert waren, nicht verfügbar. Um nun verschiedene Klassifikationsalgorithmen für die spektrale Datenverarbeitung und Szenenanalyse zu testen und zu entwickeln, sind entsprechende Trainings- und Testdaten unerlässlich.

Daher wurden die in dieser Arbeit genutzten Sensoren mit dem Snapshot-Mosaik-Sensor zusammen mit zusätzlichen Sensoren auf unterschiedlichen Fahrzeugen montiert, welche durch strukturierte und unstrukturierte Umgebungen fahren, um völlig neuartige Datensätze mit spektralen Sensordaten zu erstellen. Die beiden Kameras sind hardwareseitig synchronisiert, so sind entsprechende Sensorinformationen der aufgenommenen Szene vom sichtbaren bis in den Nahinfrarotbereich verfügbar. In Kombination mit dem ebenfalls verbauten 3-D-Laser ist eine Sensorfusion von spektralen und 3-D-Daten möglich. Dies ermöglicht theoretisch eine umfassendere Analyse ei-

ner Szene als bisher.

Der Hauptzweck dieser neuen Datensätze ist die Untersuchung der Eignung von spektralen Daten zum maschinellen Sehen, insbesondere für das Verständnis von Szenen im Rahmen des autonomen Fahrens. Ziel ist es, die Befahrbarkeitsanalyse und das autonome Fahren auch in Offroad-Szenarien zu verbessern. Die im Folgenden beschriebenen Arbeiten wurden in Teilen bereits auf einer internationalen Konferenz veröffentlicht [3]. Der folgende Abschnitt gibt einen Überblick über öffentlich verfügbare Datensätze aus verschiedenen Bereichen des maschinellen Sehens. Dabei wird eine Unterteilung in verschiedene Bereiche vorgenommen. Grundsätzlich wird im folgenden Abschnitt zwischen RGB und Hyperspektral unterschieden und im Bereich der spektralen Bildgebung zwischen satellitengestützten Daten und anderen Domänen.

5.2 STAND DER TECHNIK

5.2.1 Hyperspektrale Luftaufnahmen

Die Tabelle 5 gibt einen Überblick über bestehende Datensätze welche, auf mit Hyperspektraltechnik ausgerüsteten Satelliten oder Flugzeugen basieren. Einige Beispiele zu den Daten sind in Abbildung 46 zu sehen. Einer der ersten öffentlich verfügbaren Datensätze mit hy-

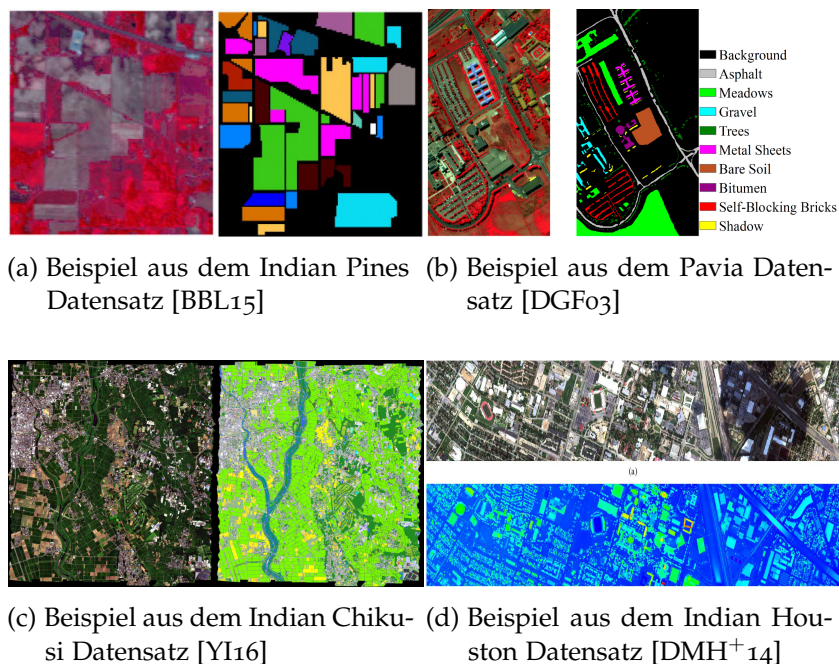


Abbildung 46: Einige Beispiele von Datensätzen mit hyperspektralen Daten, die aus der Erdumlaufbahn von Satelliten aufgenommen worden

Name	Sensorik	Klassen	Auflösung	Wellenlänge	Bänder
Indian Pines (1992) [BBL15]	AVIRIS	16	360 × 360	400-2500 nm	224
Cuprite (1995) [NAS18]	AVIRIS	5	420 × 360	400-2500 nm	224
Washington (1995) [oV18]	HYDICE	9	420 × 300	400-2500 nm	210
MoffettField (1997) [NAS18]	AVIRIS	3	512 × 614	400-2500 nm	224
Pavia (2003) [DGF03]	ROSIS-3	9	610 × 610	430-840 nm	115
Rodalquilar (2003) [BvdMvRo8]	HyMap	9	261 × 867	400-2500 nm	242
Houston (2012) [DMH ⁺ 14]	CASI	15	320 × 540	360-1005 nm	144
Chikusei (2014) [YI16]	Hyperspec	19	540 × 420	360-1002 nm	128

Tabelle 5: Übersicht verschiedener veröffentlichter hyperspektraler Bilddatensätze basierend auf Satellitendaten

perspektralen Daten stammt aus dem Jahr 1992. Dabei handelt es sich um Satellitenaufnahmen der landwirtschaftlichen Farm der Purdue University nordwestlich von West Lafayette und ist auch unter dem Begriff *Indian Pines* [BBL15] zu finden. Aufgenommen wurden die Daten mit dem (engl. *Airborne Visible InfraRed Imaging Spectrometer*) (AVIRIS)-System, welches Spektren im sichtbaren und infraroten Bereich misst. Die Daten wurden zur Unterstützung der Bodenerkundung erhoben. Die Daten bestehen aus 200 Bändern und die zugehörige Annotation besteht aus 16 Klassen. Weitere Datensätze [NAS18] basierend auf AVIRIS-Systemen wurden in den folgenden Jahren veröffentlicht. Im Jahr 2003 wurde mit dem ROSIS-Sensor in Italien in der Region Pavia ein Datensatz aufgenommen, welcher 115 Bänder umfasst und mit 9 Klassen annotiert ist.

5.2.2 Hyperspektrale Daten

Tabelle 6 zeigt eine Übersicht von Datensätzen mit hyperspektralen Daten, die nicht von Satelliten oder Flugzeugen stammen. Wiederum zeigt Abbildung 47 Beispieldaten zu den Datensätzen. Einer der ersten Datensätze mit hyperspektralen Daten, welche nicht von Flugzeugen oder Satelliten stammen, wurde 1995 veröffentlicht und bezieht sich auf die Erkennung von Vegetation in hyperspektralen Bildern, wie es von Brelstaff et al. [BPTC95] gezeigt wird. Foster et al. [FANFo6] veröffentlichte 2006 einen Datensatz mit 50 hyperspektralen Bildern von natürlichen Szenen zur Schätzung metamischer Oberflächen. Diese erscheinen dem Auge unter einer Beleuchtung gleich, werden aber unter einer anderen Beleuchtung unterschiedlich wahrgenommen. Jahre später veröffentlichte Foster et al. [FAN16] einen weiteren kleinen Datensatz, in dem Zeitraffer-Hyperspektralaufnahmen von fünf statischen Szenen mit und ohne Vegetation zu verschiedenen Tageszeitpunkten gemacht wurden. Hierbei wurden Beleuchtungsschwankungen bei Farbmessungen untersucht.

Yasuma et al. [YMIN10] präsentiert ein Framework für die Entwicklung von Kameras, die gleichzeitig einen erweiterten Dynamikbe-



(a) Beispielbild aus dem Brelstaff-Datensatz [BPTC95] (b) Beispielbild aus dem Foster-Datensatz [FANFo6]



(c) Beispielbild aus dem Nguyen-Datensatz [NPB14] (d) Beispielbild aus dem Khan-Datensatz [KMM⁺18]

Abbildung 47: Beispiele aus verschiedenen hyperspektralen Datensätzen, die nicht aus der Erdumlaufbahn aufgenommen wurden

reich und höhere spektrale Auflösungen erfassen können. Sie nahmen 31 hyperspektrale Bilder mit einem Spektrum von 400 – 700 nm von mehreren statischen Szenen mit einer Vielzahl von Materialien auf. Als Sensor wurde dazu ein abstimmbarer Filter in Kombination mit einer gekühlten CCD-Kamera verwendet. Eine umfangreichere Sammlung von 50 hyperspektralen Bildern von Innen- und Außenaufnahmen, die unter Tageslichtbeleuchtung aufgenommen wurden und 25 Bildern, die unter künstlicher und gemischter Beleuchtung aufgenommen wurden, wurde 2011 von Chakrabarti und Zickler [CZ11] veröffentlicht.

Im Jahr 2012 schlug Namin et al. [NP12] ein automatisches System zur Materialklassifizierung in natürlichen Umgebungen vor, bei dem multispektrale Bilder, bestehend aus sechs visuellen und einem nahen Infrarotband, genutzt wurden. Skauli et al. [SF13] publizierte im Jahr 2013 einen Datensatz mit hyperspektralen Daten, welcher Gesichter, statische Landschaften und Gebäude zeigt und dabei Wellenlängen von 400 – 2000 nm abdeckt. Dies erweitert die hyperspektrale Gesichtsdatenbank, welche im Spektrum des sichtbaren Lichts erfasst und von Wei Di et al. [DZZP10] veröffentlicht wurde. Die Segmen-

Name	Typ	Bilder	Auflösung	Wellenlänge	Bänder
Brelstaff(1995) [BPTC95]	Vegetation	29	256 × 256	400–700 nm	31
Brainard (1998) [VHF ⁺ 97]	Indoor	9	2000 × 2000	400-700 nm	31
Nascimento (2002) [NFFo2]	Vegetation, Urban	30	800 × 800	400-720 nm	33
Hordley (2004) [HFM04]	Lichtbox	22	N/A	400-700 nm	31
Foster (2006) [FANFo6]	Urban, Suburban	50	1344 × 1024	400-720 nm	33
Di (2010) [DZZP10]	Gesichter	25	N/A	400-720 nm	33
Yasuma (2010) [YMIN10]	Indoor, Materialien	32	512 × 512	400-700 nm	31
Chakrabarti (2011) [CZ11]	Indoor, Outdoor	50	1392 × 1040	420-720 nm	33
Skauli (2013) [SF13]	Gesichter, Outdoor	106	N/A	400-2500 nm	N/A
Nguyen (2014) [NPB14]	Indoor, Outdoor	64	N/A	400-1000 nm	31
Hirvonen (2014) [HOP ⁺ 14]	Holz	107	N/A	400-2500 nm	440
Eckhard (2015) [EEV ⁺ 15]	Outdoor	14	1392 × 1040	400-1000 nm	61
LeMoan (2015) [MGP ⁺ 15]	Indoor, Lichtbox	9	500 × 500	400-1000 nm	160
Nascimento (2016) [NAF16]	Outdoor, Vegetation	30	1344 × 1024	400-720 nm	33
Foster (2016) [FAN16]	Outdoor, Vegetation	33	1344 × 1024	400-720 nm	33
Arad (2016) [ABS16]	Vegetation, Urban	100	1392 × 1300	400-1000 nm	519
Noviyanto (2017)[NA17]	Honig	32	520 × 696	400-1000 nm	126
Zacharopoulos (2018) [ZHK ⁺ 18]	Gemälde	23	2560 × 2048	360-1150 nm	23
Nouri (2018) [NGV ⁺ 18]	Baumblätter	N/A	N/A	960-2490 nm	256
Mirashemi (2018) [Mir18]	Textilien	60	640 × 640	400-780 nm	39
Khan (2018) [KMM ⁺ 18]	div. Materialien	112	1024 × 1024	400-1000 nm	186

Tabelle 6: Überblick verschiedener bestehender hyperspektralen Datensätze, die nicht auf satellitengestützter Sensorik beruhen.

tierung von unterschiedlichen Materialien in hyperspektralen Daten wurde 2016 von Yu et al. [ZHHN16] weiter untersucht. Dort wurde eine per-Pixel Materialklassifikation genutzt, gefolgt von einer Superpixel-basierten Nachverarbeitung. Für die Auswertung kombinierten sie 51 hoch- und querformatige hyperspektrale Bilder aus mehreren der zuvor genannten Datensätze [FANFo6, YMIN10, CZ11, SF13]. Diese wurden jedoch zuvor weiter verarbeitet, um eine übergreifende Datenbasis im Bereich von 430 – 700 nm zu ermöglichen.

Bezogen auf die Snapshot-Mosaik-Technik gibt es derzeit nur zwei Datensätze, welche die in dieser Arbeit verwendeten Sensoren nutzen. Fotiadou et al. [FTT17] nutzen neuronale Netze zur Klassifikation dieser Daten. Sie schlagen eine Architektur vor, welche Methoden des Deep-Learning für eine effiziente Merkmalsextraktion nutzt und spektrale und räumliche Informationen von spektralen Szenen zusammen kodiert. Sie konzentrieren sich dabei auf die Klassifikation von spektralen Momentaufnahmen und bauen einen neuen spektralen Klassifikationsdatensatz bestehend aus Innenszenen mit 90 Bildern auf. Die Daten wurden unter verschiedenen Beleuchtungen und Blickwinkeln aufgenommen und zeigen zehn verschiedene Objekte wie Bananen, Gläser und andere Dinge.

Die Kombination von RGB- und Hyperspektraldaten wurde 2016 von Cavigelli et al. [CBMB16] auf Daten mit statischem Hintergrund und

Name	Topic	Bilder	Klassen	Sensors
MSRC (2009) [SWRC09]	Stadt	591	21	RGB
CamVid (2008) [BFC09, BSFC08]	Stadt	700	32	RGB
PascalVOC (2010) [EVGW ⁺ 10]	Stadt	11530	20	RGB
Kitti (2012) [GLU12, AAMM ⁺ 18]	Stadt	400	32	RGB (Stereo), Mono (Stereo), Laser, RTK-GPS
Ladický (2012) [LSR ⁺ 12]	Stadt	70	7	RGB (Stereo)
Daimler (2013) [SEFR13]	Stadt	500	5	RGB (Stereo)
Microsoft COCO (2014) [LMB ⁺ 14]	Divers	165482	92	RGB
Cityscapes (2016) [COR ⁺ 16]	Stadt	25000	30	RGB (Stereo)
SYNTHIA (2016) [RSM ⁺ 16]	Stadt	200000	13	RGB (Synthetisch)
ADE20K (2016) [ZZP ⁺ 16, ZZP ⁺ 17]	Indoor, Outdoor	20000	150	RGB
Mapillary Vistas (2017) [NORBK17]	Urban, Suburban	25000	66	RGB
BDD100K (2018) [YXC ⁺ 18]	Stadt	10000	40	RGB
IDD (2018) [VSN ⁺ 19]	Urban, Suburban	10000	34	RGB

Tabelle 7: Zusammenfassung verschiedener bestehender RGB-Bilddatensätze zur semantischen Szenenanalyse

einem kleinen Datensatz von 40 Bildern unter Verwendung von tiefen neuronalen Netzen ausgewertet.

5.2.3 RGB-Datensätze

Hier zeigt Tabelle 7 eine Übersicht verschiedener Datensätze, die RGB-Daten beinhalten und im Allgemeinen zur semantischen Segmentierung genutzt werden. Einige Beispiele aus den Datensätzen zeigt Abbildung 48.

Im Bereich der Bildverarbeitung mit RGB-Kameras existieren inzwischen einige etablierte und umfangreiche Datensätze. Eine der ersten Bilddatensätze zur semantischen Segmentierung war die „Cambridge-driving Labeled Video Database“ (CamVid) [BSFC08], welche 2008 veröffentlicht wurde. Dort sind ca. 700 Bilder aus einer Videosequenz von 10 Minuten annotiert. Die Kamera wurde dazu auf dem Armaturenbrett eines Autos aufgestellt und erfasste somit ein ähnliches Sichtfeld wie das des Fahrers. Schon wesentlich größer ist der 2010 veröffentlichte „PascalVOC“ [EVGW⁺10] Datensatz. Dieser besteht aus über 10 000 Bildern mit über 20 Klassen und war essentieller Bestandteil der „PASCAL Visual Object Classes (VOC) Challenge“, welche 2012 durchgeführt wurde. Der KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) Bilddatensatz [GLU12] wurde 2012 veröffentlicht. Hier sind unterschiedliche Sensoren wie Graustufen- und RGB-Kameras, 3-D-Laserscanner sowie GPS- und IMU-Einheiten auf einem Auto montiert. Der Datensatz wurde zunächst ohne semantische Annotationen veröffentlicht. Jedoch haben einige Forschungsgruppen nachträglich einige Bilddaten mit semantischen Annotationen versehen [AGLL12]. Ein weite-



(a) Beispielbild aus dem CamVid- (b) Beispielbild vom PascalVOC-
Datensatz [BFC09] Datensatz [EVGW⁺10]



(c) Beispielbild aus dem Kitti- (d) Beispiel aus dem Cityscape-
Datensatz [GLU12] Datensatz [COR⁺16]

Abbildung 48: Beispiele von Datensätzen mit Farbdaten zur semantischen Szenenanalyse

rer Benchmark-Datensatz, bestehend aus 500 Stereobildpaaren, wurde von Scharwächter et al. [SEFR13] veröffentlicht und trägt den Titel „Daimler Urban Segmentation Dataset“ (DUS). Der Datensatz liefert Annotationen, die fünf Objektklassen enthalten. Dieser Datensatz ist Teil einer Forschungsinitiative namens 6D-Vision des Automobilherstellers Daimler. Im Jahr 2016 wurde quasi eine Erweiterung dieses Datensatzes unter dem Titel „Cityscapes“ [COR⁺16] veröffentlicht. Der Datensatz basiert, wie der DUS auf Kameras, welche hinter der Windschutzscheibe montiert sind. Hier sind 30 Klassen auf 8 übergeordnete Kategorien verteilt.

Viele Deep-Learning Ansätze nutzen diesen Datensatz zurzeit als Benchmark. Mit fünfmal so vielen annotierten Bildern wurde 2017 der Mapillary Vistas Datensatz [NORBK17] veröffentlicht. Die Besonderheit hier neben der Größe des Datensatzes ist, dass die Bilder von verschiedenen Kameras stammen. Es wurden Bilder von Mobiltelefonen, Tablets, Action-Kameras und anderen Kameratypen annotiert und veröffentlicht. Derzeit ist es einer der größten und vielfältigsten offenen Datensätze mit einer geografischen Reichweite über mehrere Kontinente. Der aktuellste Datensatz zur semantischen Szenenanalyse wurde erst kürzlich im Jahr 2018 unter dem Titel „India Driving Dataset (IDD)“ [VSN⁺19] veröffentlicht. Während bspw. der Cityscapes Datensatz Straßen einer gut ausgebauten Infrastruktur wie Fahrspuren, einer kleinen Anzahl von klar definierten Kategorien für Verkehrsteilnehmer geringer Variation in der Objekt- oder Hintergrund-

gestaltung zeigt, enthält der IDD Straßenszenen in unstrukturierten Umgebungen, in denen die Infrastruktur generell schlechter ist.

5.2.4 Zusammenfassung

Die autonome Navigation entwickelt sich schnell von einem Forschungsgebiet zu einem Produkt, welches von großen Automobilherstellern unter hohem Aufwand erforscht wird. Ein wesentlicher Aspekt bei den Fortschritten der letzten Jahre liegt in der Verfügbarkeit immer größerer annotierter Datensätze. Noch vor wenigen Jahren waren Datensätze mit einigen Hundert annotierten Beispieldaten ausreichend für viele Probleme und Klassifikatoren. Jedoch hat die Einführung von größeren Datensätzen mit vielen Tausend Beispieldaten in Kombination mit der Etablierung von neuronalen Netzen zu aufsehenerregenden Durchbrüchen in vielen Bereichen der Bildverarbeitung geführt. Daher werden weltweit immer größere Bemühungen in die Entwicklung und Erstellung von umfangreichen Datensätzen gesteckt. Denn hochqualitative Datensätze spielen auf zwei Arten eine Schlüsselrolle im Fortschritt vieler Forschungsbereiche. Zum einen können so immer komplexere Klassifikatoren und neuronale Netze effektiv trainiert werden. Und zum anderen können aktuelle Fortschritte von der Forschungsgemeinschaft leichter identifiziert werden. So sind in den letzten Jahren mehrere Datensätze für die autonome Navigation verfügbar geworden, die sich jedoch auf strukturierte Fahrumgebungen und etablierte Sensorik konzentrieren.

5.3 EIGENER DATENSATZ

Soweit bekannt gibt es aktuell keinen öffentlich zugänglichen Datensatz mit annotierten spektralen Daten, welche mit den in dieser Arbeit eingesetzten Sensoren aufgenommen wurden und welche zusätzlich auf einem bewegten Fahrzeug montiert sind und dynamische Fahrscenarien erfassen. Daher wurden im Rahmen dieser Arbeit entsprechende Datensätze aufgebaut und öffentlich zur Verfügung gestellt, damit eine breite Öffentlichkeit die Verwendung von spektralen Daten in dynamischen, strukturierten und unstrukturierten Umgebungen untersuchen kann und die Forschungsergebnisse zu diesem neuen Gebiet vergleichbar sind.

Aktuell existieren mehrere spektrale Datensätze, welche frei verfügbar auf einer dafür eingerichteten Website zur Verfügung gestellt werden. Die Website bietet die spektralen Daten sowie Quellcode zum Laden und detailliertere technische Informationen. Die Tabelle 8 gibt eine Übersicht über die aktuell verfügbaren Datensätze mit annotierten spektralen Daten in Form von Hyperwürfeln. Die Benennung der Datensätze setzt sich aus dem Aufnahmedatum, dem Szenario und der enthaltenen Kameradaten zusammen. Die verschiedenen Datensätze

Name	Typ	Hyperwürfel	Klassen	Kanäle
2016Vis	Urban & Suburban	133	11	16 (VIS)
2017Nir	Urban & Suburban	90	11	25 (NIR)
CityNir	Urban	128	12	25 (NIR)
LandNir	Suburban	306	12	25 (NIR)
NirFull	Urban & Suburban	434	12	25 (NIR)
CityVis	Urban	114	12	16 (VIS)
LandVis	Suburban	308	12	16 (VIS)
VisFull	Urban & Suburban	403	12	16 (VIS)

Tabelle 8: Übersicht über verfügbare Datensätze welche spektralen Daten enthalten

aus 2018 bestehen eigentlich aus zwei Datensätzen, welche je nach Domäne (Urban/Suburban) unterteilt sind; der Zusammenhang ist in Abbildung 49 dargestellt. So kann untersucht werden, ob ein Klassifikator für eine spezielle Domäne bessere Ergebnisse liefert oder die Performance beim Training über den gesamten Datensatz besser ist.

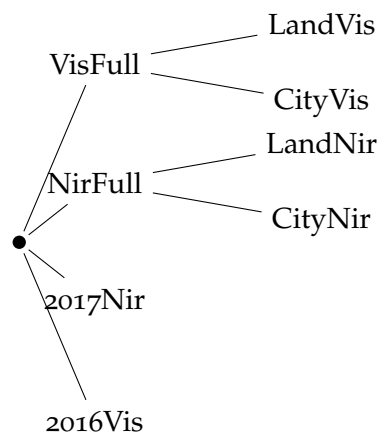
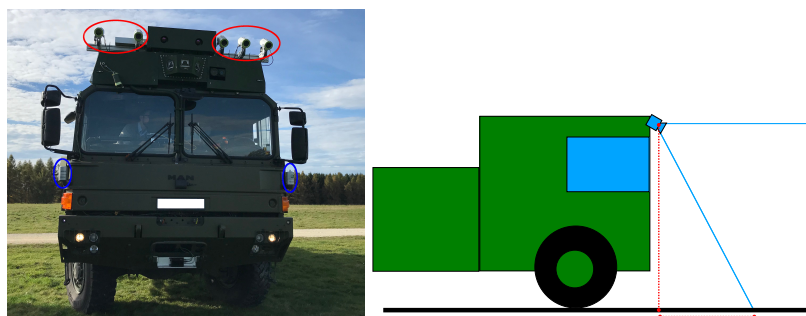


Abbildung 49: Der Graph stellt die Aufteilung und den Zusammenhang der einzelnen Datensätze dar.

Die Rohdaten der Datensätze sind verfügbar als *rosbags*¹, welche mit Hilfe der Middleware (engl. *Robot Operating System*) (ROS) [QGC⁺09] aufgezeichnet wurden. ROS ist ein Framework zur gezielten Ansteuerung von Robotikhardware und bietet verschiedene Module zur Hardwareabstraktion und Low-Level-Gerätsteuerung. Zusätzlich zu den Rohdaten sind annotierte Hyperwürfel verfügbar, die aus den *rosbags* in Form von Matlab-Dateien extrahiert und vorverarbeitet wurden, wie es in Kapitel 5.4.2 beschrieben wird. Die zur Erstellung

¹ <http://wiki.ros.org/Bags>



(a) LKW mit montierter Sensorik (b) Montageschema der Kamera auf dem LKW

Abbildung 50: Anordnung der Sensorik auf einem Versuchsträger

der Datensätze verwendete Sensorik setzt sich im Wesentlichen aus den im Kapitel 3.3 vorgestellten Sensoren zusammen.

5.4 DATENAUFNAHME



(a) Beispiel für Rohbilddaten, die von der VIS-Kamera aufgenommen wurden (b) Beispiel für Rohbilddaten, die von der NIR-Kamera aufgenommen wurden

Abbildung 51: Rohdaten der spektralen Kameras aus den veröffentlichten Datensätzen

Die Kameras wurden mit Hilfe des in Kapitel 3.3 beschriebenen Hardware-Triggers synchronisiert. So war es möglich, gleichzeitig jeweils eine Aufnahme von der VIS und der NIR-Kamera zu erhalten. Dies hat den Vorteil, dass so Informationen in einem Wellenlängenbereich von 400 – 975 nm für eine Szene abgerufen werden können. Weiterhin kann so im Folgenden auch evaluiert werden, ob VIS oder NIR Daten sich besser zur semantischen Analyse eignen.

Die Sensorik war wie in Abbildung 50 gezeigt auf dem Versuchsfahrzeug montiert. Die Blickrichtung der Kameras war in Richtung der Fahrbahn, sodass der vor dem Fahrzeug liegende Bereich abgedeckt wurde. Während der Aufnahme wurden die Daten in Form von *rosbags* auf dem Datenträger gespeichert, welche wie in Darstel-

```

path:      2018-06-28-14-08-14_8.bag
version:   2.0
duration:  60.0s
start:     Jun 28 2018 14:08:14.62 (1530187694.62)
end:       Jun 28 2018 14:09:14.57 (1530187754.57)
size:      4.1 GB
messages:  2400
compression: none [2392/2392 chunks]

types:     sensor_msgs/Image      [060021388200f6f0f447d0fcd9c64743]
           sensor_msgs/PointCloud2 [1158d486dd51d683ce2f1be655c3c181]

topics:    /M0022_nir_left/image    600 msgs   : sensor_msgs/Image
           /M0022_vis_left/image   600 msgs   : sensor_msgs/Image
           /velodyne_points1       600 msgs   : sensor_msgs/PointCloud2
           /velodyne_points2       600 msgs   : sensor_msgs/PointCloud2

```

Code 5.1: Beispiel zum Aufbau eines *rosbags* aus einem der Datensätze.

lung 5.1 aufgebaut sind. Die Daten jedes der Sensoren werden im ROS-System verteilt und in den *rosbags* serialisiert. Werden die serialisierten Daten der *rosbags* wieder ins System eingespielt, werden alle gespeicherten Daten neu im System verteilt. In diesem *rosbag* sind bei einer Laufzeit von 60 Sekunden jeweils 600 Nachrichten von zwei Kameras und zwei 3-D-Lasern gespeichert.

5.4.1 Vorverarbeitung

Während der Aufnahme liefern die in dieser Arbeit verwendeten Kameras Daten in einem verlustfreien Format mit 8 Bit, welches linear zu den Signalen ist, je Pixel; Beispiele sind in Abbildung 51 zu sehen. Die von der Kamera erfassten Rohdaten bedürfen daher einer speziellen Vorverarbeitung, um diese weiter Nutzen zu können. Um einen Hyperwürfel mit spektralen Werten aus den Rohdaten zu konstruieren, wird eine entsprechende Vorverarbeitung, wie in Kapitel 4 erläutert, angewandt.

5.4.2 Datenstruktur

Die extrahierten und rekonstruierten Daten in Form von Hyperwürfel werden als MATLAB Level 5 Mat-Dateien gespeichert. Jede Datei enthält genau eine Aufnahme einer Kamera mit der Datenstruktur wie in Tabelle 9 angegeben.

5.5 DATENEXTRAKTION

Zur Extraktion von Hyperwürfel zur weiteren Verarbeitung wurden die Daten aus den bei der Datenaufnahme gespeicherten *rosbags* extrahiert. Dazu wurde alle vier Sekunden von jeder Kamera ein vorver-

Bezeichnung	Erläuterung
image	Speichert die Rohdaten so, wie es von der Kamera aufgenommen wurden
data	Enthält die vorverarbeiteten spektralen Daten in Form eines Hyperwürfel, wie in Kapitel 4 beschrieben
label_*	Enthält die Annotationen in Form einer Maske, die den Daten während der Annotation zugewiesen wurden
wavelengths	Enthält eine Liste mit den primären Wellenlängenempfindlichkeiten für jedes Band im Hyperwürfel, so wie vom Hersteller in der Spezifikation angegeben

Tabelle 9: Datenstruktur eines Hyperwürfels

arbeiteter Hyperwürfel extrahiert. Die Zeitspanne von vier Sekunden wurde gewählt damit die Inhalte der jeweiligen Aufnahmen nicht zu ähnlich sind.

Da die Sensorik auf Fahrzeugen montiert ist, hat die Sonneneinstrahlung direkten Einfluss auf die Qualität der Datenaufnahme. Dies ist ein großer Unterschied zur Datenaufnahme in kontrollierten Szenarien, wo die Beleuchtung definiert werden kann. Daher muss sich bei der Extraktion mit Beleuchtungsänderungen und direkter Sonneneinstrahlung auf dem Sensor auseinandergesetzt werden, da die gemessenen Daten dadurch verzerrt werden können. Zwei Beispiele von solchen Aufnahmen sind in Abbildung 52 angeführt. Hier ist die Beleuchtung des Sensors zu stark bzw. zu schwach. Diese Daten sind nicht nutzbar, da das Verhalten der Sensorik laut Spezifikation des Herstellers und der Quanteneffizienz des Sensors nur bis 80 % des Maximalausschlags linear ist. Um fehlerhafte Hyperwürfel auszusortieren, wurden die Daten nach der Extraktion gefiltert und alle Hyperwürfel entfernt, bei denen mehr als 20 % der Pixel über- oder unterbelichtet sind. Die noch verbliebenen Hyperwürfel wurden dann zur Annotation verwendet.

Definition 18: Intensitätswertfilter

Fehlerhafte Hyperwürfel werden mit folgender Filterregel aussortiert:

$$f(i,j) \geq 204 \rightarrow \text{Pixel überbelichtet} \quad (39)$$

$$f(i,j) \leq 46 \rightarrow \text{Pixel unterbelichtet} \quad (40)$$

Liegen mehr als 20 % der Intensitätswerte in einem Hyperwürfel über oder unter den obigen Schwellwerten wird der Hyperwürfel verworfen.

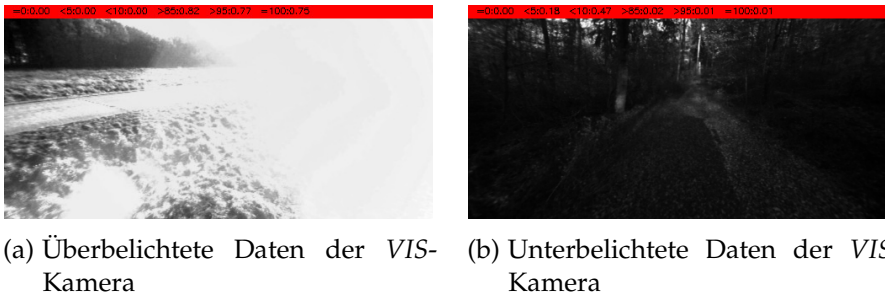


Abbildung 52: Beispiele für das Filtern von über- und unterbelichteten Daten, die mit der VIS-Kamera aufgenommen wurden

5.5.1 Beispiele

Die Abbildung 53 zeigt Plots von spektralen Werten, die mit Hilfe der Kameras aufgenommen wurden. Aufgenommen wurden mit beiden Kameras jeweils verschiedene Oberflächen wie Himmel, Straße und Vegetation. Es ist zu erkennen, dass die von den Kameras aufge-

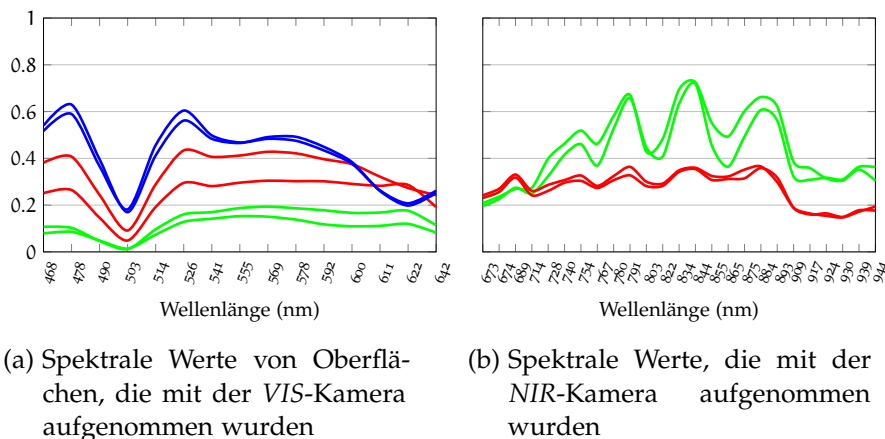


Abbildung 53: Plots der spektralen Werte von NIR-Kamera und VIS-Kamera für verschiedene Oberflächen. Himmel (blau), Straße (rot) und Vegetation (grün).

nommenen Daten für jede Oberfläche individuelle *spektrale Signaturen* zeigen. Speziell ist auch der sog. Chlorophyll-Peak der Vegetation bei 700 nm gut zu erkennen. Hier ist ein sprunghafter Anstieg der reflektierten Strahlung zu sehen. Auch sind grundsätzliche Unterschiede in den Kurven für Himmel, Straße und Vegetation in den Daten der VIS-Kamera zu erkennen. Dieses Verhalten erlaubt grundsätzlich eine Trennung der Elemente aufgrund ihrer spektralen Eigenschaften und damit auch ein gezieltes Training eines Klassifikators.

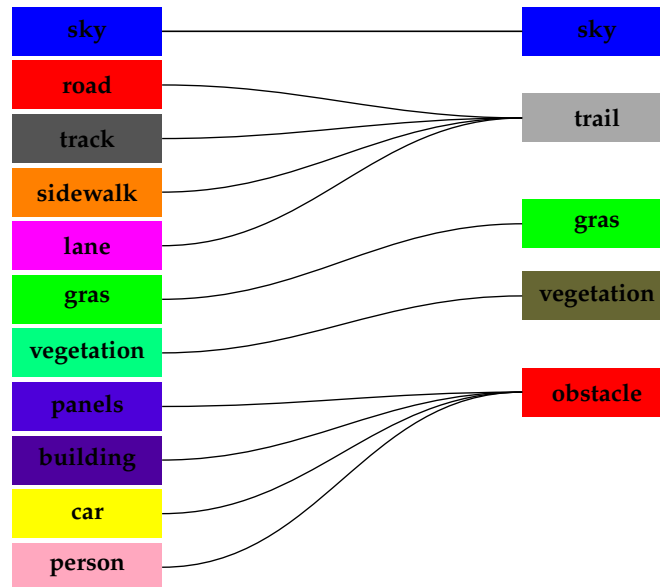
5.5.2 Annotation

Da bisher, aufgrund der Neuartigkeit der Technologie, keine Annotationssoftware existiert, die in der Lage ist, die spektralen Daten korrekt zu verarbeiten, wurde ein eigenes Tool zur Annotation der Hyperwürfel entwickelt. Durch dieses Tool kann jedem einzelnen Hyperpixel des Hyperwürfels eine definierte Klasse zugewiesen werden, was anschließend zum Training von Klassifikatoren genutzt werden kann. Die erforderlichen Annotationen zur Erstellung einer Grundwahrheit wurden von Hand durch geschultes Personal für alle extrahierten Hyperwürfel jedes Datensatzes vorgenommen.

Während der Annotation wurden nicht allen Hyperpixeln eines Hyperwürfels eine Klasse zugeordnet. Dies liegt daran, dass Grenzbereiche zwischen Materialien nicht eindeutig zuordenbar sind. Für die Annotation wurden zwei unterschiedliche Gruppen von Annotationskategorien eingeführt. Die Gruppen sind in Abbildung 54 dargestellt. Diese sind von den Annotationen anderer Datensätze aus den Bereichen der spektralen Bildverarbeitung und semantischen Szenenanalyse inspiriert. Grundsätzlich gibt es die Kategorien *Semantik* und *OffRoad*. Wobei die Kategorie *OffRoad* sich aus der *Semantik* ableitet, also quasi eine Zusammenfassung mehrerer Klassen darstellt, wie in Grafik Abbildung 54b dargestellt. Es wurde während der Annotation festgestellt, dass es durchaus auch Mehrdeutigkeiten zwischen verschiedenen Klassen/Labeln gibt und diese sich auch optisch nicht immer sauber trennen lassen. Daher wurde zusätzlich eine Klasse *undefined* eingeführt. Entsprechend den oben genannten Vorgaben wurden mehrere Hundert Hyperwürfel in verschiedene Datensätze unterteilt und annotiert. Die Kategorie *Semantik* besteht aus 11 bzw. 12 Klassen. Im Bereich des Straßenverkehrs wurden Klassen wie Straßen, Bürgersteige und Fahrbahnmarkierungen, sowie Schilder, Fahrzeuge und Passanten definiert. Ergänzt wird dies durch Gebäude, Gras und eine Klasse für die übrige Vegetation, wie Bäume oder Sträucher. Bei den Datensätzen aus dem Jahr 2018 wurde noch eine zusätzliche Klasse *Feldweg* eingeführt. Mit dieser Klasse wurden nicht asphaltierte Straßen wie Feldwege oder Schotterwege annotiert, welche in den vorherigen Datensätzen auch mit der Klasse *Straße* annotiert wurden. Das Ziel dieses Vorgehens war, eine feinere Trennung der Materialien zu ermöglichen und damit dem Klassifikator eine sauberere Datenbasis zu liefern. Wichtig ist wie bereits angesprochen auch die Klasse *undefined*, da oft nicht exakt für jedes Pixel entscheidbar ist, was darauf zu sehen ist, speziell auch in großen Entfernungen. Auch Randbereiche zwischen einzelnen Klassen, wie zwischen Gras und Feldweg, sind schwer zu unterscheiden, daher wurden solche Regionen nicht mit einer Klasse annotiert. In einer als *offRoad* bezeichneten Annotationsgruppe wurden die semantischen Klassen zusammengefasst. Wie die einzelnen Klassen zusammengefasst wurden, ist



(a) Kategorien der semantischen Annotation



(b) Kombination der semantischen Klassen zu OffRoad-Klassen























Abbildung 54: Für die Datensätze eingeführte Annotationsgruppen

in Abbildung 54b zu sehen. In einer Klasse wurden z. B. alle grundsätzlich befahrbaren Oberflächen zusammengefasst, wozu z. B. auch der Bürgersteig gehört. Weiterhin wurden alle Objekte, die Hindernisse darstellen, zu einer Klasse zusammengefasst. Durch die Reduzierung der Komplexität des Datensatzes ist diese Aufteilung eher zur Navigation außerhalb von urbanen Gegenden und auf unbefestigten Straßen geeignet. Denn die Bezeichnung der Klassen definieren implizit die Befahrbarkeit der jeweiligen Oberfläche. Gut befahrbare Bereiche sind solche, die von einem Fahrzeug befahren werden können. Darunter fallen neben Straßen auch befestigte Feldwege oder Parkflächen und auch Fußgängerwege. Als eher schlecht sind Wiesenflächen, Gehwege und Ackerflächen zu bezeichnen. Und nicht befahrbar sind solche Objekte und Bereiche in der Szene, die unter keinen Umständen befahren werden sollten. Dieser Klasse zuzuordnen sind andere Verkehrsteilnehmer und Passanten, Gebäude und Bäume. In beiden Kategorien vertreten ist die Klasse sky und diese ist natürlich eine Besonderheit. Denn sie ist zwar nicht befahrbar, aber natürlich auch kein Hindernis im klassischen Sinne.

5.5.3 Datenanalyse

Zur Auswertung der annotierten Daten wurde die Verteilung der einzelnen Klassen zur semantischen Annotation bestimmt, welche in Tabelle 10 und Tabelle 11 tabellarisch dargestellt sind. Eine graphische Darstellung der Verteilung der Klassen ist im Anhang Kapitel A zu finden. In der ersten Spalte der Tabelle ist die Farbe dargestellt, welche zur Visualisierung der jeweiligen Klasse genutzt wird. Die Spalte mit Titel *Anzahl* bezeichnet die genaue Zahl von Pixeln, welche mit der spezifischen Klasse annotiert ist. Analog beschreibt die mit *Anteil* überschriebene Spalte die relative Anzahl bezüglich aller vorhandenen Pixel. Des Weiteren sind Beispiele der erstellten Annotationen in Abbildung 55 zu sehen. Wird nun die Verteilung der jeweiligen Klassen betrachtet, ergeben sich einige diskutabile Aspekte.

Die Kameras sind frontal entlang der Straßenführung ausgerichtet, wie in Abbildung 50b dargestellt. Daher hat die Klasse *road* bzw. *track* erwartungsgemäß den höchsten Anteil an den Annotationen. Dies gilt durchweg für alle hier vorgestellten Datensätze; der Anteil schwankt zwischen knapp 30% im Datensatz 2016Vis und 46% in 2017Nir. Vor allem Werte über 40% sind kritisch. Denn eine so dominante Klasse kann zu Problemen beim späteren Training von Klassifikatoren führen. Weiterhin gibt es noch eine Besonderheit zwischen Datensätzen, die in den Jahren 2016 – 2017 und von denen, die im Jahr 2018 aufgenommen wurden. In den Jahren 2016 und 2017 hatten die Kameras eine etwas flachere Ausrichtung, sodass sie auch teilweise den Himmel aufgenommen haben. Daher ist der prozentuale Anteil der Straße am Datensatz entsprechend geringer. Im Jahr 2018 wurde der Winkel der Kameras verändert, um den blinden Bereich direkt vor dem Fahrzeug zu verkleinern und eine bessere Fusion mit den Laserdaten zu ermöglichen. Dies zeigt sich auch deutlich in der Verteilung der Klassen in den Datensätzen. Bei den entsprechenden Datensätzen liegt der Anteil der Klasse *Himmel* bei 0. Wird die Verteilung betrachtet, folgen auf *road* und *track* fast überall die Klassen *grass* und *vegetation* gefolgt von *sidewalk* und *building*. Wobei hier nochmal unterschieden werden muss zwischen Datensätzen die primär in der Stadt oder primär außerhalb der Stadt aufgezeichnet wurden. Entsprechend ist der Anteil von *grass* und *vegetation* natürlich höher bzw. niedriger. Einen sehr geringen Anteil am Datensatz macht die Klasse *person* aus. Dies liegt darin begründet, dass das Fahrzeug nicht durch Einkaufsmeilen und Fußgängerzonen gefahren ist. Abseits dieser Bereiche sind wenig Passanten unterwegs. Dieser Nachteil muss in Zukunft korrigiert werden und die Fahrrouten müssen entsprechend angepasst werden. Wird die Verteilung der zweiten Kategorie betrachtet, welche den Fokus auf Befahrbarkeit legt, so zeigt sich eine etwas ausgewogenere Verteilung der jeweiligen Klassen.

Farbe	Klasse	Anzahl	Anteil	Farbe	Klasse	Anzahl	Anteil
	undefined	2690572	14.52 %		undefined	404987	4.70 %
	road	5542273	29.92 %		road	3928854	45.56 %
	sidewalk	1046739	5.65 %		sidewalk	610520	7.08 %
	lane	187098	1.01 %		lane	105237	1.22 %
	gras	2332229	12.59 %		gras	1544335	17.91 %
	vegetation	2191965	11.83 %		vegetation	1278356	14.83 %
	panels	262804	1.42 %		panels	68196	0.79 %
	building	1438217	7.76 %		building	463402	5.37 %
	car	389275	2.10 %		car	154118	1.79 %
	person	5047	0.03 %		person	9981	0.12 %
	sky	2438001	13.16 %		sky	54716	0.63 %

(a) Datensatz: 2016Vis, 143 Hyperwürfel (b) Datensatz: 2017Nir, 90 Hyperwürfel

Tabelle 10: Tabellarische Übersicht über die Verteilung der Klassen aus der Kategorie Semantik innerhalb der jeweiligen Datensätze

5.6 FAZIT

In diesem Abschnitt wurden neuartige, synchronisierte und kalibrierte Datensätze zur Untersuchung von Problemen der autonomen Navigation in strukturierten und unstrukturierten Umgebungen vorgestellt und diskutiert. Im Gegensatz zu bestehenden Datensätzen der semantischen Segmentierung enthalten die hier diskutierten Datensätze spektrale Daten aus dynamischen Umgebungen, was erst durch die Verwendung der Snapshot-Mosaik-Technik möglich ist. Diese Datensätze setzen sich aus mehreren Stunden von Rohdaten zusammen, woraus anschließend mehrere Hundert Hyperwürfel extrahiert wurden. Dies stellt nicht nur eine neue Herausforderung für den Stand der Technik der semantischen Segmentierung dar, sondern ist zugleich ein Versuch, die spektrale Bildgebung aus anderen Domänen auf die Probleme des autonomen Fahrens zu übertragen. Neben den spektralen Informationen sind noch kalibrierte 3D-Punktwolken und teilweise Beleuchtungsinformationen eines Spektrometers in den Roh-Datensätzen verfügbar. Zusätzlich zu den extrahierten Hyperwürfeln werden Annotationen aus zwei Kategorien zur Verfügung gestellt, welche das Training von Klassifikatoren zur Szenenanalyse ermöglichen. Weiterhin wurden zu der Klassenverteilung der Annotationen Statistiken erstellt und ausgewertet und das vorhandene Ungleichgewicht diskutiert.

Farbe	Klasse	Anzahl	Anteil
■	undefined	1210018	7.8 %
■	track	9528	0.06 %
■	sidewalk	1450623	9.36 %
■	lane	269075	1.74 %
■	gras	1176500	7.59 %
■	vegetation	1277930	8.24 %
■	panels	45149	0.29 %
■	building	653493	4.22 %
■	car	337152	2.17 %
■	person	10054	0.06 %
■	sky	0	0.00 %
■	road	9063922	58.46 %

(a) Datensatz: CityNir, 178 Hyperwürfel

Farbe	Klasse	Anzahl	Anteil
■	undefined	364915	1.18 %
■	track	6353041	20.49 %
■	sidewalk	1906	0.01 %
■	lane	6427	0.02 %
■	gras	12702426	40.97 %
■	vegetation	5604057	18.07 %
■	panels	0	0.00 %
■	building	69	0.00 %
■	car	2951	0.01 %
■	person	0	0.00 %
■	sky	1416	0.00 %
■	road	5969680	19.25 %

(b) Datensatz: LandNir, 356 Hyperwürfel

Farbe	Klasse	Anzahl	Anteil
■	undefined	1560132	3.37 %
■	track	6362569	13.73 %
■	sidewalk	1422740	3.07 %
■	lane	271607	0.59 %
■	gras	13878926	29.95 %
■	vegetation	6872512	14.83 %
■	panels	45149	0.1 %
■	building	647672	1.4 %
■	car	333667	0.72 %
■	person	10054	0.02 %
■	sky	1416	0.00 %
■	road	14929692	32.22 %

(c) Datensatz: NirFull, 532 Hyperwürfel

Farbe	Klasse	Anzahl	Anteil
■	undefined	1547497	7.76 %
■	track	5408	0.03 %
■	sidewalk	1830723	9.18 %
■	lane	340707	1.71 %
■	gras	1606106	8.05 %
■	vegetation	1316167	6.60 %
■	panels	61246	0.31 %
■	building	782395	3.92 %
■	car	475377	2.38 %
■	person	15202	0.08 %
■	sky	0	0.00 %
■	road	11968332	59.99 %

(d) Datensatz: CityVis, 154 Hyperwürfel

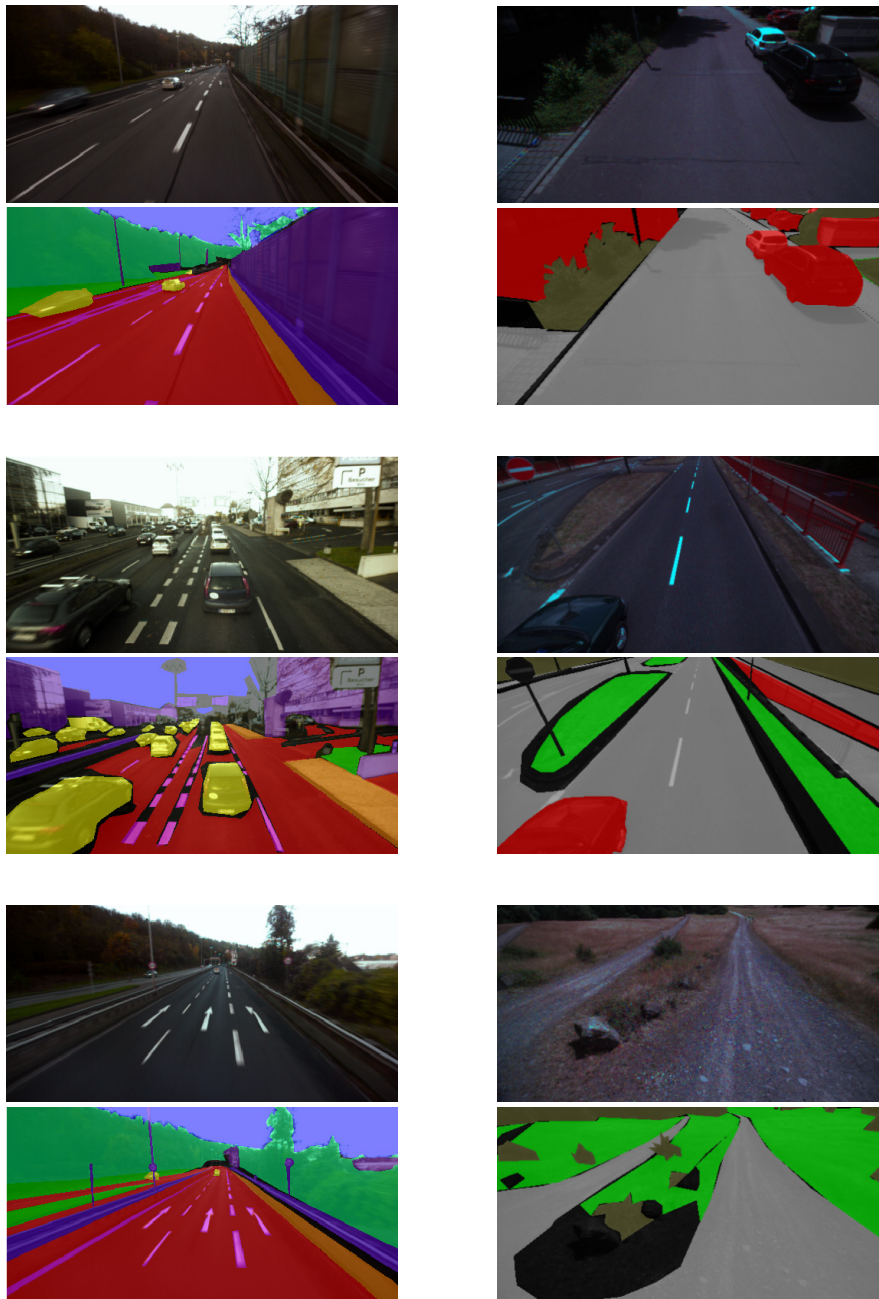
Farbe	Klasse	Anzahl	Anteil
■	undefined	743543	1.65 %
■	track	9312295	20.72 %
■	sidewalk	3302	0.01 %
■	lane	21176	0.05 %
■	gras	18140362	40.36 %
■	vegetation	8200451	18.24 %
■	panels	4547	0.01 %
■	building	94	0.00 %
■	car	3981	0.01 %
■	person	0	0.00 %
■	sky	1918	0.00 %
■	road	8518711	18.95 %

(e) Datensatz: LandVis, 347 Hyperwürfel

Farbe	Klasse	Anzahl	Anteil
■	undefined	2140609	3.30 %
■	track	9357066	14.45 %
■	sidewalk	1792005	2.77 %
■	lane	356979	0.55 %
■	gras	19816874	30.60 %
■	vegetation	9542664	14.73 %
■	panels	65793	0.10 %
■	building	777835	1.20 %
■	car	468897	0.72 %
■	person	15202	0.02 %
■	sky	1918	0.00 %
■	road	20434158	31.55 %

(f) Datensatz: VisFull, 500 Hyperwürfel

Tabelle 11: Tabellarische Übersicht über die Verteilung der Klassen aus der Kategorie Semantik innerhalb der jeweiligen Datensätze



Annotationen zur Semantik-Kategorie

Annotationen zur offRoad-Kategorie

Semantische Annotationen



OffRoad Annotationen



Abbildung 55: Beispiele von annotierten Daten aus dem veröffentlichten Datensatz

6.1 EINFÜHRUNG

Der Zustand der Vegetation auf unserem Planeten ist von substanzieller Bedeutung für das Leben auf der Erde. Drastische Veränderung der Erdvegetation haben unmittelbaren Einfluss auf die Umwelt und die Lebensqualität der Menschen. Daher begannen Forscher vor Jahrzehnten mithilfe von Satelliten und Sensoren die Vegetation zu kartieren und zu vermessen. Einer der ersten Sensoren dieser Art ist das *Advanced Very High Resolution Radiometer* (AVHRR) der *National Oceanic and Atmospheric Administration* (NOAA), welches mit 5 Kanälen Messungen vom sichtbaren Spektralbereich bis in den thermalen Infrarotbereich durchführen kann. Aus den Messungen des Lichts, welches von der Landoberfläche reflektiert wird, kann die Konzentration von Vegetation unter Verwendung eines optischen Indizes wie z. B. des normierten differenzierten Vegetationsindex (engl. *Normalized Difference Vegetation Index*) (NDVI) auf der ganzen Erde quantifiziert werden. Um die Dichte und den Zustand der Vegetation auf der Erdoberfläche zu bestimmen, müssen dazu die unterschiedlichen Wellenlängen in Bereichen des sichtbaren und nahinfraroten Lichtes gemessen werden. Die Reflexion des Lichts, welches auf ein unbewachsenes Stück Oberfläche trifft, ändert sich nur geringfügig zwischen dem sichtbaren und dem nahen Infrarotbereich. Chlorophyll, welches ein essentieller Bestandteil der Vegetation ist, absorbiert dagegen in erheblichem Maße sichtbares Licht im Bereich von 400 – 700 nm zum Zweck der Photosynthese. Licht aus dem Nahinfrarotbereich 700 – 1100 nm hingegen wird stark reflektiert. Daraus lässt sich die Regel ableiten, dass viel mehr reflektiertes Licht im Nahinfrarotbereich als im sichtbaren Bereich auf starke Vegetation schließen lässt. Ist der Anteil an reflektiertem Licht im sichtbaren und Nahinfrarotbereich ähnlich, dann ist in diesen Bereichen wahrscheinlich nur spärliche Vegetation vorhanden. Eine einfache Differenzbildung $DVI = NIR - VIS$ zwischen den Reflexionen im sichtbaren und nahen Infrarotbereich erlaubt also eine quantitative Messung von Vegetation [RJ72]. Der normierte differenzierter Vegetationsindex NDVI, ist ein Ergebnis der ersten Studien zur Fernerkundung und kann als optischer Index bezeichnet werden. Es gibt eine ganze Reihe von optischen Indizes, welche den Reflexionseffekt des Chlorophylls nutzen, um basierend auf Satellitendaten die Dichte der Vegetation auf der Erde zu bestimmen. Im Folgenden wird zunächst ein Überblick über Verschiedene dieser Indizes gegeben und anschließend wird unter-

sucht, ob sich diese Indizes auch auf Daten der Snapshot-Mosaik-Kameras anwenden und damit zur Umgebungswahrnehmung nutzen lassen.

6.2 STAND DER TECHNIK

Es existiert eine gute Literaturbasis für sog. optische Indizes, jedoch ist der Konsens, welche tatsächlich robust und tauglich sind, klein. Eine Studie, welche von Main et al. [MCM⁺11] durchgeführt wurde, betreibt hier Aufklärungsarbeit. In der Studie wurden 73 veröffentlichte Chlorophyll-Spektralindizes auf diversen Datensätzen getestet und basierend auf dem Vorhersagefehler (engl. *Root Mean Square Error*) (RMSE) bewertet.

Schon im Jahr 1965 beschäftigten sich Gates et al. [GKSW65] mit den spektralen Eigenschaften der Vegetation. Sie verglichen die spektralen Merkmale verschiedener Pflanzen zu verschiedenen Jahreszeiten miteinander und lieferten damit Grundlagenforschung zur Klassifizierung von Vegetation anhand ihres spektralen Reflexionsspektrums. Ein wichtiger Aspekt bei optischen Indizes, vor allem im Nahinfrarotbereich, ist die sog. rote Flanke (engl. *red edge*). Sie definiert einen Wertebereich mit schneller Veränderung der Reflexionseigenschaften der Vegetation im nahen Infrarotbereich des Spektrums [GGM03].

Das Chlorophyll, welches in der Vegetation enthalten ist, absorbiert das meiste Licht im sichtbaren Teil des Lichtspektrums, wie von Gates et al. beschrieben. Ab einer Wellenlänge oberhalb von 700 nm ändert sich das rapide. Der Reflexionsgrad des Lichtspektrums kann sich zwischen 680 nm und 730 nm von 5% auf 50% ändern. Die Erhöhung des Anteils an Chlorophyll und/oder Wasser in der Vegetation sorgt in der Regel dafür, dass sich die rote Flanke aufgrund der wachsenden Absorption auf längere Wellenlängen verschiebt. Die meisten der 73 untersuchten Indizes stammen aus den neunziger und Zweitausender Jahren, wobei der NDVI [RJ72] hier eine Ausnahme bildet und schon Anfang der Siebziger entwickelt wurde. Allerdings ist der Index sensitiv gegenüber atmosphärischen Effekten, was das Interesse an der Entwicklung alternativer Indizes geweckt hat. Eine Abwandlung des NDVI präsentierten Roujean und Breon 1995 mit dem *Renormalized Difference Vegetation Index* (RDVI) [RB95]. Dieser Index ist eine Kombination des (engl. *Difference Vegetation Index*) (DVI) und des (engl. *Renormalized Difference Vegetation Index*) (RDVI). Ein Jahr zuvor analysierten Carter et al. [Car94] den Zusammenhang von Pflanzengesundheit und spektralen Reflexionseigenschaften basierend auf schmalbandigen hyperspektralen Daten und veröffentlichten vier entsprechend gestaltete Indizes. Rondeaux et al. [RSB96] haben im Jahr 1996 die bekannten Schwächen des NDVI aufgegriffen und den Einfluss des Bodenuntergrunds auf die Performance der Indizes untersucht und daraus den *Optimized Soil-Adjusted Vegetation Indice* (OSA-

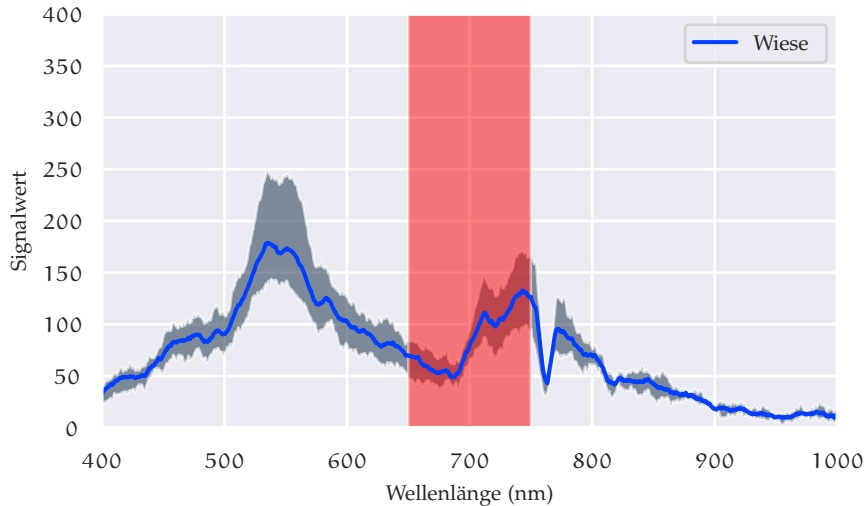


Abbildung 56: Reflektierte Spektralverteilung einer Wiese unter Sonneneinstrahlung, welche mit einem Spektrometer gemessen wurde. Der für die optischen Indizes relevante Bereich ist rot hervorgehoben.

VI) entwickelt.

Maccioni et al. zeigten 2001, dass das Verhältnis zwischen *Red Edge Position* (REP) und Chlorophyllgehalt nicht zwingend linear ist, sondern stark von der Art der Pflanze abhängt und bei einigen Arten stark variiert. Sie stellten einen im Gegensatz zum NDVI leicht veränderten Index vor. Im Jahr 2004 stellten Dash und Curran den (engl. *MERIS Terrestrial Chlorophyll Index*) (MTCI) vor, welcher speziell auf Daten des *Medium Resolution Imaging Spectrometer* (MERIS) entwickelt wurde und seit dem von der europäischen Raumfahrtagentur genutzt wird.

Cho et al. [CS06] beschrieben im Jahr 2006 eine neue Technik zur Extraktion der Position des Knickpunktes im roten Randbereich (engl. *Red Edge Position*) (REP) der spektralen Reflexionssignatur aus hyperspektralen Daten, was auch zur Abschätzung des Blattchlorophyll- oder Stickstoffgehalts verwendet wird. Diese Technik zeigte sich bei der Studie auf einigen Datensätzen den anderen überlegen und lieferte insgesamt die besten Ergebnisse. Dicht darauf folgt der 16 Jahre ältere *mREIP* Index von Miller et al. [MHW90]. Sie nutzen ebenso den Knickpunkt im roten Randbereich des Spektrums um ein auf vier Parametern basierendes invertiertes Gaussmodell, welches dann zur Klassifikation genutzt wird. Dies ist im Vergleich zu anderen Indizes ein relativ aufwendiges Verfahren.

Die Ergebnisse der Studie von Main et al. legen nahe, dass Indizes, welche die Rote-Flanke nutzen, konsistenter und robuster sind als andere Indizes. Die Mehrheit der leistungsstärksten Indizes sind einfache Verhältnis- oder normierte Differenzindizes, die auf Wellen-

längen außerhalb des Chlorophyllabsorptionszentrums im Bereich 680 – 730 nm basieren. Einer dieser Indizes ist der NDVI, welcher einer der am häufigsten genutzten Indizes ist und schematisch in Abbildung 57 dargestellt ist. Zu beachten ist noch, dass laut Cho et al. [CSSo6] die Verwendung des NDVI mit breitbandigen Sensoren nur zur grundsätzlichen Abgrenzung von Vegetation genutzt werden sollte.

Der Effekt der Roten Flanke wird in dieser Arbeit näher untersucht und auf die spektralen Daten der NIR-Kamera angewendet. Eine reflektierte Spektralverteilung ist in Abbildung 56 dargestellt und das Schema des NDVI ist in Abbildung 57 visualisiert.

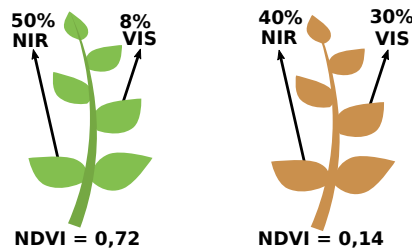


Abbildung 57: Beispiel des NDVI-Index. Links ist eine gesunde Pflanze mit einem hohen NDVI-Wert dargestellt. Der hohe Chlorophyllgehalt sorgt dafür, dass viel Licht im nahen Infrarotbereich reflektiert wird. Rechts ist eine Pflanze mit einem niedrigen NDVI-Wert dargestellt. Diese Pflanze reflektiert wenig Strahlung im nahen Infrarotbereich.

6.3 EVALUATION

Es wurden mehrere spektrale Indizes, wie in Tabelle 12 dargestellt, implementiert und als Klassifikator nutzbar gemacht.

Index	Formel (nm)	Referenz
Normalized Differenced Vegetation Index (NDVI)	$\frac{(800-670)}{(800+670)}$	[R]72]
Renormalised Difference Vegetation Index (RDVI)	$\frac{(800-670)}{\sqrt{800+670}}$	[RB95]
Optimised Soil Adjusted Vegetation Index (OSAVI)	$1.16 \cdot \frac{(800-670)}{(800+670+0.16)}$	[RSB96]
MERIS Terrestrial Chlorophyll Index (MTCI)	$\frac{(754-709)}{(709-681)}$	[DC04]
Maccioni Index (MI)	$\frac{(780-710)}{(780-680)}$	[MAM01]
Double Difference Index (DD)	$(749 - 720) - (701 - 672)$	

Tabelle 12: Auswahl verschiedener Indizes welche sich die Besonderheiten von Chlorophyll zunutze machen

Diese Indizes nutzen ausschließlich Daten der *NIR*-Kamera und arbeiten auf einem Hyperwürfel, der wie in Kapitel 4 erläutert berechnet wurde. Die Berechnung des NDVI aus den Daten eines Hyperwürfels wird hier beispielhaft erläutert

$$\text{NDVI} = \frac{\text{NIR} - \text{ROT}}{\text{NIR} + \text{ROT}} \quad (41)$$

Für *ROT* wird ein Kanal genutzt, der im Bereich zwischen 650 – 700 nm reagiert und für *NIR* ein Kanal welcher im Bereich 700 – 750 nm reagiert. Durch dieses Vorgehen ergibt sich anschließend ein Wertebereich zwischen –1 und +1. Ein Wert bis 0,2 entspricht dann Bereichen ohne Vegetation. Ein Wert nahe 1 deutet auf eine hohe Vegetationsbedeckung mit grünen Pflanzen hin. Ein Beispielergebnis dieses Vorgehens ist in Abbildung 58 und Abbildung 59 dargestellt.

Zur Evaluation wurden Experimente mit verschiedenen Indizes

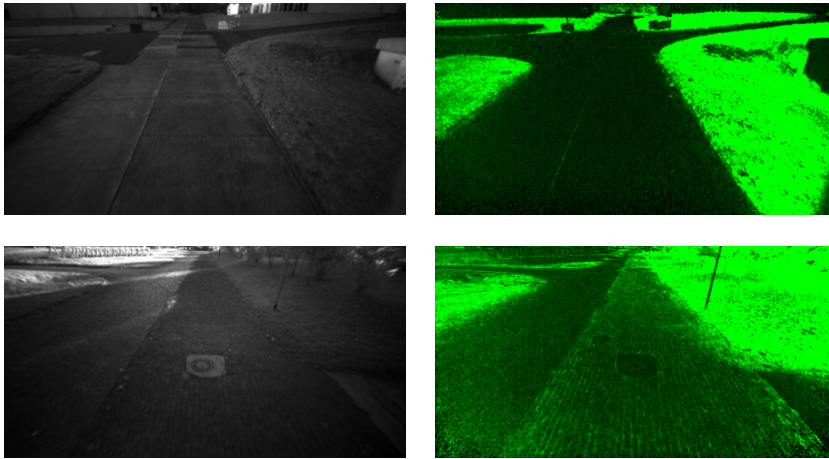


Abbildung 58: Visualisierung des NDVI-Wertes im Grünkanal. Der NDVI-Wert wird dazu von 0 – 255 skaliert und über den Grünkanal des Bildes visualisiert. Je höher der Wert desto wahrscheinlicher ist Chlorophyll enthalten wie z. B. bei Pflanzen und Gras. Je grüner der Bereich im Bild, desto höher ist der NDVI-Wert.

(vgl. Tabelle 12) durchgeführt auf einem der Datensätze, welche in Kapitel 5 dargestellt wurden. Die Ergebnisse der Experimente sind in Abbildung 60 dargestellt. Beispiele der Ergebnisse sind in Abbildung 61 und Abbildung 62 zu sehen. Sie zeigen, dass die ausgewählten Indizes vergleichbare Ergebnisse auf einem nahezu identischen Niveau erzeugen. Das lässt sich einfach damit begründen, dass sie alle auf denselben Effekt der roten Flanke setzen und diesen ausnutzen. Weiterhin zeigen die Ergebnisse des NDVI aus Abbildung 60 das unter Verwendung des Indizes quantitativ gesehen eine sehr gute Unterteilung in chlorophyllhaltige und andere Elemente möglich ist. Betrachtet man die erzeugten Ergebnisse, so bestätigt sich dieses

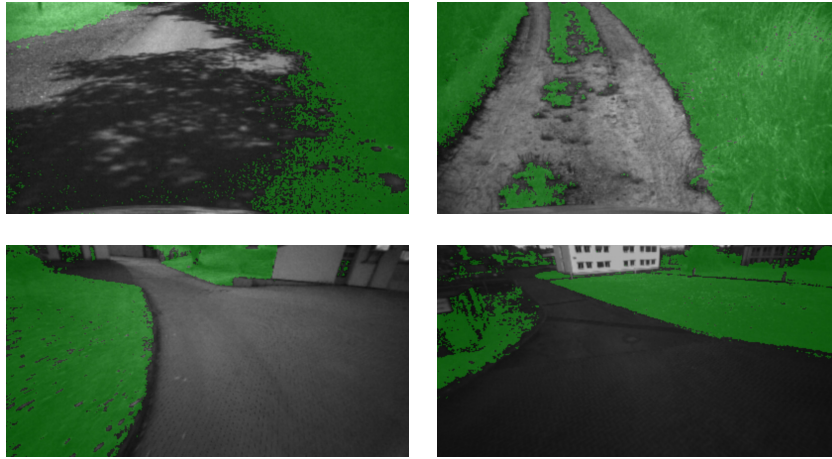


Abbildung 59: Beispielergebnisse der NDVI basierten Klassifikation. Die chlorophyllhaltigen Bereiche sind grün hervorgehoben. Diese wurden als Vegetation klassifiziert. Zur Klassifikation wurde eine einfache Schwellwertfunktion auf den NDVI-Wert angewandt.

Ergebnis. Der Index kann präzise chlorophyllhaltige Elemente wie Vegetation von anderen Elementen oder Oberflächen wie Straße trennen. Speziell in Abbildung 61 ist in Zeile zwei zu sehen, dass die kleinen, nicht annotierten Grasbüschel in der Mitte des Weges als solche klassifiziert werden konnten. Auch in Abbildung 62 zeigt sich in Zeile zwei, dass die Schilder, welche in den Büschen hängen, sauber getrennt werden.

6.4 FAZIT

In diesem Kapitel wurde die Vorhersagefähigkeit von verschiedenen optischen Indizes anhand von spektralen Daten, die mit Snapshot-Mosaik-Kameras aufgenommen wurden, untersucht. Ziel ist es dabei, die Vegetation von anderen Elementen wie z. B. einer Straße zu separieren, sodass ein Algorithmus zur autonomen Navigation sinnvoll unterstützt werden kann. Ein optischer Index hat gegenüber anderen Klassifikatoren, welche in den folgenden Kapiteln betrachtet werden, Vorteile. Zunächst ist ein optischer Index schnell und einfach zu berechnen, was ihn sehr gut auch auf schwächerer Hardware einsetzbar macht. In der Regel werden nur die Werte aus speziellen Kamerakanälen über einfache Operationen wie Subtraktion, Addition und Differenzbildung miteinander verrechnet werden müssen. Letztendlich basieren diese Indizes im Grunde auf dem Vergleich von Werten um den Bereich der roten Flanke, auch wenn die konkrete Berechnung im Einzelnen variiert. Weiterhin müssen diese Indizes nicht trainiert werden wie andere Klassifikatoren. Das heißt, es muss nicht erst ein

Datensatz aufgebaut und annotiert werden. Dementsprechend spielt die Generalisierbarkeit auch eine untergeordnete Rolle. Dieser Klassifikator ist sofort einsatzbereit und *universal* nutzbar. Beachtet werden muss natürlich, dass effektiv nicht klassifiziert wird, sondern ein Wert bestimmt wird, welcher vom Chlorophyllgehalt des Objekts abhängig ist. Somit ist nur eine binäre Klassifikation möglich. Doch betrachtet man die visuellen Ergebnisse aus der Evaluation zeigt sich, dass mit einer entsprechend montierten Kamera vor allem in Off-Road Szenarien eine einfache und effektive Unterscheidung zwischen Straße und umgebender Vegetation möglich ist.

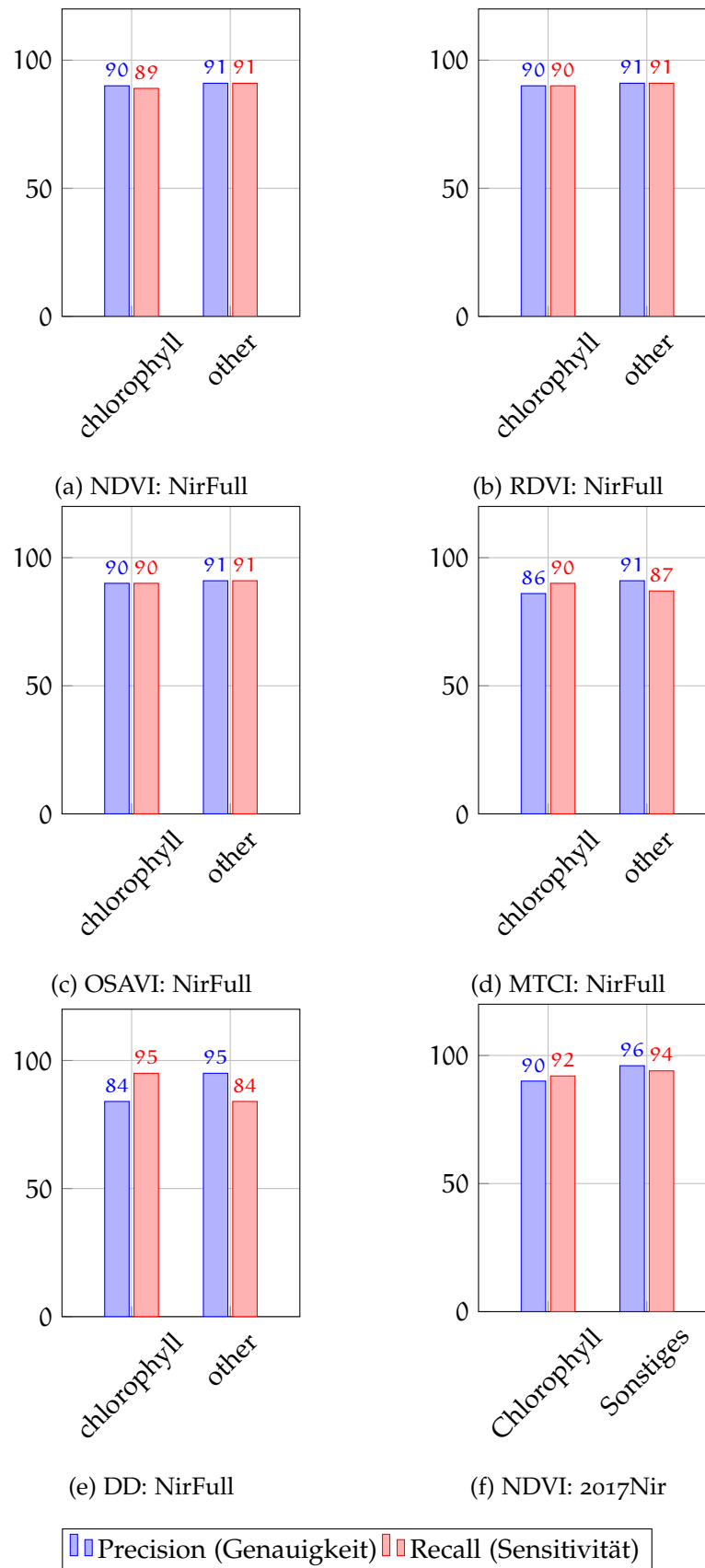


Abbildung 60: Ergebnisse der Klassifikation unter Verwendung der optischen Indizes auf den Datensätzen NirFull und 2017Nir. Die einzelnen Indizes zeigen alle sehr gute Ergebnisse auf dem Datensatz und unterscheiden sich nur geringfügig.

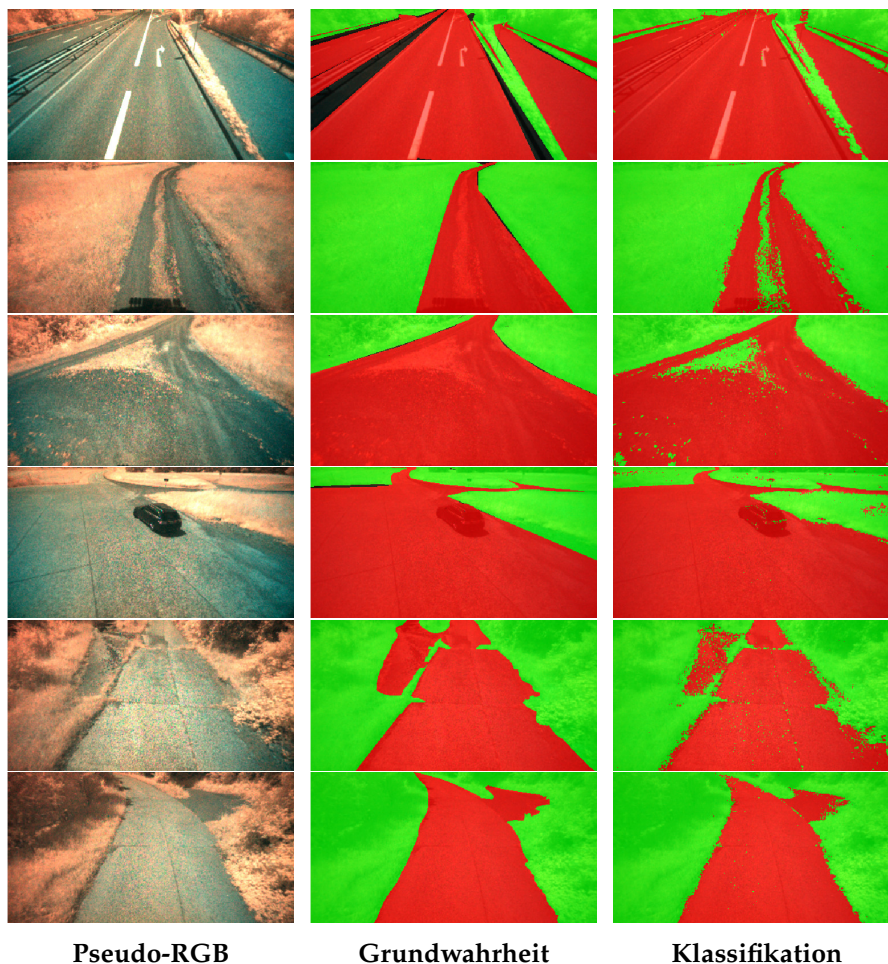


Abbildung 61: Vergleich der Klassifikationsergebnisse von NIR-Daten unter Verwendung des NDVI-Indizes auf dem Datensatz *NirFull*. Die chlorophyllhaltigen Elemente sind grün und der Rest ist rot. Die Trennung zwischen chlorophyllhaltigen Elementen wie Vegetation und Gras und anderen Elementen gelingt sehr gut. Selbst feine Bereiche mit einzelnen Grasbüscheln werden sauber von der Umgebung getrennt.

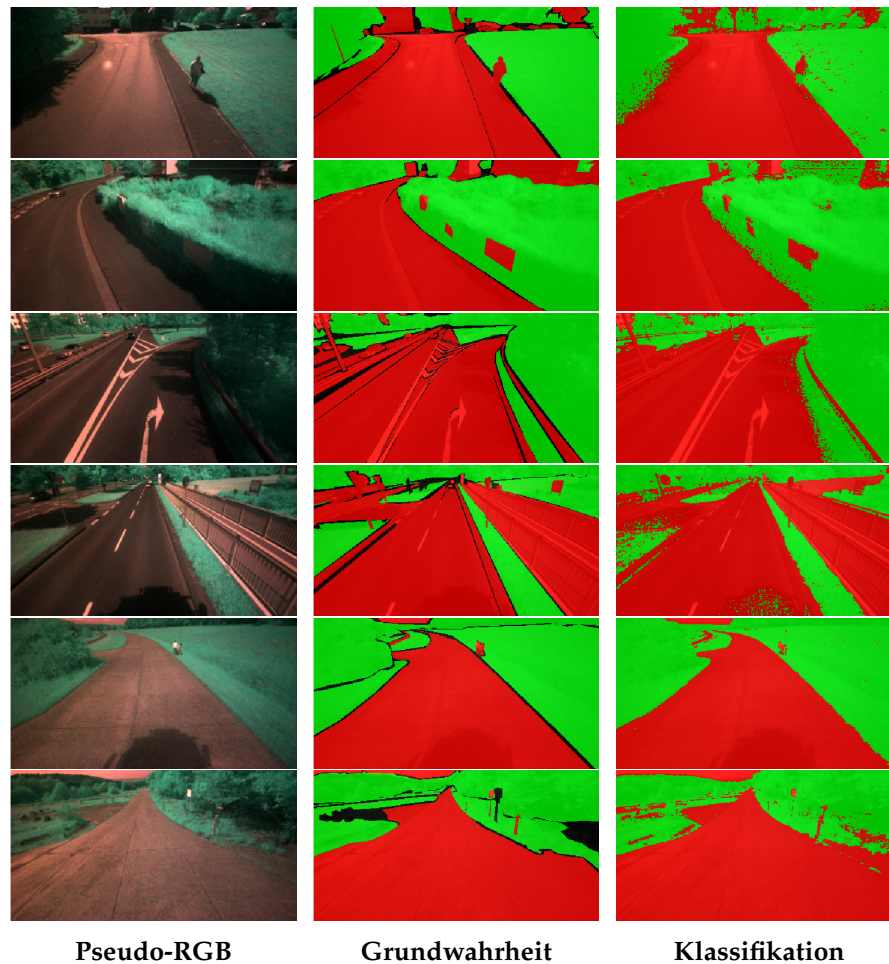


Abbildung 62: Vergleich der Klassifikationsergebnisse von *NIR*-Daten unter Verwendung des NDVI-Indizes auf dem Datensatz 2017Nir. Die chlorophyllhaltigen Elemente sind grün und die Übrigen sind rot. Auch auf diesem Datensatz ist eine saubere Trennung zwischen den Klassen zu sehen. Die Straße lässt sich beispielsweise sehr gut von den Wiesen und Büschen trennen.

ÜBERWACHTE KLASSIFIKATION

7.1 EINFÜHRUNG

Die spezifische Anwendung kann unterschiedlich sein, aber das allgemeine Ziel der spektralen Klassifikation ist es, den spektralen Daten eine definierte Anzahl von Klassen zuzuordnen. Zur Visualisierung sind Segmentierungsmasken nützlich, da sie die komplexen spektralen und räumlichen Informationen in einer definierten Anzahl von Klassen zusammenfassen. Eine Aufgabe der klassischen hyperspektralen Bildverarbeitung ist dabei die Unterscheidung verschiedener Bodentypen oder die Differenzierung von Bäumen und Gebäuden anhand ihrer spektralen Signatur. Diese Problemstellungen teilen sich die Aufgabe, eine Klassenzugehörigkeit aus einer Menge von Daten vorherzusagen. Die Bezeichnungen entsprechen verschiedenen Klassen, deren Eigenschaften für jede Domäne spezifisch ist. Die Merkmale sind typischerweise spektrale Informationen, welche von einem Satelliten- oder einem Flugzeug aufgenommen wurden. Im Gegensatz zur klassischen Analyse und Verarbeitung hyperspektraler Daten aus dem Bereich der Fernerkundung wird in diesem Abschnitt die Klassifikation spektraler Daten im Rahmen des autonomen Fahrens näher untersucht. Dazu werden die spektralen Signaturen der Oberflächen und Objekte mit überwachten Klassifikationsmethoden analysiert und trainiert, um adäquate Modelle zur Klassifikation zu berechnen. Um einen Überblick zu erhalten, welche Klassifikatoren sich für das vorliegende Szenario eignen, werden im Folgenden zunächst etablierte Algorithmen ausgewählt und entsprechend mit den vorliegenden Daten trainiert. Die im Folgenden beschriebenen Arbeiten wurden in Teilen bereits auf einer internationalen Konferenz veröffentlicht [5].

7.2 STAND DER TECHNIK

Die Klassifikation spektraler Daten wurde in den letzten 30 Jahren aktiv entwickelt und erforscht. Primärer Innovationstreiber war hier die Fernerkundung (engl. *Remote Sensing*), deren Hauptziel die Landschaftskartierung (engl. *Land-Cover Mapping*) in erheblichem Maße von der hyperspektralen Bildgebung profitiert hat. Obwohl es in der Literatur einige unbeaufsichtigte Klassifikationsalgorithmen gibt, liegt der Fokus in der Forschung überwiegend auf überwachten Klassifikationsalgorithmen. Diese sind erheblich weiter verbreitet, wie Plaza et al. [PBB⁺09] darlegen. Jedoch leiden die meisten überwachten

Klassifikatoren unter dem Hughes-Effekt [Hug68], besonders wenn hochdimensionale, in diesem Fall spektrale, Daten betrachtet werden. Der Hughes-Effekt besagt, dass bei einer festen Anzahl von Trainingsdaten die Vorhersagekraft zunächst mit Zunahme der Dimensionen der Trainingsdaten steigt und dann sinkt.

Um dieses Problem zu lösen, haben Huang et al. [HDT02], Melgani et al. [MB04], Foody [F004] und Camps-Valls et al. [CVB05] Stützvektormaschinen (engl. *Support Vektor Machine*) (SVM) mit adäquaten Kernen zur hyperspektralen Klassifikationen eingeführt. Diese Klassifikatoren setzten sich schnell durch. Stützvektormaschinen wurden ursprünglich als binärer Klassifikator [SS02] eingeführt. Um Probleme mit mehreren Klassen zu lösen, werden mehrere binäre Klassifikatoren kombiniert. Als Erweiterung wurden in der Folge Gruppen und Bündelungen (engl. *bagging*) von Klassifikatoren untersucht. Briem et al. [BBS02] legen den Fokus auf Bündelung von Klassifikatoren oder sog. *Boosting*-Verfahren wie Adaboost. Eine weitere alternative Methode präsentieren Waske et al. [WB07]. Anstatt die Klassifikationsergebnisse der Klassifikatoren zu fusionieren, werden die Ergebnisse jeder SVM-Diskriminanzfunktion in einem weiteren Fusionsprozess verwendet. Diese Fusion wird von einer weiteren SVM durchgeführt. Dieser Weg der Fusion von Klassifikatoren wurde in den folgenden Jahren von Fauvel et al. [FCB06] und Waske et al. [WvdLB⁺10] weiter entwickelt.

Im Jahr 2008 präsentierten Baofeng et al. [GGDNo8] eine Erweiterung der SVM-basierten Verfahren zur hyperspektralen Bildklassifizierung unter Verwendung von spektral gewichteten Kernen, welche verbesserte Klassifikationsergebnisse zeigten.

Ein weiterer Trend im Bereich der hyperspektralen Klassifikation war im Jahr 2010 die automatische Optimierung einer linearen Kombination von SVM Kernen deren Parameter durch den Gradientenabstieg optimiert werden, wie von Tuia et al. [TCVMK10] publiziert. Fang et al. [FLD⁺15] nutzen Superpixel in Kombination von mehreren Kernen, um die spektralen und räumlichen Informationen besser zu nutzen.

Grundsätzlich besteht eine große Diskrepanz zwischen der hohen Dimensionalität der Daten im Spektralbereich, ihrer starken Korrelation und gleichzeitig der Verfügbarkeit von annotierten Daten, die für das Training unbedingt notwendig sind. Während die Datenerfassung in der Regel recht unkompliziert ist, erweist sich die präzise und korrekte Annotation der erfassten Daten als sehr zeitaufwendig und kompliziert. Aus diesem Grund wurden semi-überwachte Techniken entwickelt, um die Generalisierbarkeit der Klassifikatoren zu erhöhen, wie von Camps-Valls et al. [CVTGC⁺11] vorgestellt. Diese semi-überwachten Methoden kombinieren die Vorzüge von überwachten und unüberwachten Methoden indem die Überwachten die Klassenzugehörigkeit definieren und die Unüberwachten globale Struktur

der Daten betrachten.

Tajudin und Landgrebe [TL00] zeigen hier im Jahr 2000 das erste Verfahren, welches nicht annotierte Daten nutzt, um die Parameter des Klassifikators unter Verwendung des EM-Algorithmus zu optimieren. Dieses Verfahren wurde ein Jahr später von Jackson und Landgrebe [JL⁺01] erweitert. Allerdings lassen sich diese Methoden nur effektiv einsetzen, wenn die Daten einer gaußschen Mischverteilung entsprechen. Daher bevorzugten spätere Forschungen Klassifikatoren, welche nicht über diese Einschränkung verfügen.

Dazu führten Bruzzone et al. [BCM06] 2006 Transductive SVMs ein, welche auf speziellen Lernalgorithmen basieren und berücksichtigen auch nicht annotierte Daten während des Trainings. Sie bestimmen die Hyperebene nach einem speziellen Prozess, welcher die nicht annotierten und die annotierten Daten zusammen integriert. Die Idee dahinter ist, eine Hyperebene zu definieren, welche sowohl die annotierten als auch die unannotierten Daten mit maximaler Spanne zwischen den Klassen trennt. Dieses Prinzip wurde von Chi und Bruzzone [CB07] durch die Vorstellung von *semisupervised* SVMs entsprechend erweitert. Ein weiterer Ansatz die Generalisierungsstärke der SVM-Klassifikatoren zu erhöhen ist die Integration von graphbasierten Modellen in den Klassifikationsprozess, wie von Camp-Valls et al. [CVMZ07] 2007 und Gómez-Chova et al. [GCCVMMCo8] 2008 gezeigt.

Weiterhin präsentierten 2010 Jun et al. [LBDP10] einen semi-überwachten Klassifikator, der unannotierte Daten basierend auf ihrer Entropie auswählt und sie dann den Trainingsdaten hinzufügt.

Einen alternativen Ansatz zur Klassifikation zeigen im Jahr 2005 Ham et al. [HYCC05] indem sie einen Random Forest mit HYPERION Daten trainieren. Im Jahr 2008 haben Chan und Paelinckx [CP08] AdaBoost und Random Forest Verfahren zur Klassifikation von Ökotopten anhand hyperspektraler Daten verwendet und damit gute Ergebnisse erzielt. In der Folge erlangte der Random Forest große Beliebtheit bei der Analyse hyperspektraler Daten. So haben Belgiu und Dragut [BD16] im Jahr 2016 einen Überblick über die Verwendung dieser Klassifikatoren in der Fernerkundung publiziert. Es zeigt sich, dass vor allem die SVMs und Random Forests im Bereich der Klassifikation von hyperspektralen Daten eine breite Anwendung finden. Daher werden diese im Folgenden näher untersucht und ihre Anwendbarkeit auf die in dieser Arbeit behandelten Daten überprüft.

7.3 PER-PIXEL KLASSIFIKATION

Die Per-Pixel-Klassifikation ist ein Prozess, bei dem jedem Pixel eines Bildes bzw. Hyperpixel eines Hyperwürfels eine Klasse zugewiesen wird. Dies ist die verbreitetste Methode, da die meisten überwachten Klassifikatoren auf dem Prinzip einer Klassenzuweisung pro Pixel

basieren. Diese überwachten Lerntechniken wie z. B. Random Forest benötigen dazu Trainingsdatensätze, welche aus einer Menge von Merkmalsvektoren bestehen, die mit einer entsprechenden Annotation versehen sind. Entsprechend sind $\Omega_k \in \Omega$ benutzerdefinierte Klassen, die normalerweise durch ganzzahlige Zahlen repräsentiert werden. Ziel des Trainings ist es, eine Funktion oder ein Modell zu lernen, welches einem bestimmten Merkmal eine entsprechende Klasse zuordnet. Dazu werden die Trainings- und Testdaten nach dem Zufallsprinzip aus jeweils einem der verfügbaren Datensätze zusammengestellt, welche in Kapitel 5 näher erläutert wurden. Gegeben eine Menge von \mathcal{N} korrespondierenden Trainingspaaren ist das Ziel des Trainings, eine Funktion zu finden, die gut genug generalisiert, sodass genaue Vorhersagen bzw. Klassifikationen für zuvor ungesehene Daten berechnet werden können.

Sei $\{\mathbf{X}^{\text{train}}, \mathbf{Y}^{\text{train}}\} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{\mathcal{N}}$ mit \mathcal{N} Trainingspaaren. Die Trainingsdaten $\mathbf{X}^{\text{train}} = \{\mathbf{x}_i \in \mathbb{R}^D, i = 1, 2, \dots, \mathcal{N}\}$ bestehen aus den Punktspektren χ_i der Hyperpixel \mathbf{p}_i^{H} eines Hyperwürfels $\mathbf{x}_i := \chi_i$ und haben die Dimension D .

Die zugehörigen Annotationen $\mathbf{y}_i \in \{1, 2, 3, \dots, |\Omega|\}$ werden durch vorher definierte Klassen Ω beschrieben. Ziel der Klassifikation ist es, ein Modell zu trainieren, dass zwischen den Eingangsproben $\mathbf{X}^{\text{train}}$ und dem Ziel $\mathbf{Y}^{\text{train}}$ eine Verbindung herstellt. Dieses gelernte Modell kann dann verwendet werden, um neue Pixel einer der Klassen zuzuordnen.

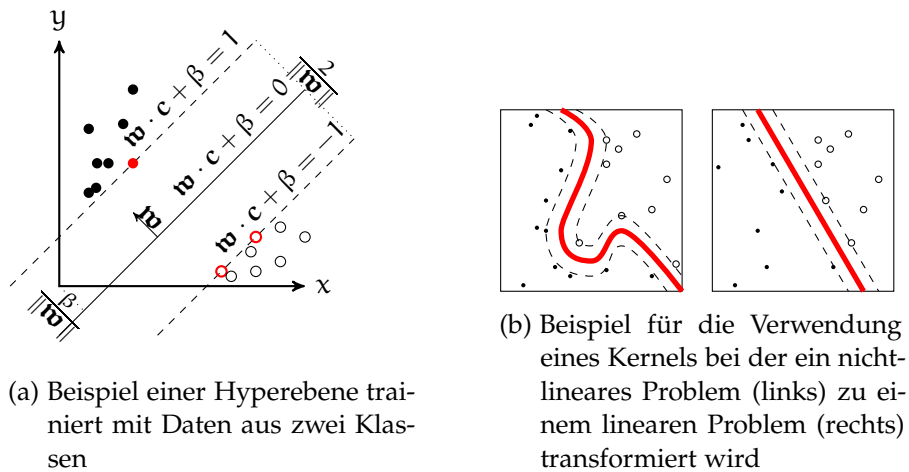
In diesem Prozess erzeugt ein Klassifikator ein Modell, welches eine Darstellung des gegebenen Problems repräsentiert, aus dem sich dann eine Klassifikation ableiten lässt. Je besser und genauer das Modell, desto besser sind auch die Ergebnisse für ungesehene Daten, was wiederum auch stark von der Qualität der Trainingsdaten abhängig ist.

Um nun zu untersuchen, ob die hier vorliegenden spektralen Daten zur Szenenanalyse geeignet sind, werden im Folgenden mehrere überwachte Klassifikatoren auf den bereits in Kapitel 5 vorgestellten spektralen Datensätzen trainiert und die Klassifikationsergebnisse ausgewertet.

7.3.1 Stützvektormaschinen

Stützvektormaschinen wurden ursprünglich entwickelt, um eine Entscheidung zwischen zwei Klassen durchzuführen. Später wurden mehrklassige Problemstellungen durch entsprechende Erweiterungen gelöst. Ziel einer SVM ist es, basierend auf Trainingsdaten eine Hyperebene zu bestimmen, die eine Trennung mit maximalem Abstand zwischen Daten zweier Klassen im Merkmalsraum definiert, wie in Abbildung 63 schematisch dargestellt. Eine Hyperebene ist dabei eine Verallgemeinerung des Terminus Ebene und bezeichnet Ebe-

nen in Räumen mit beliebig vielen Dimensionen. Die Bestimmung einer Hyperebene ist bei linear trennbaren Problemen gut möglich, da sich die Klassen im Merkmalsraum in der Regel nicht überschneiden. Gegeben sei nun ein Problem aus dem Bereich der binären Klassifi-



(a) Beispiel einer Hyperebene trainiert mit Daten aus zwei Klassen

(b) Beispiel für die Verwendung eines Kerns bei der ein nicht-lineares Problem (links) zu einem linearen Problem (rechts) transformiert wird

Abbildung 63: Schematische Darstellung der Funktionsweise einer SVM

kation. Weiterhin ist ein Trainingsdatensatz mit N Merkmalsvektoren c_i aus einem D -dimensionalen Merkmalsraum w gegeben:

$$c_i \in \mathbb{R}^D, i = 1, \dots, N \tag{42}$$

Dann ist jedem Merkmalsvektor eine Klasse $y_i \in \{-1, +1\}$ zugeordnet und es wird angenommen, dass die beiden Klassen linear trennbar sind. Dadurch ist es möglich, mindestens eine Hyperebene, definiert durch einen Normalenvektor $\mathbf{w} \in \mathbb{R}^D$ und einen Bias $\beta \in \mathbb{R}$, zu finden, welche die beiden Klassen ohne Fehler trennen kann. Entsprechend kann die Entscheidungsregel durch eine Funktion $F(c_i)$ repräsentiert werden, welche mit der Hyperebene assoziiert wird und wie folgt definiert ist:

$$F(c_i) = \mathbf{w} \cdot c_i + \beta \tag{43}$$

Um nun eine passende Hyperebene zu ermitteln, müssen \mathbf{w} und β bestimmt werden, sodass gilt

$$y_i(\mathbf{w} \cdot c_i + \beta) > 0 \tag{44}$$

Die Aufgabe des SVM-Klassifikators besteht nun darin, eine optimale Hyperebene zu finden, welche den Abstand zwischen dem nächstgelegenen Merkmalsvektor und der trennenden Hyperebene maximiert. Eine optimale Hyperebene maximiert den Abstand zwischen sich und den nächstgelegenen Merkmalsvektoren. Die Merkmalsvektoren, welche der Hyperebene am nächsten liegen, werden als Stützvektoren bezeichnet. Der Abstand zwischen diesen Stützvektoren und der

trennenden Hyperebene wird als *Margin* bezeichnet und als $\frac{1}{\|\mathbf{w}\|}$ definiert. Dies ist wichtig, da die *Margin* ein Indikator für die Generalisierungsfähigkeit des Klassifikators ist. Je größer die *Margin* desto höher ist die Generalisierbarkeit.

Die gerade beschriebene Klasse von SVMs ist allerdings nur einsetzbar, wenn die Trainingsdaten linear trennbar sind. Allerdings ist diese Bedingung bei der Klassifikation von realen Daten selten erfüllbar. Eine Möglichkeit, dieses Problem zu lösen, ist die Verwendung des *Kernel-Tricks*, dabei werden die Daten durch eine geeignete nicht lineare Transformation $\delta(\mathbf{c})$ in einen höherdimensionalen Raum abgebildet:

$$\delta(\mathbf{c}) \in \mathbb{R}^{D'} \quad (D' > D) \quad (45)$$

$$\mathbf{w} \in \mathbb{R}^{D'} \quad (46)$$

$$\beta \in \mathbb{R} \quad (47)$$

In diesem höherdimensionalen Raum ist dann wieder eine lineare Trennung der beiden Klassen möglich [Cov65]. Nun besteht das Hauptproblem in der expliziten Berechnung von $\delta(\mathbf{c})$ was durchaus komplex sein kann. Aber der Kernel-Trick bietet eine elegante und effektive Möglichkeit, derartige Probleme zu lösen. Eine detailliertere Beschreibung zu diesen Aspekten wurde von Cristianini et al. [CST⁺00] im Jahr 2000 veröffentlicht.

Ist nun ein Multiklassenproblem gegeben, wird eine Reihe von binären SVM-Klassifikatoren verwendet, um die Klassifikation durchzuführen. Dabei wird das *einer gegen alle* Prinzip angewandt, indem jede SVM ein Zweiklassen-Problem löst, bei der eine Klasse gegen die Menge der anderen Klassen antritt. Wenn eine der SVMs einem Eingangsvektor eine Klasse zuordnet, wird diese Klasse an den Vektor übergeben. In Kapitel 7.2 wurde bereits gezeigt, dass SVMs sich einer gewissen Verbreitung bei der Klassifikation von hyperspektralen Daten erfreuen.

7.3.2 *Random Forest*

Leo Breiman [Breg6] führte 1996 die Idee des „Bootstrap Aggregating“ (*Bagging*) ein. Dabei werden mehrere Versionen eines Klassifikators initialisiert und verwendet, um eine endgültige Entscheidung zu treffen, indem jeder Klassifikator eine Stimme hat und am Ende die Mehrheit der Stimmen über die Klasse entscheidet. Breiman konnte zeigen, dass mit zunehmender Anzahl der Klassifikatoren auch die Genauigkeit steigt, bis zu einem Wendepunkt, ab dem sie dann abfällt. Random Forests gehören zur Gruppe dieser Ensembleklassifizierer und verwenden eine Reihe von Klassifikatoren basierend auf Entscheidungsbäumen, um ein robustes Modell zu trainieren.

Jeder Klassifikator wird dabei auf einer eigenen Untermenge von Trainingsdaten trainiert. Dieser Ansatz ist seit der Veröffentlichung

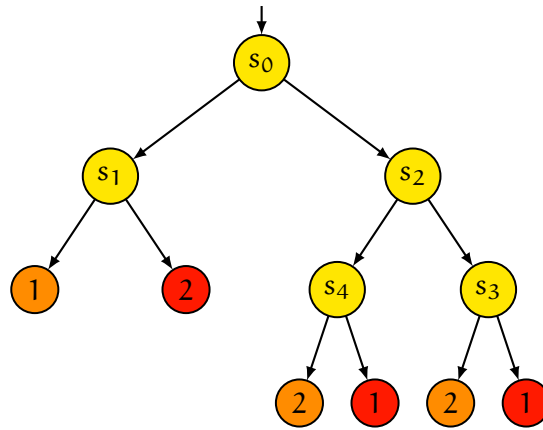


Abbildung 64: Beispiel eines einfachen Entscheidungsbaums. Dieser teilt die Daten in die Klassen 1 und 2 ein. Die Entscheidungsknoten s_0, \dots, s_4 sind gelb, die Klassenzuweisungen 1 und 2 sind in orange und rot eingefärbt.

eine weitverbreitete Methode, bei der Stichproben nach dem Zufallsprinzip gezogen und durch Ersetzen des ursprünglichen Datensatzes erzeugt werden, um eine neue Verteilung der Daten zu generieren. Dies verhindert eine Überanpassung und führt zu unterschiedlichen Mustern in den Eingabedaten. Ein einzelner Entscheidungsbaum einer solchen Ansammlung ist ein binärer Baum, der sich aus mehreren Elementen zusammensetzt, einem eindeutigen Wurzelknoten, einer Menge interner Knoten und einer Menge von Blattknoten, wie schematisch in Abbildung 64 dargestellt. Die Entscheidungen an den jeweiligen Knoten bestimmen den Nachfolgeknoten. So wird der Baum von der Wurzel über die Menge der Entscheidungen zu einem definierten Blattknoten traversiert. Diese Blattknoten enthalten dann einen Klassennamen, welcher die Klassenzugehörigkeit anzeigt. Die Entscheidung über den jeweiligen Nachfolgeknoten wird wie folgt getroffen:

$$S(\mathbf{c}, \Theta, i) = \begin{cases} 0 : \mathbf{c}_i < \Theta \\ 1 : \mathbf{c}_i \geq \Theta \end{cases} \quad (48)$$

Jeder Knoten speichert demnach den Schwellwert Θ der Schwellwertfunktion S und den Index i des Eingabevektors \mathbf{c} über den bestimmt wird. Ist das Ergebnis kleiner als der Schwellwert, so wird im linken Nachfolgeknoten fortgefahren, andernfalls im Rechten. Die Eingabedaten eines Entscheidungsbaums beim Training bestehen aus einer Menge \mathcal{C} von Merkmalsvektoren $\mathbf{c}_i \in \mathcal{C} \subset \mathbb{R}^N$ und den zugehörigen Klassen $\Omega_i \in \Omega \subset \mathbb{N}$. Die Trainingsdaten liegen am leeren Wurzelknoten an und die Unreinheit (engl. *Impurity*) J [Sha48] der

Merkmalsvektoren wird bestimmt, indem zunächst die Entropie berechnet wird:

$$J_e(\mathcal{C}) = - \sum_{i=0}^N c_i \cdot \log_2(c_i) \quad (49)$$

Anschließend wird der sog. Gini-Index bestimmt:

$$J_g(\mathcal{C}) = 1 - \sum_{i=0}^N c_i^2 \quad (50)$$

Dabei bezeichnet c_i den Anteil der Eingabevektoren mit Klasse Ω_i an der Gesamtmenge der Eingabevektoren \mathcal{C} . Das Gini-Unreinheitsmaß nimmt sein Maximum an, wenn jede Klasse in einem Knoten mit gleicher Wahrscheinlichkeit angenommen wird. Ziel ist es daher, in jedem Knoten die Eingabedaten so aufzuteilen, dass die Verunreinigung in dem Knoten minimiert wird. Das heißt, dass möglichst nur Daten einer Klasse dem Knoten zugeordnet werden. Um ermitteln zu können, ob eine Aufteilung effektiv war, wird der Informationsgewinn (engl. *information gain*) G berechnet, welcher sich aus der Unreinheit vor und nach der Aufteilung zusammensetzt:

$$G = J(\mathcal{C}) - \sum_{i=0}^N \frac{|\mathcal{C}^{(i)}|}{|\mathcal{C}|} J(\mathcal{C}^{(i)}) \quad (51)$$

Dabei ist $\mathcal{C}^{(k)}$ diejenige Teilmenge der Eingabedaten, für welche die Schwellwertfunktion dem Merkmalsvektor die Klasse Ω_k zuweist. Die Parameter, welche während des Trainings beim jeweiligen Knoten den größten Informationsgewinn ermöglichen, werden so ermittelt und gespeichert. Die Auswahl der Parameter kann dabei durch verschiedene Algorithmen erfolgen. Nach dem Festlegen der Parameter für diesen Knoten werden nun die Daten, welche an dem Knoten anliegen, unter Verwendung der Entscheidungsfunktion aufgeteilt. Für die jeweiligen Kindknoten wird dann nach demselben Prinzip vorgegangen, um die passenden Parameter zu bestimmen. An einem Knoten wird gestoppt, wenn eine vorher definierte Abbruchbedingung erfüllt wird. Das ist zum Beispiel der Fall, wenn nur noch Vektoren einer Klasse in dem Knoten vorliegen oder wenn eine maximale Baumtiefe erreicht wurde. Unter der Prämisse, dass keine zwei identischen Eingabevektoren einer unterschiedlichen Klasse zugewiesen werden, kann ein solcher Algorithmus die Eingabevektoren so lange aufteilen, bis in jedem Blattknoten nur noch eine einzige Klasse vorhanden ist und die Unreinheit somit minimiert wurde.

Bei Knoten in denen mehrere Klassen noch vorhanden sind, kann eine absolute oder relative Häufigkeit der vorhandenen Klassen bestimmt werden, um eine eindeutige Klassenzuordnung innerhalb eines Entscheidungsbaums zu erreichen. Die gesamte Menge der Entscheidungsbäume bilden dann einen Random Forest, mit denen die

generierten Teilmengen klassifiziert werden. Das Ergebnis der Klassifizierung wird am Ende durch eine Mehrheitsentscheidung basierend auf den Ausgaben der Entscheidungsbäume erzielt.

Grundsätzlich lässt sich auch eine Wahrscheinlichkeit für jede Klasse berechnen. Die Klassenwahrscheinlichkeit einer Eingabe wird nach [Bos07] durch den Stimmanteil (engl. *Average Vote*) (AV) der Bäume definiert.

$$AV(\{t_1, \dots, t_n\}, \mathbf{c}_i, \Omega) = \frac{\sum_{i=1}^n 1(\max_{\Omega'} \{t_i(\mathbf{c}_i, \Omega')\} = \Omega)}{n} \quad (52)$$

Wobei $\{t_1, \dots, t_n\}$ die einzelnen Entscheidungsbäume definiert, \mathbf{c}_i den zu klassifizierenden Merkmalsvektor und Ω_i als die jeweilige Klasse. Die Funktion $t_j(\mathbf{c}_i, \Omega')$ gibt abhängig von t_j an, mit welcher Wahrscheinlichkeit \mathbf{c}_i zur Klasse Ω' gehört. Die Funktion $1()$ gibt dann 1 zurück, wenn die Eingabe wahr ist, ansonsten 0.

7.4 TRAINING

Um nun die spektralen Daten zu klassifizieren bzw. den Eingabedaten Klassen zuzuordnen, müssen folglich geeignete Modelle unter Verwendung unterschiedlicher Klassifikatoren trainiert werden. Ein Schema einer solchen Pipeline zum maschinellen Lernen ist in Abbildung 65 dargestellt. Zum Training werden die Datensätze zufällig im Verhältnis von 80 : 20 aufgeteilt. 80% der Daten werden dabei zum Training genutzt und 20% der Daten zur Evaluation des trainierten Modells. Unter Verwendung der Daten wird ein Merkmalsraum aufgespannt, welcher zum Trainieren des Klassifikators genutzt wird. Als Ergebnis entsteht ein Modell, welches anschließend zur Klassifikation genutzt wird, um den Eingabedaten Klassen zuzuordnen.

Da zwei Kameras mit unterschiedlichen Wellenlängenempfindlichkeiten zum Einsatz kommen, sind entsprechend je Klassifikator zwei getrennte Modelle zu trainieren. Zum Training werden die annotierten Hyperwürfel der unterschiedlichen Datensätze zu Test- und Trainingsdatensätzen zusammengestellt und dann die Hyperwürfel in einzelne Hyperpixel zerlegt. So beinhaltet ein Hyperpixel das gemessene Spektrum an einer *Stelle* in der Umgebung. Als Eingabevektor zum Training erhält ein Klassifikator nun das gemessene Spektrum χ eines annotierten Hyperpixels, welches aus einem 16 bzw. 25-dimensionalen Merkmalsvektor besteht. So werden die Trainingsdaten dann genutzt, um einen Klassifikator zu trainieren und ein Modell zu generieren. Zum Testen der Klassifikationsgenauigkeit wird dem trainierten Modell des Klassifikators dann jeweils ein Hyperpixel aus einem Hyperwürfel mit dem zugehörigen Spektrum χ übergeben. Dies entspricht einer pixelbasierten Klassifikation eines Bildes, wie im nächsten Abschnitt näher erläutert. Theoretisch hätten auch mehrere Hyperpixel gemittelt werden können, was potentiell das Rau-

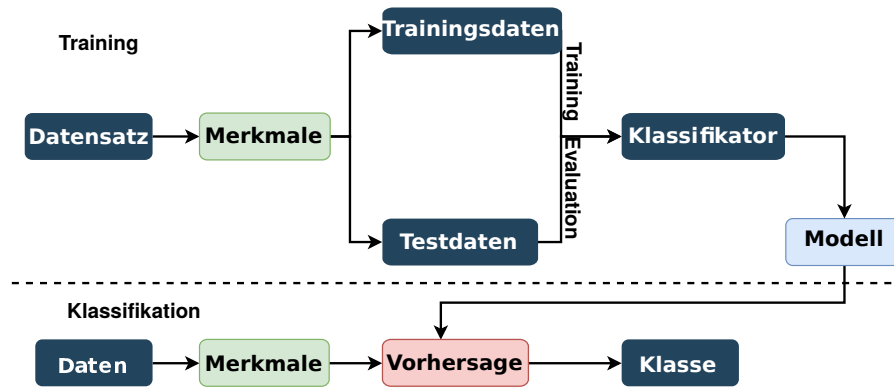


Abbildung 65: Schema der Pipeline zum maschinellen Lernen

Name	Kürzel	Quelle
Random Forest	RF	[Bre01]
Decision Tree	DT	[Qui86]
Gaussian Naive Bayes	GNB	[JL95]
Perceptron	Perc	[Gal90]
Stochastic Gradient Descent	SGD	[ZE02]
Passive Agressive	PA	[CDK ⁺ 06]
AdaBoost	Ada	[HRZZ09]

Tabelle 13: Übersicht über die hier untersuchten Klassifikatoren

schen reduziert. Allerdings stand hier die Bewertung der spektralen Informationen im Vordergrund.

7.5 EVALUATION

Zur Evaluation werden die Datensätze aus Kapitel 5 und verschiedene Klassifikatoren genutzt, um für jede Annotationsgruppe Modelle zu trainieren. Dabei werden die gemessenen Spektren χ_i der jeweiligen Hyperpixel \mathbf{p}_i^H eines Hyperwürfels als Merkmalsvektoren $\mathbf{c}_i = \chi_i$ mit $i \in \mathcal{L}_x \times \mathcal{L}_y$ interpretiert. Diese Merkmalsvektoren haben bei den *VIS*-Daten 16 und bei den *NIR*-Daten 25 Werte, womit auch die Dimension D der Merkmalsvektoren definiert ist. Mit dieser Repräsentation ist eine einfache Verwendung der zuvor beschriebenen Klassifikatoren möglich. Als Klassifikatoren wurden verschiedene aus der Literatur bekannte Verfahren genutzt, welche in Tabelle 13 dargestellt sind. Die Ergebnisse der Evaluation sind in Abbildung 66 dargestellt. Gezeigt ist für jeden Klassifikator der Precision, Recall und der Intersection Over Union Wert, deren Bedeutung in Kapitel 3.8 beschrieben sind. In Abbildung 66 sind die Ergebnisse

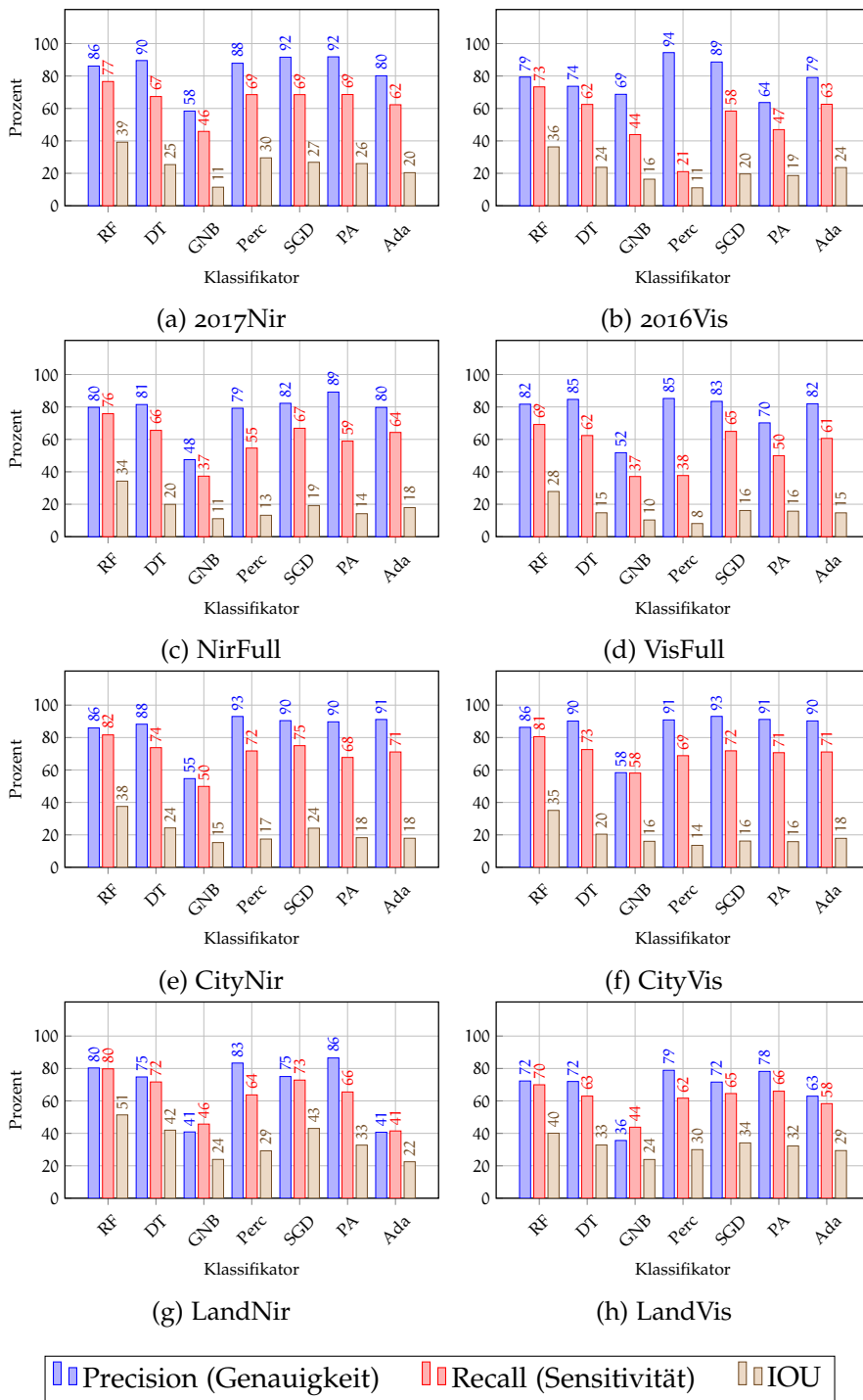


Abbildung 66: Übersicht über die Evaluationsergebnisse der verschiedenen Klassifikatoren mit semantischen Annotationen auf verschiedenen Datensätzen. Der Random-Forest (RF) zeigt auf allen Datensätzen die besten Ergebnisse.

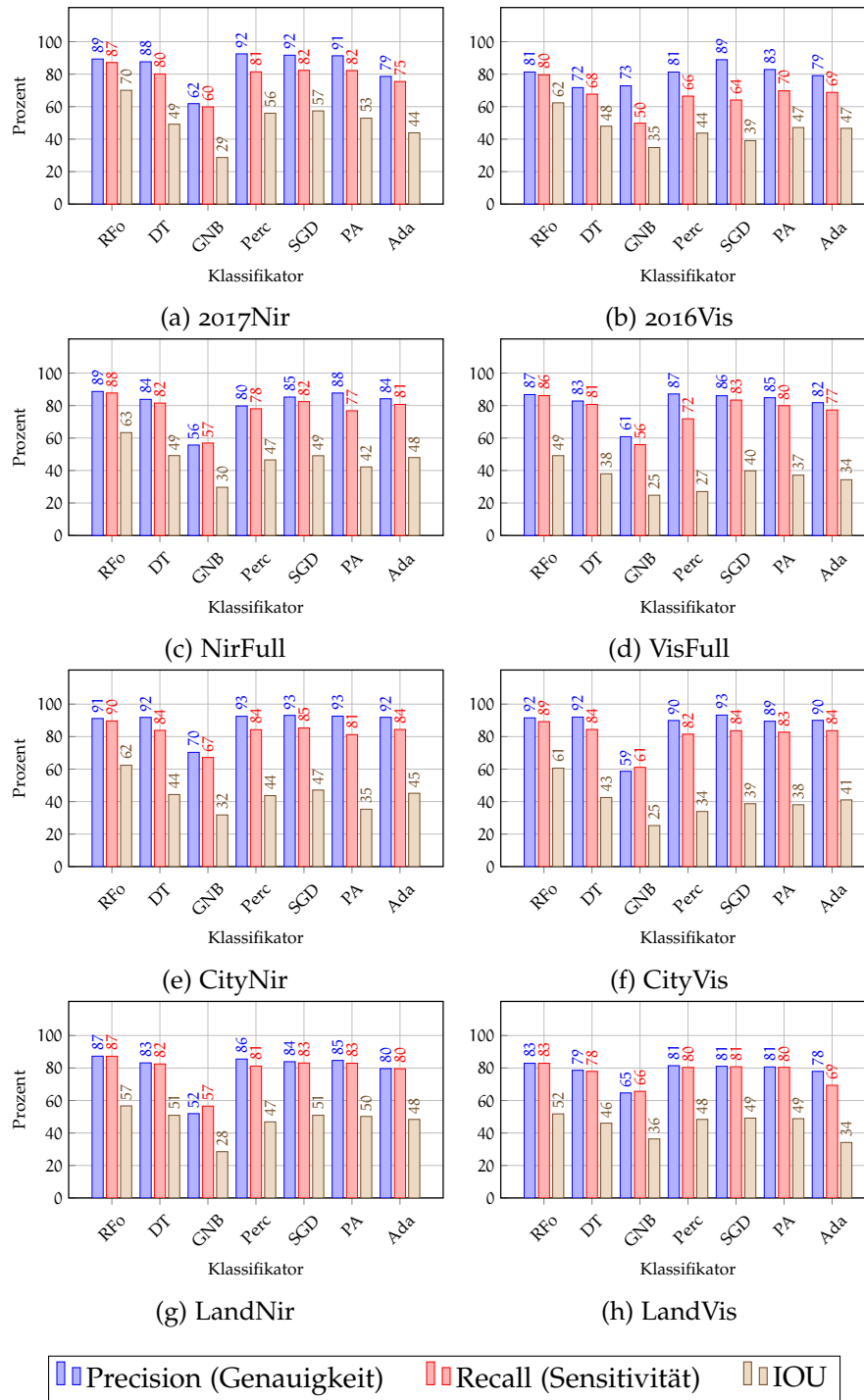


Abbildung 67: Übersicht über die Evaluationsergebnisse der verschiedenen Klassifikatoren basierend auf offRoad Annotationen auf verschiedenen Datensätzen. Der Random-Forest (RF) zeigt auf allen Datensätzen die besten Ergebnisse.

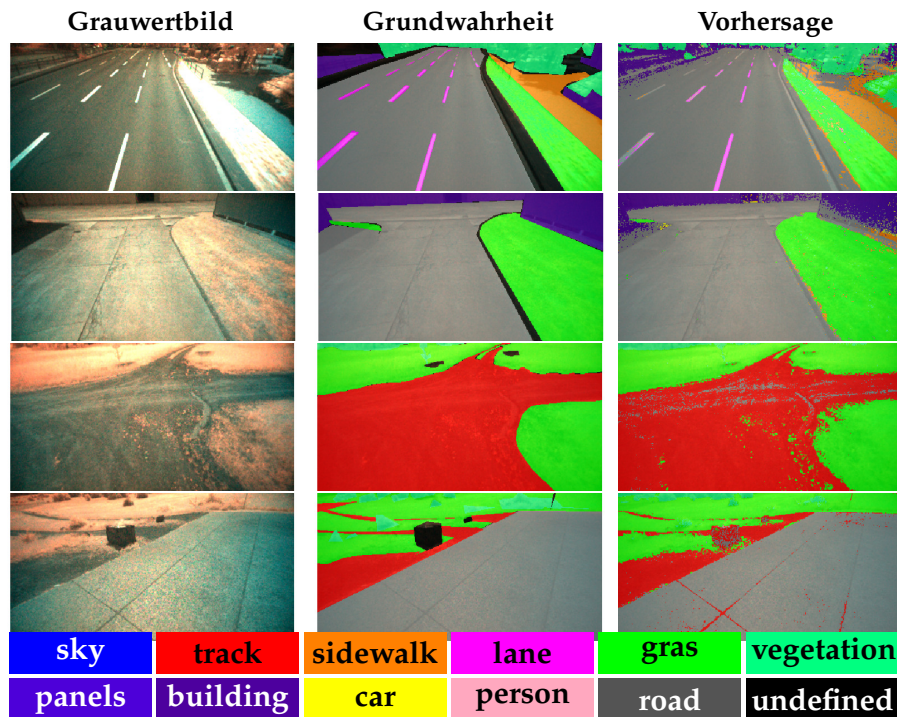


Abbildung 68: Vergleich der Klassifikationsergebnisse von *NIR*-Daten mit semantischen Annotationen. Die erste Spalte zeigt eine RGB-Darstellung der spektralen Eingabedaten, gefolgt von der Grundwahrheit und der Random Forest-Klassifikation. Die Ergebnisse der Klassifikation sind der Grundwahrheit schon sehr ähnlich. Es bestehen noch Probleme bei der Unterscheidung zwischen Straße (*street*) und Feldweg (*track*).

unter Betrachtung der semantischen Annotationen dargestellt, wobei die linke Spalte die Ergebnisse der *NIR*-Daten darstellt und die rechte Spalte die Ergebnisse der *VIS*-Daten.

7.5.1 *NIR* semantisch

Werden die Ergebnisse basierend auf den *NIR* Daten betrachtet, zeigt sich das über alle *NIR*-Datensätze hinweg der Random Forest-Klassifikator den höchsten Intersection Over Union Wert hat und auch bei der kombinierten Betrachtung der Precision und Recall Werte zeigt der Random Forest-Klassifikator die besten Ergebnisse. Beim 2017Nir-Datensatz ist es der Perceptron-Klassifikator welcher die zweitbesten Ergebnisse liefert. Bei den übrigen Datensätzen stehen der Decision Tree- und der Stochastic Gradient Descent-Klassifikator nahezu gleichauf an zweiter Stelle. Am schlechtesten schneidet, bei Be-

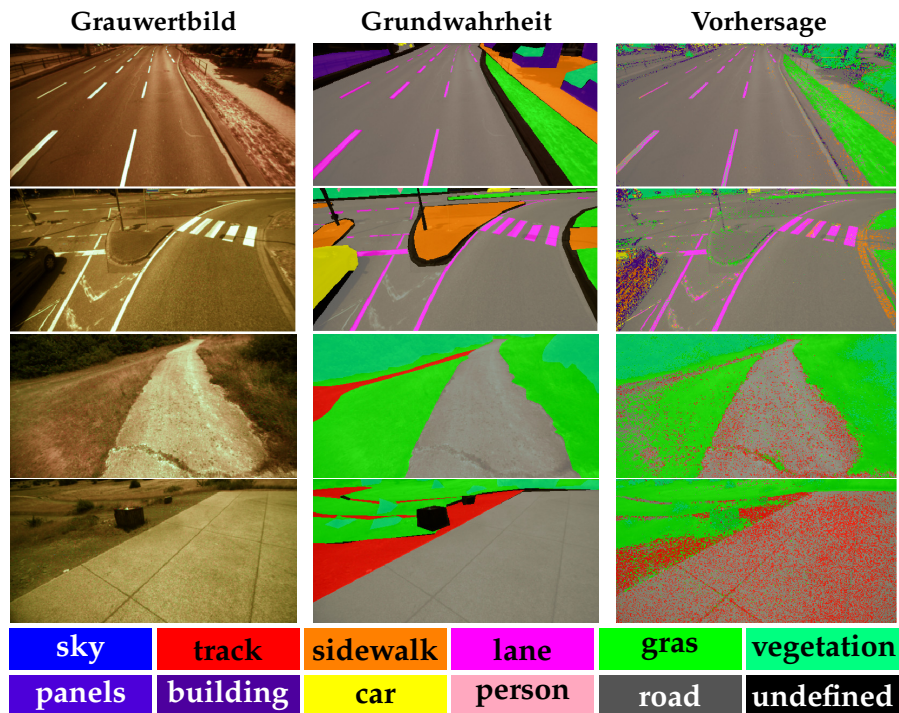


Abbildung 69: Vergleich der Klassifikationsergebnisse von VIS-Daten mit semantischen Annotationen. Die erste Spalte zeigt eine Grauwertdarstellung der spektralen Eingabedaten, gefolgt von der Grundwahrheit und der Random Forest-Klassifikation (Vorhersage). Im Vergleich zur Klassifikation mit NIR-Daten ist hier die Trennung zwischen Feldweg (*track*) und Straße (*street*) nicht so sauber.

trachtung aller Datensätze, der Gaussian Naive Bayes Klassifikator ab, welcher insgesamt wesentlich schlechtere Ergebnisse zeigt als der Random Forest-Klassifikator. Die Werte für Precision und Recall sind überwiegend sehr ähnlich. Das ist ein Indiz dafür, dass der Klassifikator keine systematischen Fehler bei der Klassifikation macht.

Beispiele der semantischen Segmentierung mit dem Random Forest-Klassifikator auf den jeweiligen Datensätzen sind in Abbildung 68 und Abbildung 69 dargestellt.

7.5.2 VIS semantisch

Auch bei den VIS-Datensätzen werden mit einem Random Forest-Klassifikator über alle Datensätze hinweg die besten Ergebnisse erzielt. Im Vergleich zu den NIR-Daten sind die Ergebnisse aber etwas schlechter. Dieser Trend zeigt sich bei allen Klassifikatoren, wenn beide Datensätze betrachtet werden. Weiterhin wird deutlich das vor

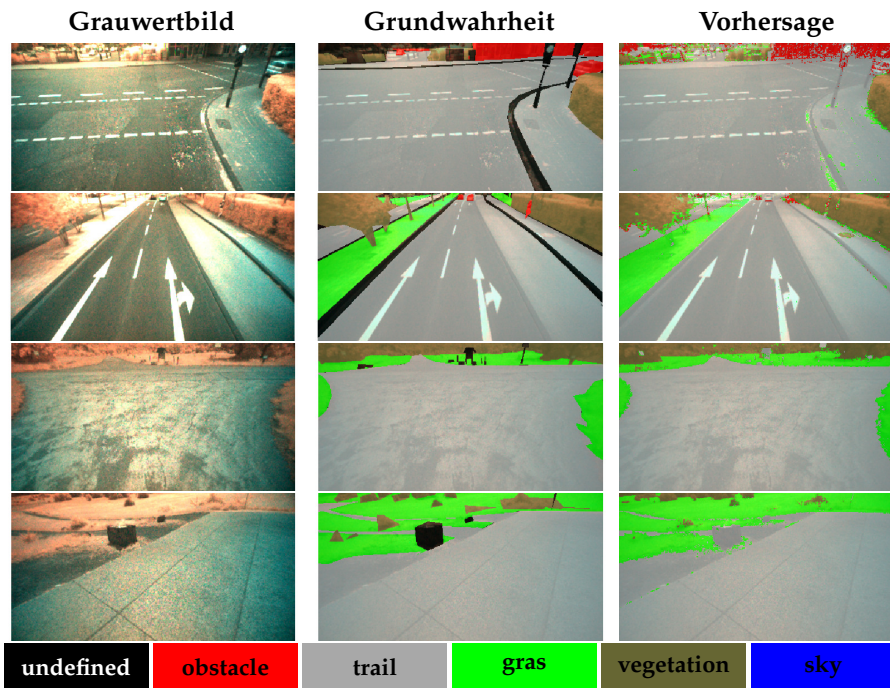


Abbildung 70: Vergleich der Klassifikationsergebnisse von *NIR*-Daten mit offRoad-Annotationen. Die erste Spalte zeigt eine RGB-Darstellung der spektralen Eingabedaten, gefolgt von der Grundwahrheit und der Random Forest-Klassifikation. Es sind wenige Unterschiede zwischen Grundwahrheit und Vorhersage zu erkennen.

allem der Recall Wert im Verhältnis zum Precision Wert bei den *VIS*-Daten viel schlechter ist. Auch hier teilen sich der Decision Tree- und der Stochastic Gradient Descent-Klassifikator bei allen *VIS*-Datensätzen den zweiten Platz. Bis auf den Datensatz 2018LandVis, liefert der Perceptron-Klassifikator bei allen anderen Datensätzen die schlechtesten Ergebnisse.

7.5.3 *NIR offRoad*

Die Ergebnisse der Klassifikation von spektralen Daten unter Verwendung der offRoad Annotation sind in Abbildung 67 dargestellt. Wie zuvor sind in der linken Spalte die *NIR*-Datensätze und in der rechten Spalte die *VIS*-Datensätze dargestellt. Wie bei der semantischen Annotation zeigt der Random Forest-Klassifikator hier über alle *NIR*-Datensätze hinweg die besten Ergebnisse. Die Precision und Recall Werte sinken hierbei nie unter 85 und der Intersection Over Union Wert beträgt mindestens 57. Diese Werte sind durchweg höher als bei der semantischen Annotation. Dies liegt darin begründet, dass die Anzahl der Klassen nur ungefähr halb so groß ist. Hier werden

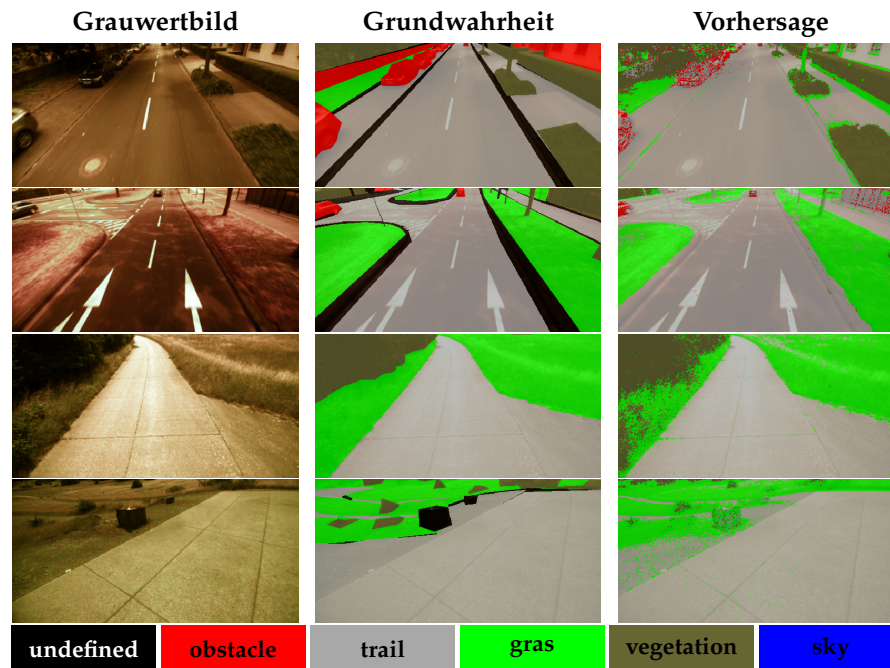


Abbildung 71: Vergleich der Klassifikationsergebnisse von *VIS*-Daten mit *offRoad*-Annotationen. Die erste Spalte zeigt eine RGB-Darstellung der spektralen Eingabedaten, gefolgt von der Grundwahrheit und der Random Forest-Klassifikation. Die Trennung von Vegetation (*vegetation*) und Gras (*grass*) ist nicht so sauber wie bei der Nutzung von *NIR*-Daten.

Klassen wie z. B. Bürgersteig und Straße, welche aufgrund ihrer Zusammensetzung spektral nur schwer zu trennen sind, zusammengelegt. Dies macht es dem Klassifikator natürlich wesentlich einfacher, ein geeignetes Modell zu lernen. Die besten Ergebnisse werden auf dem 2017Nir Datensatz erzielt. Beispielhafte Ergebnisse der Klassifikation sind in Abbildung 70 dargestellt.

7.5.4 *VIS offRoad*

Ein ähnliches Bild wie bei den *NIR*-Datensätzen zeigt sich auch bei den *VIS*-Datensätzen. Im Vergleich zur semantischen Annotation sind die Ergebnisse für Precision, Recall und Intersection Over Union wesentlich besser. Der Intersection Over Union Wert bei Datensatz 2018VisFull liegt hier bei 49 während er bei der semantischen Annotation nur bei 28 liegt. Insgesamt werden die besten Ergebnisse auf dem 2016Vis Datensatz erreicht. Ein paar Beispiele dieser Klassifikation sind in Abbildung 71 zu sehen.

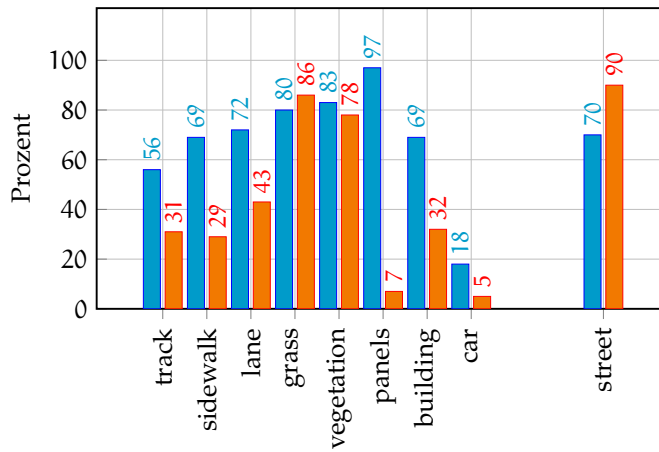
7.5.5 Dimensionsreduktion

Weiterhin wurde untersucht, ob sich die PCA-Methode nutzen lässt, um die Dimension der Daten effektiv zu reduzieren. Dazu wurde ein Random Forest-Klassifikator jeweils mit Daten der *NIR*-Kamera trainiert. In zwei Fällen wurden die Daten dafür zuvor mittels PCA auf 3 bzw. 7 Dimensionen reduziert. Dazu wurden jeweils nach der Anwendung der PCA nur die ersten 3 bzw. 7 Komponenten verwendet. Die Ergebnisse der Tests sind in Abbildung 72 dargestellt. Basierend auf den vorliegenden Daten zeigt sich, dass die Dimensionsreduktion mittels PCA keine Vorteile bei der Klassifikation bietet. Je niedriger die Dimensionalität der Daten, desto schlechter sind auch die Klassifikationsergebnisse des Klassifikators nach dem Training. Offensichtlich ist die PCA keine geeignete Methode, um die spektralen Daten der hier untersuchten Kameras zu komprimieren. Dies stimmt mit den Ergebnissen von Cheriyyadat [CB03] überein. Da die PCA auf den gesamten Datenraum angewendet wird, versucht sie, einige der globalen Eigenschaften des Datenraums zu optimieren und ignoriert einige der lokalen Eigenschaften, welche bei der Diskriminierung von Klasse unterstützen können.

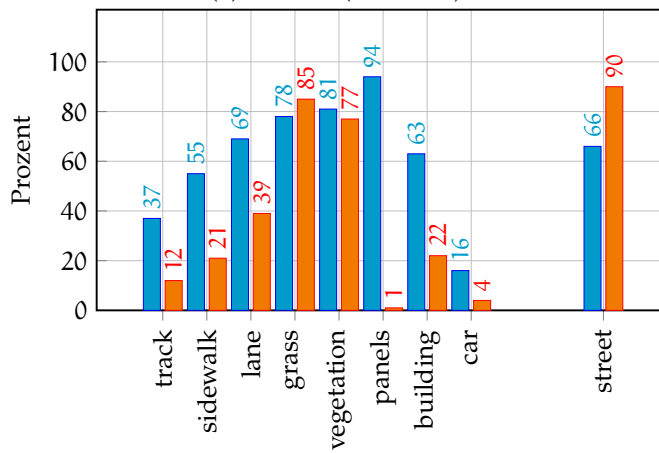
7.6 FAZIT

Die semantische Annotation stellt im Gegensatz zur *offRoad*-Annotation eine größere Herausforderung für die getesteten Algorithmen dar, da die Anzahl der Klassen fast doppelt so hoch ist. Die feinere Unterteilung der Szene in Klassen hat folgerichtig auch eine Reduktion der Daten pro Klasse zur Folge. Die Klassifikatoren entscheiden sich auffällig oft für die Klasse Straße oder Feldweg. Diese kommt natürlich auch am häufigsten in den Trainingsdaten vor. Daher ist zu erwarten, dass ein Klassifikator sich bei Mehrdeutigkeiten in den Daten für diese Klasse entscheidet. Die durchgeführten Evaluationen implizieren, dass von den getesteten Klassifikatoren der Random Forest-Klassifikator für die Klassifikation spektraler Daten in Kombination mit den Snapshot-Kameras am besten geeignet ist. Der Random Forest-Klassifikator liefert auf den getesteten Datensätzen durchweg gute Ergebnisse sowohl für die *NIR*-Kamera als auch für die *VIS*-Kamera. Aufgrund seiner Struktur kann der Random Forest-Klassifikator sehr gut parallelisiert und effektiv beschleunigt werden. Ein weiteres Ergebnis der Evaluationen ist, dass auch hier die Ausgewogenheit des Datensatzes mitentscheidend für die Qualität der Klassifikationsergebnisse ist. Diese vielversprechenden Ergebnisse sind ein erstes Beispiel für die Leistungsfähigkeit der neuartigen Sensorik und ihre Eignung zur Szenenanalyse, z. B. beim autonomen Fahren. Um die pixelweise Klassifikation zu verbessern, wird im Fol-

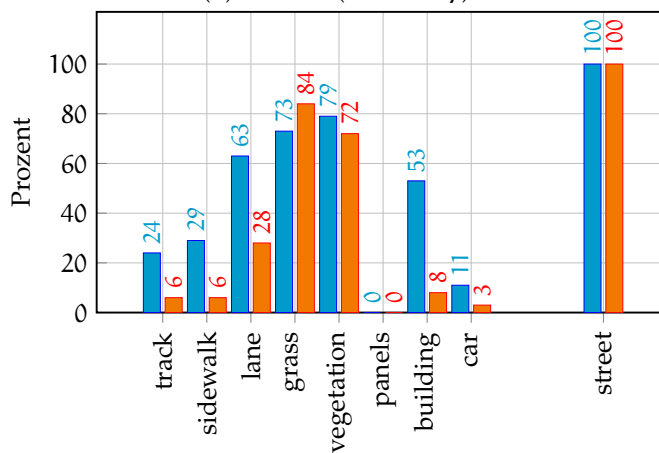
genden versucht, diese Ergebnisse mit einem bedingten Zufallsfeld (engl. *Conditional Random Field*) (CRF) zu kombinieren.



(a) NirFull (RF RAW)



(b) NirFull (RF PCA 7)



(c) NirFull (RF PCA 3)

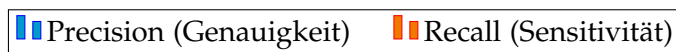


Abbildung 72: Übersicht über die Evaluationsergebnisse für NIR-Daten und semantische Annotationen unter Verwendung der PCA im Vergleich zur Verwendung der Rohdaten. Es zeigt sich, dass die Verwendung der PCA keine Vorteile bei der Klassifikation ermöglicht. Bis auf *street* zeigen alle Klassen schlechtere Ergebnisse.

8.1 EINFÜHRUNG

Viele Klassifikatoren behandeln hyperspektrale Daten als eine Reihe von spektralen Messungen und berücksichtigen keine zusätzlichen räumlichen Informationen. Folglich werden die Daten nur aufgrund ihrer spektralen Informationen klassifiziert. Diese Ansätze verwerfen Informationen, welche sich aus dem Kontext der benachbarten Pixel ergeben. Daher scheinen kombinierte spektrale und räumliche Klassifikationstechniken sinnvoll, um diesen Nachteil zu kompensieren und Unsicherheiten bei der Klassifikation zu reduzieren. So kann auch zwischen verschiedenen Strukturen unterschieden werden, welche aus den gleichen Materialien bestehen. Weiterhin lässt sich durch die Verwendung von weiteren Merkmalen auch dem Problem der Beleuchtungsänderung begegnen, da die Struktur bzw. Textur einer Oberfläche beleuchtungsunabhängig ist. Im Folgenden Abschnitt wird daher untersucht, wie die zuvor genutzten Klassifikatoren entsprechend erweitert werden können, um eine präzisere Klassifikation zu ermöglichen. Dazu werden etablierte, überwachte Klassifikatoren genutzt, um Modelle zu trainieren, welche auf zuvor berechneten Merkmalen basieren. So werden nicht nur ausschließlich die rohen spektralen Werte zur Klassifikation herangezogen. Die im Folgenden beschriebenen Arbeiten wurden in Teilen bereits auf einer internationalen Konferenz veröffentlicht [1].

8.2 STAND DER TECHNIK

Obwohl es in der Literatur einige unüberwachte Klassifikationsalgorithmen gibt, fokussiert sich diese Arbeit auf die überwachte Klassifikation. Diese hat einen höheren Verbreitungsgrad, wie Plaza et al. [PBB⁺09] zeigen und ist in der Lage den Daten explizite Klassen zuzuweisen. Unüberwachte Algorithmen können aus Ermangelung von Informationen lediglich ein Clustering vornehmen. Die Standardverfahren zur bildbasierten Umgebungswahrnehmung unter zusätzlicher Verwendung von erweiterten Merkmalen definieren sich durch die Aufnahme von RGB-Bildern, und anschließende Klassifikation durch zuvor trainierte Modelle, wie es z. B. Chetan et al. [CKJ10] vorstellen. Sie verwendeten Farbinformationen und lokale Binärmuster (engl. *Local Binary Patterns*) (LBP) in Kombination mit unterschiedlichen überwachten Klassifikatoren. Die Klassifikation von hyperspektralen Daten zeigt weiterhin einige wichtige Herausforderungen.

rungen auf. Es besteht eine große Diskrepanz zwischen der hohen Dimensionalität der Daten im Spektralbereich, ihrer starken Korrelation und der Verfügbarkeit von annotierten Daten, die für das Training unbedingt notwendig ist. Eine weitere Herausforderung ist die richtige Kombination und Integration von räumlichen und spektralen Informationen, um die Vorteile beider Domänen zu nutzen. In verschiedenen Untersuchungen von Li et al. [LBDP12] wurde beobachtet, dass Klassifikationsergebnisse verbessert werden können, indem räumliche Informationen parallel zu den Spektraldaten genutzt werden. So wurden verschiedene Anstrengungen unternommen, um kontextsensitive Informationen in Klassifikatoren für hyperspektrale Daten [PBB⁺09] zu integrieren. So hat Landgrebe [Lan05] als einer der ersten mit dem *Extraction and Classification of Homogeneous Objects* (ECHO) Klassifikator ein Verfahren vorgestellt, welches räumliche und spektrale Informationen kombiniert. Auch Fauvel et al. [FBCSo8] fusionieren morphologische und hyperspektrale Daten zur Verbesserung der Klassifikationsergebnisse.

So hat sich die Erkenntnis durchgesetzt, dass die kombinierte Nutzung von räumlichen und spektralen Informationen erhebliche Vorteile bietet. Um nun Kontextinformationen in kernelbasierte Klassifikatoren zu integrieren, kann ein Pixel gleichzeitig sowohl in der spektralen Domäne als auch in der räumlichen Domäne durch eine entsprechende Merkmalsextraktion definiert werden. Kontextbezogene Merkmale werden z. B. durch die Berechnung des Mittelwerts oder der Standardabweichung eines Bereichs pro Spektralband erzeugt.

Dies führt zu einer Familie von neuen Kernel basierten Methoden für die hyperspektrale Datenklassifikation, welche von Camps-Valls et al. [CVGCM⁺06] publiziert und mit Hilfe einer SVM implementiert wurden. Brown et al. [BS11] nutzten die PCA zur Reduzierung der Dimensionalität und schlagen eine Erweiterung des bekannten SIFT-Deskriptors [Low99] vor, der als multispektraler SIFT (MSIFT) bezeichnet wird. Salamati et al. [SLC11] untersuchte verschiedene Kombinationen von SIFT und Spektralinformationen, um die Erkennungsgenauigkeit zu verbessern. So zeigte sich, dass die Kombination aus Textur- und Farbinformationen, die aus dem sichtbaren und nahem Infrarotbereich gewonnen werden, sehr gute Ergebnisse erzielen.

Ein alternativer Ansatz zur Kombination von kontextuellen und spektralen Informationen ist die Verwendung von Markov-Netzwerken (engl. *Markov Random Fields*) (MRF). Sie nutzen die probabilistische Korrelation benachbarter Label [TFCB10] aus. Wird speziell die Literatur zur hyperspektralen Klassifikation mittels terrestrischer Spektralabbildung, bei der die Daten nicht von einer Erdumlaufbahn oder einem Flugzeug erfasst wurden, betrachtet, so finden sich hier wenige Arbeiten zu diesem Themenbereich. Ein Beispiel ist die Vegetationserkennung in hyperspektralen Daten, wie sie von Bradley et al. [BUB07] demonstriert wurde, die gezeigt hat, dass die Verwendung

des NDVI die Klassifikationsgenauigkeit verbessert.

Namin et al. [NP12] schlägt ein automatisches System zur Materialklassifizierung in natürlichen Umgebungen vor, indem es multispektrale Bilder, bestehend aus sechs visuellen und einem NIR-Band, verwendet. Auf den sieben Bändern wurden dann lokale Merkmale berechnet, um die Klassifikation robuster gegenüber Beleuchtungsänderungen zu machen.

8.3 EXTRAKTION VON MERKMALEN

Hauptziel dieses Abschnitts ist die Klassifikation von spektralen Daten mit k -Bändern unter Verwendung von räumlichen und spektralen Informationen. Dazu werden verschiedene Merkmale untersucht, die eine räumliche Beziehung zwischen einzelnen Hyperpixeln herstellen.

8.3.1 *Superpixel*

Zur Erzeugung der verschiedenen Merkmale werden die Hyperwürfel zunächst mit dem SLIC-Superpixel-Algorithmus [ASS⁺12] segmentiert, welcher im Rahmen einer betreuten Abschlussarbeit so erweitert wurde, dass er auf Hyperwürfeln arbeiten kann. Dieses Verfahren verbindet Pixel zu einem Segment, basierend auf der Entfernung im Farb- bzw. Bildraum. Für die Nutzung von RGB-Bildern, wird von den Autoren die Konvertierung in den CIE-LAB-Farbraum empfohlen, um die menschliche Wahrnehmung bei der Messung der Farbähnlichkeit zu modellieren. Im vorliegenden Fall wird der euklidische Abstand der Signalwerte verschiedener Punkte als Maß für die Ähnlichkeit herangezogen. Die so durchgeführte Segmentierung sorgt für homogene Klassifikationsergebnisse. Ein repräsentatives Segmentationsergebnis wird in Abbildung 73 dargestellt.

8.3.2 *Merkmalsextraktion*

Um zu untersuchen, ob bestimmte Merkmale die Klassifikation verbessern, müssen die zu untersuchenden Merkmale zunächst aus den in Superpixel aufgeteilten Hyperwürfel extrahiert werden. Nur so können zusätzliche Informationen gewonnen werden.

Eine große Fehlerquelle bei der Szenenanalyse ist die variable Beleuchtung der Szene, denn die Sensorik misst Daten, die sich mit der Variation der Beleuchtung bzw. Sonneneinstrahlung ändern. Um diese Varianz zu kompensieren, werden hier die Bänder einzeln normalisiert, so dass ein normiertes Spektrum als Merkmal genutzt werden kann. Das hyperspektrale Gegenstück zur log-chromatischen Darstellung nach Finlayson et al. [FHLDo6] von RGB-Bildern wird verwendet, um den Einfluss der Szenenbeleuchtung und anderer Unregelmä-

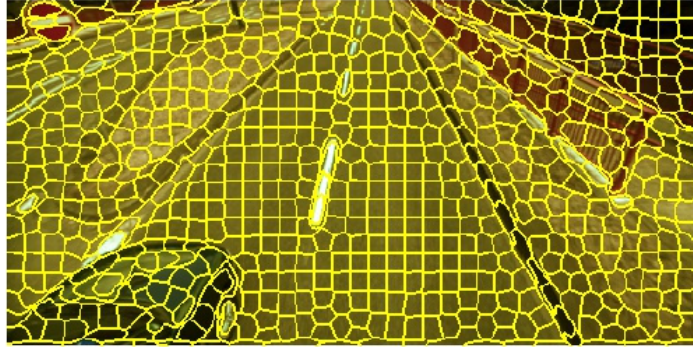


Abbildung 73: Pseudo-RGB Darstellung der spektralen Daten, die mit der Segmentierungsmaske überlagert ist. Die Kantenlänge ist auf 15 Pixel eingestellt, so dass die Segmente jeweils ca. 225 Pixel umfassen.

ßigkeiten zu kompensieren. Im RGB-Fall erfolgt die Normalisierung der Werte eines Pixels an der Position (i,j) durch das geometrische Mittel $\bar{\mathbf{B}}$ der drei Kanäle bzw. der Bänder \mathbf{B} .

$$\bar{\mathbf{B}} = \sqrt[3]{\prod_{n=1}^3 \mathbf{B}_n} \quad (53)$$

Im hyperspektralen Fall ist die Anzahl der Bänder viel höher, was zu numerischen Instabilitäten führen kann, wobei die Berechnung der Wurzeln höherer Ordnung hier als Ursache zu sehen ist. Stattdessen wird die Normalisierung durch die Summe der \mathcal{N}_λ der Werte des gemessenen Spektrums verteilt über die Bänder

$$\bar{\mathbf{B}} = \sum_{n=1}^{\mathcal{N}_\lambda} \mathbf{B}_n \quad (54)$$

entsprechend zu [GSWoo] erreicht. Das sog. normierte Spektrum für einen Hyperpixel an der Stelle (i,j) wird durch die Normierung der einzelnen Spektralbänder \mathbf{B}_k wie folgt berechnet

$$\hat{\mathbf{B}}_k(i,j) = \frac{\mathbf{B}_k(i,j)}{\bar{\mathbf{B}}(i,j)} \quad (55)$$

Zusätzlich wird die Summe aller Komponenten des Spektrums $\bar{\mathbf{B}}$ zum Merkmalsvektor hinzugefügt, um dessen Helligkeit darzustellen.

Um den Merkmalsraum mit den im Bildraum vorhandenen Texturen

zu erweitern, wird eine Gabor-Filterbank [NWO96] eingesetzt. Durch entsprechende Parametrisierung wird eine Menge gleichmäßig verteilter Filterkerne in Orientierung, Maßstab und Frequenz erzeugt. Für die verfügbaren Daten wurden sechs Orientierungen für die Kernel mit jeweils sechs Kombinationen aus Skala und Frequenz verwendet. In natürlichen Umgebungen sind Texturen, die das gleiche Material oder die gleiche Oberfläche darstellen, oft unterschiedlich ausgerichtet, z. B. können Grashalme zur Seite geneigt sein, anstatt gerade zu stehen. Daher werden Filterbänke mit unterschiedlicher Ausrichtung kombiniert. Bei sechs Skalen, die für jede Orientierung verwendet werden, würden sechs Merkmale pro Spektralband dem Merkmalsvektor hinzugefügt. Bei der großen Anzahl verfügbarer Spektralbänder in hyperspektralen Daten würde es daher zu einer zu hohen Dimensionalität kommen, wenn die Merkmale pro Band berechnet würden. Dies ist aber gar nicht notwendig, da lediglich Texturmerkmale gewonnen werden sollen. Daher werden die Gabor-Merkmale auf Grauwertbildern berechnet, anstatt die Filterbank auf jedes Band einzeln anzuwenden. Die dafür zugrunde liegende Grauwertdarstellung setzt sich aus dem Mittelwert der einzelnen Bänder für jedes Hyperpixel zusammen.

Um ein abstrakteres Texturelement zu erhalten, werden zusätzlich die Welligkeit und Granularität berücksichtigt. In Kombination mit den Graustufen-Merkmalen von Gabor kann so eine Vielzahl von möglichen Textureigenschaften im Merkmalsraum dargestellt werden. Zusätzlich kann Domänenwissen genutzt werden, um die Klassifikationsgenauigkeit zu verbessern. Die Bildsituation für ein fahrendes Fahrzeug zeigt konstante Bedingungen, z. B. ist der Himmel typischerweise im oberen Teil des Bildes, während befahrbares Gelände dort nicht zu finden ist. Entsprechend kann die Einbeziehung von Informationen auf *Gravitationsbasis* die Klassifikation unterstützen, indem sie diese Bedingungen modelliert. So werden die Wahrscheinlichkeiten aller möglichen Klassenzugehörigkeiten für jede y -Koordinate aus den verfügbaren Grundwahrheiten extrahiert und in einer Lookup-Tabelle gespeichert. Bei der Merkmalsextraktion werden die Wahrscheinlichkeiten für individuelle Klassenzugehörigkeiten für die jeweilige y -Koordinate der Bildposition aus der Lookup-Tabelle ausgelesen und ebenso im Merkmalsvektor gespeichert. Dieses Merkmal ist natürlich auf das beschriebene Szenario eines autonomen Bodenfahrzeugs beschränkt. Andere Situationen, wie z. B. die Flugsteuerung einer Drohne, weisen unterschiedliche vertikale Anordnungen von Szenenkomponenten auf. Dennoch wurde es in dieser Arbeit verwendet, um zu untersuchen, wie es sich auf die Klassifikationsgenauigkeit der anderen Merkmale auswirkt.

Zur Berechnung werden die vorgeschlagenen Merkmale aus den Superpixeln des segmentierten Hyperwürfels extrahiert und über die Pixel des Superpixel-Segments gemittelt. Die Samples für die Gabor-

basierten und die gravitationsbasierten Texturmerkmale werden aus der Mitte des jeweiligen Superpixel-Segments entnommen. Das Zentrum \mathbf{p}^c wird mit Hilfe des ersten Moments

$$\mathbf{p}^c = \frac{1}{N} \sum_{i=0}^N \mathbf{p}_i \quad (56)$$

für die N Pixel \mathbf{p}_i des Segments anhand der zweidimensionalen Koordinaten bestimmt, aus denen sich das Segment zusammensetzt. Dies kann auch als Massenmittelpunkt verstanden werden. Für das normierte Spektrum wird das Merkmal des Segments nicht aus dem Zentrum extrahiert, sondern für jedes Pixel berechnet und dann der Mittelwert der Pixel für jedes Spektralband zurückgegeben. So wird für jedes Segment ein 32-dimensionaler Merkmalsvektor generiert, wenn alle Merkmale genutzt werden. Dazu gehören die normierten Spektren, die aus den 16 normierten Bändern, bei der Nutzung der VIS-Daten bestehen. Zusätzlich wird die Summe aller Bänder berechnet und gespeichert. Beide Gabor-Merkmale tragen jeweils sechs Elemente bei, da sechs Skalen und Frequenzen in der Filterbank zur Verfügung stehen. Das gravitationsbasierte Merkmal liefert für jede zu erkennende Klasse einen Wert, die sich in diesem Fall auf vier Werte summieren.

8.4 EVALUATION

Der Klassifikator, der für die Auswertung verwendet wird, ist der Random Forest [Bre01], welcher sich im vorherigen Kapitel als erfolgreich erwiesen hat.

Zum Training des Klassifikators wurde ein Datensatz aus den in Kapitel 5 beschriebenen Datensätzen verwendet und zum Extrahieren von Merkmalen genutzt. Die Annotation setzt sich in diesem Fall aus 4 Klassen zusammen, welche die Befahrbarkeit der Oberfläche beschreiben.

Im Folgenden ist eine Übersicht der Merkmale dargestellt welche einzeln als Merkmalsvektor untersucht und zum Training genutzt wurden:

Untersuchte Merkmale als Merkmalsvektor:

- A Einfaches Spektrum (Raw)
- B Normiertes Spektrum
- C Graustufen-Gabor
- D Welligkeit und Granularität
- E Normiertes Spektrum und gravitationsbasiertes Merkmal

F Kombination aller vorherigen Merkmale

Von Namin et al. [NP12] publizierte Merkmale:

G Grauwertmatrix (GLCM)

H Grauwertmatrix (GLCM) and einfaches Spektrum

I Einfaches Spektrum, Standardabweichung im Segment, GLCM und Fourier Merkmale

Um die Merkmale zu vergleichen, werden zusätzlich ein paar Merkmale aus der Publikation von Namin et al. verwendet. Diese zeigten gute Ergebnisse und lassen sich grundsätzlich auch auf die hier vorliegenden Daten übertragen. Die Klassifikationsergebnisse auf dem Datensatz für die oben genannten Merkmale sind in Tabelle 14 dargestellt. Die jeweils höchsten Werte pro Klasse sind farblich hervorgehoben. In Abbildung 74 sind die Klassifikationsergebnisse der

Sensitivität

Klasse	Merkmalsvektor								
	A	B	C	D	E	F	G	H	I
sky	82.02%	83.95%	75.24%	75.11%	89.17%	89.65%	80.45%	81.42%	81.94%
drivable	81.18%	84.48%	69.53%	69.58%	86.09%	85.76%	67.39%	75.08%	74.70%
rough	66.41%	66.31%	49.67%	52.34%	66.68%	64.48%	59.34%	54.78%	56.11%
obstacle	67.97%	70.26%	65.39%	66.38%	74.01%	74.56%	63.44%	64.58%	64.52%
Mittelwert	74.40	76.25	64.96	65.85	78.99	78.61	67.66	68.97	69.32

Genauigkeit

Klasse	Merkmalsvektor								
	A	B	C	D	E	F	G	H	I
sky	97.22%	97.67%	94.18%	92.63%	98.26%	98.15%	88.42%	94.55%	94.22%
drivable	89.63%	93.88%	70.65%	74.12%	97.10%	97.20%	84.21%	78.79%	78.58%
rough	76.16%	79.72%	65.39%	67.02%	83.75%	83.64%	64.99%	73.76%	73.37%
obstacle	88.03%	90.12%	81.39%	82.12%	95.41%	95.09%	80.36%	81.93%	83.23%
Mittelwert	87.76	90.35	77.90	78.97	93.63	93.52	79.50	82.26	82.35

Tabelle 14: Ergebnisse der Klassifikation für den ausgewählten Datensatz. Die besten Ergebnisse pro Klasse sind jeweils grün hervorgehoben. Insgesamt zeigen die Merkmalsvektoren E und F die besten Ergebnisse. Dies zeigt, dass die Verwendung des normierten Spektrum sich vorteilhaft auswirkt.

verschiedenen Merkmale am Beispiel eines Hyperwürfels dargestellt. Hier zeigt sich in Abbildung 74c die Fehlinterpretation von Schatten als Hindernis und die Spiegelung des Himmels auf der Motorhaube wird fälschlicherweise als Himmel klassifiziert. Wenn ein gemittelttes Spektrum der Superpixel-Segmente als Merkmal verwendet wird, scheitert die Klassifizierung in manchen Bereichen. Die Klassifikation

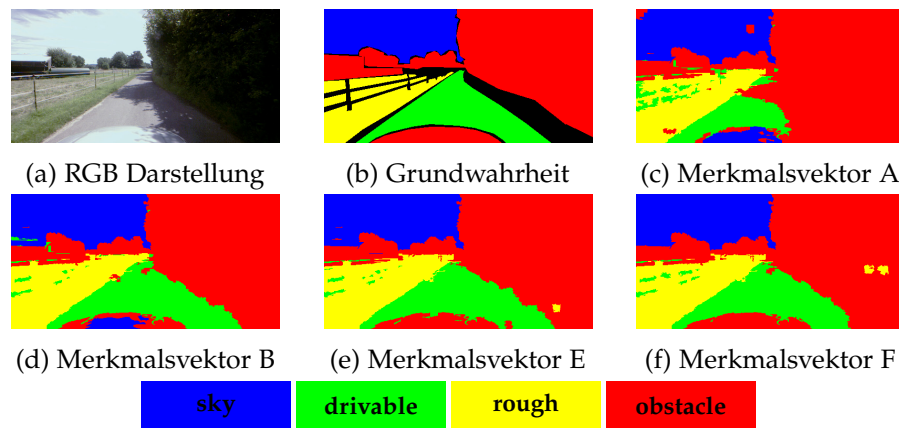


Abbildung 74: Vergleich der Klassifikation anhand von einfachen Spektren, normierten Spektren und verschiedenen Merkmalskombinationen. Im Vergleich zu 74c zeigt 74e wesentlich bessere Ergebnisse. Der Schatten wird bspw. nicht mehr falsch klassifiziert.

unter Verwendung des normalisierten Spektrums ist in dem Bild, das in Abbildung 74d gezeigt wird, wesentlich besser. Schattierte Bereiche werden in den meisten Fällen korrekt erkannt und auch andere Details stimmen besser mit der Grundwahrheit als bei der Verwendung des reinen Spektrums überein. Die Genauigkeit des trainierten Klassifikators mit normiertem Spektrum liegt bei 90%, während derjenige, der die einfachen Spektren der Segmente verwendet, nur eine Genauigkeit von 88% auf den verfügbaren Daten erreicht.

Die anderen vorgeschlagenen Merkmale zeigen eher schlechte Performance, wenn sie ausschließlich verwendet werden. Die Graustufen-Gabor-Merkmale erreichen so eine Erkennungsrate von 78%. Die Welligkeits- und Granularitätsmerkmale schneiden ähnlich ab mit einer Genauigkeit von 79%. Es zeigt sich, dass in Kombination mit dem normierten Spektrum bessere Klassifikationsergebnisse erzielt werden können. Die Kombination aus normiertem Spektrum und dem gravitationsbasierten Merkmal zeigt mit 93,63% im Vergleich mit allen Merkmalskombinationen die besten Ergebnisse im Bezug auf die Genauigkeit.

Zum Vergleich der vorgeschlagenen Merkmale mit Merkmalen aus der Literatur wurden einige der in [NP12] beschriebenen Merkmale zum Trainieren eines Klassifikators genutzt. Die Genauigkeit dieser Merkmale ist mit 79,50% und 82,26% schlechter als diejenige, welche mit einfachen Spektren erreicht wurde. Dafür gibt es mehrere Gründe: Schatten wurden in diesem Datensatz nicht wie in der Arbeit von Namin extra behandelt, indem sie im Training mit einer eigenen Klasse versehen wurden.

Die in Abbildung 75 gezeigten Ergebnisse bestätigen, dass die meis-

ten Schattenbereiche daher als Hindernisse klassifiziert wurden. Des Weiteren ist die Funktion zur Vegetationserkennung [BUBo7] nicht verfügbar, da die verwendete VIS-Kamera Infrarotlicht nicht erkennt. Dieser Ansatz war mit ausschlaggebend für die guten Ergebnisse, die in der Literatur erzielt wurden. Daher kann natürlich nicht geschlossen werden, dass die Merkmale von Namin grundsätzlich schlechter sind, vielmehr eignen sie sich in dem vorliegenden Fall nicht so gut wie in anderen Szenarien. So konnte laut den Ergebnissen die Berechnung des GLCM für alle Bänder keine relevanten Zusatzinformationen liefern. Ergebnisse der Klassifikation unter Verwendung der Merkmale von Namin et al. sind in Abbildung 75 visualisiert. Hier sind auch deutlich Artefakte auf der rechten Seite der Motorhaube des Fahrzeugs zu sehen.

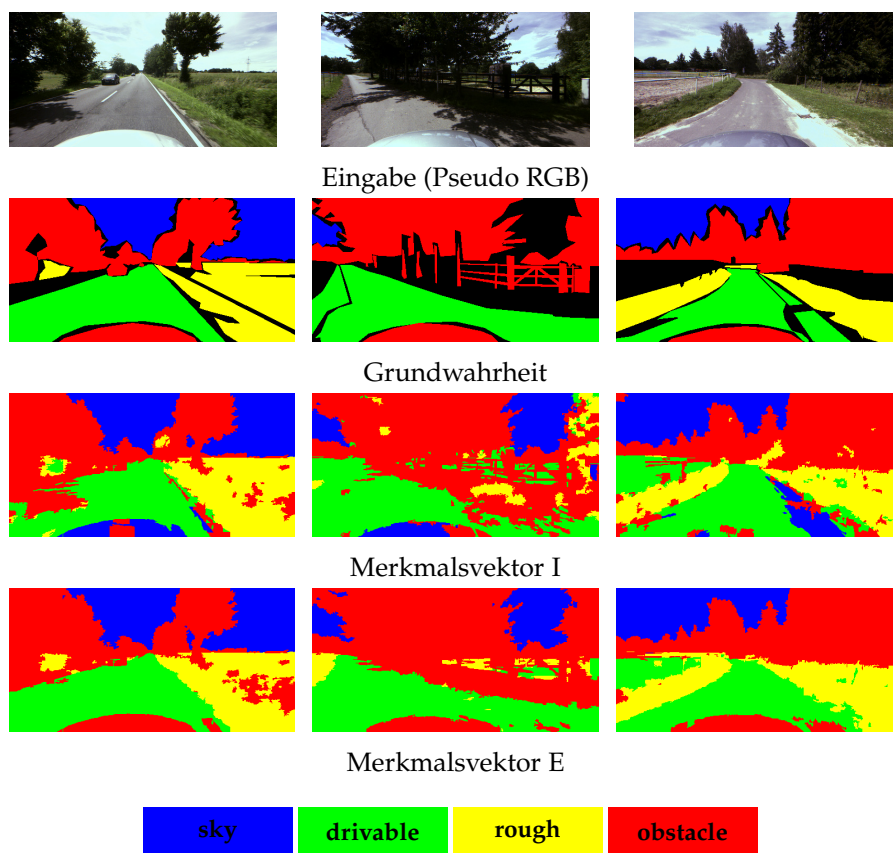


Abbildung 75: Vergleich der Klassifikationsergebnisse. Die erste Zeile zeigt eine RGB-Darstellung der spektralen Eingabedaten. Die folgenden Zeilen jeweils die Grundwahrheit, den von [NP12] vorgeschlagenen Merkmalsvektor und die Kombination aus normalisierten Spektren und dem in dieser Arbeit vorgestellten gravitationsbasierten Merkmal. Der Merkmalsvektor E liefert insgesamt die besseren Ergebnisse.

8.5 FAZIT

In diesem Abschnitt wurden sowohl spektrale als auch räumliche Informationen für die Klassifizierung von spektralen Daten untersucht. Die untersuchten Merkmale sind leicht zu extrahieren, was ihre Verwendung in der Online-Szenenanalyse ermöglicht. Basierend auf den erfassten spektralen Daten konnten Straßen- oder befahrbare Bereiche präzise von nicht befahrbaren Bereichen wie unwegsamem Gelände oder Hindernissen unterschieden werden.

Dies könnte die Leistung der Analyse von spektralen Daten zur Umgebungswahrnehmung verbessern. Insbesondere die Verwendung des normierten Spektrums verbesserte die Gesamtgenauigkeit der Klassifikation und wird im weiteren Verlauf der Untersuchung weiter verwendet. Obwohl die vorgeschlagenen Gabor-Textur-Merkmale die Genauigkeit nicht ausreichend verbessern, um ihre Verwendung zu rechtfertigen, könnten diese Textureigenschaften in anderen Szenarien, oder auch auf den normalisierten Daten, durchaus noch mächtiger sein.

9.1 EINFÜHRUNG

Das Klassifikationsproblem besteht in der Regel aus zwei Teilen. Dem Extrahieren von Merkmalen aus den Eingabedaten und dem Labeln der extrahierten Merkmale durch einen bestimmten Klassifikator. So berücksichtigen viele Algorithmen der Literatur neben dem Problem der expliziten Modellierung hyperspektraler Daten nur spektrale Variationen von Pixeln. Sie ignorieren so räumliche Korrelationen und behandeln jedes Pixel unabhängig. Neben den spektralen Korrelationen haben sich die räumlichen Korrelationen aber als sehr nützlich für die Bildanalyse sowohl in der Fernerkundung als auch in der Bildverarbeitung erwiesen.

Hier werden die räumlichen Korrelationen auch als Kontextinformationen oder räumliche Muster bezeichnet. Basierend auf räumlichen Mustern ist es plausibel anzunehmen, dass benachbarte Pixel mit hoher Wahrscheinlichkeit zur selben Klasse gehören.

Um dies zu nutzen, sind Ansätze zur Verwendung von Kontextinformationen wesentlich durch zwei Strategien bestimmt. In der ersten Variante werden die Kontextinformationen in die extrahierten Merkmale integriert. Bei der zweiten Variante werden diese Informationen durch die intrinsische Struktur des Klassifikators modelliert.

In diesem Abschnitt wird die Integration von Kontextinformationen in die Klassifikation der spektralen Daten zur Szenenanalyse diskutiert. Dazu wird eine Klassifikationspipeline erläutert, die sowohl spektrale als auch Kontextinformationen verwendet, um ein konsistentes Segmentierungsergebnis zu erzielen. Dazu wird zunächst ein Random Forest-Klassifikator eingesetzt, um eine initiale per-Pixel-Klassifikation zu erhalten, welche dann als Eingabe für ein angepasstes, vollständig verbundenes bedingtes Zufallsfeld (engl. *Fully Connected Conditional Random Field*) (FCRF) dient. Dieses ermittelt paarweise Potentiale auf allen Pixelpaaren und kann so die Segmentierungsergebnisse verbessern. Die im Folgenden beschriebenen Arbeiten wurden in Teilen bereits auf einer internationalen Konferenz veröffentlicht [2].

9.2 STAND DER TECHNIK

Eines der Standardverfahren zur bildbasierten Szenenanalyse wird definiert, indem reguläre RGB-Bilder erfasst werden und versucht wird, verschiedene Klassen zu identifizieren, wie Chetan et al.

[CKJ10] und andere zeigen. Sie benutzten Farbinformationen und weitere Merkmale wie lokale binäre Muster (LBP) und trainieren verschiedene überwachte Klassifikatoren. Wie zuvor beschrieben gibt es zwei Wege Kontext in den Prozess zu integrieren, der eine ist, spezielle Merkmale zu entwickeln. Die Geschichte der Kontextmerkmale beginnt 1973 mit Haralick [HS⁺73]. Die Haralick Merkmale werden zur Quantifizierung von Bilddaten anhand von Texturinformationen verwendet. Das grundlegende Konzept zur Berechnung von Haralick-Textur-Merkmalen basiert auf Graustufen Matrizen (GLCM) welche auf den Nachbarn eines Zentrumspixels berechnet werden. Fortan wurden weitere Merkmale vor allem auch im Bereich der hyperspektralen Datenanalyse entwickelt [KEK90, SA99, RDFZ04, UBo4]. Beim zweiten Ansatz werden passende Klassifikatoren genutzt, um Kontext zu integrieren. Hier haben sich z. B. MRFs etabliert, wie von Elia et al. [DPS03], Salzenstein und Collet [SCo6], sowie Neher et al. [NS05] demonstrieren. Weiterhin präsentierten Galleguillos et al. [GRBo8] im Jahr 2008 ein Verfahren, welches gleichzeitig Textur- und Kontextinformationen erfasst. Diese Informationen werden in ein CRF integriert, welches wie ein MRF zusätzlich auch die Nachbarschaftsinformationen mit berücksichtigt. Shotton et al. [SWRC09] führten im Jahr 2009 eine neuartige Methode zur effizienten Erkennung und semantischen Szenenanalyse aus Bilddaten ein. So schlugen sie neue Funktionen zur Merkmalsextraktion vor, die Layout-, Textur- und Kontextinformationen erfassen und in einem Merkmalsvektor bündeln, welcher als *Texton* bezeichnet wird. Die Klassifikation ist außerdem zusätzlich in ein CRF integriert, was die Klassifikationsgenauigkeit erhöht.

Eine segmentweise Szenenanalyse für urbane Straßenszenen, die eine Superpixel-Darstellung verwendet, wird von Ess et al. [EMGVG09] im selben Jahr vorgeschlagen. Fulkerson et al. [FVS09] zeigten eine Methode, um Objekte in Bildern zu identifizieren und zu lokalisieren. Auch hier werden Superpixel als Grundeinheit zur Segmentierung genutzt. Dazu wird dann ein Klassifikator basierend auf Histogrammen lokaler Merkmale trainiert, welche aus den Superpixeln berechnet wurden. Schließlich wird das Segmentationsergebnis durch den Einsatz eines bedingten Zufallsfelds CRF im Superpixel-Diagramm noch weiter verbessert. Ein Ansatz, bei dem lokale Deskriptoren wie SIFT [Low99], Local Binary Patterns [HW90, OPH96] (LBP) verwendet und mit zusätzlichen Bildinformationen angereichert werden, wurde von Carreira et al. [CCBS12] vorgestellt. Dazu werden zunächst Superpixel ähnliche Freiformbereiche definiert, auf denen dann Merkmale berechnet und anschließend klassifiziert werden. Das führte zu guten Ergebnissen bei der Pascal VOC 2011.

Einen Schritt weiter gingen 2013 Wojek et al. [WWR⁺13]. Sie entwickeln ein 3-D-Szenenverständnis von städtischen Verkehrsszenen unter Verwendung eines probabilistischen Szenenmodells und ei-

ner Monokularkamera. Das System führt monokulare 3-D-Szenenrekonstruktionen in realistischen Verkehrsszenen durch und ermöglicht so eine zuverlässigere Erkennung von Objekten. Im selben Jahr kombinieren Scharwaechter et al. [SEFR13] Stixel und ein spezielles Klassifikationsschema zur semantischen Segmentierung auf Graustufen- und Tiefendaten aus einem Stereokamerasystem. Liu et al. [LLS15] zeigen einen Ansatz als Kombination von CRF und SVM zur semantischen Segmentierung, welche auf Merkmalen basiert, die aus einem vortrainierten neuronalen Faltungsnetzwerk zur Bildsegmentierung stammen. Im Gegensatz zu vorherigen Ansätzen werden hier nicht mehr die statischen Merkmale wie SURF, SIFT oder LBP genutzt, sondern mit zuvor trainierten Merkmalen gearbeitet. Das CRF stellt wiederum sicher, dass am Ende ein homogenes Ergebnis entsteht, bei dem auch die Nachbarschaftsinformationen mit betrachtet werden. Im Jahr 2016 gingen Chen et al. [CPK⁺16] einen Schritt weiter und ein System mit der Bezeichnung *Deeplab* vor, das trainierte Netzwerke zur bildbasierten semantischen Szenensegmentierung verwendet. Sie kombinierten dazu neuronale Faltungsnetze und vollständig verbundene bedingte Zufallsfelder für detaillierte und homogene Segmentierungsergebnisse.

Fast alle Algorithmen zur semantischen Szenenanalyse verwenden RGB-Bilddaten zur Klassifikation. Jedoch hat in den letzten Jahren die hyperspektrale Bildgebung und Klassifikation zusätzliches Interesse auf sich gezogen. Daneben gibt es eine Reihe von Algorithmen, welche für die hyperspektrale Datenverarbeitung und -analyse geeignet erscheinen. In den Jahren 2008 [ZW08] und 2010 [ZW10] präsentierten Zhong und Wang erste Ansätze ein CRF zur Klassifikation von hyperspektralen Satellitendaten zu nutzen. So war es möglich, räumliche und spektrale Abhängigkeiten gleichzeitig zu modellieren. Cavigelli et al. [CBMB16] analysierte das Potential hyperspektraler Kameras in Kombination mit tiefen neuronalen Netzen zur semantischen Klassifikation. Sie analysierten die Daten von spektralen Kameras auf einem kleinen Datensatz mit statischem Hintergrund. Eine Kombination aus Adaboost, *Rotation-Forest* und CRF wird von Li et al. [LXS⁺15] genutzt, um hyperspektrale Satellitendaten zu klassifizieren. Hier fungieren die Daten der Klassifikatorkombination als unäres Potential für das CRF welches in einem abschließenden Verfeinerungsschritt verwendet wird. Alam et al. [AZL16] haben sich im Jahr 2016 von Zheng et al. [ZJRP⁺15] inspirieren lassen und analysierten hyperspektrale Satellitendaten durch eine Kombination von Faltungsnetzwerken (CNN) und CRF. Dazu erzeugen sie zunächst Superpixel, die auf spektralen und räumlichen Informationen der Pixel entlang der Bänder eines Hyperwürfels basieren. Anschließend werden die Superpixel mit einem Faltungsnetzwerk klassifiziert. Die Ausgabe des Klassifikators dient dann als Eingabe für weitere spezielle Schichten des CNN. Diese Schichten emulieren das Verhalten eines

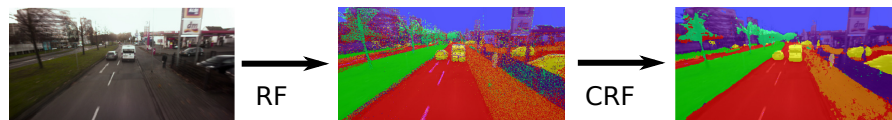


Abbildung 76: Pipeline zur kontextsensitiven Klassifikation. Nach der Klassifikation der Daten mittels eines Random Forest (RF) Klassifikators wird ein Conditional Random Field (CRF) genutzt, um das Ergebnis der Klassifikation zu verbessern.

CRF. Somit kann die komplette Klassifikationspipeline in einem neuronalen Netz abgebildet werden. Ebenso haben Alam et al. [AZL⁺19] mit Faltungsnetzen datenbasierte Merkmale gelernt und formulierten dann ein CRF bei denen die unären und paarweisen Potentiale auf den gelernten Merkmalen basieren.

Speziell die bedingten Zufallsfelder haben bei der Klassifikation von Daten in den letzten Jahren größere Beliebtheit erlangt, da sie in der Lage sind, kontextuelle Informationen zu erfassen und damit z. B. Segmentierungsergebnisse zu verbessern.

9.3 SZENENANALYSE

Der bisher verwendete Random Forest-Klassifikator führt eine per-Pixel-Klassifikation durch, entsprechend unterliegen die Ergebnisse einem gewissen Rauschen. Die grundlegende Idee, welche in diesem Abschnitt erläutert wird, ist es, die Klassifikationsergebnisse, welche von einem Random Forest geliefert werden, als Eingabe für eine vollständig verbundenes CRF zu nutzen. Dieses führt dann einen zweiten Klassifikationsschritt durch und glättet die Ergebnisse, wie in Abbildung 76 dargestellt. Dazu wird zunächst im Folgenden das CRF näher erläutert.

9.3.1 Bedingtes Zufallsfeld

Das bedingte Zufallsfeld (CRF) wurde im Jahr 2001 von Lafferty et al. [LMP01] publiziert. Es ist ein probabilistischer Ansatz zur Klassifikation und Segmentierung strukturierter Daten. Es kombiniert die Vorteile von Klassifikation und grafischer Modellierung, indem es die Fähigkeit zur kompakten Modellierung multivariater Daten mit der Fähigkeit kombiniert, eine große Anzahl von Eingabefunktionen für die Vorhersage zu nutzen. So sei X ein Zufallsfeld und seien Y gege-

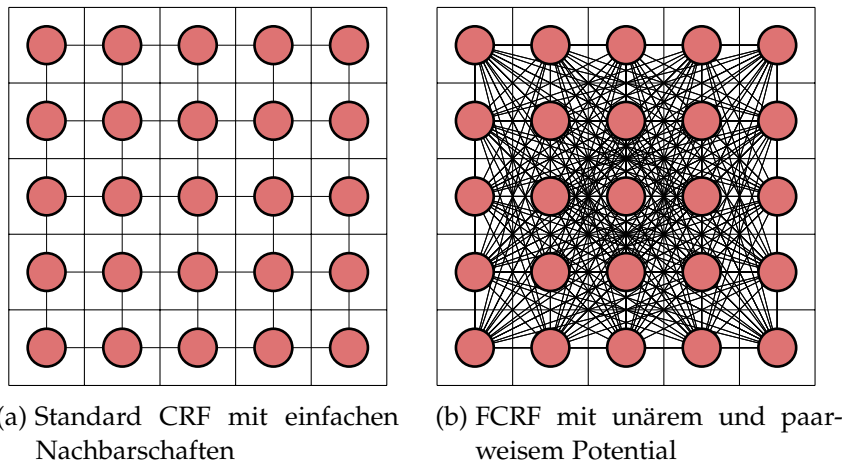


Abbildung 77: Schematische Beschreibung verschiedener CRFs auf einem Pixelgitter. Das Gittermuster definiert die Pixelstruktur eines Bildes. Die Knoten des darüberliegenden Graphen repräsentieren jeweils das Pixel und die zugehörigen Kanten modellieren Nachbarschaftsbeziehungen zwischen den Pixeln, auf denen dann Potentiale definiert werden können. Über die Kanten sind alle Pixel innerhalb der Gitterstruktur direkt miteinander verbunden.

bene Werte. So definiert $P(\mathbf{X}|\mathbf{Y})$ ein bedingtes Zufallsfeld unter der Bedingung:

$$P(\mathbf{X}|\mathbf{Y}) = \frac{1}{Z(\mathbf{X})} \prod_C \exp(Q(C, \mathbf{X})) \quad (57)$$

Mit C als eine Clique¹ über \mathbf{X} und $Z(\mathbf{X})$ als einer Normierungskonstante, so dass die Summe von $P(\mathbf{X}|\mathbf{Y})$ über alle möglichen Instanzen des Zufallsfeldes 1 ergibt. Weiterhin definiert $\exp(Q(C, \mathbf{X}))$ eine Potentialfunktion.

Nun ist eine Verteilung für das Zufallsfeld gesucht, welche am besten die gegebenen Daten beschreibt. Dies entspricht derjenigen Verteilung mit der maximalen informationstheoretischen Entropie. Dies entspricht:

$$\hat{\mathbf{X}} = \operatorname{argmax}_{\mathbf{X}} P(\mathbf{X}|\mathbf{Y}; \xi) \quad (58)$$

Mit ξ als Gewichtungsfaktor der Potentialfunktion, welcher aus den Daten geschätzt wird.

¹ Eine Clique ist eine Teilmenge von Knoten in einem Graph, welche alle vollständig über Kanten verbunden sind.

Bei der Bildklassifikation werden kontextuelle Informationen aus benachbarten Klassen (Cliques) verwendet, was die Menge der Informationen erhöht, die das Modell für eine gute Vorhersage nutzen kann. Dazu werden Glattheitsterme definiert, welche die Gleichheit von Annotation benachbarter Pixel annehmen, und auch kontextuelle Beziehungen zwischen Objekten modellieren können. Das führt zu einer Energiefunktion, welche minimiert werden muss.

Die Energie wird dabei als eine Art Kostenfunktion betrachtet. Durch die Zuordnung der wahrscheinlichsten Klasse zu jedem Pixel kann eine geringere Energie bzw. niedrigere Kosten und damit eine höhere Genauigkeit erzielt werden.

CRF-Modelle zur Segmentierung bestehen aus zwei Arten von Potentialen. Die unären Potentiale sind auf einzelnen Pixeln oder Bildpatches definiert und spezifizieren die Kosten einer Klassenzuordnung, die mit der des Klassifikators nicht übereinstimmt. Unär bedeutet, dass nur die Klassenzuordnung des einzelnen Pixels zu jeder Zeit berücksichtigt wird. Die paarweisen Potentialen sind auf benachbarten Pixeln oder Patches definiert. Sie spezifizieren die Kosten für zwei ähnliche Pixel z. B. Nachbarpixel oder Pixel, die mit ähnlicher Farbe unterschiedliche Klassenzuordnungen besitzen. Das Ergebnis aus dieser Kombination ist eine Nachbarschaftsstruktur, die in einem CRF kodiert ist, wie in Abbildung 77a dargestellt. Die Grundstruktur eines klassischen CRF ist jedoch sehr begrenzt in der Fähigkeit, weitreichende Zusammenhänge und Beziehungen zu modellieren. Dadurch werden Kanten von Objekten in der Regel zu stark geglättet, wie Abbildung 78b zeigt. Um die Segmentierung bzw. Szenenanalyse zu verbessern, haben Galleguillos et al. [GRBo8] und Rabinovich et al. [RVG⁺07] das ursprüngliche CRF-Framework um eine hierarchische Konnektivität und paarweise Relationen erweitert. Diese Konnektivität führt dazu, dass alle Knoten, welche die Pixel des Bildes darstellen, in einem Graphen miteinander verbunden sind, wie in Abbildung 77b visualisiert. Dies ermöglicht präzisere Segmentierung von Objekten, wie in Abbildung 78d dargestellt. Aber die Komplexität der Inferenz in vollständig verbundenen CRFs ist sehr hoch, was die Einsatzmöglichkeiten einschränkt. Zur Verfeinerung der pixelbasierten Klassifikation wird daher in dieser Arbeit eine Implementierung eines vollständig verbundenen bedingten Zufallsfelds (Fully Connected CRF) genutzt, wie es von Krähenbühl et al. [KK11, KK13] veröffentlicht wurde. Dort wird ein vollständig verbundenes CRF verwendet, welches paarweise Potentiale auf allen Pixelpaaren im Bild aufbaut. Die dazu entwickelte Implementierung liefert einen hocheffizienten Inferenzalgorithmus für vollständig verbundene CRF-Netze.

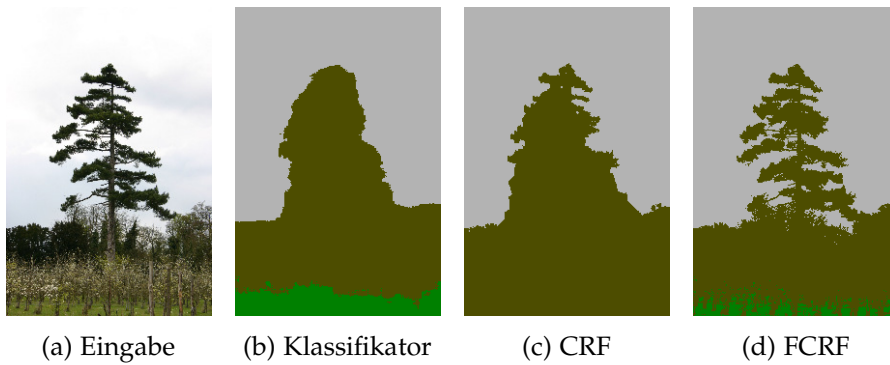


Abbildung 78: Vergleich der pixelbasierten Klassifikation mit verschiedenen Graph-Modellen. Hier zeigt 78b die Ausgabe des Klassifikators bei dem die feinen Strukturen des Baumes verloren gegangen sind. In einem Nachverarbeitungsschritt wurde ein einfaches CRF genutzt, um die Strukturen wiederherzustellen, dies ist in 78c zu sehen. Einige der Strukturen des Baumes wurden rekonstruiert, jedoch ein Großteil des Baumes noch nicht. Das Ergebnis unter Verwendung eines FCRF ist in 78d dargestellt. Dank den umfassenden Nachbarschaftsinformationen konnte die Struktur des Baumes rekonstruiert werden. Bildquelle: [KK11]

9.3.2 Fully Connected CRF

Nach [KK11] ist ein bedingtes Zufallsfeld CRF \mathbf{X} definiert über eine Menge $\mathbf{X} = \{X_1, \dots, X_n\}$ von Variablen, deren Domäne durch eine Menge von Klassen spezifiziert ist $\mathbf{Y} = \{y_1, \dots, y_k\}$. In der Bildverarbeitung definiert dabei eine Menge $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ von Pixeln ein Eingabebild \mathbf{f} mit der Pixelmenge n als Eingabedaten. Weiterhin definiert \mathbf{X} die mögliche per-Pixel Klassenzuordnung des Bildes. Ein CRF (\mathbf{f}, \mathbf{X}) wird nun durch eine Gibbs-Verteilung wie folgt beschrieben:

$$P(\mathbf{X}|\mathbf{f}) = \exp\left(-\sum_{c \in \mathbf{C}_G} \phi_c(\mathbf{X}_c|\mathbf{f})\right) \quad (59)$$

Dabei wird ein Graph auf \mathbf{X} definiert, bei der jede Clique C ein Potential ϕ_c definiert. Weiterhin wird die Gibbs-Energie E einer Klassenzuordnung $\mathbf{x} \in \mathbf{Y}^n$ als

$$E(\mathbf{x}|\mathbf{f}) = \sum_{c \in \mathbf{C}_G} \phi_c(\mathbf{x}_c|\mathbf{f}) \quad (60)$$

definiert. So ergibt sich die Maximum-a-posteriori Klassifikation (Labeling):

$$\hat{\mathbf{x}} = \operatorname{argmax}_{\mathbf{x} \in \mathbf{Y}^n} P(\mathbf{x}|\mathbf{f}) \quad (61)$$

Bei einem vollständig verbundenen CRF ist im Gegensatz zum klassischen CRF die Clique C_G auf dem kompletten Graphen definiert. So wird die korrespondierende Gibbs-Energie E der Klassenzuordnungen \mathbf{x} für die Eingabedaten \mathbf{f} wie folgt definiert:

$$E(\mathbf{x}) = \sum_i \underbrace{\psi_u(\mathbf{p}_i)}_{\text{Unärer Term}} + \sum_{i < j} \underbrace{\psi_p(\mathbf{p}_i, \mathbf{p}_j)}_{\text{Paarweiser Term}}$$

mit i und j im Bereich von 1 bis $n = \mathcal{N}_x \cdot \mathcal{N}_y$. Hier definiert $\psi_u(\mathbf{p}_i) = -\log P(\mathbf{p}_i)$ das unäre Potential, wobei $P(\mathbf{p}_i)$ die Zuordnungswahrscheinlichkeit einer Klasse für Pixel \mathbf{p}_i ist. Diese wird unabhängig für jedes Pixel von einem Klassifikator berechnet, welcher eine Wahrscheinlichkeitsverteilung über die verfügbaren Klassen anhand von Bildmerkmalen erzeugt. Die paarweisen Potentiale

$$\sum_i \sum_{i < j} \psi_p(\mathbf{p}_i, \mathbf{p}_j) \quad (62)$$

werden als Mischungen von Gauß-Kernen im Merkmalsraum wie folgt modelliert:

$$\psi_p(\mathbf{p}_i, \mathbf{p}_j) = \mu(\mathbf{p}_i, \mathbf{p}_j) \sum_{m=1}^K w^{(m)} k^{(m)}(\mathbf{c}_i, \mathbf{c}_j)$$

wobei $k^{(m)}$ einen Gauß-Kern und $w^{(m)}$ eine Linearkombination von Gewichten definiert. Die Funktion $\mu(\mathbf{p}_i, \mathbf{p}_j)$ spezifiziert eine einfache Klassen-Kompatibilitätsfunktion, sie führt einen Strafterm für benachbarte Pixel mit ähnlichem Signalwert ein, denen aber unterschiedliche Klassen zugeordnet wurden. Die Vektoren \mathbf{c}_i und \mathbf{c}_j sind Merkmalsvektoren in einem konstruierten Merkmalsraum. Für Multi-Klassen-Probleme werden zwei Kernel-Potentiale definiert, welche die Merkmalsvektoren \mathbf{c}_i und \mathbf{c}_j für die Pixel \mathbf{p}_i und \mathbf{p}_j enthalten.

$$k(\mathbf{c}_i, \mathbf{c}_j) = w^{(1)} \underbrace{\exp\left(-\frac{|\mathbf{p}_i^p - \mathbf{p}_j^p|^2}{2\sigma_\alpha^2} - \frac{|\mathbf{p}_i^v - \mathbf{p}_j^v|^2}{2\sigma_\beta^2}\right)}_{\text{Erscheinungsbild}} + w^{(2)} \underbrace{\exp\left(-\frac{|\mathbf{p}_i^p - \mathbf{p}_j^p|^2}{2\sigma_\phi^2}\right)}_{\text{Glattheit}}$$

Ein Kern spezifiziert in diesem Beispiel die Wahrscheinlichkeit, dass Pixel mit der gleichen Farbe (für ein RGB-Bild) zur gleichen Klasse gehören. Hier definiert als die Positionen \mathbf{p}_i^p und \mathbf{p}_j^p sowie die Intensitätswerte \mathbf{p}_i^I und \mathbf{p}_j^I der jeweiligen Pixel. Weiterhin kontrollieren σ_α^2 und σ_β^2 Nähe und Ähnlichkeit der beiden Pixel. Der Glattheitskern hingegen entfernt kleine, isolierte Bereiche. Das Besondere an einem

vollständig verbundenen CRF ist, dass die paarweisen Potentiale auf allen Pixelpaaren im Bild erzeugt werden. So hat dieser paarweise Term eine Form, die eine effiziente Inferenz ermöglicht, sodass ein vollständig verbundener Graph genutzt werden kann.

9.3.3 Eigener Ansatz

Um eine semantische Szenenanalyse durchführen zu können, muss ein geeignetes Modell mit Hilfe eines Random Forest-Klassifikators trainiert werden. Da in dieser Arbeit zwei Kameras mit unterschiedlichen Wellenlängenempfindlichkeiten untersucht werden, müssen zwei getrennte Modelle trainiert werden. Wie bereits in Kapitel 3.7 erwähnt, bildet ein vorverarbeitetes Bild einen Hyperwürfel \mathbf{f}^H mit einem gemessenem Spektrum χ das aus 16 bzw. 25 Bändern für jeden Hyperpixel \mathbf{p}^H besteht. Für das Training werden die annotierten Hyperwürfel wie in Kapitel 7 zunächst in einzelne Hyperpixel \mathbf{p}^H zerlegt. Zum Training wird dann das gemessene Spektrum $\bar{\chi}_i$ von \mathbf{p}_i^H mit normalisierten Bändern als Merkmalsvektor $\mathbf{c}_i = \bar{\chi}_i$ verwendet, wie in Kapitel 8 erläutert.

Dies entspricht dann einer per-Pixel-Klassifikation des Hyperwürfels. Da die Daten nur Pixel für Pixel von der Random Forest klassifiziert werden, unterliegen die Ergebnisse für einen kompletten Hyperwürfel einem gewissen Rauschen, da keine Nachbarschaftsinformationen verwendet werden. Zur Nutzung der spektralen Daten im Zusammenhang mit dem FCRF wurde die Implementation von Krähenbühl entsprechend angepasst und erweitert. So wird zunächst, wie zuvor, der Random Forest-Klassifikator genutzt, um die oben beschriebenen Klassen-Zuordnungswahrscheinlichkeiten $P(\mathbf{p}_i^H)$ pro Hyperpixel gemäß Term (52) zu bestimmen. Der Klassifikator berechnet so eine Wahrscheinlichkeitsverteilung über die verfügbaren Klassen, welche dann als Eingabe für das unäre Potential $\psi_u(\mathbf{p}_i^H)$ genutzt werden. Für den paarweisen Term $\psi_p(\mathbf{p}_i^H, \mathbf{p}_j^H)$ wird ein sechsdimensionaler Merkmalsraum w^6 bestehend aus Pixelposition, einem lokalen binären Mustermerkmal (LBP) sowie, im Falle der VIS-Kamera, ein Pseudo-RGB Wert. Für die NIR-Kamera werden drei ausgewählte Kanäle im Bereich um die rote Kante definiert, welche an das FCRF übergeben werden.

9.4 EVALUATION

Wie zuvor wird die Evaluation auf den bereits erwähnten Datensätzen durchgeführt. Zur Evaluation werden die semantischen und offRoad Annotationen genutzt, um die Verwendung von spektralen Daten für das Verständnis semantischer Szenen zu untersuchen.

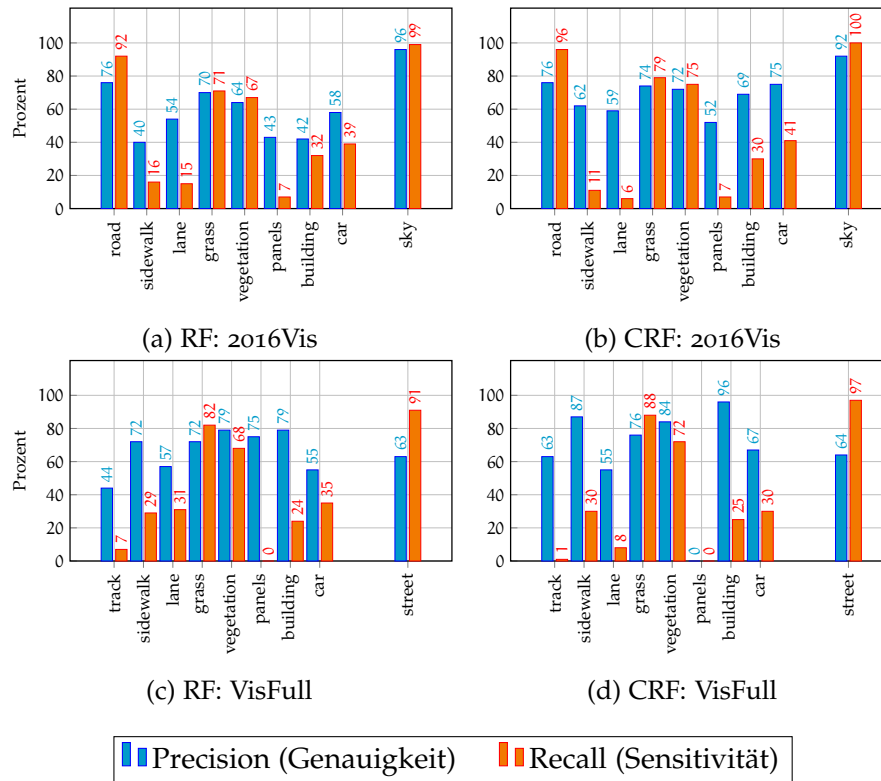


Abbildung 79: Übersicht über die Evaluationsergebnisse für VIS-Daten und semantische Annotationen. Vor allem die Klassen *grass*, *vegetation* und *street* profitieren von der Verwendung eines CRF.

9.4.1 VIS semantisch

Die Ergebnisse der Experimente mit semantisch annotierten VIS-Daten werden in Abbildung 79 und mit einigen Beispielen in Abbildung 83 dargestellt. Werden die hier vorliegenden Ergebnisse betrachtet, so zeigt sich das bei fast allen Klassen die Genauigkeit durch die Verwendung des CRF erhöht werden konnte. So konnte wohl vor allem die Anzahl der Falsch Positiven gesenkt werden. Des Weiteren zeigt sich, dass auch die Sensitivität erhöht werden konnte. Vor allem die Klassen, welche große Flächen abdecken wie Straße und Wiese, haben davon profitiert. Klassen mit *feinen* Elementen wie Straßenmarkierungen haben teilweise niedrigere Sensitivitätswerte. Werden die Beispielbilder betrachtet, bestätigt sich dieser Eindruck vor allem bei Betrachtung des Beispiels in der dritten Spalte. Hier ist gut zu sehen, dass Straße und Wiese sauberer voneinander getrennt sind, sowie die Autos an der Kreuzung einheitlicher auch als Auto erkannt wurden.

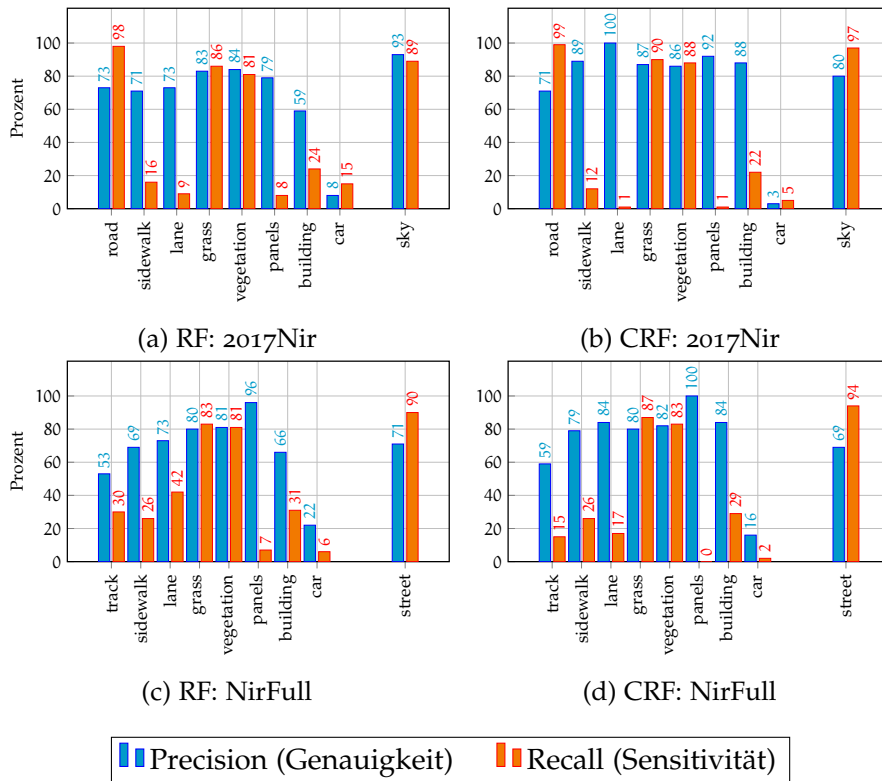


Abbildung 80: Übersicht über die Evaluationsergebnisse für NIR-Daten und semantische Annotationen. Die Klassen *grass*, *vegetation* und *street* profitieren von der Verwendung eines CRF.

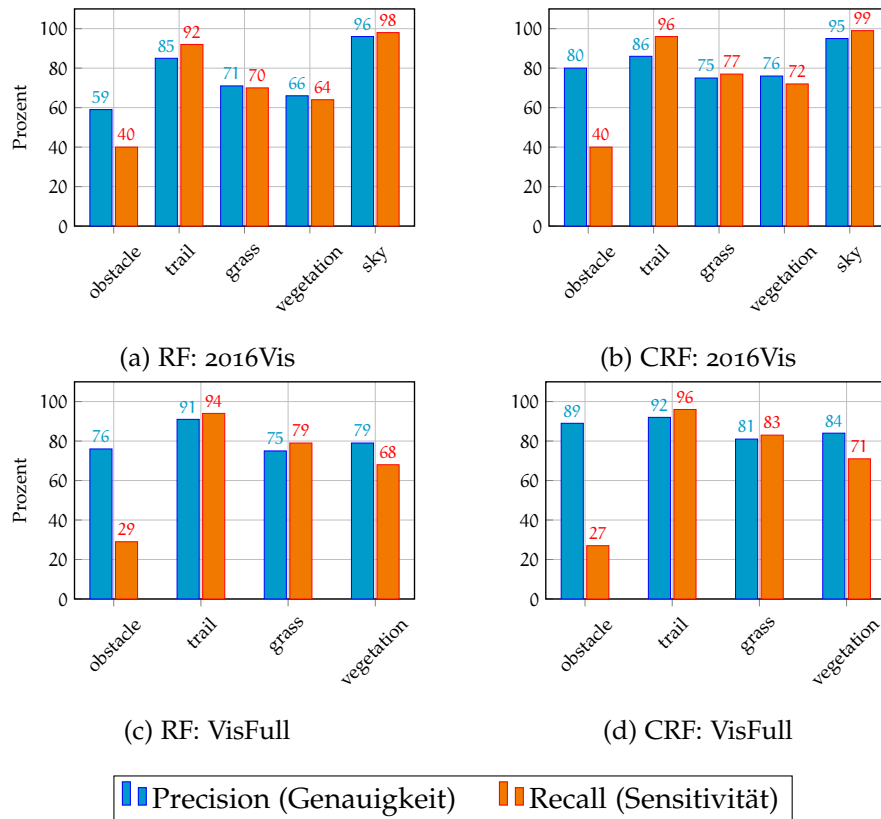


Abbildung 81: Übersicht über die Evaluationsergebnisse für VIS-Daten und offRoad Annotationen. Vor allem die Klasse *obstacle* profitiert von der Verwendung eines CRF.

9.4.2 NIR *semantisch*

Die Ergebnisse der Experimente mit semantisch annotierten NIR-Daten werden in Abbildung 80 sowie mit einigen Beispielen in Abbildung 84 dargestellt. Bei den Ergebnissen hier zeigt sich ein ähnliches Bild wie bei den semantischen VIS Daten, die Genauigkeit steigt durchgehend teils erheblich. Allerdings sinkt hier teilweise die Sensitivität bei *kleinen* Elementen der Szene wie z. B. den Fahrbahnmarkierungen. Das kommt daher, dass die Fahrbahnmarkierung im Vergleich zur Straße kaum vorhanden sind. Das wurde in Kapitel 5 bei der Analyse der Datensätze diskutiert. Werden die dargestellten Beispiele betrachtet, zeigt sich aber auch hier, dass durch die Verwendung des CRF eine saubere Trennung der Oberflächen in der Szene möglich ist.

9.4.3 VIS *offRoad*

Die Ergebnisse der Experimente mit semantisch annotierten VIS-Daten werden in Abbildung 81 und Abbildung 85 dargestellt. Auch

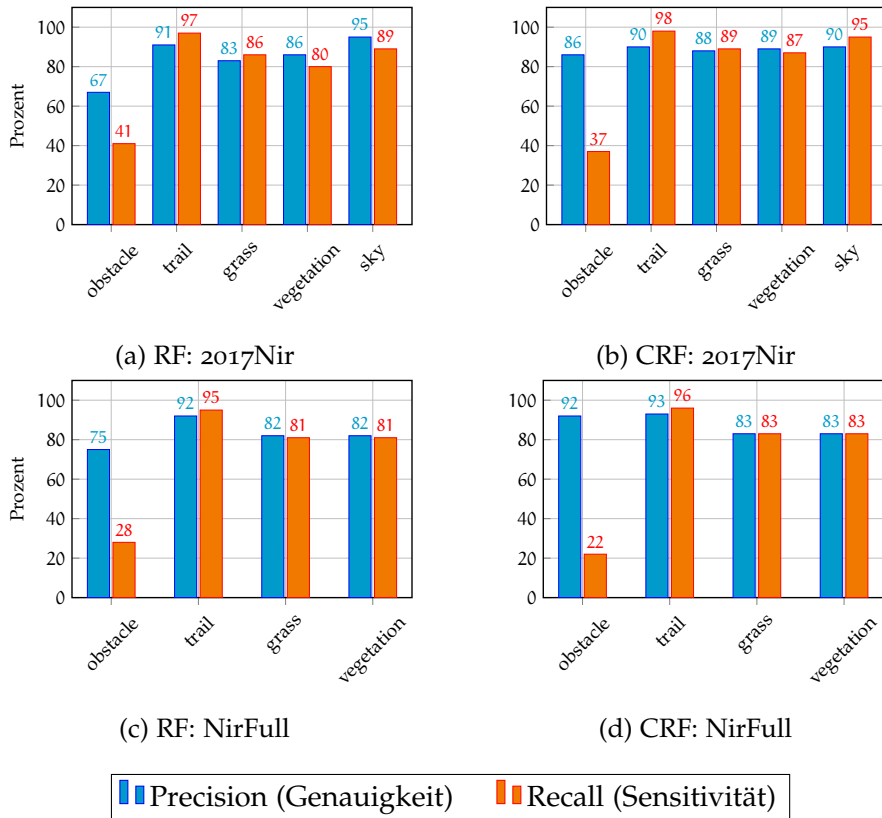


Abbildung 82: Übersicht über die Evaluationsergebnisse für NIR-Daten und offRoad Annotationen. Alle Klassen erzielen bessere Ergebnisse durch die Verwendung eines CRF.

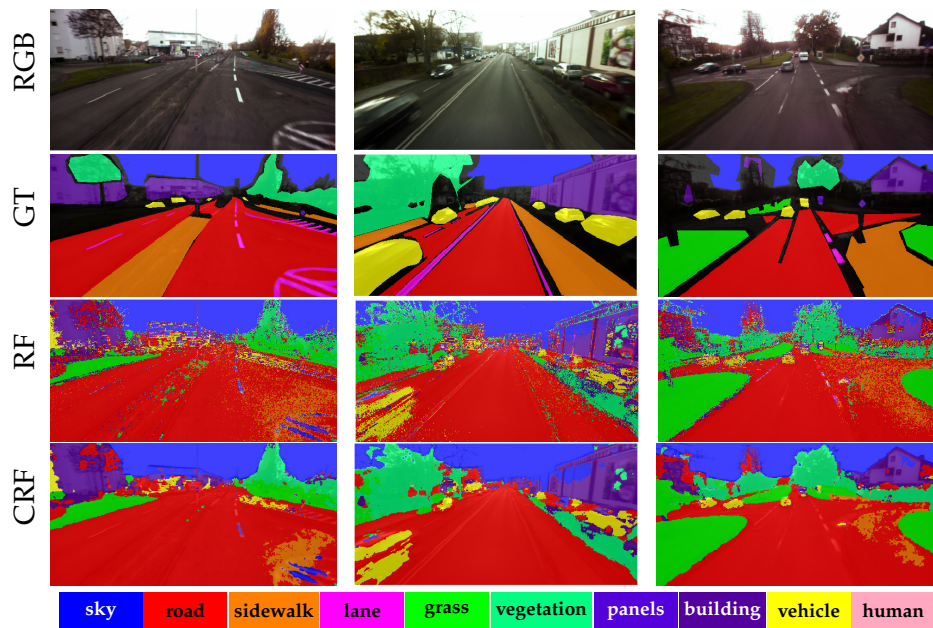


Abbildung 83: Vergleich der Klassifikationsergebnisse von *VIS*-Daten (2016Vis) mit semantischen Annotationen. Vor allem die Vegetation und Straße werden nun sauberer getrennt und klassifiziert.

hier zeigt sich das die Verwendung des CRF bei allen Klassen zu erhöhten Genauigkeitswerten führt, selbiges gilt hier auch für die Sensitivität. Da bei dieser Annotation aufgrund der reduzierten Anzahl an Klassen quasi keine Klassen mit wenigen Datenpunkten vorhanden sind, ist die Gefahr auch geringer, das Bereiche in der Szene marginalisiert werden. Werden die Beispiele betrachtet, ist direkt zu sehen, dass das CRF eine bessere Trennung zwischen den Oberflächen ermöglicht sowie die Oberflächen an sich auch besser klassifiziert werden.

9.4.4 NIR *offRoad*

Die Ergebnisse der Experimente mit semantisch annotierten *VIS*-Daten werden in Abbildung 82, Abbildung 86 und Abbildung 87 dargestellt. Bei den Ergebnissen hier ist direkt zu erkennen, dass der (engl. *Random Forest*) (RF)-Klassifikator schon sehr gute Ergebnisse erzielt. Doch werden auch hier fast durchgehend die Ergebnisse durch den Einsatz des CRF verbessert. Dies zeigt sich auch bei der Betrachtung der Beispiele, das Klassifikationsergebnis des CRF ist teilweise nahezu identisch zur Grundwahrheit und ermöglicht so eine Weiterverwendung zur Navigation.

Unter Berücksichtigung der Ergebnisse der CRF-Klassifikation ist zu erkennen, dass die semantische Szenenanalyse durch hinzufügen

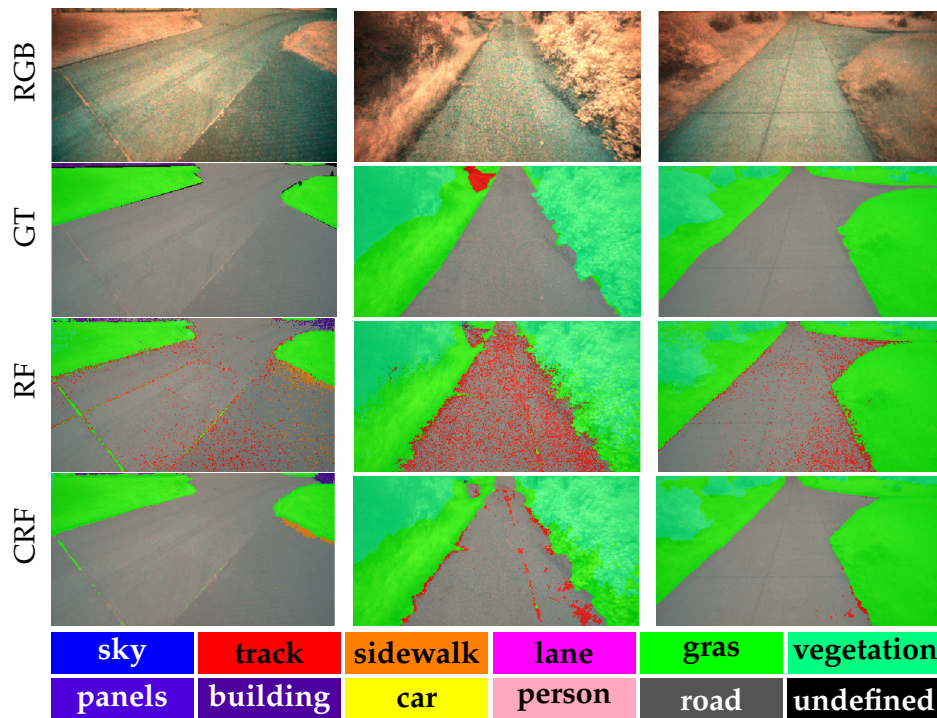


Abbildung 84: Vergleich der Klassifikationsergebnisse von *NIR*-Daten (NirFull) mit semantischen Annotationen. Hier sind vor allem suburbane Szenen zu sehen. In der Mitte ist zu sehen, wie vor allem Fehler bei der Trennung von *road* und *track* korrigiert werden.

von Kontextinformationen verbessert wird. Im Vergleich mit der per-Pixel-Klassifikation ist zu sehen, dass viele Ausreißer im Bereich der Straße und anderen Oberflächen entfernt wurden. Betrachtet man die Ergebnisse, so ist zu erkennen, dass sich durch die Verwendung eines CRF die Unterscheidung zwischen Vegetation und Gras verbessert. Dies ist eine wichtige Voraussetzung, um autonome Navigation im Gelände und abseits von Straßen zu ermöglichen. Darüber ist so eine zuverlässige Klassifikation von befahrbaren Straßen möglich. Die Ergebnisse zeigen, dass die Verwendung von vollständig verbundenen CRFs die Leistung der Klassifikation in der Szenensegmentierung erhöhen kann.

9.5 FAZIT

In diesem Abschnitt wurde eine Pipeline zur spektralen Szenenanalyse vorgestellt, welche eine per-Pixel-Klassifikation mit vollständig verbundenen CRFs kombiniert. Die Verwendung eines CRF ermöglicht die Integration von Kontextinformationen in den Klassifikationsprozess, was eine lokale Konsistenz ermöglicht. Die untersuchte Kom-

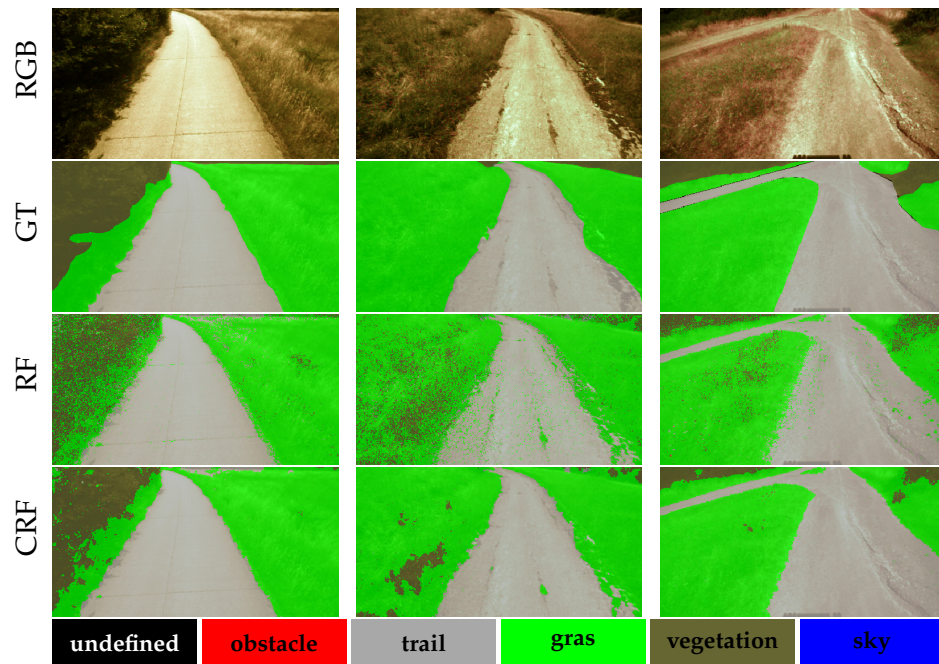


Abbildung 85: Vergleich der Klassifikationsergebnisse von *VIS*-Daten (VisFull) mit offRoad Annotationen. Durch die Verwendung des CRF werden über alle Klassen hinweg bessere Ergebnisse erzielt. Vor allem die Klassifikation von *obstacle* Bereichen verbessert sich erheblich.

bination aus spektralen Daten und dichter Konnektivität auf Pixel-Ebene führt zu einer genaueren Szenenanalyse, wie die Experimente nahe legen. So ist eine Nutzung der Daten zur Unterstützung des autonomen Fahrens möglich.

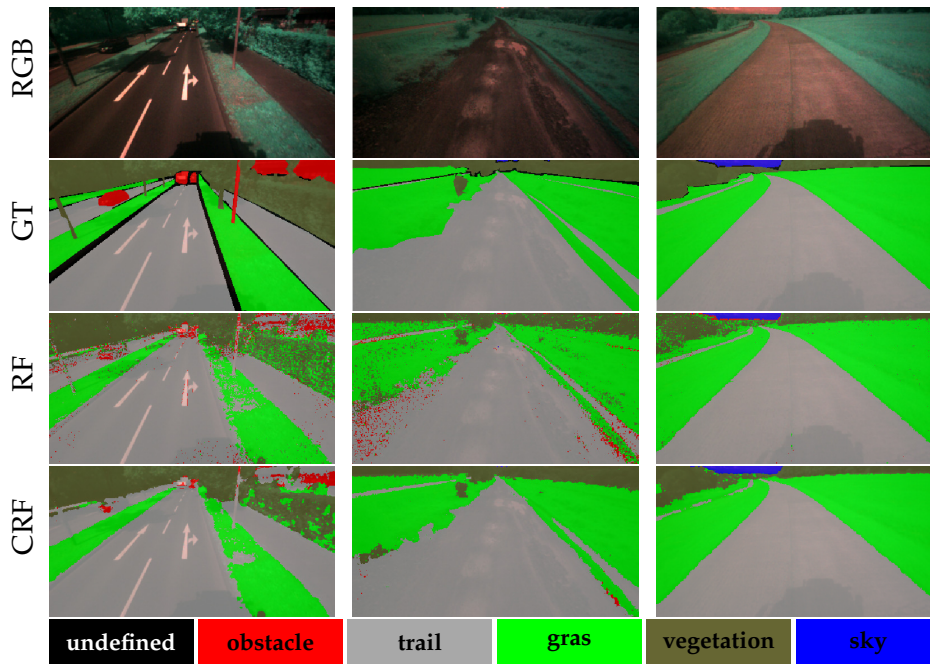


Abbildung 86: Vergleich der Klassifikationsergebnisse von *NIR*-Daten (2017Nir) mit offRoad Annotationen. Die Trennung der Klassen *trail*, *grass* und *vegetation* verbessert sich erheblich.

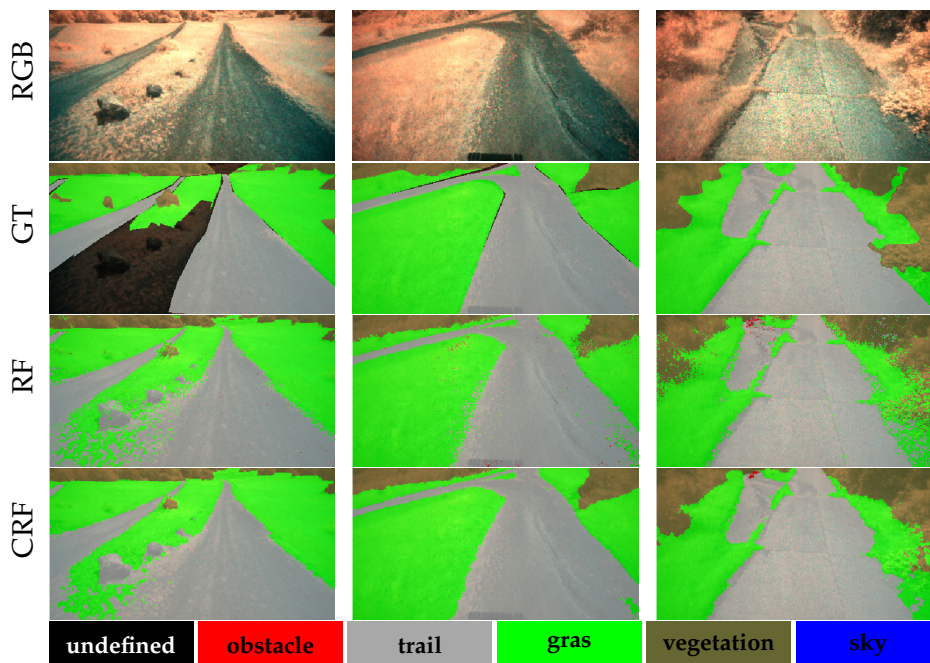


Abbildung 87: Vergleich der Klassifikationsergebnisse von *NIR*-Daten (NirFull) mit offRoad Annotationen. Die Trennung der Klassen *vegetation* und *grass* verbessert sich erheblich.

NEURONALE NETZE

10.1 EINFÜHRUNG

Die Szenenanalyse ist eines der Kernprobleme der Bildverarbeitung, welches praktische Anwendungen wie z. B. autonomes Fahren und Augmented Reality hat. Das Problem wurde in der Vergangenheit mit verschiedenen traditionellen Techniken des maschinellen Sehens und maschinellen Lernens gelöst. Doch trotz der Popularität dieser Methoden haben ihnen (engl. *Deep Learning*) Techniken wie neuronale Netze mittlerweile auch in diesem Gebiet den Rang abgelaufen. Im Wesentlichen ist Deep-Learning auch ein Teil in der Gruppe des maschinellen Lernens, welcher auf dem Lernen von Datenrepräsentationen und nicht auf aufgabenspezifischen Algorithmen basiert. Es werden Eingabedaten verschiedenster Art durch Schichten und Neuronen so gefiltert, dass eine den Daten zugrunde liegende Struktur gelernt werden kann.

Ähnlich wie bei einem menschlichen Gehirn verwenden neuronale Netze dabei mehrschichtige Filter, bei denen jede Schicht von der vorherigen Schicht lernt und ihre Ergebnisse dann an die nächste Schicht weitergibt. So werden die Daten über die einzelnen Schichten weitergereicht und verarbeitet, bis die letzte Schicht eine finale Ausgabe der definierten Fragestellung liefert. Dabei kann der Ausgabewert abhängig von der Netzarchitektur und den Eingabedaten kontinuierlich, binär oder kategorisch sein. Um eine sinnvolle Ausgabe zu produzieren, muss das neuronale Netz mit annotierten Daten trainiert werden. Und so erlaubt es das Netz im Idealfall auch Daten zu klassifizieren, die zuvor noch nicht gesehen wurden. In diesem Abschnitt werden verschiedene Neuronale Netze aus dem Bereich des Deep-Learning vorgestellt und diskutiert. Des Weiteren wird untersucht, ob sie zur semantischen Segmentierung mit spektralen Daten trainiert werden können.

Eine der beliebtesten Arten von neuronalen Netzen sind sog. Faltungsnetze (engl. *Convolutional Neural Network*) (CNN). Das CNN faltet gelernte Merkmale mit Eingangsdaten und verwendet dazu 2D-Faltungen.

10.2 STAND DER TECHNIK

10.2.1 Faltungsnetze (CNN)

Ein Faltungsnetz CNN ist ein spezielles neuronales Netzwerk. Es nutzt sog. Faltungsschichten (engl. *convolutional layers*), diese speziellen Schichten *falten* die Eingabedaten und generieren so Merkmale, welche für die Vorhersage und Klassifikation genutzt werden können. Schichten mit Faltungen haben gegenüber den Faltungen der klassischen Bildverarbeitung den Vorteil, dass die Parameter der Faltung während des Trainings gelernt werden. Fast alle CNN-Architekturen folgen den gleichen allgemeinen Konstruktionsprinzipien. Nacheinander werden verschiedene Faltungsschichten auf die Eingabe angewendet, um z. B. die räumlichen Dimensionen periodisch zu verkleinern und gleichzeitig die Anzahl der Merkmalskarten (engl. *feature maps*) zu erhöhen. Während die klassischen Netzwerkarchitekturen einfach aus gestapelten Faltungsschichten (engl. *stacked convolutional layers*) bestehen, gehen moderne Architekturen neue Wege bei der Konstruktion der Faltungsschichten, so dass ein effizienteres Lernen möglich ist. Die meisten modernen Architekturen basieren auf wiederholbaren Einheiten, welche im gesamten Netzwerk verwendet werden. Die hier aufgeführten Architekturen dienen gemeinhin als Vorlagen, welche angepasst und modifiziert werden, um verschiedene Aufgaben des maschinellen Sehens zu lösen. Sie werden üblicherweise als Merkmalsextraktoren genutzt, um Aufgaben aus den Bereichen der Objekterkennung, Klassifizierung, Segmentierung und anderen zu erfüllen. Im Folgenden werden die relevantesten Architekturen der letzten Jahre vorgestellt:

LENET-5 Das LeNet-5-Modell von Lecun et al. [LBBH98] wurde 1998 entwickelt, um handschriftliche Ziffern für die Postleitzahlerkennung zu identifizieren. Dieses Modell führte erstmals Faltungen in einem neuronalen Netzwerk ein. Die LeNet5-Architektur, in Ab-

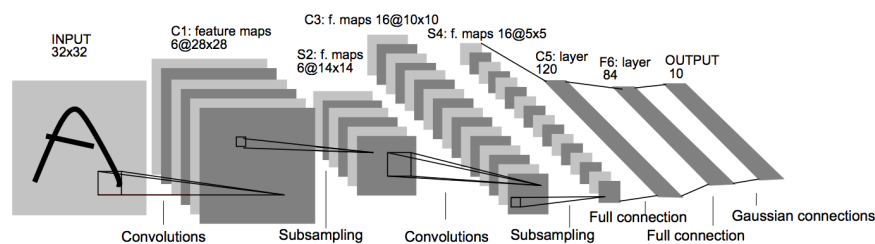


Abbildung 88: Architektur des LeNet-5, ein Faltungsnetzwerk zur Nummernerkennung. Jede Ebene ist eine Merkmalskarte (engl. *Feature-Map*). Quelle: [LBBH98]

bildung 88 dargestellt, war grundlegend neu. Es wurden Faltungen

mit lernbaren Parametern genutzt, um ähnliche Merkmale an mehreren Stellen im Bild mit wenigen Parametern zu extrahieren. Lecun erklärte die Verwendung einzelner Pixel des Bildes als separate Eingabefunktionen in der ersten Schicht seien nicht sinnvoll, da die Pixel im Bild stark räumlich korreliert sind. Daher wurden folgende Neuerungen eingeführt:

- Subsampling unter Verwendung des räumlichen Mittelwertes (engl. *pooling*)
- Faltung, um räumliche Merkmale zu extrahieren
- Nichtlinearität der Aktivierungsfunktion in Form von Tangens hyperbolicus (Tanh) oder Sigmoid
- Feste Abfolge von 3 Schichten:
 - Faltung (engl. *convolution*)
 - Bündelung (engl. *pooling*)
 - Nichtlinearität (engl. *non-linearity*)

Eine Pooling-Schicht dient dazu, nur die relevantesten Informationen an die nächsten Schichten weiter zu geben, und so abstraktere Repräsentation des Inhalts zu realisieren und gleichzeitig die Anzahl der Parameter eines Netzes zu reduzieren.

ALEXNET Die AlexNet Architektur wurde 2012 von Krizhevsky et al. [KSH12] publiziert. Diese Netzarchitektur ist der von LeNet-5 sehr ähnlich, obwohl dieses Modell wesentlich größer ist. Die Erkenntnisse aus LeNet-5 wurden in ein größeres neuronales Netzwerk skaliert, das genutzt werden kann, um komplexere Objekte und Objekthierarchien zu lernen. Der Beitrag dieser Arbeit ist:

- Einführung von Dropout, um einzelne Neuronen während des Trainings zu deaktivieren, was einer Überanpassung (engl. *overfitting*) des Modells entgegen wirkt
- Verwendung von gleichgerichteten Lineareinheiten (engl. *linear units (ReLU)*) als Nichtlinearitäten
- Überlappendes Max-Pooling

VGG Im Jahr 2014 stellten Simonyan et al. [SZ14] die VGG-Netze vor. Hier wurden zum ersten Mal in jeder Faltungsschicht wesentlich kleinere 3×3 -Filter verwendet. Die Erkenntnis aus der VGG-Architektur ist, dass kombinierte 3×3 -Faltungen in Folge denselben Effekt wie größere Filterkerne haben. Folglich verwenden VGG-Netze mehrere Sequenzen von 3×3 -Faltungsschichten, um komplexe Merkmale darzustellen. Diese Erkenntnis findet auch Einzug in neuere Netzwerkarchitekturen. Ein Nachteil ist, dass dieses Vorgehen zu einer

Vielzahl von Parametern führt, welche trainiert werden müssen und was entsprechende Rechenkapazität voraussetzt.

INCEPTION/GOOGLNET Im Jahr 2014 vorgestellt wurde das Inception-Netzwerk von Szegedy et al. [SLJ⁺15]. Das Modell besteht aus einer Grundeinheit, die als *Inception Cell* bezeichnet wird. In dieser Einheit werden eine Reihe von Faltungen mit verschiedenen Skalierungen durchgeführt und anschließend die Ergebnisse zusammenfasst. Es werden 1×1 -Faltungsblöcke verwendet, um die Dimension der Eingangsdaten zu reduzieren. Bevor die Daten dann an rechenintensive (hochdimensionale) Faltungsschichten weitergegeben werden, wird so die Anzahl der Merkmale etwa um das Vierfache reduziert. Dies führt zu großen Einsparungen bei den Rechenkosten. Im Vergleich zu AlexNet und VGG kommt diese Architektur auch mit wesentlich weniger Operationen aus, was insgesamt zu einem sehr effizienten Netzwerkdesign führt.

INCEPTION V2 Ioffe et al. [IS15] führte die Inception V2 Architektur ein. Neu ist hier die Batch-Normalisierung. Diese berechnet den Mittelwert und die Standardabweichung aller Merkmalskarten (engl. *Feature-Maps*) einer Schicht und normiert diese. Dies entspricht dem sog. (engl. *whitening*) der Daten. Somit liegen alle Daten im gleichen Wertebereich und haben einen Mittelwert von Null.

RESNET Eine weitere Evolution der Architekturen wurde von He et al. [HZRS16] präsentiert. Grundsätzlich soll ein neuronales Netzwerk bei gegebenen Eingabedaten x eine zugrunde liegende Funktion $H(x)$ finden, welche den Daten eine Klasse zuordnet. Laut He et al. ist es ein allgemein anerkanntes Prinzip, dass tiefere Netzwerke in der Lage sind, komplexere Funktionen und Darstellungen der Eingabedaten zu erlernen, die zu einer besseren Leistung führen sollten. Es wurde jedoch immer wieder festgestellt, dass ab einem gewissen Punkt das Hinzufügen zusätzlicher Schichten letztendlich einen negativen Einfluss auf die Leistung hatte.

Dieser Fakt wird von den Autoren als Degradationsproblem bezeichnet. Eine der Ursachen dafür ist der Effekt des verschwindenden Gradienten (engl. *vanishing gradient*), bei der mit zunehmender Tiefe die Gradienten gegen Null gehen. Bessere Parameterinitialisierungstechniken und Batch-Normalisierung ermöglichen zwar eine Konvergenz tieferer Netzwerke, diese korrelieren aber oft mit einer höheren Fehlerrate als flachere Architekturen. Die Autoren schlagen daher eine Lösung für dieses Degradationsproblem vor, indem sie Residualblöcke einführen, in denen die Zwischenschichten des Blocks eine spezielle Residual-Funktion lernen. Diese Funktion dient als eine Instanz eines Verfeinerungsschrittes, wodurch die Eingabedaten für höherwertige Merkmale angepasst werden. Weiterhin wurden Verknüpfun-

gen hinzugefügt, welche es ermöglichen, einzelne Ebenen zu *überspringen*. Dadurch können bestimmte Schichten beim Lernen *übersprungen* werden. So kann das Modell, lernen welche Schichten *nützlich* sind und welche nicht.

MOBILENETS Im Jahr 2017 stellten Howards et al. [HZC⁺17] die MobileNets-Architektur vor. Diese Architektur verwendet separierbare Faltungen, um die Anzahl der Parameter zu reduzieren. Die separierbare Faltung führt unabhängig voneinander Faltungen in räumlichen und kanalbezogenen Bereichen durch. Durch diese Faktorisierung der Faltungen werden die Rechenkosten erheblich reduziert.

RESNEXT Die ResNeXt-Architektur [XGD⁺17] ist eine Erweiterung des tiefen Restnetzwerks, das den Standard-ResBlock durch einen ersetzt, der eine „split-transform-merge“-Strategie (d.h. verzweigte Pfade innerhalb einer Zelle) nutzt. Anstatt Faltungen über die gesamte Input-Feature-Map durchzuführen, wird die Eingabe des Blocks in eine Reihe von tieferen (Kanal-)Dimensionsdarstellungen projiziert.

DENSENET Im Jahr 2016 stellten Huang et al. die DenseNet Architektur [HLW17] vor. Sie erklären, dass es nützlich sein kann, auf

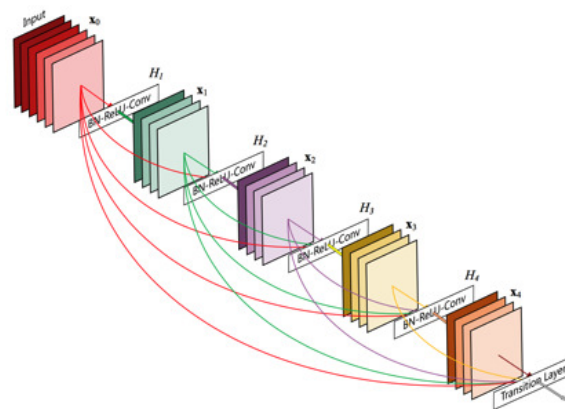


Abbildung 89: Schema der DenseNet Architektur. Jede Schicht erhält zusätzlich Merkmalskarten aus vorherigen Schichten als Eingabe. Quelle:[HLW17]

Feature-Maps aus früheren Zeitschichten im Netzwerk zu verweisen. Somit wird die Feature-Map jeder Schicht mit der Eingabe jeder nachfolgenden Schicht innerhalb eines dichten Blocks verknüpft, wie in Abbildung 89 dargestellt. Dies ermöglicht es späteren Ebenen innerhalb des Netzwerks, die Funktionen aus früheren Ebenen direkt

zu nutzen und die Wiederverwendung von Funktionen innerhalb des Netzwerks zu fördern. Die Autoren erklären, dass *verkettende* Feature-Maps, die von verschiedenen Schichten gelernt wurden, die Variation in der Eingabe der nachfolgenden Schichten erhöhen und die Effizienz verbessern. Im Vergleich zu ResNet-Modellen sollen DenseNets eine bessere Leistung bei geringerer Komplexität erzielen.

Mishkin et al. [MSM17] haben 2017 eine systematische Bewertung der CNN-Architekturen vorgenommen. Basierend auf ihren Ergebnissen führen unter anderem folgende Elemente zu besseren Ergebnissen:

- ReLU mit Batch-Normalisierung
- Nutzung einer Kombination von (engl. *average*) und (engl. *max pooling*)-Schichten
- Nutzung von vollständig verbundene Schichten als Faltungsschichten und Mittelung der Vorhersagen für die endgültige Entscheidung
- Sauberkeit des Datensatzes, im Sinne der Annotationen, ist wichtiger als die Größe

10.3 SEMANTISCHE SEGMENTIERUNG

In diesem Abschnitt wird beschrieben, wie CNN-Netze zur semantischen Bildsegmentierung genutzt werden können. Die Definition der Bildsegmentierung wurde bereits in Kapitel 3.2 vorgenommen. Ein naiver Ansatz zur Konstruktion einer neuronalen Netzwerkarchitektur zur semantischen Segmentierung besteht darin, einfach eine definierte Anzahl von Faltungsschichten zu stapeln und am Ende eine Segmentierungsmaske auszugeben. Dieses Netzwerk erlernt durch die sukzessive Transformation von Merkmalen eine Transformation der Eingabedaten auf die entsprechende Segmentierungsmaske. Solch eine Architektur ist allerdings recht rechenintensiv, da die volle Bildauflösung im gesamten Netzwerk beibehalten wird.

Ein Standard-Ansatz von Architekturen zur Bildsegmentierung ist es daher, einer Encoder/Decoder-Struktur zu folgen, bei der die räumliche Auflösung der Eingabe reduziert wird. Der Encoder ist in der Regel ein vortrainiertes Klassifizierungsnetzwerk wie VGG oder ResNet, gefolgt von einem Decodernetzwerk, wie in Abbildung 90 dargestellt. Die Aufgabe des Decoders besteht darin, die vom Encoder auf niedriger Auflösung gelernten Merkmale auf die höhere Auflösung zu projizieren, um eine per-Pixel-Klassifikation zu erhalten. Im Gegensatz zur reinen Bildklassifikation, bei der das Endergebnis des Netzwerks nur aussagt, was im Bild zu sehen ist, unterliegt die semantische Segmentierung einem höheren Aufwand. Hier erfolgt nicht nur eine Entscheidung auf Pixelebene, sondern es wird auch ein Mechanismus

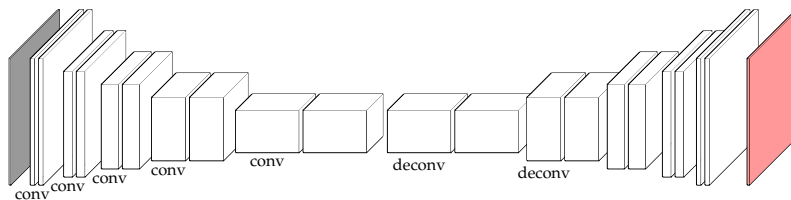


Abbildung 90: Schema für eine Standard-Architektur zur semantischen Segmentierung mit mehreren verbundenen Schichten, welche ein *Downsampling* durch Faltungen (engl. *convolution*) und anschließend ein *Upsampling* durch Entfaltungen (engl. *deconvolution*) durchführen. Die jeweiligen Schichten sind durch Boxen dargestellt die Dimension der Boxen visualisiert die Dimension der Daten.

benötigt, um die in verschiedenen Phasen des Encoders erlernten Merkmale auf die volle Auflösung anzuwenden und zu projizieren. Verschiedene Ansätze verwenden unterschiedliche Mechanismen als Teil des Dekodiermechanismus.

10.3.1 Vollständig gefaltete Netze

Der Ansatz, ein *vollständig gefaltetes* (engl. *fully convolutional*) Netzwerk zu semantischen Segmentierung zu verwenden, wurde im Jahr 2014 von Long et al. [LSD17] veröffentlicht. Die Autoren schlagen vor, bestehende, gut untersuchte Architekturen wie AlexNet zur Klassifikation anzupassen. Somit dienen diese als Encoder-Modul des Netzwerks, an das anschließend ein Decoder-Modul mit transponierten Faltungsschichten angefügt wird. So werden die groben Merkmalskarten in eine Segmentierungsmaske mit voller Auflösung überführt. Problematisch ist jedoch, dass das Encoder-Modul die Auflösung der Eingabedaten um den Faktor 32 reduziert. Somit hat das Decoder-Modul Probleme, akkurate Segmentierungen zu erzeugen. Um dieses Problem zu beheben, wird die Ausgabe des Encoder-Moduls stufenweise hochskaliert. Zusätzlich werden sog. „Skip-Verbindungen“ aus vorherigen Schichten hinzugefügt und die jeweiligen Merkmalskarten zusammengefasst. Diese Skip-Verbindungen von früheren Schichten im Netzwerk liefern die notwendigen Details, um akkurate Begrenzungen der jeweiligen Objekte zu rekonstruieren.

Etwas später optimieren Ronneberger et al. [RFB15] die vollständig gefaltete Architektur vor allem durch die Erweiterung der Kapazität des Decoder-Moduls des Netzwerks. Sie schlagen die U-Net-Architektur vor, welche aus einem kontrahierenden Pfad zur Erfassung des Kontextes und einem symmetrisch expandierenden Pfad besteht, der so eine präzise Lokalisierung ermöglicht. Das U-Net-Modell, welches in Abbildung 91 dargestellt ist, setzt sich aus einer Reihe von

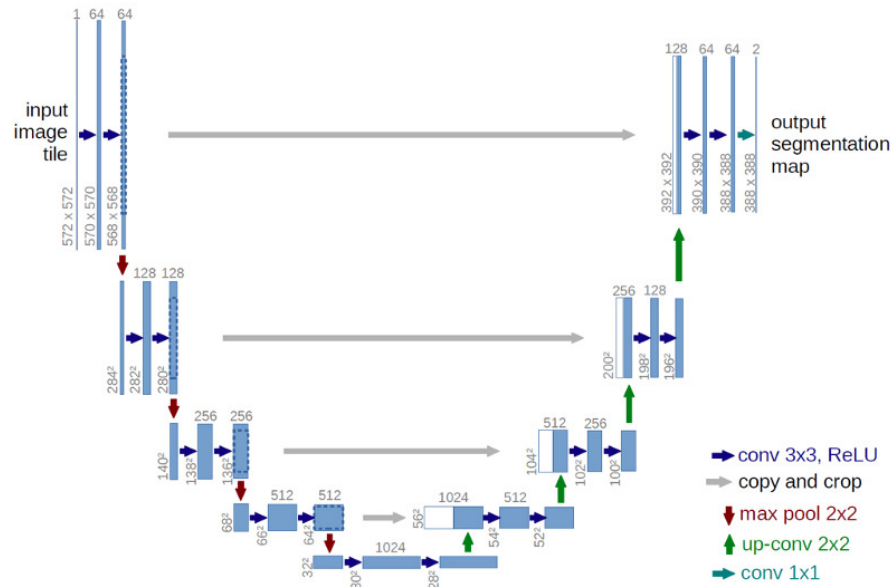


Abbildung 91: Eine Darstellung der UNet-Architektur. Jede blau Box repräsentiert eine mehrkanalige Merkmalskarte. Die Anzahl der Kanäle steht oberhalb der Box. Die räumliche Dimension ist darunter angegeben. Die Pfeile symbolisieren verschiedene Operationen. Die linke Seite stellt den kontrahierenden Pfad dar, in dem die räumliche Auflösung reduziert wird. Die rechte Seite zeigt den expandierenden Pfad in dem die Auflösung wieder erhöht wird und mit hochauflösenden Merkmalen des kontrahierenden Pfades kombiniert wird. Quelle: [RFB15]

Faltungsoperationen für jeden Block zusammen. Diese Netzwerkar-chitektur war in der Folge sehr beliebt und wurde entsprechend modifiziert auf diverse Segmentierungsprobleme angewendet. Beispielsweise tauschen Drozdal et al. [DVC⁺16] die gestapelten Faltungsblöcke gegen Residual-Blöcke aus. Diese Blöcke führen kurze Skip-Verbindungen innerhalb des Blocks ein. Sie verbessern die Qualität der Segmentierung an Objektkanten und ermöglichen die direkte Übertragung von Informationen aus frühen hochauflösenden Schichten in tiefere Schichten des Netzwerks.

Normalerweise werden zur Verbesserung der Objektkanten bedingte Zufallsfelder (CRF) auf die Ausgabedaten von pixelbasierten Klassifikatoren angewendet, um konsistentere Ergebnisse zu erzielen, wie bereits in Kapitel 9 näher erläutert. Weiterhin finden sich diese Ansätze auch in der Literatur wieder, wie bei Chen et al. [CPSA17], wo ein Neuronales Netz mit einem CRF kombiniert wird. Doch diese separaten Nachverarbeitungsschritte werden zunehmend durch spezielle Netzwerkarchitekturen ersetzt, welche die *Mittelwertbildung* von CRFs entsprechend innerhalb der Netzarchitektur approximie-

ren. Als Beispiel sind hier die Arbeiten von Zheng et al. [ZJRP⁺15] und Badrinarayanan et al. [BKC15] zu nennen, welche als Erste die Eigenschaften eines CRF in der Netzarchitektur nachgebildet haben. Weiterentwickelt wurde dieses Konzept von Jegou et al. [JDV⁺16] welche dichte (engl. *dense*) Blöcke, die der U-Net Architektur folgen, implementieren. Die dichten Blöcke eignen sich gut zur semantischen Segmentierung, da sie Low-Level Merkmale aus frühen Schichten mit High-Level Merkmalen aus späteren Schichten kombinieren können, was eine effiziente Wiederverwendung von Merkmalen ermöglicht.

10.3.2 Architektur-Übersicht

Im Folgenden ist in Tabelle 15 eine Übersicht von verschiedenen Netzwerkarchitekturen zur semantischen Segmentierung gegeben, welche auf den zuvor genannten Konzepten basieren. Die aufgelisteten Architekturen stellen nur eine kleine Auswahl der aktuell verfügbaren Architekturen dar. Sie zeigen auf aktuellen Datensätzen wie z. B. dem Cityscapes Datensatz gute Ergebnisse (IoU-Score) und wurden daher ausgewählt, um zu untersuchen, ob sie mit spektralen Daten trainiert werden können.

Name	IoU Cityscapes	HySpec	Jahr (Publikation)	Referenz
Encoder-Decoder (SegNet)	57,0	✓	2015	[BKC15]
DenseNet	NA	✓	2016	[JDV ⁺ 16]
PSPNet	81,2	✗	2016	[ZSQ ⁺ 16]
RefineNet	73,6	✗	2016	[LMSR17]
FRRN	71,8	✓	2016	[PHML17]
MobileUNet	NA	✓	2017	[HZC ⁺ 17]
DeepLabv3	82,1	✗	2017	[CPSA17]
GCN	80,5	✗	2017	[PZY ⁺ 17]
AdapNet	63,8	✗	2017	[VVDB17]
ICNet	70,6	✗	2017	[ZQS ⁺ 17]
DenseASPP	80,6	✗	2018	[YYZ ⁺ 18]
BiSeNet	68,4	✗	2018	[YWP ⁺ 18]
DDSC	NA	✗	2018	[BP18]

Tabelle 15: Übersicht aktueller und etablierter Netzarchitekturen zur semantischen Segmentierung mit IoU-Score (Stand: 05/2019)

Die meisten Netzarchitekturen wurden in erster Linie für die Nutzung und Anwendung von RGB-Daten konzipiert, wie auch Garcia et al. [GGEOE⁺18] in ihrem Review über verschiedene Architekturen erläutern. Für diese Netze existieren etablierte Datensätze mit großen Mengen an annotierten Daten wie in Tabelle 7 in Kapitel 5 aufgelistet. Gleichzeitig stehen für viele Netze wie z. B. PSPNet, ResNet und Deeplab, die auf Millionen von RGB-Bildern trainiert wurden, vor-

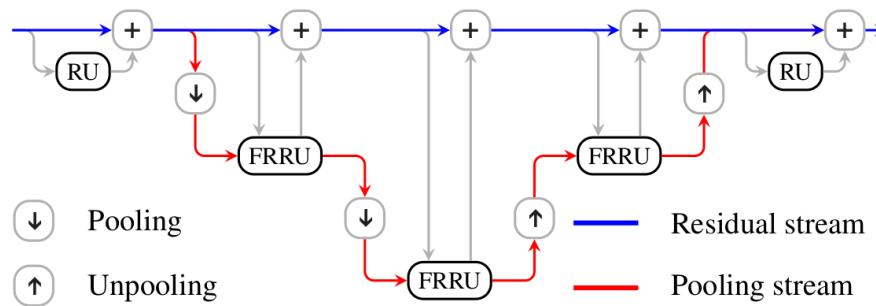


Abbildung 92: Beispiel einer FRRN-Netzwerkstruktur aus [HZRS16]. Der Pooling-Stream (rot) durchläuft eine Folge von Pooling- und Unpooling-Einheiten, während der verbleibende Stream (blau) in voller Bildauflösung bleibt.

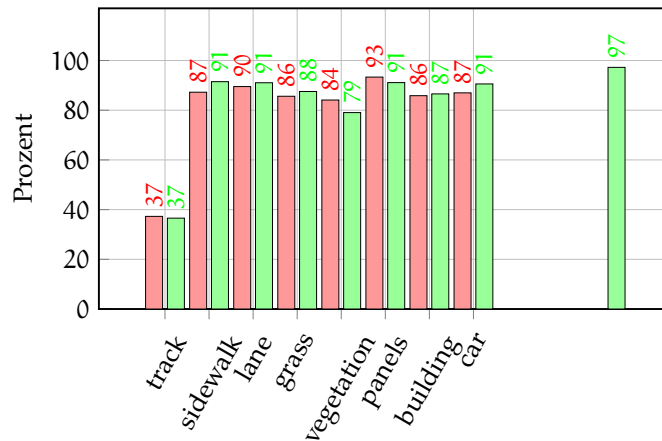
trainierte Modelle zum Download [22] zur Verfügung.

Leider sind nicht alle Architekturen in der Lage, Daten mit mehr als drei Kanälen ohne umfangreiche Strukturänderungen zu verarbeiten. Es gibt jedoch ein paar Ausnahmen, welche grundsätzlich mit hochdimensionalen Daten trainiert werden können. Diese sind entsprechend in Tabelle 15 gekennzeichnet.

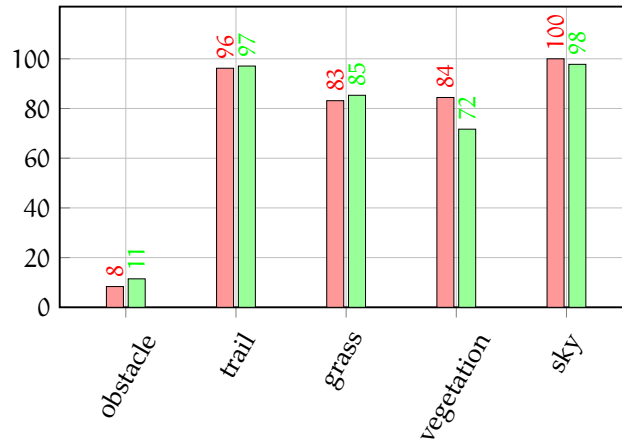
Dementsprechend wurden im Folgenden die vier gekennzeichneten Netzwerke mit den spektralen Datensätzen trainiert und die Ergebnisse ausgewertet. Dazu wurden an den ausgewählten Netzarchitekturen Anpassungen an der Eingabeschicht vorgenommen, sodass die spektralen Datensätze direkt genutzt werden können.

Zu den modifizierten Netzen gehört auch eine Netzarchitektur, welche 2017 von Pohlen et al. [PHML17] veröffentlicht wurde. Die von Pohlen et al. vorgeschlagenen neuartigen Netzwerkarchitekturen heißen *Full-Resolution Residual-Networks* (FRRNs) und wurden speziell für die semantische Segmentierung in Straßenszenen konzipiert. Sie zeigen die gleichen überlegenen Trainingsqualitäten wie die zuvor erwähnten Residual-Netzwerke (ResNet), verfügen aber über einen *Residual*- und einen *Pooling*-Stream. Dies wird durch die Notwendigkeit motiviert, Netzwerke zur Verfügung zu stellen, welche gleichzeitig gute High-Level-Features für die Erkennung und gute Low-Level-Features für die Lokalisierung berechnen können. Durch die Kombination zweier Verarbeitungsströme sind FRRNs in der Lage, beide Arten von Merkmalen gleichzeitig zu berechnen. Die Merkmale des Residualstroms werden durch Addition aufeinanderfolgender Residuen berechnet, während die Merkmale des anderen Stroms das direkte Ergebnis einer Folge von Faltungs- und Pooling-Operationen sind, die auf die Eingabedaten angewendet werden.

10.4 EVALUATION VON NETZARCHITEKTUREN



(a) FRRN-Architektur mit semantischer Annotation auf dem NirFull-Datensatz. Hier zeigt sich das die Nutzung von NIR-Daten vor allem bei der Klassifikation von Vegetation Vorteile bietet.



(b) FRRN-Architektur mit offRoad Annotation. Auch hier ist sehr deutlich der Vorsprung der NIR-Daten bei der Vegetation zu sehen.



Abbildung 93: Übersicht über die Evaluationsergebnisse der FRRN-Architektur mit semantischer und offRoad Annotation für den VisFull- und NirFull-Datensatz. Bei beiden Annotationen zeigt sich, dass vor allem die Klassifikation von Vegetation durch die Verwendung von NIR-Daten profitiert.

Da spektrale Daten analysiert werden sollen, können leider keine vortrainierten Netze verwendet werden. Deshalb müssen die zu untersuchenden Netzwerke von Grund auf neu mit nur begrenzten Daten

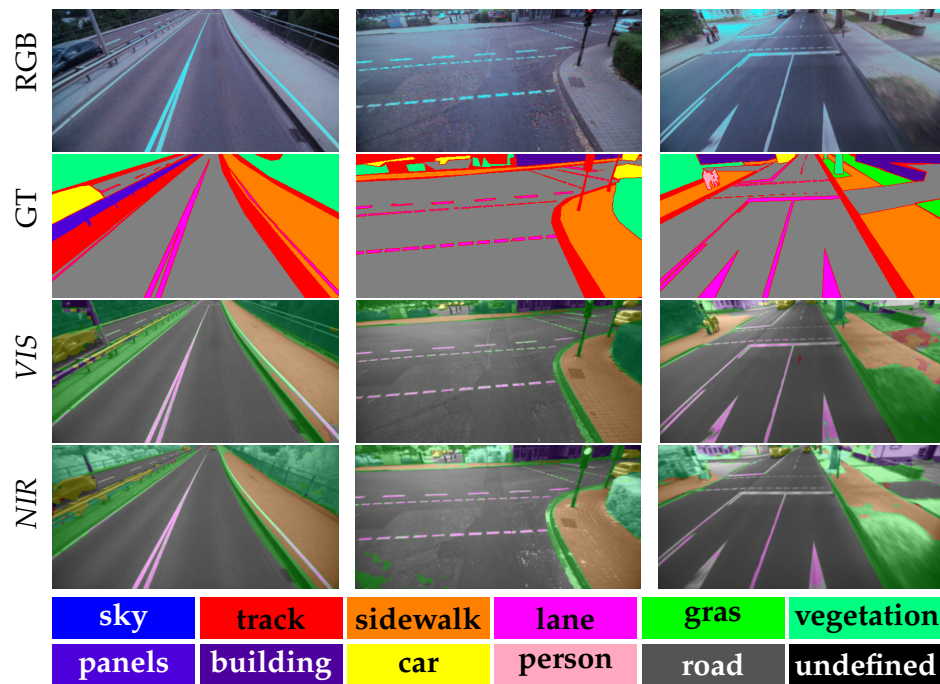


Abbildung 94: Vergleich der Klassifikationsergebnisse der Netzarchitekturen auf *NIR*- und *VIS*-Daten mit semantischer Annotation. Vor allem bei den *NIR*-Daten zeigt sich eine saubere Trennung zwischen Vegetation und anderen Bereichen. Auch konnten die Fahrbahnmarkierungen sauber abgegrenzt werden.

trainiert werden. Jede Schicht in einem neuronalen Netzwerk berechnet eine Funktion F , womit die Ausgabe x_n der n -ten Schicht berechnet wird als

$$x_n = F(x_{n-1}; \omega_n) \quad (63)$$

mit ω_n als die der jeweiligen Schicht zugehörigen Gewichtsmatrix. Die Eingabeschicht eines Netzwerks ist als Schicht von unveränderlichen Neuronen modelliert, welche mit einer Folgeschicht verbunden ist, somit ist eine Berechnung und Analyse von Daten möglich. Die Ausgabe des neuronalen Netzes ist in diesem Fall eine Klassenzuordnung einer Klasse $\Omega_i \in \Omega$ für jeden Hyperpixel. Die Klassen werden als positive Ganzzahlen definiert und durch einzelne Neuronen in der Ausgabeschicht repräsentiert.

Die Eingabeschicht der passenden Netzwerke wurde zum Training und der Evaluation entsprechend parametrisiert, so dass vollständige und die gemessenen Spektren Hyperwürfel mit 16 bzw. 25 Bändern als Eingabe $x_i = \mathbf{f}^H$ für die jeweiligen Netze dienen können.

Im Folgenden werden Ergebnisse des Netzwerks, das auf verschiedenen Datensätzen trainiert wurde, dargestellt und diskutiert.

Da die vorgestellten spektralen Datensätze in puncto Größe nicht mit

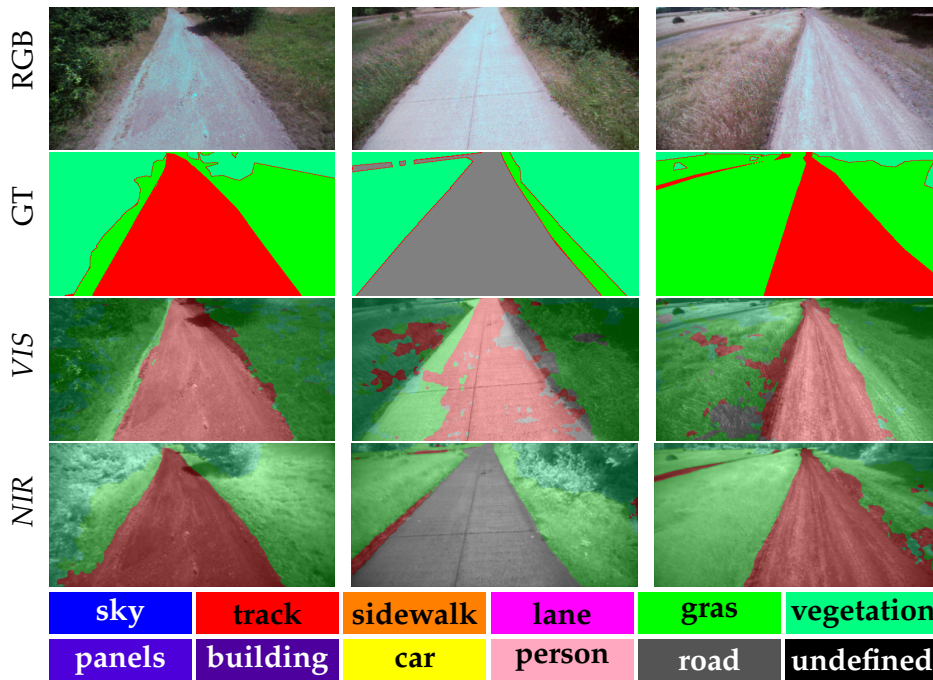


Abbildung 95: Vergleich der Klassifikationsergebnisse der Netzarchitekturen auf *NIR*- und *VIS*-Daten mit semantischer Annotation. Hier sind Szenen in suburbanem Gelände gezeigt. Hier sind verbreitet Feldwege oder Schotterpisten anzutreffen. Vor allem diese Wege konnten unter Verwendung der *NIR*-Daten besser von Straßen abgegrenzt werden, dies zeigen die vorletzte und die letzte Zeile sehr gut.

Datensätzen wie Cityscapes mithalten können, wird Datenaugmentation genutzt, um eine Überanpassung zu minimieren und den Datensatz künstlich zu vergrößern. Die hier verwendete Methode zur Datenaugmentation ist vertikales Spiegeln. Das Netzwerk wurde für 250 Iterationen mit einer Batchgröße von 4 Hyperwürfeln trainiert, indem ein Cross-Entropie-Verlust unter Verwendung des ADAM-Optimierers [KB15] minimiert wurde. Die Ergebnisse der Evaluation sind in Tabelle 16 dargestellt. Es zeigt sich sehr deutlich, dass bei nahezu allen Datensätzen auf beiden Annotationsgruppen die FRRN-Netzarchitektur die besten Ergebnisse erzielt.

Die Validierung ergab bspw. für den NirFull-Datensatz einen IoU-Score von 0,58 bei der semantischen und 0,53 bei der offRoad-Annotation. Dies sind gute Ergebnisse, wenn man bedenkt, dass nur sehr begrenzte Daten verfügbar sind. Der VisFull-Datensatz erreicht einen ähnlichen IoU-Score von 0,57 bei der semantischen und 0,47 bei der offRoad-Annotation. Somit ist zunächst auf diesen Datensätzen kein großer Unterschied zwischen *NIR* und *VIS*-Daten festzustellen. Interessant ist, dass bei der offRoad-Annotation die Werte etwas schlech-

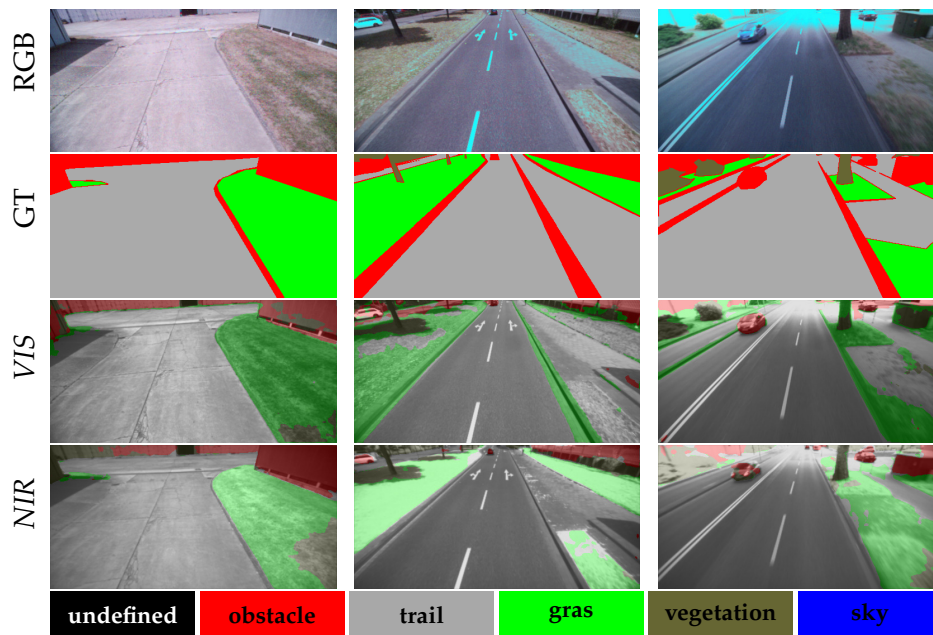


Abbildung 96: Vergleich der Klassifikationsergebnisse der Netzarchitekturen auf *NIR*- und *VIS*-Daten mit offRoad-Annotation. Hier sind die Grenzen zwischen den einzelnen Objekten der Szene nicht so sauber wie bei der semantischen Annotation.

ter sind. Dies widerspricht zunächst der Intuition, da die Daten weniger Klassen zugeordnet werden müssen.

Werden die Ergebnisse der anderen Datensätze verglichen, ist zu erkennen, dass bei den Datensätzen, welche im Feld aufgenommen wurden dieser Effekt zu beobachten ist. Die Datensätze in der Stadt folgen der Intuition und zeigen bessere Ergebnisse bei der offRoad-Annotation. Dies legt den Schluss nahe, dass die Zusammenführung der kleinen Klassen Gebäude, Auto, Mensch und Panel zu Problemen führt. Denn die Klassen sind im Datensatz vom Feld quasi nicht vorhanden und erhöhen somit letztendlich nur den Intra-Klassenabstand. Dies ist sowohl bei den *VIS*-Daten als auch bei den *NIR*-Daten zu beobachten.

Einige Beispielbilder werden in Abbildung 94 und Abbildung 95 dargestellt. Insgesamt kann die Segmentierung von befahrbaren Straßen gegenüber Gras und Vegetation sehr zuverlässig durchgeführt werden. Es kann auch zwischen befahrbarem Gras und unpassierbarer Vegetation unterscheiden werden. Hier hilft auch die Chlorophyllkonzentration, die sich besonders gut im nahen Infrarotbereich nachweisen lässt, wie auch schon in Kapitel 6 beschrieben.

Semantische Annotation				offRoad Annotation			
Datensatz	Netz	IOU	Accuracy	Datensatz	Netz	IOU	Accuracy
CityVis	FRRN	0.47	0.79	CityVis	FRRN	0.58	0.84
	MobileUNet	0.35	0.73		MobileUNet	0.49	0.81
	Encoder-Decoder	0.43	0.77		Encoder-Decoder	0.55	0.83
	FC-DenseNet	0.21	0.65		FC-DenseNet	0.38	0.75
LandVis	FRRN	0.37	0.55	LandVis	FRRN	0.24	0.48
	MobileUNet	0.39	0.57		MobileUNet	0.24	0.48
	Encoder-Decoder	0.39	0.57		Encoder-Decoder	0.24	0.48
	FC-DenseNet	0.3	0.52		FC-DenseNet	0.21	0.46
VisFull	FRRN	0.57	0.78	VisFull	FRRN	0.47	0.76
	MobileUNet	0.51	0.73		MobileUNet	0.45	0.74
	Encoder-Decoder	0.54	0.78		Encoder-Decoder	0.47	0.75
	FC-DenseNet	0.26	0.49		FC-DenseNet	0.31	0.59
CityNir	FRRN	0.48	0.78	CityNir	FRRN	0.59	0.84
	MobileUNet	0.37	0.72		MobileUNet	0.5	0.81
	Encoder-Decoder	0.43	0.75		Encoder-Decoder	0.55	0.83
	FC-DenseNet	0.22	0.65		FC-DenseNet	0.41	0.76
LandNir	FRRN	0.65	0.83	LandNir	FRRN	0.55	0.85
	MobileUNet	0.61	0.79		MobileUNet	0.5	0.81
	Encoder-Decoder	0.64	0.81		Encoder-Decoder	0.53	0.84
	FC-DenseNet	0.46	0.66		FC-DenseNet	0.38	0.71
NirFull	FRRN	0.57	0.79	NirFull	FRRN	0.53	0.8
	MobileUNet	0.49	0.73		MobileUNet	0.49	0.79
	Encoder-Decoder	0.54	0.77		Encoder-Decoder	0.51	0.81
	FC-DenseNet	0.34	0.56		FC-DenseNet	0.31	0.6

Tabelle 16: Ergebnisse der Klassifikation unter Nutzung verschiedener Netzarchitekturen. Das beste Ergebnis je Datensatz ist grün hervorgehoben und das beste Ergebnis je Annotation ist rot hervorgehoben. Bei der Nutzung von NIR-Daten zeigt eindeutig der suburbane Datensatz die besten Ergebnisse. bei den VIS-Daten ist es der urbane Datensatz. Die Ergebnisse der NIR-Daten sind insgesamt besser, als die der VIS-Daten.

10.5 FAZIT

In diesem Abschnitt wurden Neuronale Netze aus dem Bereich des Deep-Learnings vorgestellt und diskutiert. Ausgewählte Netz-Architekturen wurden mit spektralen Daten trainiert, um zu analysieren, ob die trainierten Modelle zur Szenenanalyse genutzt werden können. Grundsätzlich konnten von den ausgewählten Architekturen vier Netze mit spektralen Daten trainiert werden. Werden die Ergebnisse der jeweiligen Architekturen betrachtet, lässt sich feststellen, dass die Unterschiede zwischen FRRN, MobileUNet und Encoder-Decoder relativ gering ausfallen. Wobei sich die FRRN-Architektur meist durchgesetzt hat. Die Kombination aus Residual- und Pooling-Stream scheint vielversprechend zu sein. Als negativer Ausreißer lässt sich hier das FC-DenseNet sehen, welches in den meisten Fäl-

len keine gute Repräsentation der Daten erlernen konnte. Basierend auf den vorliegenden Daten scheinen die ResNet-Architekturen der DenseNet-Architektur überlegen zu sein. Dies ist aber noch mit weiteren Evaluationen tiefer zu ergründen. Werden die Ergebnisse etwas genauer betrachtet, zeigt sich, dass die trainierten Netz-Architekturen durchweg Probleme mit Klassen wie Mensch, Fahrbahnmarkierung, Schild und Auto hatten. Dies lässt sich mit der äußerst geringen Datengrundlage für die jeweiligen Klassen begründen, denn die Klassen mit niedrigen Klassifikationsraten entsprechen auch denen mit einer niedrigen Repräsentation in den Datensätzen. Die Klassen erhalten dadurch beim Training wahrscheinlich nur ein geringes Gewicht und werden folglich weniger und schlechter trainiert. Sie werden auch tendenziell vernachlässigt weil der Klassifikator eine höhere Unsicherheit bei anderen Klassen in Kauf nehmen muss um die kleinen Klassen zu berücksichtigen. Werden die spektralen Eigenschaften berücksichtigt, sind speziell Menschen schwierig zu fassen, da sie aufgrund von unterschiedlicher Kleidung diverse spektrale Reflexionsspektren aufweisen können. Wird das noch in Beziehung gesetzt zum Raumbedarf im Bildraum, ergibt sich eine ungünstige Konstellation aus komplexer Repräsentation und niedriger Datenlage. Der Abstand zwischen Datenpunkten innerhalb einer Klasse kann hier sehr groß sein. Ansonsten setzt sich das Bild aus den vorherigen Kapiteln fort. Straße und Vegetation werden vor allem in den NIR-Daten sehr sauber getrennt.

Grundsätzlich ist festzuhalten, dass sich vorhandene Netz-Architekturen gut zur Szenenanalyse mit spektralen Daten eignen. Allerdings sind die vorliegenden Datensätze noch zu klein, um die Fähigkeiten der Netze effektiv zu nutzen, speziell auch im Hinblick auf die hohe Dimensionalität der Daten. Denn bei Neuronalen Netzen hängt die Generalisierungsfähigkeit direkt von der Größe und Qualität des zum Training verwendeten Datensatzes ab. Um dem Problem der hohen Dimensionalität entgegenzuwirken, können Methoden aus dem Bereich des unüberwachten Lernens verwendet werden um die Dimension der Daten effektiv zu reduzieren.

DATENKOMPRESSION

11.1 EINLEITUNG

Die effektive Nutzung moderner Techniken und Algorithmen zur Klassifikation von vor allem hochdimensionalen Daten bedingt die Verfügbarkeit umfangreicher, annotierter Datensätze. Sollen nun Deep-Learning-Techniken zur Analyse von spektralen Daten genutzt werden, sind umfangreiche Datensätze notwendig. Die im Rahmen dieser Arbeit aufgebauten Datensätze sind allerdings, im Vergleich zu etablierten Datensätzen (vgl. Kapitel 5), noch relativ klein. Sie bestehen aus einigen Hundert annotierten Hyperwürfeln. Google und Co. nutzen Tausende bzw. Millionen von annotierten Daten für ihre Netzarchitekturen.

Eine Konsequenz von zu kleinen Datensätzen ist eine relative Überanpassung des Klassifikators an die Trainingsdaten, was in der Folge zu einem Modell führt, welches schlecht generalisiert. Wie in Kapitel 5 bereits ausgeführt, können die etablierten Datensätze nicht genutzt werden, um Klassifikatoren zur Analyse von spektralen Daten zu befähigen. Allerdings ist die Etablierung von umfangreichen Datensätzen kein triviales Problem, da die Datenannotation nach wie vor eine komplexe, auf Menschen basierende Arbeit ist, welche anfällig für Fehler ist. Speziell die Generierung von semantischen Datensätzen zur Segmentierung mit Annotationen auf Pixelebene ist aufwendig und teuer. Daher müssen Verfahren entwickelt und analysiert werden, die mit einer begrenzten Menge an Daten trainiert werden können oder aus den begrenzten Daten Datenrepräsentationen generieren die eine bessere Generalisierung des trainierten Modells erlauben.

Unter Verwendung von weiteren Konzepten des Deep Learnings, wird in diesem Abschnitt ein Klassifikationssystem vorgestellt welches auf Dimensionsreduktion und unüberwacht trainierten Merkmalen basiert. Dazu wird zunächst ein Autoencoder (AE) mit benutzerdefinierten Regularisierungen, die sich auf die Modellierung des latenten Raums und nicht auf den Rekonstruktionsfehler konzentrieren, konstruiert. Dies ermöglicht es, eine dimensionsreduzierte Darstellung der spektralen Daten zu lernen, welche einen neuen Merkmalsraum aufspannt. Die erlernte Darstellung wird als Eingabe für die, auf neuronalen Netzen basierte, Klassifikation von spektralen Daten verwendet. Die im Folgenden beschriebenen Arbeiten wurden in Teilen bereits auf einer internationalen Konferenz veröffentlicht [6].

11.2 STAND DER TECHNIK

Die Standardverfahren im Bereich der bildbasierten Szenensegmentierung sind, wie bereits in Kapitel 10 erwähnt, gekennzeichnet durch die Nutzung von RGB-Daten und den Versuch, anhand dieser verschiedene Klassen zu differenzieren. Dies beschreiben z. B. auch Chetan et al. [CKJ10], Thoma [Tho16] in ihren Studien zur bildbasierten Segmentierung.

Klassifikatoren zur Analyse spektraler Daten nutzen hingegen zu meist wesentlich mehr Kanäle. Allerdings existieren grundsätzliche Probleme bei der Klassifikation von spektralen Daten, wie die begrenzte Anzahl von annotierten Daten und die große räumliche Variabilität der spektralen Signaturen wie auch schon Camps-Valls und Bruzzone in ihrer Arbeit von 2005 [CVB05] dokumentieren. So ist die Auswahl bzw. Generierung geeigneter Merkmale bei der spektralen Klassifikation ebenso wichtig wie das Design des Klassifikators selbst. Daraus leitet sich das Forschungsgebiet der Merkmalsgewinnung auch (engl. *Feature-Mining*) [DTTB07, JKC13] genannt ab. Dabei wird versucht, aus gegebenen Eingabedaten sinnvolle Merkmale für eine bestimmte Anwendung zu ermitteln bzw. abzuleiten. Das Gebiet lässt sich grundsätzlich in drei Kategorien der spektralen Darstellung aufteilen, welche als Merkmalsauswahl, Merkmalsextraktion und Hybridmethode bezeichnet werden. Bei der Merkmalsauswahl werden die ursprünglichen Wellenlängen beibehalten, wogegen bei Verfahren der Merkmalsextraktion wie z. B. der Hauptkomponentenanalyse (PCA) [Rico6] die ursprünglichen Messungen in einen neuen Teilraum überführt werden. Weiterhin zu dieser Kategorie gehören auch Techniken wie die Dimensionstransformation [JL99, HC94, BKL02] und die Bandauswahl [SHS12, CDSA99, SB01]. Hybride Methoden verwenden sowohl die Merkmalsauswahl als auch die Extraktion zum Feature-Mining, z. B. wählen Forscher Bänder auf der Grundlage von Expertenerfahrung aus, und anschließend wird eine Merkmalsextraktion auf die ausgewählten Bänder angewendet. Für die Darstellung der räumlichen Eigenschaften werden Form, Textur und das sog. spektrale Profil im Bereich der spektralen Klassifikation umfassend genutzt.

Um mit der räumlichen Variabilität der spektralen Daten umzugehen, versuchen einige Ansätze, räumliche Informationen zu berücksichtigen, wie sie von Tarabalka et al. [TBC09] und Plaza et al. [PPM09] vorgeschlagen werden.

Wie bereits in Kapitel 10 beschrieben, werden zunehmend auch Techniken aus dem Bereich des Deep-Learning, wie z. B. neuronale Netze, zur Analyse und Verarbeitung hochdimensionaler Daten genutzt. So wurden auch in diesem Bereich Konzepte und Verfahren zur Dimensionsreduktion und dem unüberwachten Lernen von spektralen und räumlichen Merkmalen entwickelt. Im Jahr 2016 wurde von

Chen et al. [CJL⁺16] ein regularisiertes 3-D-CNN-basiertes Merkmals-extraktionsmodell zur Extraktion effizienter spektraler Merkmale mit einer räumlichen Dimension eingeführt. Darüber hinaus kombinierten Ghamisi et al. [GCZ16] ein CNN mit einem Partikelschwarm-Optimierungsalgorithmus, um iterativ die informativsten Bänder für das Training von spektralen Daten auszuwählen.

Weiterhin wurden unüberwachte Algorithmen entwickelt, die zunächst getrennt von der eigentlichen Klassifikation die Daten reduzieren. Beim unüberwachten Merkmalslernen werden relevante Merkmale aus nicht annotierten Daten extrahiert, Eingaberedundanzen erkannt und entfernt, um nur Schlüsselaspekte der Daten in robusten und aussagekräftigen Darstellungen zu erhalten. So stellten Zhao et al. [CZ]15 einen mehrskaligen, gestapelten Autoencoder vor, um eine effiziente Merkmalsdarstellung aus nicht annotierten Daten in Kombination mit einem linearen SVM zur spektralen Datenklassifikation zu lernen. Später wurde dieses Verfahren durch Tao et al. [TPLZ15] verbessert, der einen AE vorschlug, welcher eine reduzierte Merkmalsdarstellung lernt, die dazu neigt, effektiver und diskriminierender für die Klassifikation zu sein. Im Jahr 2018 präsentierten Wang et al. [WZWZ18] ein neuartiges spektrales Klassifikations-Framework, bei dem das unüberwachte Lernen von Merkmalen und die eigentliche Klassifikation der Daten separat modelliert werden. Die Nutzung von Deep-Learning-Methoden zur Dimensionsreduzierung von spektralen Daten scheint ein vielversprechender Ansatz. Wie auch beim Problem der Klassifikation stellt sich die Frage der richtigen Architektur und der Parametrisierung.

11.3 AUTOENCODER ZUR DIMENSIONSREDUKTION

Ein Autoencoder [HSo6, VLL⁺10] ist ein symmetrisches neuronales Netzwerk, das es erlaubt, die Eigenschaften und die Struktur von gegebenen Daten, unbeaufsichtigt zu lernen. Es nimmt eine Eingabe $\mathbf{x} \in \mathbb{R}^D$ und ordnet sie einer latenten Repräsentation $\mathbf{h} \in \mathbb{R}^M$ unter Verwendung einer nichtlinearen Abbildung $\mathbf{h} = F(\boldsymbol{\omega}\mathbf{x} + \boldsymbol{\beta})$ zu. Dabei definiert $\boldsymbol{\beta}$ einen Bias-Vektor und $\boldsymbol{\omega}$ eine Gewichtsmatrix, welche trainiert werden muss. Weiterhin definiert F eine nichtlineare Aktivierungsfunktion wie z. B. eine Sigmoid-Funktion. Entsprechend wird eine Umkehrabbildung $\mathbf{y} = F(\boldsymbol{\omega}'\mathbf{h} + \boldsymbol{\beta}')$ verwendet, um die Eingabedaten \mathbf{x} aus der latenten Darstellung \mathbf{h} mit $\boldsymbol{\omega}' = \boldsymbol{\omega}^T$ zu rekonstruieren. Die aus der latenten Repräsentation rekonstruierten Daten werden mit \mathbf{y} bezeichnet. Wenn $M < D$ gilt, wird der AE *undercomplete* genannt und lernt eine niedrigdimensionale komprimierte Darstellung, welche die wichtigsten Merkmale der ursprünglichen Datenverteilung darstellt. Der definierte Lernprozess minimiert dazu den Rekonstruktionsfehler

$$l = \frac{1}{n} \sum_{i=1}^n (x - y)^2 \quad (64)$$

aus Eingabedaten \mathbf{x} und rekonstruierten Daten \mathbf{y} . Als Nebeneffekt der Kompression wird ein latenter Raum konstruiert, in den die Daten transformiert werden.

Wenn der AE linear aufgebaut ist und die Verlustfunktion als mittlerer quadratischer Fehler wie in Gleichung 64 definiert ist, konstruiert der AE einen Teilraum, wie er auch durch eine Hauptkomponentenanalyse PCA konstruiert werden kann. Die Abbildung 97a zeigt einen einfachen AE, welcher sich aus einem Single-Layer Decoder und einem Single-Layer Encoder [GBC16] zusammensetzt. In der Mitte ist die sog. versteckte Schicht (engl. *Hidden Layer*) zu sehen, welche auch als (engl. *Bottleneck Layer*) bezeichnet wird. An dieser Stelle gehen die Daten in den latenten Raum über.

Der Aufbau von tiefen AE-Netzwerken mit mehreren versteckten

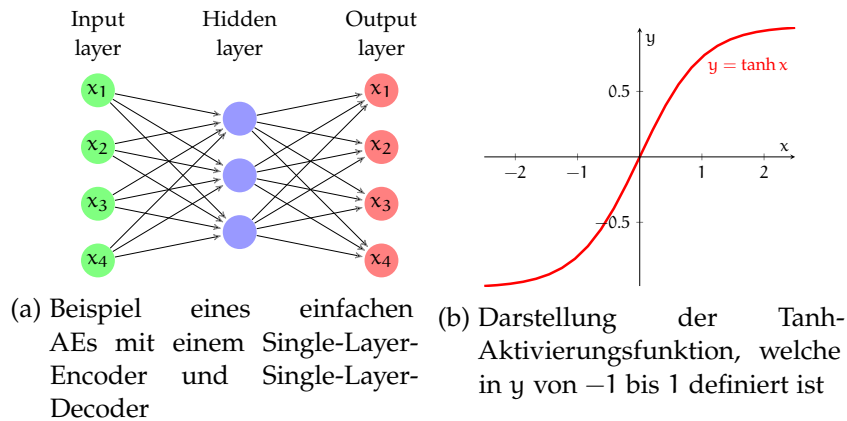


Abbildung 97: Darstellung eines einfachen AE und einer Tanh-Aktivierungsfunktion

Schichten kann viele Vorteile bieten, wie z. B. höhere Robustheit gegenüber Rauschen, Erfassung nichtlinearer Zusammenhänge und generell eine bessere Funktionsannäherung.

Das leitet sich auch indirekt aus dem universellen Approximator-Theorem [Cyb89] ab. Dieses garantiert, dass ein neuronales Netzwerk mit mindestens einer versteckten Schicht eine Approximation jeder Funktion mit beliebiger Genauigkeit darstellen kann. Voraussetzung dafür ist, dass genügend versteckte Einheiten implementiert sind. Sogenannte regularisierte AE verwenden eine Verlustfunktion, die nicht nur darauf abzielt, dass das gelernte Modell Eingabedaten in eine Ausgabe kopiert. Im Gegensatz zur üblichen Parametrisierung und Auslegung, bei der die Minimierung des Rekonstruktionsfehlers im Vordergrund steht, liegt der Fokus hier auf der Konstruktion eines konditionierten latenten Raumes, der eine komprimierte Darstellung

der spektralen Daten ermöglicht. Dazu wurden verschiedene Maßnahmen ergriffen, die im Folgenden erläutert werden.

11.3.1 Autoencoder-Design

Landgrebe [Lano2] beschrieb im Jahr 2002 die hyperspektrale Datenverarbeitung als ein hochdimensionales Signalverarbeitungsproblem. Dort erläuterte er, dass der hochdimensionale Raum zumeist leer ist, was bedeutet, dass die multivariaten Daten in der Regel in einer niedrigeren Dimensionsstruktur vorliegen. Daher können hochdimensionale Daten zur Analyse auf einen niedrigdimensionalen Raum projiziert werden, ohne signifikante Informationen zu verlieren.

Um nun die Dimension der vorliegenden spektralen Daten effektiv zu reduzieren, wird ein Autoencoder mit einem speziellen latenten Raum konstruiert. Um das gesteckte Ziel zu erreichen, wird der latente Raum durch die Einführung von speziellen Regeln explizit konditioniert. Dies ist ein Vorgehen, welches auch als Repräsentationslernen (engl. *Representation Learning*) [BCV13] bezeichnet wird.

Dazu muss auch die Batchgröße¹ der Daten berücksichtigt werden, da sie die Form der Datenverteilung bestimmt [Ben12]. AE werden in der Regel mit Mini-Batches und stochastischer Gradientenabsenkung (engl. *Stochastic Gradient Descent*) (SGD) trainiert. Die Mini-Batchgröße hängt von der jeweiligen Aufgabe ab. Kleine Werte führen zu schnellen Änderungen der trainierten Gewichte, während größere Werte mehr Daten berücksichtigen und das Netzwerk langsamer verändern.

Im hier vorliegenden Anwendungsfall muss die Mini-Batchgröße groß genug sein, um eine statistische Aussage treffen zu können, aber klein genug, um eine geeignete Lösung zu finden. Um die latente Verteilung zu definieren, wird eine strukturierte Verlustfunktion (engl. *structured loss*) \mathcal{S} genutzt, die wie folgt definiert ist:

$$\mathcal{S} = G(\bar{\mathbf{x}} - \mu^*) + G(\sigma_{\mathbf{x}} - \sigma^*)$$

wobei G die Summe der quadrierten Elemente über die erste Dimension eines Tensors² und $\sigma_{\mathbf{x}}$ den Mittelwert und die Standardabweichung $\bar{\mathbf{x}}$ von \mathbf{x} definiert. Das Symbol σ^* bezeichnet die gewünschte Standardabweichung und analog dazu bezeichnet μ^* den gewünschten Mittelwert. Die so gewählte Formulierung der Verlustfunktion soll erreichen, dass die statistischen Eigenschaften für jede latente Dimension vorhanden sind und nicht nur für den gesamten Raum

¹ Der Trainingsdatensatz wird in Teilmengen auch Batches genannt zerlegt. Die Batchgröße ist die Anzahl der Trainingsdaten, welche in einem einzigen Batch enthalten sind.

² Ein Tensor ist eine Verallgemeinerung eines Vektors oder einer Matrix und wird hier als multidimensionales Array verstanden.

gelten. Außerdem wird ein Gewichtsverlust (engl. *weight decay*) eingeführt, womit während des Trainings in jeder Epoche die Gewichte mit einem Faktor von weniger als 1 multipliziert werden

$$W = \sum_i^n ||w_i||$$

um eine Überanpassung (engl. *overfitting*) des Netzes zu unterbinden. So begrenzt der Term das Wachstum der Netze Gewichte und reduziert gleichzeitig die Anzahl der freien Gewichtsparameter. Dies führt zu einem definierten Modell. Die Gesamtkosten des Trainings sind definiert als $c = \alpha_0 \cdot l + \alpha_1 \cdot S + \alpha_2 \cdot W$ mit $\alpha_0 \gg \alpha_1 \gg \alpha_2$. Diese so gewählte Konstellation zwingt den AE während des Trainings, zunächst den Rekonstruktionsfehler l zu minimieren und sobald dieses Problem mit ausreichender Genauigkeit gelöst ist, wird die Verteilung der latenten Darstellung optimiert. Dies wird durch die zuvor definierte Verlustfunktion S erzwungen, welche statistische Vorgaben macht.

Insgesamt ist das Wachstum der Gewichte begrenzt und kann im Laufe der Zeit sogar reduziert werden, wenn keine weitere Verbesserung des Modells möglich ist.

Unter Verwendung dieses Konzepts wird ein relativ einfacher AE mit drei Encoder- und drei Decoder-Schichten aufgebaut, wie in Abbildung 98 dargestellt. Jede Schicht hat eine hyperbolische Tangens-Aktivierung wie in Abbildung 97b dargestellt. Die Eigenschaft dieser Funktion ist, dass negativen Eingabedaten stark negativ und Werte bei 0 auch nahe 0 abgebildet werden. Dadurch werden die Werte der Ausgaben jeder Schicht, auch die der Netzwerkausgänge in einem definierten Bereich begrenzt. Damit die letzten Gewichte nicht unkontrolliert wachsen, werden die Eingabedaten im Bereich von -0.5 bis 0.5 skaliert. In Bezug auf diese Aktivierung ist der latente Raum auf $\mu^* = 0$ und $\sigma^* = 0.1$ konditioniert.

11.4 EVALUATION

Im Folgenden werden die mit den entwickelten AE durchgeführten Experimente erläutert und die Ergebnisse diskutiert.

Um das zuvor definierte AE-Netzwerk zu trainieren, wurden mehr als 1 000 000 000 000 Hyperpixel aus den vorhandenen Datensätzen extrahiert. Der AE erhält dazu einen aus einem Hyperwürfel extrahierten einzelnen Hyperpixel mit 16 (*VIS*) oder 25 (*NIR*) Werten als Eingabe. Um einen festen Wertebereich zu definieren, werden die Hyperpixel entsprechend normiert, so dass jeder Kanal einen Wertebereich zwischen -0.5 und 0.5 hat. Daraus folgt, dass der Mittelwert etwa 0 ist. So wird dem latenten Raum durch die Normierung bereits eine bestimmte Struktur vorgegeben. Diese Normalisierung pro Kanal ist wichtig, da eine Normalisierung über den gesamten Vek-

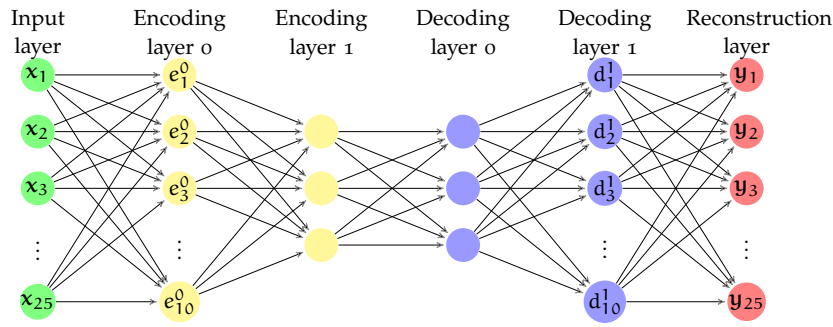


Abbildung 98: Beispielinstantz eines AEs für spektrale NIR-Daten mit 25 Kanälen. Der latente Raum (Encoding Layer 1) hat die Dimension drei und wird im Prinzip in *Encoding Layer 1* erreicht. An dieser Stelle kann der Autoencoder dann auch getrennt werden, um ihn zur Kompression zu nutzen.

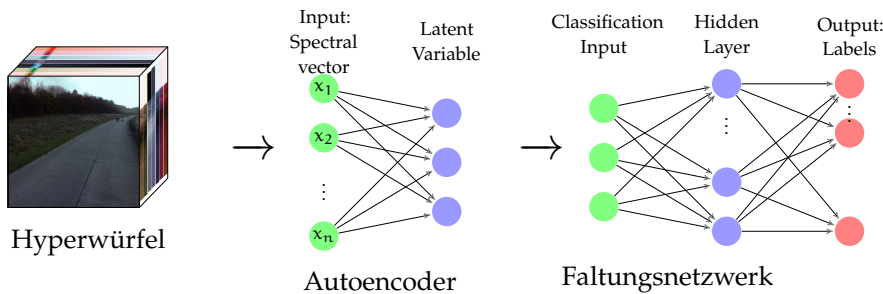
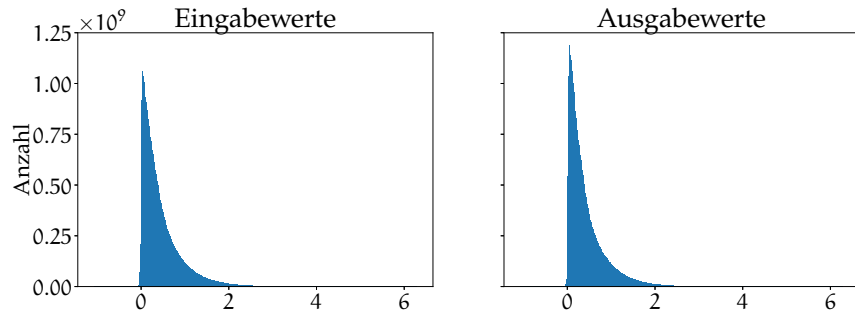


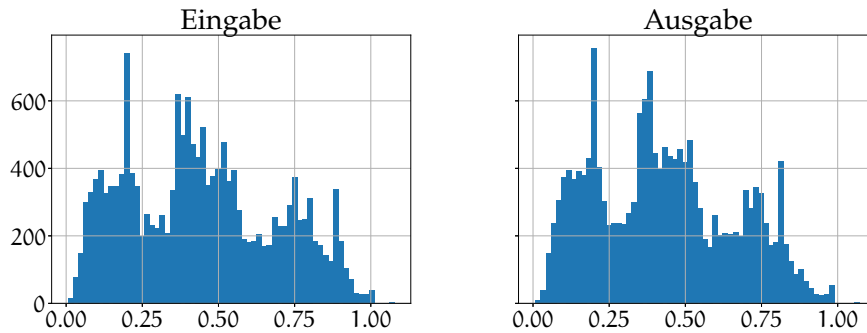
Abbildung 99: Schema des vorgestellten Frameworks, welches aus zwei Teilen besteht: Ein spektraler Hyperpixelvektor als Eingabe für den AE. Der latente Raum des AEs wird wieder zu einem Bild mit 3 Kanälen zusammengesetzt und dient als Eingabe für das Faltungsnetzwerk mit einer Ausgangsschicht zur semantischen Segmentierung.

tor zu einer schlechten Konditionierung der einzelnen Kanäle führen kann. Das Training wurde mit einer Batchgröße von 100 000 und einer Lernrate von 0.001 durchgeführt. Die Trainingsergebnisse für den AE basierend auf *NIR*-Daten werden in Abbildung 100 visualisiert.

Ein Überblick über die gesamte Eingangs- und Ausgangswertverteilung zeigt, dass der AE die Eingabedaten sehr gut komprimieren und rekonstruieren konnte. Hier sind nur geringe Abweichungen in der Verteilung zu erkennen. In Abbildung 100b zeigt ein Histogramm mit 64 Bins die Verteilung der Mittelwerte aus Eingabe- und Ausgabedaten an. Dies zeigt auch, dass das AE-Netzwerk nur wenige Fehler bei der Rekonstruktion der gegebenen Daten macht. Daraus folgt, dass es nur marginale Unterschiede zwischen Eingabedaten und Ausgabedaten gibt. So lässt sich schlussfolgern, dass der AE einen sehr



(a) Gesamte Verteilung der Eingangs- und Ausgangsdaten



(b) Histogramm mit der Verteilung der Mittelwerte

Abbildung 100: Ergebnisse des AE-Trainings basierend auf 1 000 000 000 000 spektralen Vektoren der NIR-Kamera. Die Daten der Ein- und Ausgabe sind sehr ähnlich, daraus lässt sich folgern, dass der Autoencoder eine effektive Kompression ohne signifikante Datenverluste ermöglicht.

effizienten dreidimensionalen Merkmalsraum der 25-dimensionalen Eingangsdaten aufgebaut hat. Dies spricht in der Konsequenz für die Präzision des entworfenen AEs unter Verwendung der gegebenen Daten.

11.4.1 Fazit

In diesem Abschnitt wurde ein AE-Netzwerk zur Dimensionsreduktion von spektralen Daten erläutert. Zur besseren Charakterisierung der Hyperpixel im Spektralraum, wurde das unüberwachte Trainieren eines tiefen AE mit zusätzlichen Regularisierungstermen, die sich auf die Modellierung von latenten Daten konzentrieren, eingeführt. Dieser neue Merkmalsraum ermöglicht den Einsatz von etablierten Methoden und Netzwerken des Deep Learning, welche bereits den Stand der Technik bei RGB-Daten darstellen. Die Ergebnisse und Rekonstruktionsfehler des trainierten AEs legen eine vielversprechende Robustheit und Übertragbarkeit der erlernten Merkmale nahe.

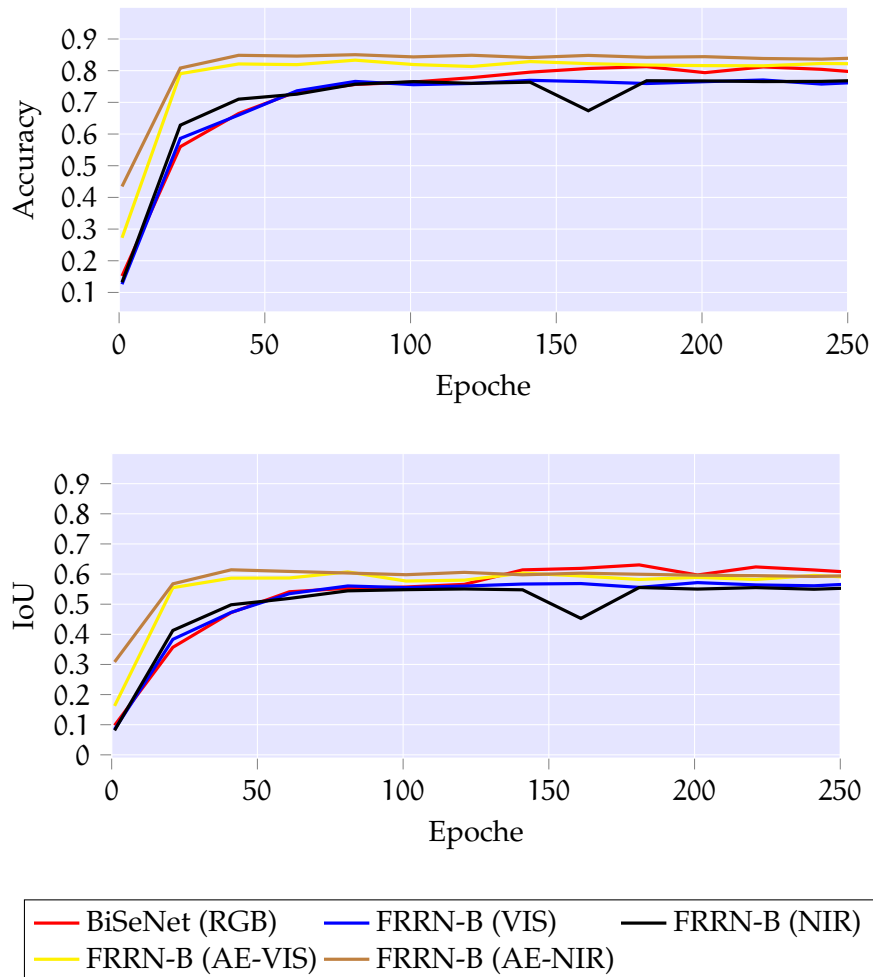


Abbildung 101: Vergleich der Klassifikationsergebnisse verschiedener Architekturen und Ausgangsdaten auf dem Vis-Full bzw.. NirFull Datensatz. Die Netze, welche mit Autoencoder-Daten trainiert wurden (braun und gelb), zeigen schon in früheren Epochen bessere Ergebnisse.

11.5 INTEGRATION IN SEMANTISCHE SEGMENTIERUNG

Im vorherigen Abschnitt wurde ein AE mit Regularisierungen zur Modellierung des latenten Raums von spektralen Daten trainiert. Somit kann dieser AE genutzt werden, um die hochdimensionalen Daten auf einen niedrigeren Raum zu projizieren. Diese dimensionsreduzierte Repräsentation der Daten kann dann als Eingabe für andere Algorithmen wie z. B. Faltungsnetze genutzt werden. Wo der AE Merkmale in der spektralen Dimension der Daten gelernt hat, kann nun durch Verwendung von Faltungsnetzen die räumliche Dimension der Daten betrachtet werden.

Um ein breites Spektrum an Netzarchitekturen untersuchen zu können, wurde der AE so parametrisiert, dass er einen dreidimensionalen latenten Raum erzeugt und wie zuvor beschrieben trainiert. So können die bereits im Kapitel 10 vorgestellten Netzarchitekturen genutzt werden.

Dazu wurde ein Framework wie in Abbildung 99 dargestellt realisiert. Vom zuvor trainierten AE wird der Decoder-Abschnitt abgetrennt, so dass der dreidimensionale latente Raum der Daten die Ausgabe-schnittstelle repräsentiert. An diese Schnittstelle können nun diverse Faltungsnetzwerke angedockt werden, welche dann die komprimierten Daten als Eingabe bekommen. Als Ausgabe liefern diese dann eine Segmentierungsmaske welche, die Klassenzuweisung der einzelnen Pixel darstellt. Vor dem Training des Netzwerks wird der AE-Anteil so parametrisiert, dass die bereits gelernten Gewichte fixiert sind. So werden nur die Gewichte der Faltungsnetze während des Trainings optimiert. Da die hierfür aufgebauten Datensätze in Umfang, Größe und Variabilität nicht mit anderen Datensätzen mithalten können wird hier auch *Data Augmentation* genutzt, um den Datensatz künstlich zu vergrößern. Das Netze wurde für 250 Epochen mit einer Batchgröße von 1 trainiert.

Um eine Aussage über die Nutzung von spektralen Daten und die Qualität der einzelnen Algorithmen zu treffen, wurden fünf synchronisierte Datenrepräsentationen mit verschiedenen Netzen trainiert:

- *VIS* mit 16 Bändern
- *NIR* mit 25 Bändern
- Pseudo-RGB aus *VIS* erzeugt (vgl. Kapitel 4.4)
- AE-*VIS* mit drei Dimensionen durch Autoencoder reduziert
- AE-*NIR* mit drei Dimensionen durch Autoencoder reduziert

Wie bereits zuvor erläutert, sind nicht alle diskutierten Netzwerkar-chitekturen in der Lage, Daten mit mehr als drei Kanälen ohne umfangreiche Strukturänderungen zu verarbeiten. Jedoch können die erzeugten RGB-Daten und die mittels AE auf drei Dimensionen reduzierten Daten als Eingabe für alle Netze dienen. Dementsprechend

wurden vier Netzarchitekturen mit spektralen (*VIS*, *NIR*) Daten trainiert und die anderen mit RGB und komprimierten Autoencoder Daten. Die Ergebnisse sind in Tabelle 17 dargestellt.

Weiterhin sind in Abbildung 103 die Klassifikationsergebnisse für jede Klasse dargestellt. Die besten Ergebnisse zeigt das BiSeNet in Kom-

Semantik (Autoencoder)				Semantik (Raw)			
Datensatz	Netz	IoU	Accuracy	Datensatz	Netz	IoU	Accuracy
AE-NirFull	FRRN	0.6	0.81	VisFull	FRRN	0.57	0.78
	MobileUNet	0.52	0.79		MobileUNet	0.51	0.73
	Encoder-Decoder	0.51	0.79		Encoder-Decoder	0.54	0.78
	FC-DenseNet	0.2	0.42		FC-DenseNet	0.26	0.49
	DenseASPP	0.32	0.51	NirFull	FRRN	0.57	0.79
	BiSeNet	0.67	0.87		MobileUNet	0.49	0.73
	DeepLabV3	0.26	0.48		Encoder-Decoder	0.54	0.77
	RefineNet	0.17	0.38		FC-DenseNet	0.34	0.56
	GCN	0.1	0.2	RGBFull	FRRN-A	0.6	0.79
	AdapNet	0.43	0.68		MobileUNet	0.51	0.75
PSPNet	0.35	0.61	Encoder-Decoder		0.53	0.77	
FRRN	0.63	0.84	FC-DenseNet		0.27	0.44	
MobileUNet	0.56	0.82	DenseASPP		0.29	0.51	
Encoder-Decoder	0.56	0.84	BiSeNet		0.63	0.81	
FC-DenseNet	0.26	0.45	DeepLabV3		0.21	0.38	
DenseASPP	0.28	0.48	RefineNet		0.26	0.41	
AE-VisFull	BiSeNet	0.66	0.86		GCN	0.35	0.59
DeepLabV3	0.23	0.41	AdapNet		0.34	0.5	
RefineNet	0.24	0.48	PSPNet	0.41	0.62		
GCN	0.1	0.22					
AdapNet	0.46	0.7					
PSPNet	0.39	0.63					

Tabelle 17: Klassifikationsergebnisse zur semantischen Segmentierung. Die besten Ergebnisse je Modalität sind grün hervorgehoben. Das BiSeNet zeigt in Kombination mit dem Autoencoder auf dem NirFull-Datensatz (AE-NirFull) die besten Ergebnisse. Hier ist der IoU-Wert mit 0.67 auch höher als im Vergleich mit den reinen RGB-Daten wo ein IoU-Wert von 0.63 erreicht wurde.

bination mit dem Autoencoder mit einem IOU-Score von 0,67 auf dem NirFull-Datensatz. Diese Netzarchitektur konnte zuvor nicht mit den spektralen Daten verwendet werden, da das Netzwerk für drei Kanäle konstruiert wurde. Werden die Ergebnisse des BiSeNet über alle Datensätze betrachtet zeigt sich, die komprimierten Daten der RGB-Rekonstruktion überlegen sind. Die BiSeNet-Architektur zeigt auf dem NirFull- und VisFull-Datensatz in Kombination mit dem Autoencoder die besten Werte. Die Ergebnisse sind auch besser als bei der direkten Anwendung auf den unverarbeiteten Daten mit anderen Netzarchitekturen. Dies spricht für die positiven Eigenschaften des Autoencoders. Werden die Werte über alle Datensätze und Netzwerke betrachtet, so ist zu erkennen, dass die spektralen Daten leichte Vorteile gegenüber den klassischen RGB-Daten bieten. Das zeigt

auch noch mal deutlich die Abbildung 101. Hier ist zu sehen, dass die Netzwerke, welche mit Autoencoder-Daten trainiert werden, schon in früheren Epochen bessere Ergebnisse erzielen. Werden weiterhin die Ergebnisse der einzelnen Klassen in Abbildung 103 betrachtet, zeigt sich das vor allem die *NIR*-Daten und vor allem bei Vegetation und Grasflächen sehr gute Ergebnisse erzielen. Beispielhafte Annotationen sind in Abbildung 104 und Abbildung 102 dargestellt. Diese bestätigen die zuvor beschriebenen Ergebnisse bezüglich der *NIR*-Daten und der Vegetation. Hier ist auch gut zu sehen, dass die Grenzen zwischen Straße und Gras speziell auf den Autoencoder-Daten sauberer verlaufen als z. B. bei den RGB-Daten

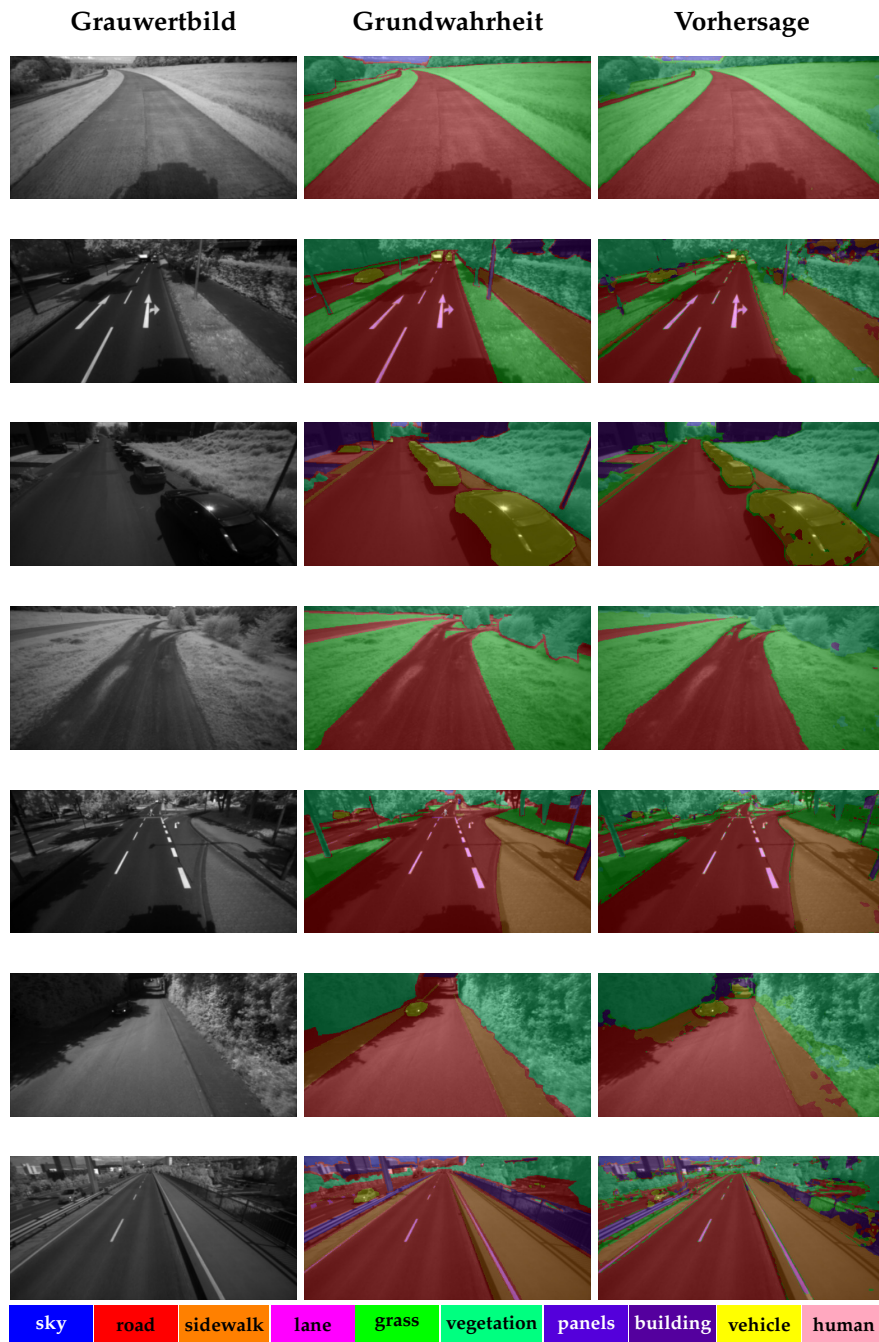


Abbildung 102: Vergleich der Klassifikationsergebnisse von *NIR*-Daten mit semantischen Annotationen. Die Zeilen zeigen jeweils eine Grauwertdarstellung der spektralen Eingabedaten, der Grundwahrheit und der semantischen Klassifikation basierend auf dem Autoencoder und der BiSeNet-Architektur. Die jeweiligen Szenenelemente werden gut wiedergegeben und eine saubere Trennung zwischen Fahrbahn und dem Rest ist zu erkennen.

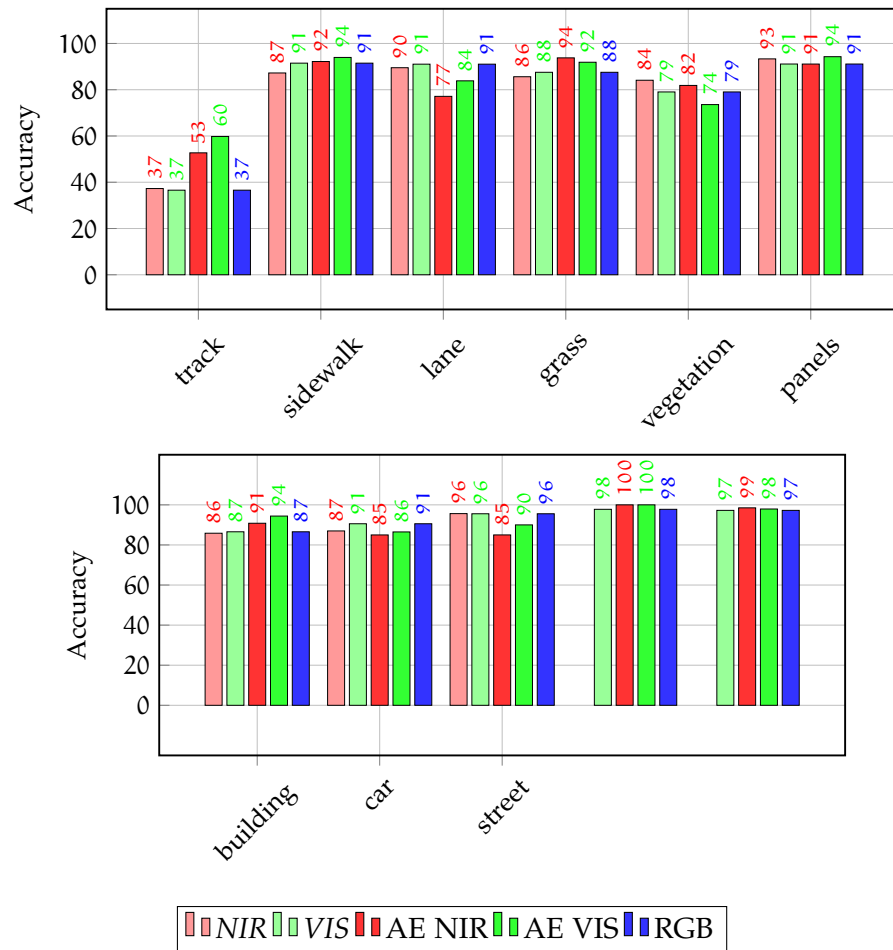


Abbildung 103: Evaluationsergebnisse der besten Netzarchitekturen auf NIR, VIS und RGB-Daten. Die Kombination aus Autoencoder und NIR-Daten (rot) liefert in den meisten Fällen die besten Ergebnisse. Diese Kombination performt in fast allen Fällen besser als die RGB-Daten (blau).

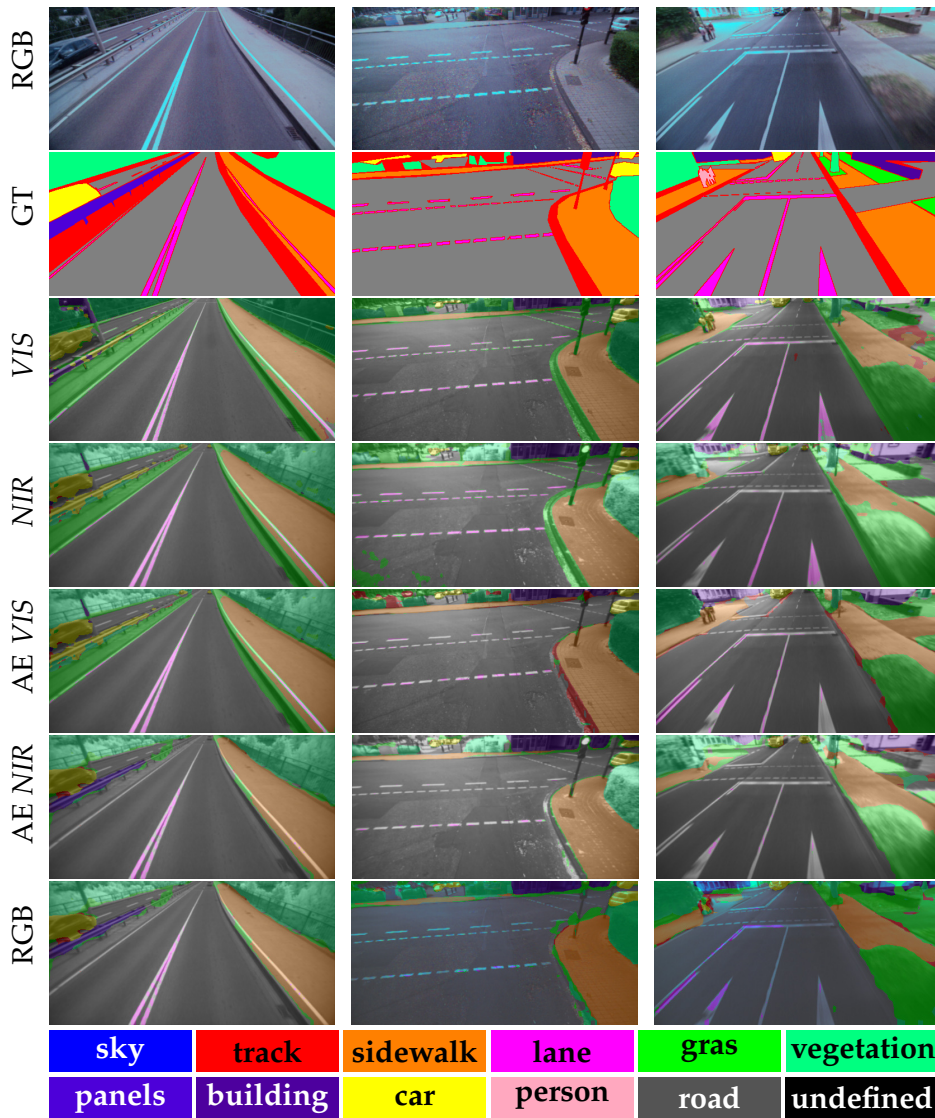


Abbildung 104: Vergleich der Klassifikationsergebnisse der Netzarchitekturen auf *NIR*- und *VIS*-Daten. Je Spalte ist ein Beispiel gezeigt und in jeder Zeile das Ergebnis einer anderen Datenquelle. Im Vergleich zu RGB sind vor allem die Ergebnisse der Autoencoderdaten präzise bei der Trennung von Straße und anderen Szenenelementen.

11.6 ZUSAMMENFASSUNG

In diesem Kapitel wurde ein Framework vorgestellt, welches das Lernen von Merkmalen in der spektralen und räumlichen Dimension ermöglicht, um spektrale Daten zu analysieren. Dazu wurden unüberwachte und überwachte Deep-Learning-Methoden kombiniert.

Um jedes Hyperpixel im Spektralbereich besser zu charakterisieren, wurde ein regularisierter AE genutzt, welcher sich auf die Modellierung des latenten Raumes und nicht nur auf die Rekonstruktion der Daten konzentriert. Dieser neu aufgespannte latente Raum ermöglicht dann den Einsatz von weiteren Deep-Learning-Methoden und Netzwerken, die bereits zum State-of-the-Art bei der Analyse von RGB-Daten gehören. So wurden in einem zweiten Schritt etablierte Netzarchitekturen zur mit einem vortrainierten AE kombiniert, um neben der spektralen auch die räumliche Dimension der Daten in den Lernprozess mit einzubeziehen. Die Experimente wurden auf den bereits vorgestellten Datensätzen durchgeführt. Die Ergebnisse und Rekonstruktionsfehler des trainierten Autoencoders zeigen vielversprechende Robustheit und Übertragbarkeit der erlernten Merkmale. Die durchgeführten Ergebnisse deuten darauf hin, dass die Kombination aus Autoencoder-Netzwerk und Faltungsnetzen zu einer präzisen Klassifikation auf Pixelebene führt. So zeigen die mit dem Autoencoder trainierten Netzwerke nach weniger Epochen schon bessere Ergebnisse. Damit erscheint eine Dimensionsreduktion speziell bei begrenzter Menge an verfügbaren Daten als sinnvoll.

FAZIT

In diesem Abschnitt wird eine Zusammenfassung der Ergebnisse der Arbeit gegeben, wobei im ersten Abschnitt speziell auf die Vor- und Nachteile der hyperspektralen Daten gegenüber klassischen RGB-Daten eingegangen wird.

12.1 ZUSAMMENFASSUNG UND BEWERTUNG

Die in dieser Dissertation untersuchte Sensorik der Firma Ximea stellt neuartige Bildgebungstechnik, welche zuvor so nicht zur Verfügung stand, dar. Dementsprechend muss geprüft werden, ob etablierte Techniken und Verfahren der hyperspektralen Bildverarbeitung genutzt werden können. Die Programmierschnittstelle von Ximea liefert zunächst nur Rohdaten, aus denen mittels einer speziellen Vorverarbeitung spektrale Werte berechnet werden müssen. Diese spektralen Werte haben entsprechend der Anzahl an Kanälen eine hohe Dimensionalität kombiniert mit einer hohen Korrelation. Die neuartige Sensortechnik wurde verwendet, um spektrale Informationen von strukturierter und unstrukturierter Umgebung aufzunehmen. Die generierten Daten wurden annotiert, um Datensätze aufzubauen, welche der semantischen Szenenanalyse dienen. Unter Verwendung von verschiedenen Methoden und Algorithmen wurden die vorliegenden Daten analysiert und klassifiziert. So sollte die Frage beantwortet werden, ob die zusätzlichen spektralen Informationen einen Vorteil im Bereich der semantischen Szenenanalyse und der Terrainklassifikation bieten. Beim überwachten Lernen und speziell beim Einsatz von neuronalen Netzen sind große Mengen an annotierten Daten notwendig, um adäquate Modelle trainieren zu können. Leider ist die Verfügbarkeit annotierter Daten stark begrenzt und beschränkt sich fast ausschließlich auf Luftaufnahmen. Daher wurden im Rahmen dieser Arbeit zunächst eigene Datensätze aufgebaut und veröffentlicht.

Zunächst wurden auf den vorgestellten Datensätzen verschiedene Klassifikatoren aus dem Bereich der klassischen Bildverarbeitung getestet. Dazu wurden die Daten zuerst pixelweise klassifiziert. Hier zeigte ein Random-Forest-Klassifikator die besten Ergebnisse. Bei den vielversprechendsten Klassifikatoren konnte ein positiver Zusammenhang zwischen der Anzahl der spektralen Bänder und den erzielten Ergebnissen festgestellt werden. Der Random-Forest zeigte besonders auf NIR-Daten sehr überzeugende Ergebnisse, was unter anderem auf die speziellen Absorptionseigenschaften von Chlorophyll zurückzuführen ist. Die zuerst untersuchten Verfahren haben auf ei-

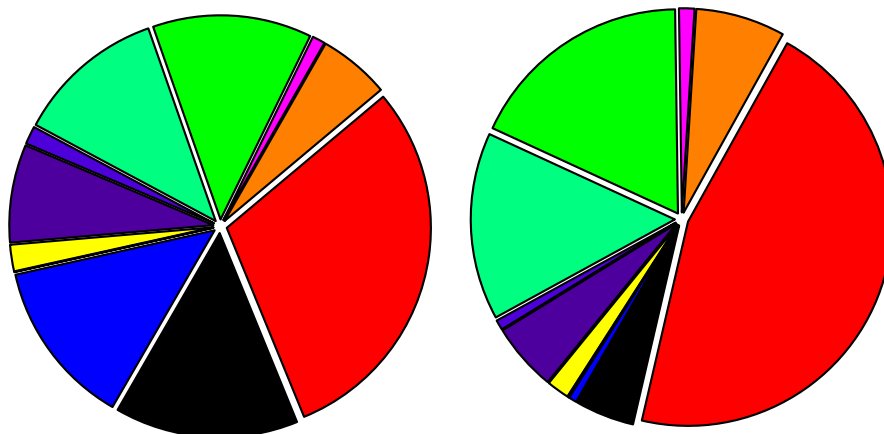
ner per-Pixel-Basis lediglich die spektralen Informationen zur Klassifikation genutzt. Im Weiteren Vorgehen wurden neben einer Normalisierung der spektralen Informationen auch verschiedene Merkmale untersucht, welche auch die räumlichen Informationen wie z. B. Texturen nutzen. Hier konnte vor allem die bandweise Normalisierung der spektralen Informationen bessere Ergebnisse erzielen. Im Weiteren Verlauf wurde erfolgreich ein graphbasiertes Verfahren mit einem Random-Forest-Klassifikator kombiniert, um räumliche Informationen in den Klassifikationsprozess mit einzubeziehen und die Nachteile der per-Pixel-Klassifikation zu kompensieren. Die Erstellung von Datensätzen ist grundsätzlich extrem aufwendig, deswegen standen im Rahmen dieser Dissertation nur eine begrenzte Menge an Daten zur Verfügung. Vermutlich lassen sich mit größeren Datensätzen zum Training, welche auch unterschiedlichere Beleuchtungssituationen und Jahreszeiten abdecken, noch bessere Ergebnisse erzielen. Speziell die in den letzten Kapiteln untersuchten Neuronalen Netze könnten von erweiterten Datensätzen profitieren und bessere Ergebnisse erzielen als bisher. Aufgrund der begrenzten Anzahl an annotierten Daten kann das volle Potential dieser Klassifikatoren noch nicht ausgeschöpft werden, da die Komplexität und Tiefe der Neuronalen Netze durch die Daten indirekt beschränkt wird. Die Kombination aus der eingeführten Autoencoder-Architektur und etablierten Deep-Learning-Architekturen kann hier punkten und führt zu einer guten Klassifikationsleistung auf Pixelebene, wie die Experimente zeigen.

Werden die Ergebnisse insgesamt betrachtet, so lässt sich bezogen auf die Daten der VIS-Kamera aus den erreichten Ergebnissen ableiten, dass eine semantische Szenenanalyse von spektralen Daten durchaus profitieren kann. Die erreichten Vorteile lassen das komplexe Problem der Szenenanalyse jedoch nicht trivial werden. Grundsätzlich können unter Verwendung der spektralen Informationen im sichtbaren Bereich bessere Ergebnisse gegenüber RGB-Daten erzielt werden, demgegenüber steht aber auch ein erheblicher Mehraufwand an Vorverarbeitung und Komplexität. Die Beobachtungen legen auch nahe, dass die Natur mit der Selektion von drei spektralen Empfindlichkeiten (RGB) eine gute Auswahl getroffen zu haben scheint. Werden die Daten der NIR-Kamera betrachtet, zeigen sich hier mehr Vorteile gegenüber klassischen RGB-Kameras. Unter Verwendung des Nahinfrarotbereichs können sich beispielsweise aufgrund der spektralen Reflexionseigenschaften von Pflanzen erhebliche Verbesserungen bei der semantischen Analyse ergeben. Durch die Nutzung von speziellen Indizes besteht die Möglichkeit, Vegetation effektiv ohne Training und beleuchtungsunabhängig von anderen Materialien zu trennen, was durchaus Vorteile beim autonomen Navigieren vor allem in unstrukturierter Umgebung bringen kann. Rückblickend lässt sich feststellen, dass die NIR-Kamera das größere Potential bietet. Für wei-

tere Untersuchungen sollte noch mal der Fokus auf die Vorverarbeitung und die Integration von Beleuchtungsinformationen gelegt werden. Dazu wäre wahrscheinlich eine zweite Kamera desselben Typs notwendig, welche nur Beleuchtungsinformationen aufnimmt. So wäre es tatsächlich möglich, spektrale Reflexionsspektren zu berechnen, welche dann evtl. robuster bei der Klassifikation sind. Weiterhin sollte auf jeden Fall versucht werden, die vorhandenen Datensätze weiter zu vergrößern, um die Vorteile der Neuronalen Netze in diesem Bereich besser nutzen zu können.

STATISTIK DER DATENSÄTZE

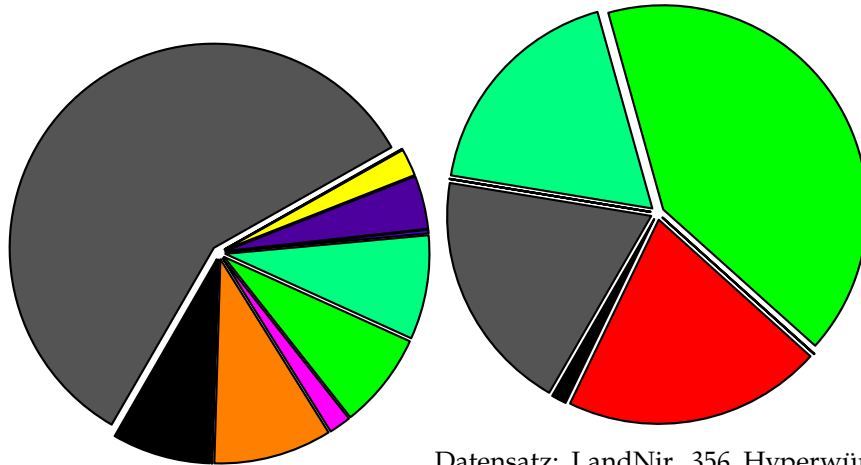
Die folgenden Grafiken zeigen Visualisierungen der statistischen Zusammensetzung der erstellten Datensätze.



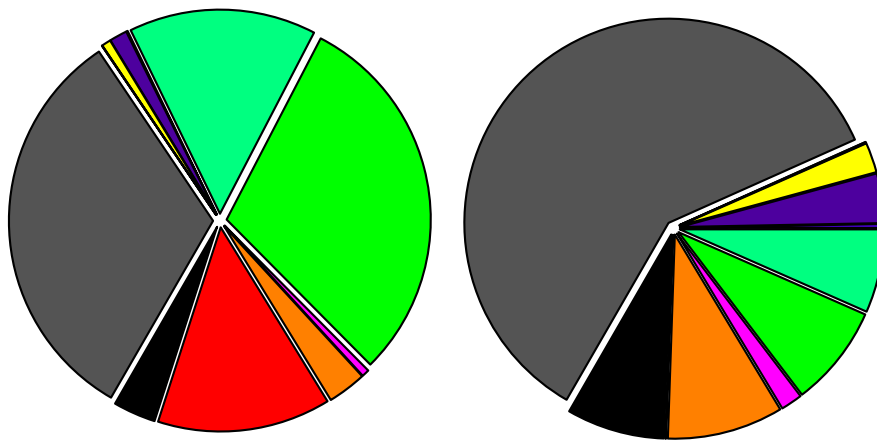
Datensatz: 2016Vis, 143 Hyperwürfel Datensatz: 2017Nir, 90 Hyperwürfel

Abbildung 105: Diagramme zur Verteilung der verschiedenen Klassen aus der Kategorie Semantik innerhalb der Datensätze

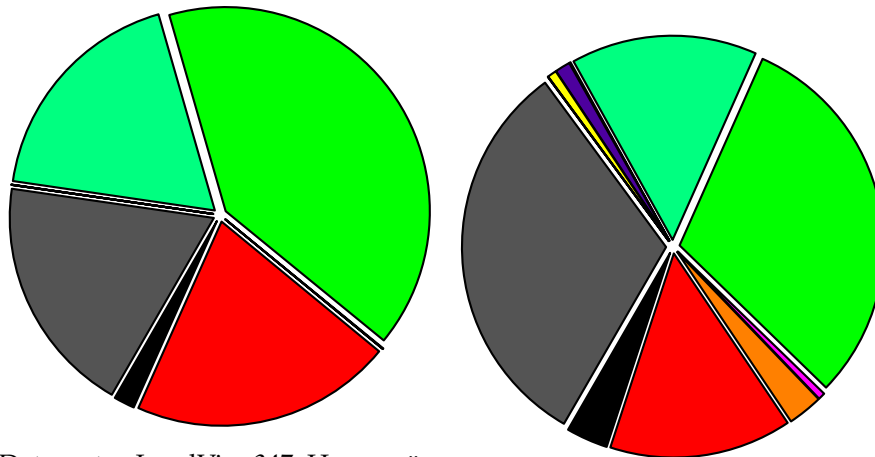




Datensatz: CityNir, 178 Hyperwürfel Datensatz: LandNir, 356 Hyperwürfel



Datensatz: NirFull, 532 Hyperwürfel Datensatz: CityVis, 154 Hyperwürfel



Datensatz: LandVis, 347 Hyperwürfel Datensatz: VisFull, 500 Hyperwürfel

Abbildung 106: Diagramme zur Verteilung der verschiedenen Klassen aus der Kategorie Semantik innerhalb der Datensätze

sky	track	sidewalk	lane	gras	vegetation
panels	building	car	person	road	undefined

ABBILDUNGSVERZEICHNIS

Abbildung 1	Beispiele für unstrukturierte Umgebungen	3
Abbildung 2	Beispiel semantischer Klassifikation	8
Abbildung 3	Beispieldaten unterschiedlicher Technolo- gien	9
Abbildung 4	Veröffentlichungen zur Spektraltechnik . .	13
Abbildung 5	Beispiele von Satellitenaufnahmen	14
Abbildung 6	FHM Am Beispiel einer Gauss-Glocke . . .	16
Abbildung 7	Landsat 3 Satellit	17
Abbildung 8	Überblick über Sensorsysteme	18
Abbildung 9	Beispiel zur semantischen Segmentierung	24
Abbildung 10	Überblick über Techniken zur spektralen Bildgebung	25
Abbildung 11	Überblick über Techniken zur hyperspek- tralen Bildgebung	26
Abbildung 12	Darstellung der beiden Kameramodelle von Ximea	29
Abbildung 13	Schema der Makropixel	30
Abbildung 14	Schema von Layout und Aufbau der Ma- kropixel	30
Abbildung 15	Fahrzeug mit Kameras	31
Abbildung 16	Arduino-Nano als Auslöser-Board	32
Abbildung 17	Sensor-Plattform mit montierter Sensorik .	33
Abbildung 18	Darstellung des sichtbaren Lichts	37
Abbildung 19	Ausschnitt des globalen Standardspektrums	38
Abbildung 20	Spektrale Hellempfindlichkeitskurve . . .	39
Abbildung 21	Empfindlichkeitsspektren der menschi- chen Rezeptoren	39
Abbildung 22	Beispiele zur Metamerie	40
Abbildung 23	Schematische Darstellung der Experimen- te von Wright und Guild	41
Abbildung 24	Strahldichten $\bar{r}, \bar{g}, \bar{b}$ der Primärlichtquellen $\mathcal{R}, \mathcal{B}, \mathcal{G}$	42
Abbildung 25	Strahldichten $\bar{x}, \bar{y}, \bar{z}$ der virtuellen Primär- lichtquellen $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	43
Abbildung 26	Interferenz	48
Abbildung 27	Vielstrahlinterferenz	49
Abbildung 28	Quanteneffizienz der CMOS-Sensoren . . .	51
Abbildung 29	Transmissionskurve eines spektralen Filters	52
Abbildung 30	Daten der 16 Spektralfilter	53
Abbildung 31	Verhalten der 25 Spektralfilter	54
Abbildung 32	Illustration Hyperwürfel	57

Abbildung 33	Schematische Darstellung eines Hyperwürfels	58
Abbildung 34	Klassifikationsfälle	59
Abbildung 35	Schema von Intersection over Union	61
Abbildung 36	Vorverarbeitung	66
Abbildung 37	Makropixel-Zuschnitt	67
Abbildung 38	Vignetten	68
Abbildung 39	Plots der Vignetten	68
Abbildung 40	Platte zum Weißabgleich	69
Abbildung 41	Schematische Darstellung des Hyperwürfels	71
Abbildung 42	Abtastung an Objektkanten	73
Abbildung 43	Geometrische Korrektur	73
Abbildung 44	Korrekturkurve	75
Abbildung 45	Beispiele zur Transformation	77
Abbildung 46	Datensätze mit hyperspektralen Daten	80
Abbildung 47	Beispiele aus verschiedenen hyperspektralen Datensätzen	82
Abbildung 48	Datensätzen mit Farbdaten zur semantischen Szenenanalyse	85
Abbildung 49	Test-Hierarchie 1	87
Abbildung 50	Anordnung der Sensorik auf einem Versuchsträger	88
Abbildung 51	Rohdaten der spektralen Kameras	88
Abbildung 52	Filterung der Daten	91
Abbildung 53	Plots der spektralen Werte von <i>NIR</i> -Kamera und <i>VIS</i> -Kamera	91
Abbildung 54	Für die Datensätze eingeführte Annotationsgruppen	93
Abbildung 55	Beispiele von annotierten Daten aus dem veröffentlichten Datensatz	97
Abbildung 56	Reflektierte Spektralverteilung einer Wiese	101
Abbildung 57	Beispiel NDVI-Index	102
Abbildung 58	Visualisierung des NDVI-Wertes im Grünkanal	103
Abbildung 59	Ergebnisse der NDVI basierten Klassifikation	104
Abbildung 60	Ergebnisse der optischen Indizes	106
Abbildung 61	Vergleich der NDVI-Klassifikationsergebnisse	107
Abbildung 62	Vergleich der <i>NIR</i> -Klassifikationsergebnisse	108
Abbildung 63	Schematische Darstellung der Funktionsweise einer SVM	113
Abbildung 64	Entscheidungsbaum	115
Abbildung 65	Schema der Pipeline zum maschinellen Lernen	118
Abbildung 66	Ergebnisse zur semantischen Annotation	119
Abbildung 67	Ergebnisse zur offRoad Annotation	120

Abbildung 68	Ergebnisse auf <i>NIR</i> -Daten	121
Abbildung 69	Ergebnisse auf <i>NIR</i> -Daten	122
Abbildung 70	Vergleich der Klassifikationsergebnisse von <i>NIR</i> -Daten	123
Abbildung 71	Ergebnisse auf <i>VIS</i> -Daten	124
Abbildung 72	Ergebnisse zu <i>VIS</i> -Daten	127
Abbildung 73	Pseudo-RGB Darstellung der spektralen Daten	132
Abbildung 74	Vergleich der Klassifikation anhand von einfachen Spektren	136
Abbildung 75	Vergleich der Klassifikationsergebnisse . .	137
Abbildung 76	Schema kontextsensitive Klassifikation . .	142
Abbildung 77	Beispiele zu CRF	143
Abbildung 78	Vergleich der Klassifikation mit verschie- denen Graph-Modellen	145
Abbildung 79	Ergebnisse zu <i>VIS</i> -Daten	148
Abbildung 80	Ergebnisse der semantischen Annotation .	149
Abbildung 81	Ergebnisse zur offRoad-Annotation	150
Abbildung 82	Ergebnisse zur offRoad-Annotation	151
Abbildung 83	Klassifikationsergebnisse <i>VIS</i> -Daten	152
Abbildung 84	Klassifikationsergebnisse <i>NIR</i> -Daten	153
Abbildung 85	Klassifikationsergebnisse <i>VIS</i> -Daten offRoad	154
Abbildung 86	Vergleich der Klassifikationsergebnisse . .	155
Abbildung 87	Klassifikationsergebnisse <i>NIR</i> offRoad . .	155
Abbildung 88	Architektur des LeNet-5	158
Abbildung 89	Schema der DenseNet Architektur	161
Abbildung 90	Beispiel einer Architektur zur semanti- schen Segmentierung	163
Abbildung 91	Darstellung der UNet-Architektur	164
Abbildung 92	Beispiel einer FRRN-Netzwerkstruktur . .	166
Abbildung 93	Ergebnisse der FRRN-Architektur	167
Abbildung 94	Netzarchitekturergebnisse Semantik	168
Abbildung 95	Netzarchitekturergebnisse Semantik sub- urban	169
Abbildung 96	Netzarchitekturergebnisse offRoad	170
Abbildung 97	Darstellung eines einfachen AE	176
Abbildung 98	Beispielinstanz eines AEs	179
Abbildung 99	Schema des vorgestellten Frameworks . . .	179
Abbildung 100	Ergebnisse des AE-Trainings	180
Abbildung 101	Vergleich der AE Klassifikationsergebnisse	181
Abbildung 102	Vergleich der Ergebnisse von <i>NIR</i> -Daten .	185
Abbildung 103	Ergebnisse zu <i>VIS</i> -Daten	186
Abbildung 104	Vergleich der Netzarchitekturergebnisse .	187
Abbildung 105	Diagramme zur Verteilung der Klassen . .	193
Abbildung 106	Diagramme zur Verteilung der verschiede- nen Klassen	194

TABELLENVERZEICHNIS

Tabelle 1	Stufen der Autonomie	2
Tabelle 2	Übersicht der Spektralsensoren	19
Tabelle 3	Spezifikation der Kameras	29
Tabelle 4	Relevante physikalische Größen	37
Tabelle 5	Hyperspektrale Bilddatensätze auf Satellitendaten	81
Tabelle 6	Überblick von hyperspektralen Datensätzen	83
Tabelle 7	Auflistung verschiedener RGB-Bilddatensätze	84
Tabelle 8	Übersicht über Datensätze mit spektralen Daten	87
Tabelle 9	Datenstruktur eines Hyperwürfels	90
Tabelle 10	Verteilung der Klassen	95
Tabelle 11	Verteilung der Klassen	96
Tabelle 12	Auswahl verschiedener Indizes	102
Tabelle 13	Übersicht der untersuchten Klassifikatoren	118
Tabelle 14	Ergebnisse versch. Merkmale	135
Tabelle 15	Übersicht von Netzarchitekturen zur semantischen Segmentierung	165
Tabelle 16	Ergebnisse der Netzerkevaluation	171
Tabelle 17	Ergebnisse semantische Segmentierung . .	183

MATHEMATISCHE SYMBOLE

FORMELZEICHEN

\mathbf{B}	Ein spektrales Band.
\bar{b}	Strahldichte Blau.
β	Bias.
β	Bias-Vektor.
\mathfrak{B}	CIE Gewichtungskurven.
\mathcal{B}	B-Primärvalenz.
\mathcal{C}	Menge von Merkmalsvektoren.
c	Merkmalsvektor.
c	Kosten des Trainings.
\mathcal{C}	Funktion der Korrekturkurve.
χ	Spektrum.
C	Clique.
c	Element der Clique.
D	Dimension.
d	Abstand der Spiegelplatten.
δ	Nichtlineare Transformationsfunktion.
Δ_s	Gangunterschied (Wegdifferenz) kohärenter Wellen..
E	Bestrahlungsstärke.
\mathcal{E}	Spektrale Empfindlichkeit des k-ten Bands der Kamera.
E	Gibbs-Energie.
η	Quanteneffizienz (QE).
F	Funktion.
\mathfrak{F}	Menge von Bildern.
f	Bild.
\bar{g}	Strahldichte Grün.
G	Informationsgewinn.
\mathcal{G}	G-Primärvalenz.

H	Mapping Funktion.
h	Latente Repräsentation.
m	Harmonische Ordnung.
I	Intensität.
i	Laufvariable.
J	Unreinheit [Sha48].
j	Laufvariable.
K	Kameraverstärkung.
k	Gauß-Kernel.
k	Normierungsfaktor zur Spektralverteilung.
L	Strahldichte.
λ	Wellenlänge.
\mathcal{L}	Bereich/Domäne.
\mathbf{l}	Rekonstruktionsfehler.
Loc	Ortsbereich eines Bildes.
l	Rekonstruktionsfehler.
\mathcal{M}	Seitenlänge eines Makropixels.
n	Hilfsvariable.
n	Brechungsindex.
N_x	Dimension der X-Achse.
N_y	Dimension der Y-Achse.
N_λ	Dimension der Z-Achse.
Ω	Eine Klasse.
ω	Gewichtsmatrix.
$\mathbf{\Omega}$	Menge von Klassen.
P	Funktion der Wahrscheinlichkeit.
\mathbf{p}^H	Hyperpixel.
Φ	Strahlungsfluss.
ψ	Potential.

\mathbf{p}	Pixel.
Q	Strahlungsenergie.
R	Funktion zum Reflexionsfaktor.
\bar{r}	Strahldichte Rot.
ρ	Funktion zum Reflexionsgrad.
\mathcal{R}	R-Primärvalenz.
S	Entscheidungsfunktion.
\mathcal{S}	Loss.
\mathcal{G}	Funktion zur spektralen Leistungsverteilung (SPD).
$T_{H \rightarrow RGB}$	Transformation spektraler Daten in ein RGB-Bild.
Θ	Schwellwert.
θ	Einfallswinkel.
V	Funktion zur spektrale Hell-Empfindlichkeit.
Val	Wertebereich eines Bildes.
v	Wert im Bild.
w	Merkmalsraum.
\mathbf{W}	Menge von Gewichtsverlusten..
w	Gewichtsverlust.
\mathbf{w}	Stützvektor.
\bar{x}	Strahldichte X.
\mathbf{X}	Menge von Variablen in einem CRF.
x	Klassenzuordnung in einem CRF.
x	Eingabedaten.
\mathcal{X}	X-Primärvalenz.
X	Variable in einem CRF.
\mathbf{y}	Ausgabedaten.
\bar{y}	Strahldichte Y.
\mathcal{Y}	Y-Primärvalenz.
y	Label.
\mathbf{Y}	Menge von Labeln.

\bar{z}	Strahldichte Z.
ζ	Z-Primärvalenz.

AKRONYME

- AE Autoencoder. 173, 175–180, 182, 188, 197
- AVIRIS (engl. *Airborne Visible InfraRed Imaging Spectrometer*). 81
- CIE (engl. *Commission Internationale de l’Eclairage*). 40, 41, 45
- CMF (engl. *Color Matching Functions*). 41
- CNN (engl. *Convolutional Neural Network*). 157, 158, 162, 175
- CRF (engl. *Conditional Random Field*). 126, 140–154, 164, 197, 203
- DVI (engl. *Difference Vegetation Index*). 100
- FCRF (engl. *Fully Connected Conditional Random Field*). 139, 143, 147
- FWHM (engl. *Full Width at Half Maximum*). 15, 50
- IMEC Interuniversity Microelectronics Centre. 28
- IoU (engl. *Intersection over Union*). 60, 61
- LBP (engl. *Local Binary Patterns*). 129, 140
- MRF (engl. *Markov Random Fields*). 130, 140
- MTCI (engl. *MERIS Terrestrial Chlorophyll Index*). 101, 102
- NDVI (engl. *Normalized Difference Vegetation Index*). 99–103, 107, 108, 131
- PCA (engl. *Principal Component Analysis*). 19, 125, 127, 130, 176
- QE Quanteneffizienz. 50
- RDVI (engl. *Renormalized Difference Vegetation Index*). 100, 102
- REP (engl. *Red Edge Position*). 101
- RF (engl. *Random Forest*). 152
- RMSE (engl. *Root Mean Square Error*). 100
- ROS (engl. *Robot Operating System*). 87
- SGD (engl. *Stochastic Gradient Descent*). 177
- SPD (engl. *Spectral Power Distribution*). 37
- SVM (engl. *Support Vektor Machine*). 110, 111, 113, 114, 130, 141, 196

LITERATURVERZEICHNIS

- [A⁺10] E. M. V. Association et al. Emva standard 1288, standard for characterization of image sensors and cameras. *Release*, 3:29, 2010.
- [AAMM⁺18] H. Abu Alhaja, S. K. Mustikovela, L. Mescheder, A. Geiger und C. Rother. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. Band 126, Seiten 961–972, September 2018. DOI: 10.1007/s11263-018-1070-x. ISSN: 1573-1405.
- [ABS16] B. Arad und O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In *Computer Vision – ECCV 2016*, Seiten 19–34. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-46478-7_2.
- [AGLL12] J. M. Alvarez, T. Gevers, Y. LeCun und A. M. Lopez. Road scene segmentation from a single image. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato und C. Schmid (Editoren), *Computer Vision – ECCV 2012*, Seiten 376–389, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN: 978-3-642-33786-4.
- [AHLZT18] H. Aasen, E. Honkavaara, A. Lucieer und P. Zarco-Tejada. Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows. *Remote Sensing*, 10(7):1091, jul 2018. DOI: 10.3390/rs10071091.
- [ASS⁺12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua und S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, November 2012. DOI: 10.1109/tpami.2012.120.
- [ATG⁺16] P. Agrawal, K. Tack, B. Geelen, B. Masschelein, P. M. A. Moran, A. Lambrechts und M. Jayapala. Characterization of VNIR hyperspectral sensors with monolithically integrated optical filters. *Electronic Imaging*, 2016(12):1–7, Februar 2016. DOI: 10.2352/issn.2470-1173.2016.12.imse-280.
- [AZL⁺19] F. I. Alam, J. Zhou, A. W. Liew, X. Jia, J. Chanusot und Y. Gao. Conditional random field and deep feature learning for hyperspectral image clas-

- sification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3):1612–1628, March 2019. DOI: 10.1109/TGRS.2018.2867679.
- [AZL16] F. I. Alam, J. Zhou, A. W.-C. Liew und X. Jia. CRF learning with CNN features for hyperspectral image segmentation. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, Julio 2016. DOI: 10.1109/igarss.2016.7730798.
- [BBF00] M. Bertozzi, A. Broggi und A. Fascioli. Vision-based intelligent vehicles: State of the art and perspectives. *Robotics and Autonomous Systems*, 32(1):1 – 16, 2000. DOI: [https://doi.org/10.1016/S0921-8890\(99\)00125-6](https://doi.org/10.1016/S0921-8890(99)00125-6). ISSN: 0921-8890.
- [BBL15] M. Baumgardner, L. Biehl und D. Landgrebe. 220 Band aviris hyperspectral image data set: June 12, 1992 Indian pine test site 3. 2015. DOI: 10.4231/r7rx991c.
- [BBS02] G. J. Briem, J. A. Benediktsson und J. R. Sveinsson. Multiple classifiers applied to multisource remote sensing data. *IEEE transactions on geoscience and remote sensing*, 40(10):2291–2299, Januar 2002. DOI: 10.1109/TGRS.2002.802476.
- [BCM06] L. Bruzzone, M. Chi und M. Marconcini. A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11):3363–3373, 2006. DOI: 10.1109/TGRS.2006.877950.
- [BCV13] Y. Bengio, A. Courville und P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, August 2013. DOI: 10.1109/tpami.2013.50.
- [BD16] M. Belgiu und L. Drăguț. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24 – 31, 2016. DOI: 10.1016/j.isprsjprs.2016.01.011. ISSN: 0924-2716.
- [Ben12] Y. Bengio. Practical recommendations for gradient-based training of deep architectures. In *Lecture Notes in Computer Science*, Seiten 437–478. Springer Berlin Heidelberg, 2012. DOI: 10.1007/978-3-642-35289-8_26.
- [BFC09] G. J. Brostow, J. Fauqueur und R. Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2):88–97, Januar 2009. DOI: 10.1016/j.patrec.2008.04.

- 005.
- [BK16] F. B. Balcik und A. K. Kuzucu. DETERMINATION OF LAND COVER/LAND USE USING SPOT 7 DATA WITH SUPERVISED CLASSIFICATION METHODS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W1:143–146, Oktober 2016. DOI: 10.5194/isprs-archives-xlii-2-w1-143-2016.
- [BKC15] V. Badrinarayanan, A. Kendall und R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *The Computing Research Repository (CoRR)*, abs/1511.00561, 2015. DOI: 10.1109/TPAMI.2016.2644615.
- [BKL02] L. M. Bruce, C. H. Koger und J. Li. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on geoscience and remote sensing*, 40(10):2331–2338, 2002. DOI: 10.1109/TGRS.2002.804721.
- [Bla01] T. Blaschke. What’s wrong with pixels? some recent developments interfacing remote sensing and gis. *GIS – Zeitschrift für Geoinformationssysteme*, 6:12–17, 2001.
- [Boa98] J. W. Boardman. Leveraging the high dimensionality of aviris data for improved sub-pixel target unmixing and rejection of false positives: mixture tuned matched filtering. In *Summaries of the seventh JPL Airborne Geoscience Workshop, JPL Publication, 1998*, Band 97, Seiten 55–56. NASA Jet Propulsion Laboratory, 1998.
- [Bos07] H. Bostrom. Estimating class probabilities in random forests. In *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*. IEEE, Dezember 2007. DOI: 10.1109/icmla.2007.64.
- [BP18] P. Bilinski und V. Prisacariu. Dense decoder shortcut connections for single-pass semantic segmentation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Juni 2018. DOI: 10.1109/cvpr.2018.00690.
- [BPTC95] G. J. Brelstaff, A. Párraga, T. Troscianko und D. Carr. Hyperspectral camera system: acquisition and analysis. In *Geographic Information Systems, Photogrammetry, and Geological/Geophysical Remote Sensing*, Band 2587, Seiten 150–160. International Society for Optics and Photonics, 1995. DOI: 10.1117/12.226819.
- [Bre96] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, August 1996. DOI: 10.1007/

- bf00058655.
- [Bre01] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oktober 2001. DOI: 10.1023/A:1010933404324. ISSN: 1573-0565.
- [BS11] M. Brown und S. Sússtrunk. Multi-spectral sift for scene category recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seiten 177–184. Institute of Electrical and Electronics Engineers (IEEE), 2011. DOI: 10.1109/CVPR.2011.5995637.
- [BSFCo8] G. J. Brostow, J. Shotton, J. Fauqueur und R. Cipolla. Segmentation and recognition using structure from motion point clouds. In *Lecture Notes in Computer Science*, Seiten 44–57. Springer Berlin Heidelberg, 2008. DOI: 10.1007/978-3-540-88682-2_5.
- [BUBo7] D. M. Bradley, R. Unnikrishnan und J. Bagnell. Vegetation detection for driving in complex environments. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, April 2007. DOI: 10.1109/robot.2007.363836.
- [BvdMvRo8] E. Bedini, F. van der Meer und F. van Ruitenbeek. Use of HyMap imaging spectrometer data to map mineralogy in the rodalquilar caldera, southeast Spain. *International Journal of Remote Sensing*, 30(2):327–348, November 2008. DOI: 10.1080/01431160802282854.
- [C⁺99] I. E. Commission et al. IEC 61966-2-1, 1999. *Multimedia systems and equipment—Colour measurements and management—Part*, Seiten 2–1, 1999.
- [C⁺14] S. O.-R. A. V. S. Committee et al. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE Standard J*, 3016:1–16, 2014.
- [Cam11] J. Campbell. *Introduction to remote sensing*. Guilford Press, New York, N.Y, 2011. ISBN: 978-1-60918-176-5.
- [Car94] G. A. Carter. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *Remote sensing*, 15(3):697–703, 1994. DOI: 10.1080/01431169408954109.
- [CB03] A. Cheriyyadat und L. M. Bruce. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, Band 6, Seiten 3420–3422. Institute of Electrical and Electro-

- tics Engineers (IEEE), 2003. DOI: 10.1109/IGARSS.2003.1294808.
- [CB07] M. Chi und L. Bruzzone. Semisupervised classification of hyperspectral images by SVMs optimized in the primal. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1870–1880, 2007. DOI: 10.1109/TGRS.2007.894550.
- [CBJNo4] L. Christensen, B. Bennedsen, R. Jørgensen und H. Nielsen. Modelling nitrogen and phosphorus content at early growth stages in spring barley using hyperspectral line scanning. *Biosystems Engineering*, 88(1):19 – 24, 2004. DOI: <https://doi.org/10.1016/j.biosystemseng.2004.02.006>. ISSN: 1537-5110.
- [CBMB16] L. Cavigelli, D. Bernath, M. Magno und L. Benini. Computationally efficient target classification in multispectral image data with deep neural networks. In *SPIE Security+ Defence*, Seiten 99970L–99970L. International Society for Optics and Photonics, 2016. DOI: 10.1117/12.2241383.
- [CCBS12] J. Carreira, R. Caseiro, J. Batista und C. Sminchisescu. Semantic segmentation with second-order pooling. *Computer Vision–ECCV 2012*, Seiten 430–443, 2012. DOI: 10.1007/978-3-642-33786-4_32.
- [CDK⁺06] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz und Y. Singer. Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 7(Mar):551–585, 2006. ISSN: 1532-4435.
- [CDSA99] C.-I. Chang, Q. Du, T.-L. Sun und M. L. Althouse. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE transactions on geoscience and remote sensing*, 37(6):2631–2641, 1999. DOI: 10.1109/36.803411.
- [cho14] Woodhead publishing series in textiles. In A. K. R. Choudhury (Editor), *Principles of Colour and Appearance Measurement*, Seiten ix – xv. Woodhead Publishing, 2014. DOI: <https://doi.org/10.1016/B978-0-85709-229-8.50013-9>. ISBN: 978-0-85709-229-8.
- [CIE32] C. CIE. Commission internationale de l’éclairage proceedings, 1931. *Cambridge University Press Cambridge*, 1932.
- [CIE17] CIE. *Multispectral image formats*. CIE Central Bureau, Vienna, 2017. ISBN: 978-3-902842-10-7.
- [CJL⁺16] Y. Chen, H. Jiang, C. Li, X. Jia und P. Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks.

- IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016. DOI: 10.1109/TGRS.2016.2584107.
- [CJS⁺98] T. Cocks, R. Jenssen, A. Stewart, I. Wilson und T. Shields. The hymaptm airborne hyperspectral sensor: The system, calibration and performance. In *Proceedings of the 1st EARSeL workshop on Imaging Spectroscopy*, Seiten 37–42. EARSeL, 1998.
- [CKJ10] J. Chetan, M. Krishna und C. Jawahar. Fast and spatially-smooth terrain classification using monocular camera. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, Seiten 4060–4063. Institute of Electrical and Electronics Engineers (IEEE), 2010. DOI: 10.1109/ICPR.2010.987. ISBN: 978-1-4244-7541-4.
- [Com17] W. Commons. File:landsat 3.jpg — wikimedia commons, the free media repository, 2017. [Online; accessed 2-October-2019].
- [Com18] W. Commons. File:hyperspectralcube.jpg — wikimedia commons, the free media repository, 2018. [Online; accessed 2-October-2019].
- [COR⁺16] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth und B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Seiten 3213–3223, 2016. DOI: 10.1109/CVPR.2016.350.
- [Cov65] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE transactions on electronic computers*, (3):326–334, 1965. DOI: 10.1109/PGEC.1965.264137.
- [CPo8] J. C.-W. Chan und D. Paelinckx. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6):2999 – 3011, 2008. DOI: 10.1016/j.rse.2008.02.011. ISSN: 0034-4257.
- [CPK⁺16] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy und A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *The Computing Research Repository (CoRR)*, abs/1606.00915, 2016. DOI: 10.1109/TPAMI.2017.2699184.
- [CPSA17] L. Chen, G. Papandreou, F. Schroff und H. Adam. Rethinking atrous convolution for semantic image

- segmentation. *The Computing Research Repository (CoRR)*, abs/1706.05587, 2017.
- [CS06] M. A. Cho und A. K. Skidmore. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote sensing of environment*, 101(2):181–193, 2006. DOI: 10.1016/j.rse.2005.12.011.
- [CSK⁺93] R. CLARK, G. SWAYZE, T. KING, A. GALLAGHER und W. CALVIN. The us geological survey, digital spectral reflectance library. In *JPL, Summaries of the 4th Annual JPL Airborne Geoscience Workshop.*, Band 1, 1993.
- [CSS06] M. A. Cho, I. M. Sobhan und A. K. Skidmore. Estimating fresh grass/herb biomass from hmap data using the red edge position. In *Remote Sensing and Modeling of Ecosystems for Sustainability III*, Band 6298, Seite 629805. International Society for Optics and Photonics, 2006. DOI: 10.1117/12.681640.
- [CST⁺00] N. Cristianini, J. Shawe-Taylor et al. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000. DOI: 10.1017/CB09780511801389.
- [CVB05] G. Camps-Valls und L. Bruzzone. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362, 2005. DOI: 10.1109/TGRS.2005.846154.
- [CVGCMM⁺06] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés und J. Calpe-Maravilla. Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 3(1):93–97, 2006. DOI: 10.1109/LGRS.2005.857031.
- [CVMZ07] G. Camps-Valls, T. V. B. Marsheva und D. Zhou. Semi-supervised graph-based hyperspectral image classification. *IEEE transactions on Geoscience and Remote Sensing*, 45(10):3044–3054, 2007. DOI: 10.1109/TGRS.2007.895416.
- [CVTGC⁺11] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez und J. Malo. Remote sensing image processing. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 5(1):1–192, 2011. DOI: 10.2200/S00392ED1V01Y201107IVM012.
- [Cyb89] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989. DOI: 10.1007/BF02551274.

- [CZ11] A. Chakrabarti und T. Zickler. Statistics of real-world hyperspectral images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, Seiten 193–200. Institute of Electrical and Electronics Engineers (IEEE), 2011. DOI: 10.1109/CVPR.2011.5995660.
- [CZ]15] Y. Chen, X. Zhao und X. Jia. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392, 2015. DOI: 10.1109/JSTARS.2015.2388577.
- [DBVG09] M. Dalponte, L. Bruzzone, L. Vescovo und D. Gianelle. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment*, 113(11):2345 – 2355, 2009. DOI: <https://doi.org/10.1016/j.rse.2009.06.013>. ISSN: 0034-4257.
- [DC04] J. Dash und P. Curran. The meris terrestrial chlorophyll index. 2004. DOI: 10.1080/0143116042000274015.
- [DC]⁺15] K. Degraux, V. Cambareri, L. Jacques, B. Geelen, C. Blanch und G. Lafruit. Generalized inpainting method for hyperspectral image acquisition. In *Image Processing (ICIP), 2015 IEEE International Conference on*, Seiten 315–319. Institute of Electrical and Electronics Engineers (IEEE), 2015. DOI: 10.1109/ICIP.2015.7350811.
- [DGo6] J. Davis und M. Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*, Seite 233–240, New York, NY, USA, 2006. Association for Computing Machinery. DOI: 10.1145/1143844.1143874. ISBN: 1595933832.
- [DGF03] F. Dell’Acqua, P. Gamba und A. Ferrari. Exploiting spectral and spatial information for classifying hyperspectral data in urban areas. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS’03. Proceedings. 2003 IEEE International*, Band 1, Seiten 464–466. Institute of Electrical and Electronics Engineers (IEEE), 2003. DOI: 10.1109/IGARSS.2003.1293810.
- [DMH⁺14] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama et al. Hyperspectral and LiDAR data fusion: Outcome of the 2013 grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2405–

- 2418, 2014. DOI: 10.1109/JSTARS.2014.2305441.
- [DPS03] C. D'Elia, G. Poggi und G. Scarpa. A tree-structured markov random field model for bayesian image segmentation. *IEEE Transactions on Image Processing*, 12(10):1259–1273, Oct 2003. DOI: 10.1109/TIP.2003.817257.
- [DTTB07] P. Dollár, Z. Tu, H. Tao und S. Belongie. Feature mining for image classification. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Seiten 1–8. Institute of Electrical and Electronics Engineers (IEEE), 2007. DOI: 10.1109/CVPR.2007.383046.
- [Dudo1] R. Duda. *Pattern classification*. Wiley, New York, 2001. ISBN: 978-0-471-05669-0.
- [DVC⁺16] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury und C. Pal. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*, Seiten 179–187. Springer, 2016. DOI: 10.1007/978-3-319-46976-8_19.
- [DZZP10] W. Di, L. Zhang, D. Zhang und Q. Pan. Studies on hyperspectral face recognition in visible spectrum with feature band selection. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(6):1354–1361, 2010. DOI: 10.1109/TSMCA.2010.2052603.
- [EEV⁺15] J. Eckhard, T. Eckhard, E. M. Valero, J. L. Nieves und E. G. Contreras. Outdoor scene reflectance measurements using a Bragg-grating-based hyperspectral imager. *Applied Optics*, 54(13):D15–D24, 2015. DOI: 10.1364/AO.54.000D15.
- [EMGVG09] A. Ess, T. Müller, H. Grabner und L. J. Van Gool. Segmentation-based urban traffic scene understanding. In *BMVC*, Band 1, Seite 2, 2009. DOI: 10.5244/C.23.84.
- [EVGW⁺10] M. Everingham, L. Van Gool, C. K. Williams, J. Winn und A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. DOI: 10.1007/s11263-009-0275-4.
- [Fab99] C. Fabry. Theorie et applications d'une nouvelle methods de spectroscopie intereferentielle. *Ann. Chim. Ser. 7*, 16:115–144, 1899.
- [FAN16] D. H. Foster, K. Amano und S. M. Nascimento. Time-lapse ratios of cone excitations in natural scenes. *Vision research*, 120:45–60, 2016. DOI: 10.1016/j.visres.2015.03.012.

- [FANFo6] D. H. Foster, K. Amano, S. M. Nascimento und M. J. Foster. Frequency of metamerism in natural scenes. *Josa a*, 23(10):2359–2372, 2006. DOI: 10.1364/JOSAA.23.002359.
- [FBCSo8] M. Fauvel, J. A. Benediktsson, J. Chanussot und J. R. Sveinsson. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, 2008. DOI: 10.1109/TGRS.2008.922034.
- [FCBo6] M. Fauvel, J. Chanussot und J. A. Benediktsson. Decision fusion for the classification of urban remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):2828–2838, 2006. DOI: 10.1109/TGRS.2006.876708.
- [FHLDo6] G. D. Finlayson, S. D. Hordley, C. Lu und M. S. Drew. On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):59–68, 2006. DOI: 10.1109/TPAMI.2006.18.
- [FLD⁺15] L. Fang, S. Li, W. Duan, J. Ren und J. A. Benediktsson. Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels. *IEEE Transactions on Geoscience and Remote Sensing*, 53(12):6663–6674, Dec 2015. DOI: 10.1109/TGRS.2015.2445767.
- [Fo004] G. M. Foody. Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5):627–633, 2004. DOI: 10.14358/PERS.70.5.627.
- [FPo2] D. A. Forsyth und J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002. ISBN: 0130851981.
- [FTT17] K. Fotiadou, G. Tsagkatakis und P. Tsakalides. Deep convolutional neural networks for the classification of snapshot mosaic hyperspectral imagery. *Electronic Imaging*, 2017(17):185–190, 2017. DOI: 10.2352/ISSN.2470-1173.2017.17.COIMG-445.
- [FVS09] B. Fulkerson, A. Vedaldi und S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *Computer Vision, 2009 IEEE 12th International Conference on*, Seiten 670–677. Institute of Electrical and Electronics Engineers (IEEE), Institute of Electrical and Electronics Engineers (IEEE), 2009. DOI: 10.1109/ICCV.2009.5459175.
- [Gal90] S. Gallant. Perceptron-based learning algorithms. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 1(2):179 – 191, 1990.

- DOI: 10.1109/72.80230.
- [GBC16] I. Goodfellow, Y. Bengio und A. Courville. *Deep Learning*. MIT Press, Cambridge, Massachusetts, 2016. ISBN: 978-0262035613. <http://www.deeplearningbook.org>.
- [GBSC88] A. A. Green, M. Berman, P. Switzer und M. D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, Jan 1988. DOI: 10.1109/36.3001.
- [GCCVMMCo8] L. Gómez-Chova, G. Camps-Valls, J. Muñoz-Mari und J. Calpe. Semisupervised image classification with laplacian support vector machines. *IEEE Geoscience and Remote Sensing Letters*, 5(3):336–340, 2008. DOI: 10.1109/LGRS.2008.916070.
- [GCZ16] P. Ghamisi, Y. Chen und X. X. Zhu. A self-improving convolution neural network for the classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 13(10):1537–1541, 2016. DOI: 10.1109/LGRS.2016.2595108.
- [GGDN08] B. Guo, S. R. Gunn, R. I. Damper und J. D. Nelson. Customizing kernel functions for SVM-based hyperspectral image classification. *IEEE Transactions on Image Processing*, 17(4):622–629, 2008. DOI: 10.1109/TIP.2008.918955.
- [GGM03] A. A. Gitelson, Y. Gritz und M. N. Merzlyak. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of plant physiology*, 160(3):271–282, 2003. DOI: 10.1078/0176-1617-00887.
- [GGEO⁺18] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez und J. Garcia-Rodriguez. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, 70:41–65, September 2018. DOI: 10.1016/j.asoc.2018.05.018.
- [GK07] J. Grehn und J. Krause. *Metzler Physik SII*, chapter 3, Seiten 132–133. Schroedel Verlag GmbH, Braunschweig, 4. Auflage, August 2007. ISBN: 3507107104.
- [GKSW65] D. M. Gates, H. J. Keegan, J. C. Schleter und V. R. Weidner. Spectral properties of plants. *Applied optics*, 4(1):11–20, 1965. DOI: 10.1364/AO.4.000011.
- [GLU12] A. Geiger, P. Lenz und R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark sui-

- te. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, Seiten 3354–3361. Institute of Electrical and Electronics Engineers (IEEE), 2012. DOI: 10.1109/CVPR.2012.6248074.
- [GME02] C. Gueymard, D. Myers und K. Emery. Proposed reference irradiance spectra for solar energy systems testing. *Solar energy*, 73(6):443–467, 2002. DOI: 10.1016/S0038-092X(03)00005-7.
- [GOC⁺07] A. Gowen, C. O'Donnell, P. Cullen, G. Downey und J. Frias. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. *Trends in Food Science and Technology*, 18(12):590 – 598, 2007. DOI: <https://doi.org/10.1016/j.tifs.2007.06.001>. ISSN: 0924-2244.
- [Gol10] D. B. Goldman. Vignette and exposure calibration and compensation. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2276–2288, 2010. DOI: 10.1109/TPAMI.2010.55.
- [Gra53] H. Grassmann. Zur theorie der farbenmischung. *Annalen der Physik*, 165(5):69–84, 1853. DOI: 10.1002/andp.18531650505.
- [GRBo8] C. Galleguillos, A. Rabinovich und S. Belongie. Object categorization using co-occurrence, location and appearance. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Seiten 1–8. Institute of Electrical and Electronics Engineers (IEEE), Institute of Electrical and Electronics Engineers (IEEE), 2008. DOI: 10.1109/CVPR.2008.4587799.
- [GSW00] T. Gevers, H. Stokman und J. v. d. Weijer. Colour constancy from hyper-spectral data. In *Proceedings of the British Machine Vision Conference*, Seiten 30.1–30.10. BMVA Press, 2000. DOI: 10.5244/C.14.30.
- [GTL14] B. Geelen, N. Tack und A. Lambrechts. A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic. In *Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VII*, Band 8974, Seite 89740L. International Society for Optics and Photonics, International Society for Optics and Photonics (SPIE), März 2014. DOI: 10.1117/12.2037607.
- [GVSR85] A. F. Goetz, G. Vane, J. E. Solomon und B. N. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153, 1985. DOI: 10.1126/science.228.4704.1147. ISSN: 0036-8075.
- [GYMo6] Y. Garini, I. T. Young und G. McNamara. Spectral imaging: Principles and applications. *Cytometry*

- Part A*, 69A(8):735–747, 2006. DOI: 10.1002/cyto.a.20311.
- [Har97] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 19(6):580–593, 1997. DOI: 10.1109/34.601246.
- [HB20] A. M. Hafiz und G. M. Bhat. A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval*, 9(3):171–189, Jul 2020. DOI: 10.1007/s13735-020-00195-x. ISSN: 2192-662X.
- [HC94] J. C. Harsanyi und C.-I. Chang. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Transactions on geoscience and remote sensing*, 32(4):779–785, 1994. DOI: 10.1109/36.298007.
- [HDT02] C. Huang, L. Davis und J. Townshend. An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4):725–749, 2002. DOI: 10.1080/01431160110040323.
- [HFM04] S. Hordley, G. Finalyson und P. Morovic. A multispectral image database and its application to image rendering across illumination. In *Image and Graphics (ICIG'04), Third International Conference on*, Seiten 394–397. Institute of Electrical and Electronics Engineers (IEEE), 2004. DOI: 10.1109/ICIG.2004.10.
- [HK13] N. Hagen und M. W. Kudenov. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):090901–090901, 2013. DOI: 10.1117/1.OE.52.9.090901.
- [HLW17] G. Huang, Z. Liu und K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seiten 2261–2269, Julio 2017. DOI: 10.1109/CVPR.2017.243.
- [HOP⁺14] T. Hirvonen, J. Orava, N. Penttinen, K. Luostarinen, M. Hauta-Kasari, M. Sorjonen und K.-E. Peiponen. Spectral image database for observing the quality of nordic sawn timbers. *Wood science and technology*, 48(5):995–1003, 2014. DOI: 10.1007/s00226-014-0655-y.
- [HP11] R. W. G. Hunt und M. R. Pointer. *Measuring Colour*. John Wiley & Sons, Ltd, September 2011. DOI: 10.1002/9781119975595.

- [HRZZ09] T. Hastie, S. Rosset, J. Zhu und H. Zou. Multi-class adaboost. *Statistics and its Interface*, 2(3):349–360, 2009. DOI: 10.4310/SII.2009.v2.n3.a8.
- [HS⁺73] R. M. Haralick, K. Shanmugam et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973. DOI: 10.1109/TSMC.1973.4309314.
- [HS06] G. E. Hinton und R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006. DOI: 10.1126/science.1127647.
- [HSF94] P. M. Hubel, D. Sherman und J. E. Farrell. A comparison of methods of sensor spectral sensitivity estimation. *Color and Imaging Conference*, 1994(1):45–48, 1994. ISSN: 2166-9635.
- [Hua02] K. Huang. A synergistic automatic clustering technique (syntract) for multispectral image analysis. *Photogrammetric Engineering and Remote Sensing*, 68:33–40, 01 2002.
- [Hug68] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63, 1968. DOI: 10.1109/TIT.1968.1054102.
- [HW90] D.-C. He und L. Wang. Texture unit, texture spectrum, and texture analysis. *IEEE transactions on Geoscience and Remote Sensing*, 28(4):509–512, 1990. DOI: 10.1109/TGRS.1990.572934.
- [HYCG05] J. Ham, Yangchi Chen, M. M. Crawford und J. Ghosh. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501, March 2005. DOI: 10.1109/TGRS.2004.842481.
- [HZC⁺17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto und H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *The Computing Research Repository (CoRR)*, abs/1704.04861, 2017.
- [HZRS16] K. He, X. Zhang, S. Ren und J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Seiten 770–778, 2016. DOI: 10.1109/CVPR.2016.90.
- [IEE] IEEE. IEEE standard letter designations for radar-frequency bands. DOI: 10.1109/ieeestd.2003.94224.

- [IS15] S. Ioffe und C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In F. Bach und D. Blei (Editoren), *Proceedings of the 32nd International Conference on Machine Learning*, Band 37 of *Proceedings of Machine Learning Research*, Seiten 448–456, Lille, France, 07–09 Jul 2015. Proceedings of Machine Learning Research (PMLR).
- [ISO89] Thermal insulation – Heat transfer by radiation – Physical quantities and definitions. Standard, International Organization for Standardization, Geneva, CH, März 1989.
- [Jaco2] P. Jaccard. Lois de distribution florale dans la zone alpine. *Bull Soc Vaudoise Sci Nat*, 38:69–130, 1902. DOI: 10.5169/seals-266762.
- [JDS⁺16] J. Jablonski, C. Durell, T. Slonecker, K. Wong, B. Simon, A. Eichelberger und J. Osterberg. Best practices in passive remote sensing VNIR hyperspectral system hardware calibrations. In D. P. Bannan (Editor), *Hyperspectral Imaging Sensors: Innovative Applications and Sensor Standards 2016*, Band 9860, Seiten 13 – 42. International Society for Optics and Photonics, SPIE, 2016. DOI: 10.1117/12.2224022.
- [JDV⁺16] S. Jégou, M. Drozdal, D. Vázquez, A. Romero und Y. Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *The Computing Research Repository (CoRR)*, abs/1611.09326, 2016. DOI: 10.1109/CVPRW.2017.156.
- [Jen15] J. R. Jensen. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Prentice Hall Press, Upper Saddle River, NJ, USA, 4th. Auflage, 2015. ISBN: 013405816X, 9780134058160.
- [JKC13] X. Jia, B.-C. Kuo und M. M. Crawford. Feature mining for hyperspectral image classification. *Proceedings of the IEEE*, 101(3):676–697, 2013. DOI: 10.1109/JPROC.2012.2229082.
- [JL95] G. H. John und P. Langley. Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, UAI'95*, Seiten 338–345. Morgan Kaufmann Publishers Inc., 1995. ISBN: 1-55860-385-9.
- [JL99] L. O. Jimenez und D. A. Landgrebe. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2653–2667, 1999. DOI: 10.1109/36.803413.

- [JL⁺01] Q. Jackson, D. A. Landgrebe et al. An adaptive classifier design for high-dimensional data analysis with a limited training data set. *IEEE Transactions on Geoscience and Remote Sensing*, 39(12):2664–2679, 2001. DOI: 10.1109/36.975001.
- [Jä12] B. Jähne. *Digitale Bildverarbeitung*. Springer Berlin Heidelberg, 2012. DOI: 10.1007/978-3-642-04952-1.
- [KB15] D. P. Kingma und J. Ba. Adam: A method for stochastic optimization. 2015.
- [KEK90] T. Kusaka, H. Egawa und Y. Kawata. Classification of the spot image using spectral and spatial features of primitive regions that have nearly uniform color. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):749–752, July 1990. DOI: 10.1109/TGRS.1990.573011.
- [KGC01] S. Kumar, J. Ghosh und M. M. Crawford. Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1368–1379, July 2001. DOI: 10.1109/36.934070.
- [KHM16] J. Klein, B. Hill und D. Merhof. Multispectral imaging: aberrations and acquisitions from different viewing positions. 2016.
- [KK11] P. Kraehenbuehl und V. Koltun. Efficient inference in fully connected crfs with Gaussian edge potentials. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira und K. Q. Weinberger (Editoren), *Advances in Neural Information Processing Systems 24*, Seiten 109–117. Curran Associates, Inc., 2011. ISBN: 978-1-61839-599-3.
- [KK13] P. Kraehenbuehl und V. Koltun. Parameter learning and convergent inference for dense random fields. In S. Dasgupta und D. McAllester (Editoren), *Proceedings of the 30th International Conference on Machine Learning*, Band 28 of *Proceedings of Machine Learning Research*, Seiten 513–521, Atlanta, Georgia, USA, Juni 2013. PMLR.
- [KLB⁺93] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon und A. Goetz. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2):145 – 163, 1993. DOI: 10.1016/0034-4257(93)90013-N. ISSN: 0034-4257. Airbone Imaging Spectrometry.

- [KMM⁺18] H. A. Khan, S. Mihoubi, B. Mathon, J.-B. Thomas und J. Y. Hardeberg. Hytexila: High resolution visible and near infrared hyperspectral texture images. *Sensors*, 18(7):2045, 2018. DOI: 10.3390/s18072045.
- [KSH12] A. Krizhevsky, I. Sutskever und G. E. Hinton. Image-net classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou und K. Q. Weinberger (Editoren), *Advances in Neural Information Processing Systems 25*, Seiten 1097–1105. Curran Associates, Inc., 2012.
- [Lano2] D. Landgrebe. Hyperspectral image data analysis. *IEEE Signal processing magazine*, 19(1):17–28, 2002. DOI: 10.1109/79.974718.
- [Lano5] D. A. Landgrebe. *Frontmatter*. John Wiley and Sons, Ltd, 2005. DOI: 10.1002/0471723800.fmatter. ISBN: 9780471723806.
- [LBBH98] Y. LeCun, L. Bottou, Y. Bengio und P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. DOI: 10.1109/5.726791.
- [LBDP10] J. Li, J. M. Bioucas-Dias und A. Plaza. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4085–4098, 2010. DOI: 10.1109/TGRS.2010.2060550.
- [LBDP12] J. Li, J. M. Bioucas-Dias und A. Plaza. Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3):809–823, 2012. DOI: 10.1109/TGRS.2011.2162649.
- [LFOM15] L. Lopez-Fuentes, G. Oliver und S. Massanet. Revisiting image vignetting correction by constrained minimization of log-intensity entropy. In *International Work-Conference on Artificial Neural Networks*, Seiten 450–463. Springer, 2015. DOI: 10.1007/978-3-319-19222-2_38.
- [LGG⁺14] A. Lambrechts, P. Gonzalez, B. Geelen, P. Soussan, K. Tack und M. Jayapala. A CMOS-compatible, integrated approach to hyper-and multispectral imaging. In *Electron Devices Meeting (IEDM), 2014 IEEE International*, Seiten 10–5. Institute of Electrical and Electronics Engineers (IEEE), 2014. DOI: 10.1109/IEDM.2014.7047025.

- [LLS15] F. Liu, G. Lin und C. Shen. Crf learning with CNN features for image segmentation. *Pattern Recognition*, 48(10):2983–2992, 2015. DOI: 10.1016/j.patcog.2015.04.019.
- [LMB⁺14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár und C. L. Zitnick. Microsoft coco: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele und T. Tuytelaars (Editoren), *Computer Vision – ECCV 2014*, Seiten 740–755, Cham, 2014. Springer International Publishing. ISBN: 978-3-319-10602-1.
- [LMP01] J. D. Lafferty, A. McCallum und F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, Seiten 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN: 1-55860-778-1.
- [LMSR17] G. Lin, A. Milan, C. Shen und I. D. Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seiten 5168–5177, 2017. DOI: 10.1109/CVPR.2017.549.
- [Low99] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Band 2, Seiten 1150–1157 vol.2, Sep. 1999. DOI: 10.1109/ICCV.1999.790410.
- [LSD17] J. Long, E. Shelhamer und T. Darrell. Fully convolutional networks for semantic segmentation. Band 39, Seiten 640–651, April 2017. DOI: 10.1109/TPAMI.2016.2572683.
- [LSR⁺12] L. Ladický, P. Sturges, C. Russell, S. Sengupta, Y. Bastanlar, W. Clocksin und P. H. S. Torr. Joint optimization for object class segmentation and dense stereo reconstruction. *International Journal of Computer Vision*, 100(2):122–133, November 2012. DOI: 10.1007/s11263-011-0489-0. ISSN: 1573-1405.
- [LW07] D. Lu und Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, März 2007. DOI: 10.1080/01431160600746456.
- [LWTG14] P.-J. Lapray, X. Wang, J.-B. Thomas und P. Gouton. Multispectral filter arrays: Recent advances and practical implementation. *Sensors*, 14(11):21626–

- 21659, 2014. DOI: 10.3390/s141121626.
- [LXS⁺15] F. Li, L. Xu, P. Siva, A. Wong und D. A. Clausi. Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2427–2438, 2015. DOI: 10.1109/JSTARS.2015.2414816.
- [LY07] C. P. Lo und A. K. W. Yeung. *Concepts and Techniques of Geographic Information Systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2007. DOI: 10.1080/1365881031000111173. ISBN: 013149502X.
- [MAM01] A. Maccioni, G. Agati und P. Mazzinghi. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. *Journal of Photochemistry and Photobiology B: Biology*, 61(1-2):52–61, 2001. DOI: 10.1016/S1011-1344(01)00145-2.
- [MB04] F. Melgani und L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790, 2004. DOI: 10.1109/TGRS.2004.831865.
- [MCM⁺11] R. Main, M. A. Cho, R. Mathieu, M. M. O’Kennedy, A. Ramoelo und S. Koch. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6):751–761, 2011. DOI: 10.1016/j.isprsjprs.2011.08.001.
- [MGB⁺11] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke und Q. Weng. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5):1145–1161, Mai 2011. DOI: 10.1016/j.rse.2010.12.017.
- [MGP⁺15] S. L. Moan, S. T. George, M. Pedersen, J. Blahová und J. Y. Hardeberg. A database for spectral image quality. In M.-C. Larabi und S. Triantaphillidou (Editoren), *Image Quality and System Performance XII*, Band 9396, Seiten 225 – 232. International Society for Optics and Photonics, SPIE, 2015. DOI: 10.1117/12.2080760.
- [MHW90] J. Miller, E. Hare und J. Wu. Quantitative characterization of the vegetation red edge reflectance 1. an inverted-gaussian reflectance model. *Remote Sensing*, 11(10):1755–1773, 1990.

- [Mir18] A. Mirhashemi. Introducing spectral moment features in analyzing the spectex hyperspectral texture database. *Machine Vision and Applications*, 29(3):415–432, 2018. DOI: 10.1007/s00138-017-0892-9.
- [MMNM07] N. MEMARSADEGHI, D. M. MOUNT, N. S. NETANYAHU und J. L. MOIGNE. A FAST IMPLEMENTATION OF THE ISODATA CLUSTERING ALGORITHM. *International Journal of Computational Geometry & Applications*, 17(01):71–103, Februar 2007. DOI: 10.1142/s0218195907002252.
- [MSA⁺78] E. J. Milton, M. E. Schaeppman, K. Anderson, M. Kneubühler und N. Fox. Progress in field spectroscopy. *Remote Sensing of Environment*, 113:S92–S109, 1978.
- [MSM17] D. Mishkin, N. Sergievskiy und J. Matas. Systematic evaluation of convolution neural network advances on the imagenet. *Comput. Vis. Image Underst.*, 161(C):11–19, August 2017. DOI: 10.1016/j.cviu.2017.05.007. ISSN: 1077-3142.
- [NA17] A. Noviyanto und W. H. Abdullah. Honey dataset standard using hyperspectral imaging for machine learning problems. In *Signal Processing Conference (EUSIPCO), 2017 25th European*, Seiten 473–477. Institute of Electrical and Electronics Engineers (IEEE), 2017. DOI: 10.23919/EUSIPCO.2017.8081252.
- [NAF16] S. M. Nascimento, K. Amano und D. H. Foster. Spatial distributions of local illumination color in natural scenes. *Vision Research*, 120:39–44, 2016. DOI: 10.1016/j.visres.2015.07.005.
- [NAS18] NASA. *AVIRIS Data*, 1995 (accessed August 4, 2018).
- [NFF02] S. M. Nascimento, F. P. Ferreira und D. H. Foster. Statistics of spatial cone-excitation ratios in natural scenes. *JOSA A*, 19(8):1484–1490, 2002. DOI: 10.1364/JOSAA.19.001484.
- [NGV⁺18] M. Nouri, N. Gorretta, P. Vaysse, M. Giraud, C. Germain, B. Keresztes und J.-M. Roger. Near infrared hyperspectral dataset of healthy and infected apple tree leaves images for the early detection of apple scab disease. *Data in brief*, 16:967–971, 2018. DOI: 10.1016/j.dib.2017.12.043.
- [NMNA13] I. Nishidate, T. Maeda, K. Niizeki und Y. Aizu. Estimation of melanin and hemoglobin using spectral reflectance images reconstructed from a digital rgb image by the wiener estimation method. *Sensors*, 13(6):7902–7915, Jun 2013. DOI: 10.3390/s130607902.

- ISSN: 1424-8220.
- [NORBK17] G. Neuhold, T. Ollmann, S. Rota Bulo und P. Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE International Conference on Computer Vision*, Seiten 4990–4999, 2017. DOI: 10.1109/ICCV.2017.534.
- [NP12] S. T. Namin und L. Petersson. Classification of materials in natural scenes using multi-spectral images. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Seiten 1393–1398. Institute of Electrical and Electronics Engineers (IEEE), 2012. DOI: 10.1109/IROS.2012.6386074.
- [NPB14] R. M. Nguyen, D. K. Prasad und M. S. Brown. Training-based spectral reconstruction from a single rgb image. In *European Conference on Computer Vision*, Seiten 186–201. Springer, 2014. DOI: 10.1007/978-3-319-10584-0_13.
- [NS05] R. Neher und A. Srivastava. A bayesian mrf framework for labeling terrain using hyperspectral imaging. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1363–1374, June 2005. DOI: 10.1109/TGRS.2005.846865.
- [NWO96] G. A. Naghdy, J. Wang und P. O. Ogunbona. Texture analysis using Gabor wavelets. In B. E. Rogowitz und J. P. Allebach (Editoren), *Human Vision and Electronic Imaging*, Band 2657, Seiten 74 – 85. International Society for Optics and Photonics, SPIE, 1996. DOI: 10.1117/12.238703.
- [OPH96] T. Ojala, M. Pietikäinen und D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996. DOI: 10.1016/0031-3203(95)00067-4.
- [oV18] T. A. C. of Virginia. *Washington DC Mall*, 1995 (accessed August 4, 2018).
- [PBB⁺09] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri et al. Recent advances in techniques for hyperspectral image processing. *Remote sensing of environment*, 113:S110–S122, 2009. DOI: 10.1016/j.rse.2007.07.028.
- [PBS⁺03] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso und S. L. Carman. Hyperion, a space-based imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1160–1173, 2003. DOI: 10.1109/TGRS.2003.815018.

- [PF99] A. Perot und C. Fabry. On the application of interference phenomena to the solution of various problems of spectroscopy and metrology. *The Astrophysical Journal*, 9:87, 1899. DOI: 10.1086/140557.
- [PHML17] T. Pohlen, A. Hermans, M. Mathias und B. Leibe. Full-resolution residual networks for semantic segmentation in street scenes. *arXiv preprint*, 2017. DOI: 10.1109/CVPR.2017.353.
- [PK10] J. Pasher und D. J. King. Multivariate forest structure modelling and mapping using high resolution airborne imagery and topographic information. *Remote Sensing of Environment*, 114(8):1718–1732, August 2010. DOI: 10.1016/j.rse.2010.03.005.
- [PPM09] A. Plaza, J. Plaza und G. Martin. Incorporation of spatial constraints into spectral mixture analysis of remotely sensed hyperspectral data. In *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*, Seiten 1–6. Institute of Electrical and Electronics Engineers (IEEE), 2009. DOI: 10.1109/MLSP.2009.5306202.
- [Pri15a] L. Priese. *Computer Vision: Einführung in die Verarbeitung und Analyse digitaler Bilder*, chapter 4, Seiten 61–65. Springer-Verlag, 2015. DOI: 10.1007/978-3-662-45129-8. ISBN: 978-3-662-45128-1.
- [Pri15b] L. Priese. *Computer Vision: Einführung in die Verarbeitung und Analyse digitaler Bilder*, chapter 3, Seiten 33–37. Springer-Verlag, 2015. ISBN: 978-3-662-45128-1.
- [PWS⁺01] D. R. Peddle, H. P. White, R. J. Soffer, J. R. Miller und E. F. Ledrew. Reflectance processing of remote sensing spectroradiometer data. *Computers & geosciences*, 27(2):203–213, 2001.
- [PZY⁺17] C. Peng, X. Zhang, G. Yu, G. Luo und J. Sun. Large kernel matters - improve semantic segmentation by global convolutional network. *The Computing Research Repository (CoRR)*, abs/1703.02719, 2017. DOI: 10.1109/CVPR.2017.189.
- [QCK⁺13] J. Qin, K. Chao, M. S. Kim, R. Lu und T. F. Burks. Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, 118(2):157–171, 2013. DOI: 10.1016/j.jfoodeng.2013.04.001.
- [QGC⁺09] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler und A. Ng. Ros: an open-source robot operating system. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA)*

- Workshop on Open Source Robotics*, Kobe, Japan, Mai 2009.
- [Qui86] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986. DOI: 10.1007/BF00116251.
- [Ray02] S. F. Ray. *Applied photographic optics: Lenses and optical systems for photography, film, video, electronic and digital imaging*. Focal Press, 2002. DOI: 10.4324/9780080499253.
- [RB95] J.-L. Roujean und F.-M. Breon. Estimating par absorbed by vegetation from bidirectional reflectance measurements. *Remote sensing of Environment*, 51(3):375–384, 1995. DOI: 10.1016/0034-4257(94)00114-3.
- [RBB99] M. Rast, J. L. Bezy und S. Bruzzi. The ESA medium resolution imaging spectrometer MERIS a review of the instrument and its mission. *International Journal of Remote Sensing*, 20(9):1681–1702, Januar 1999. DOI: 10.1080/014311699212416.
- [RBZ⁺93] L. J. Rickard, R. W. Basedow, E. F. Zalewski, P. R. Silverglate und M. Landers. Hydice: An airborne system for hyperspectral imaging. In *Imaging Spectrometry of the Terrestrial Environment*, Band 1937, Seiten 173–179. International Society for Optics and Photonics, 1993.
- [RDFZ04] G. Rellier, X. Descombes, F. Falzon und J. Zerubia. Texture feature analysis using a gauss-markov model in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1543–1551, July 2004. DOI: 10.1109/TGRS.2004.830170.
- [RFB15] O. Ronneberger, P. Fischer und T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells und A. F. Frangi (Editoren), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Seiten 234–241, Cham, 2015. Springer International Publishing. DOI: 10.1007/978-3-319-24574-4_28. ISBN: 978-3-319-24574-4.
- [RFRB10] R. L. Rich, L. Frelich, P. B. Reich und M. E. Bauer. Detecting wind disturbance severity and canopy heterogeneity in boreal forest by coupling high-spatial resolution satellite imagery and field data. *Remote Sensing of Environment*, 114(2):299 – 308, 2010. DOI: 10.1016/j.rse.2009.09.005. ISSN: 0034-4257.
- [RGC⁺98] D. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer und R. Green. Mapping chaparral in the

- santa monica mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, 65(3):267 – 279, 1998. DOI: 10.1016/S0034-4257(98)00037-6. ISSN: 0034-4257.
- [Rico06] J. A. Richards. *Remote sensing digital image analysis*, Band 3. Springer, 2006. DOI: 10.1007/978-3-662-02462-1.
- [RJ72] J. Rouse Jr. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. 1972.
- [RKAJo8] E. Reinhard, E. A. Khan, A. O. Akyuz und G. Johnson. *Color imaging: fundamentals and applications*. AK Peters/CRC Press, 2008. DOI: 10.1201/b10637.
- [RSB96] G. Rondeaux, M. Steven und F. Baret. Optimization of soil-adjusted vegetation indices. *Remote sensing of environment*, 55(2):95–107, 1996. DOI: 10.1016/0034-4257(95)00186-7.
- [RSM⁺16] G. Ros, L. Sellart, J. Materzynska, D. Vazquez und A. M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Juni 2016. DOI: 10.1109/CVPR.2016.352.
- [RVG⁺07] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora und S. Belongie. Objects in context. 2007. DOI: 10.1109/ICCV.2007.4408986.
- [SA99] P. C. Smits und A. Annoni. Updating land-cover maps by using texture information from very high-resolution space-borne imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1244–1254, May 1999. DOI: 10.1109/36.763282.
- [SB01] S. B. Serpico und L. Bruzzone. A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1360–1367, 2001. DOI: 10.1109/36.934069.
- [SB02] G. Sharma und R. Bala. *Digital color imaging handbook*. CRC press, 2002.
- [SB03] G. A. Shaw und H. K. Burke. Spectral imaging for remote sensing. *Lincoln laboratory journal*, 14(1):3–28, 2003.
- [SB10] W. S. Stiles und J. M. Burch. Npl colour-matching investigation: final report (1958). *Journal of Modern Optics*, January 1959:1–26, November 2010. DOI: 10.1080/713826267.

- [SCo6] F. Salzenstein und C. Collet. Fuzzy markov random fields versus chains for multispectral image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1753–1767, Nov 2006. DOI: 10.1109/TPAMI.2006.228.
- [scho7a] *Translation of CIE 1931 Resolutions on Colorimetry*. John Wiley and Sons, Ltd, 2007. DOI: 10.1002/9780470175637.ch1. ISBN: 9780470175637.
- [Scho7b] R. Schowengerdt. *Remote sensing : models and methods for image processing*. Academic Press, Burlington, MA, 2007. ISBN: 9780123694072.
- [SEFR13] T. Scharwächter, M.ENZWEILER, U. Franke und S. Roth. Efficient multi-cue scene segmentation. In *German Conference on Pattern Recognition*, Seiten 435–445. Springer, 2013. DOI: 10.1007/978-3-642-40602-7_46.
- [SF13] T. Skauli und J. E. Farrell. A collection of hyperspectral images for imaging systems research. In *Digital Photography*, Seite 86600C, 2013. DOI: 10.1117/12.2007097.
- [Sha48] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948. DOI: 10.1063/1.3067010.
- [She95] C. Sheppard. Approximate calculation of the reflection coefficient from a stratified medium. *Pure and Applied Optics: Journal of the European Optical Society Part A*, 4(5):665, 1995. DOI: 10.1088/0963-9659/4/5/018.
- [SHS12] F. Samadzadegan, H. Hasani und T. Schenk. Simultaneous feature selection and SVM parameter determination in classification of hyperspectral imagery using ant colony optimization. *Canadian Journal of Remote Sensing*, 38(2):139–156, 2012. DOI: 10.5589/m12-022.
- [SLC11] N. Salamati, D. Larlus und G. Csurka. Combining visible and near-infrared cues for image categorisation. In *Proc. of the 22nd British Machine Vision Conference (BMVC 2011)*., number EPFL-CONF-169247, Seiten 1–11, 2011. DOI: 10.5244/C.25.49.
- [SLJ]⁺15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke und A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Seiten 1–9, 2015. DOI: 10.1109/CVPR.2015.7298594.

- [SRHK03] K. Segl, S. Roessner, U. Heiden und H. Kaufmann. Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58:99–112, 06 2003. DOI: 10.1016/S0924-2716(03)00020-0.
- [SS00] A. Stockman und L. T. Sharpe. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737, 2000. DOI: 10.1016/S0042-6989(00)00021-3.
- [SS02] B. Schölkopf und A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002. DOI: 10.7551/mitpress/4175.001.0001.
- [SWRC09] J. Shotton, J. Winn, C. Rother und A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 81(1):2–23, 2009. DOI: 10.1007/s11263-007-0109-1.
- [SZ14] K. Simonyan und A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [TBC09] Y. Tarabalka, J. A. Benediktsson und J. Chanussot. Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2973–2987, 2009. DOI: 10.1109/TGRS.2009.2016214.
- [TBEG17] D. R. Thompson, J. W. Boardman, M. L. Eastwood und R. O. Green. A large airborne survey of earth’s visible-infrared spectral dimensionality. *Optics express*, 25(8):9186–9195, 2017. DOI: 10.1364/OE.25.009186.
- [TCVMK10] D. Tuia, G. Camps-Valls, G. Matasci und M. Kanevski. Learning relevant image features with multiple-kernel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3780–3791, 2010. DOI: 10.1109/TGRS.2010.2049496.
- [TFCB10] Y. Tarabalka, M. Fauvel, J. Chanussot und J. A. Benediktsson. SVM-and mrf-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 7(4):736–740, 2010. DOI: 10.1109/LGRS.2010.2047711.

- [THKS88] C. Thorpe, M. H. Hebert, T. Kanade und S. A. Shafer. Vision and navigation for the carnegie-mellon navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(3):362–373, May 1988. DOI: 10.1109/34.3900.
- [Tho16] M. Thoma. A survey of semantic segmentation. *CoRR*, abs/1602.06541, 2016.
- [TJGT16] G. Tsagkatakis, M. Jayapala, B. Geelen und P. Tsakalides. Non-negative matrix completion for the enhancement of snapshot mosaic multispectral imagery. *Electronic Imaging*, 2016(12):1–6, 2016. DOI: 10.2352/ISSN.2470-1173.2016.12.IMSE-277.
- [TLoo] S. Tadjudin und D. A. Landgrebe. Robust parameter estimation for mixture model. *IEEE Transactions on Geoscience and Remote Sensing*, 38(1):439–445, 2000. DOI: 10.1109/36.823939.
- [TLSH12] N. Tack, A. Lambrechts, P. Soussan und L. Haspelagh. A compact, high-speed, and low-cost hyperspectral imager. In *Silicon Photonics VII*, Band 8266, Seite 8266oQ. International Society for Optics and Photonics, 2012. DOI: 10.1117/12.908172.
- [TPLZ15] C. Tao, H. Pan, Y. Li und Z. Zou. Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geoscience and Remote Sensing Letters*, 12(12):2438–2442, 2015. DOI: 10.1109/LGRS.2015.2482520.
- [TT16] G. Tsagkatakis und P. Tsakalides. A self-similar and sparse approach for spectral mosaic snapshot recovery. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, Seiten 341–345, Oktober 2016. DOI: 10.1109/IST.2016.7738248.
- [UBo4] C. Unsalan und K. L. Boyer. Classifying land development in high-resolution satellite imagery using hybrid structural-multispectral features. *IEEE Transactions on Geoscience and Remote Sensing*, 42(12):2840–2850, Dec 2004. DOI: 10.1109/TGRS.2004.835224.
- [Vago7] F. Vagni. Survey of hyperspectral and multispectral imaging technologies. *RTO Technical Report*, 2007.
- [vdPD93] H. van der Piepen und R. Doerffer. Rosis - ein abbildendes spektrometer für die biosphärenforschung. In R. Winter und W. Markwitz (Editoren), 9. *Nutzerseminar des Deutschen Fernerkundungsdatenzentrums der DLR, Oberpfaffenhofen, 14.-15.9.1992*, DLR-Mitt, Seiten 48–50, 1993. LIDO-Berichtsjahr=1993,

- monograph_id=93-08,.
- [VHF⁺97] P. L. Vora, M. L. Harville, J. E. Farrell, J. D. Tietz und D. H. Brainard. Image capture: synthesis of sensor responses from multispectral images. In G. B. Beretta und R. Eschbach (Editoren), *Color Imaging: Device-Independent Color, Color Hard Copy, and Graphic Arts II*, Band 3018, Seiten 2 – 11. International Society for Optics and Photonics, SPIE, 1997. DOI: 10.1117/12.271577.
- [VLL⁺10] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio und P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, Dezember 2010. ISSN: 1532-4435.
- [VSN⁺19] G. Varma, A. Subramanian, A. Namboodiri, M. Chandraker und C. Jawahar. Idd: A dataset for exploring problems of autonomous navigation in unconstrained environments. Seiten 1743–1751, 2019. DOI: 10.1109/WACV.2019.00190.
- [VVDB17] A. Valada, J. Vertens, A. Dhall und W. Burgard. Adapnet: Adaptive semantic segmentation in adverse environmental conditions. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, Seiten 4644–4651. Institute of Electrical and Electronics Engineers (IEEE), 2017. DOI: 10.1109/ICRA.2017.7989540.
- [WB07] B. Waske und J. A. Benediktsson. Fusion of support vector machines for classification of multisensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3858–3866, 2007. DOI: 10.1109/TGRS.2007.898446.
- [Wil88] M. Williams. Prometheus-the european research programme for optimising the road transport system in europe. In *IEE Colloquium on Driver Information*, Seiten 1/1–1/9, Dec 1988.
- [WS00] G. Wyszecki und W. S. Stiles. *Color science*. Wiley-Interscience, 2000. ISBN: 0471399183.
- [WvdLB⁺10] B. Waske, S. van der Linden, J. A. Benediktsson, A. Rabe und P. Hostert. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7):2880–2889, 2010. DOI: 10.1109/TGRS.2010.2041784.
- [WWR⁺13] C. Wojek, S. Walk, S. Roth, K. Schindler und B. Schiele. Monocular visual scene understanding: Understanding multi-object traffic scenes. *IEEE tran-*

- sactions on pattern analysis and machine intelligence, 35(4):882–897, 2013. DOI: 10.1109/TPAMI.2012.174.
- [WZWZ18] C. Wang, L. Zhang, W. Wei und Y. Zhang. When low rank representation based hyperspectral imagery classification meets segmented stacked denoising auto-encoder based spatial-spectral feature. *Remote Sensing*, 10(2):284, 2018. DOI: 10.3390/rs10020284.
- [XGD⁺17] S. Xie, R. Girshick, P. Dollar, Z. Tu und K. He. Aggregated residual transformations for deep neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [XR99] Xiuping Jia und J. A. Richards. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):538–542, Jan 1999. DOI: 10.1109/36.739109.
- [YI16] N. Yokoya und A. Iwasaki. Airborne hyperspectral data over chikusei. *Tech. Rep. SAL-2016-05-27*, 2016.
- [YK15] H. W. Yoon und R. N. Kacker. Guidelines for radiometric calibration of electro-optical instruments for remote sensing. Mai 2015. DOI: 10.6028/nist.hb.157.
- [YMIN10] F. Yasuma, T. Mitsunaga, D. Iso und S. K. Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010. DOI: 10.1109/TIP.2010.2046811.
- [YNI⁺15] K. Yoshida, I. Nishidate, T. Ishizuka, S. Kawauchi, S. Sato und M. Sato. Multispectral imaging of absorption and scattering properties of in vivo exposed rat brain using a digital red-green-blue camera. *Journal of Biomedical Optics*, 20(5):1 – 15, 2015. DOI: 10.1117/1.JBO.20.5.051026.
- [YWP⁺18] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu und N. Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In V. Ferrari, M. Herbert, C. Sminchisescu und Y. Weiss (Editoren), *Computer Vision – ECCV 2018*, Seiten 334–349, Cham, 2018. Springer International Publishing. ISBN: 978-3-030-01261-8.
- [YXC⁺18] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan und T. Darrell. BDD100K: A diverse driving video database with scalable annotation tooling. *The Computing Research Repository (CoRR)*, abs/1805.04687, 2018.

- [YYZ⁺18] M. Yang, K. Yu, C. Zhang, Z. Li und K. Yang. Denseaspp for semantic segmentation in street scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Juni 2018. DOI: 10.1109/CVPR.2018.00388.
- [ZE02] B. Zadrozny und C. Elkan. Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, Seiten 694–699. ACM, 2002. DOI: 10.1145/775047.775151.
- [ZHHN16] Y. Zhang, C. P. Huynh, N. Habili und K. N. Ngan. Material segmentation in hyperspectral images with minimal region perimeters. *IEEE International Conference on Image Processing*, September 2016. DOI: 10.1109/ICIP.2016.7532474.
- [ZHK⁺18] A. Zacharopoulos, K. Hatzigiannakis, P. Karamaoynas, V. M. Papadakis, M. Andrianakis, K. Melessanaki und X. Zabulis. A method for the registration of spectral images of paintings and its evaluation. *Journal of Cultural Heritage*, 29:10–18, 2018. DOI: 10.1016/j.culher.2017.07.004.
- [ZJRP⁺15] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang und P. H. Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, Seiten 1529–1537, 2015. DOI: 10.1109/ICCV.2015.179.
- [ZQS⁺17] H. Zhao, X. Qi, X. Shen, J. Shi und J. Jia. ICnet for real-time semantic segmentation on high-resolution images. *The Computing Research Repository (CoRR)*, abs/1704.08545, 2017. DOI: 10.1007/978-3-030-01219-9_25.
- [ZSQ⁺16] H. Zhao, J. Shi, X. Qi, X. Wang und J. Jia. Pyramid scene parsing network. *The Computing Research Repository (CoRR)*, abs/1612.01105, 2016. DOI: 10.1109/CVPR.2017.660.
- [ZWo8] P. Zhong und R. Wang. Learning sparse crfs for feature selection and classification of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12):4186–4197, 2008. DOI: 10.1109/TGRS.2008.2001921.
- [ZW10] P. Zhong und R. Wang. Learning conditional random fields for classification of hyperspectral images. *IEEE transactions on image processing*, 19(7):1890–1907, 2010. DOI: 10.1109/TIP.2010.2045034.

- [ZZP⁺16] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso und A. Torralba. Semantic understanding of scenes through the ade20k dataset. *arXiv preprint arXiv:1608.05442*, 2016. DOI: 10.1007/s11263-018-1140-0.
- [ZZP⁺17] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso und A. Torralba. Scene parsing through ade20K dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. DOI: 10.1109/CVPR.2017.544.

INTERNETQUELLEN

- [@1] Arduino. Arduino-Nano-Board, 2019. <https://store.arduino.cc/arduino-nano>. Zuletzt abgerufen am 06.11.2019.
- [@2] Statistisches Bundesamt und Kraftfahrt-Bundesamt. Bestand an Kraftfahrzeugen und Schienenfahrzeugen für die Jahre 2015 bis 2019, *DeStatis*, 2019. <https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Transport-Verkehr/Unternehmen-Infrastruktur-Fahrzeugbestand/Tabellen/fahrzeugbestand.html>. Zuletzt abgerufen am 20.03.2012.
- [@3] Golem Media GmbH. VW investiert 2,6 Milliarden Dollar in autonomes Fahren, 2019. <https://www.golem.de/news/argo-ai-vw-investiert-2-6-milliarden-dollar-in-autonomes-fahren-1907-142537.html>. Zuletzt abgerufen am 20.03.2012.
- [@4] Handelsblatt GmbH. Autonome Autos und Flugzeuge: Die Zukunft lässt auf sich warten, 2019. <https://www.handelsblatt.com/technik/digitale-revolution/digitale-revolution-autonome-autos-und-flugzeuge-die-zukunft-laesst-auf-sich-warten/24194826.html>. Zuletzt abgerufen am 20.03.2012.
- [@5] Handelsblatt GmbH. Das vollkommen autonome Fahren wird vorerst nicht kommen, 2019. <https://www.handelsblatt.com/politik/deutschland/hohe-kosten-das-vollkommen-autonome-fahren-wird-vorerst-nicht-kommen/24597246.html?ticket=ST-67088122-A1VQau3qFuafvMqfrBnf-ap1>. Zuletzt abgerufen am 20.03.2012.
- [@6] Handelsblatt GmbH. Milliarden fürs autonome Fahren – Bosch wagt sich ins KI-Duell mit Google, 2019. <https://www.handelsblatt.com/unternehmen/industrie/autozulieferer-milliarden-fuers-autonome-fahren-bosch-wagt-sich-ins-ki-duell-mit-google/23926612.html>. Zuletzt abgerufen am 30.01.2019.
- [@7] Darpa Mil. DARPA Grand Challenge 2004, 2004. <https://www.darpa.mil/about-us/timeline/-grand-challenge-for-autonomous-vehicles>. Zuletzt abgerufen am 20.03.2012.
- [@8] Darpa Mil. DARPA Urban Challenge 2007, 2007. <https://www.darpa.mil/about-us/timeline/-grand-challenge-for-autonomous-vehicles>. Zuletzt abgerufen am 20.03.2012.
- [@9] NASA. Landsat 1, 1972 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-1/>.
- [@10] NASA. Landsat 2, 1975 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-2/>.

- [@11] NASA. Landsat 3, 1978 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-3/>.
- [@12] NASA. Landsat 4, 1982 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-4/>.
- [@13] NASA. Landsat 5, 1984 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-5/>.
- [@14] NASA. Landsat 6, 1993 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-6/>.
- [@15] NASA. Landsat 7, 1999 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-7/>.
- [@16] NASA. Landsat 8, 2013 (accessed September 8, 2019). <https://landsat.gsfc.nasa.gov/landsat-8/>.
- [@17] Web of Science. Web of science, 2020 (accessed June 10, 2020). <https://wcs.webofknowledge.com>.
- [@18] Pixabay. Pixabay1, 2019. <https://pixabay.com/de/photos/pr%C3%A4rie-pfad-feld-landschaftlich-1246633/>. Zuletzt abgerufen am 04.11.2019.
- [@19] Pixabay. Pixabay2, 2019. <https://pixabay.com/de/photos/stra%C3%9Fe-weg-pfad-wald-holz-schlamm-791160/>. Zuletzt abgerufen am 04.11.2019.
- [@20] Pixabay. Pixabay3, 2019. <https://pixabay.com/de/photos/baum-natur-wald-waldweg-feldweg-3095703/>. Zuletzt abgerufen am 04.11.2019.
- [@21] Pixabay. Pixabay4, 2019. <https://pixabay.com/de/photos/fr%C3%BChling-rapsfeld-wolken-wetter-4503686/>. Zuletzt abgerufen am 04.11.2019.
- [@22] Google / Tensorflow. Tensorflow models, 2020 (accessed September 8, 2019). <https://github.com/tensorflow/models/tree/master/official/>.
- [@23] CSRU Texas. AVIRIS Data CSRU Texas, 2019. <http://www.csr.utexas.edu/projects/rs/hrs/process.html>. Zuletzt abgerufen am 04.11.2019.

CURRICULUM VITÆ

PERSÖNLICHE DATEN

Christian Winkens

AUSBILDUNG

Universität Koblenz-Landau Promotionsstudium - FB 4 Thema Dissertation: Klassifikation hyperspektraler Daten zur Befahrbarkeitsanalyse Abschluss: Dr. rer. nat.	2011 - 2021
Universität Koblenz-Landau Studium Computervisualistik Abschluss: Diplom-Informatiker	04/2006 - 04/2011
Gymnasium im Kannenbäckerland Allgemeine Hochschulreife	08/1996 - 03/2005

25. April 2021