Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 807

Joachim-Christoph Niemeyer

Verwendung von Kontext zur Klassifikation luftgestützter Laserdaten urbaner Gebiete

München 2017

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 | ISBN 978-3-7696-5219-2

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 336, Hannover 2017

Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 807

Verwendung von Kontext zur Klassifikation luftgestützter Laserdaten urbaner Gebiete

Von der Fakultät für Bauingenieurwesen und Geodäsie

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation

Vorgelegt von

Dipl.-Ing. Joachim-Christoph Niemeyer

Geboren am 26.09.1984 in Heidelberg

München 2017

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 ISBN 978-3-7696-5219-2

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 336, Hannover 2017

Adresse der DGK:

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 331 – 288 1685 ● Telefax +49 – 331 – 288 1759

E-Mail post@dgk.badw.de ● http://www.dgk.badw.de

Prüfungskommission:

Vorsitzende: Prof. Dr.-Ing. Monika Sester

Referent: Prof. Dr.-Ing. Christian Heipke

Korreferenten: Prof. Dr.-Ing. Uwe Sörgel (Universität Stuttgart)
apl. Prof. Dr.-Ing. Claus Brenner
apl. Prof. Dr. techn. Franz Rottensteiner

Tag der mündlichen Prüfung: 14.07.2017

© 2017 Bayerische Akademie der Wissenschaften, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325 ISBN 978-3-7696-5219-2

Kurzfassung

Auf Flugzeugen montierte Laserscanner können innerhalb kürzester Zeit große Gebiete der Erdoberfläche dreidimensional erfassen. Eine wichtige Aufgabe besteht in der Klassifikation der anfallenden Daten, welche als Punktwolken bezeichnet werden. Jeder Laserpunkt wird dabei einer Objektklasse zugeordnet. Die semantisch angereicherten Daten sind für weiterführende Applikationen erforderlich, wie etwa für die Erstellung dreidimensionaler Stadtmodelle.

Diese Arbeit befasst sich mit der automatischen Zuordnung dieser Punktwolken zu Objektklassen unter Ausnutzung von räumlicher Kontextinformation. Die Zielsetzung besteht in der Entwicklung eines neuen Klassifikationsansatzes, welcher Kontext anhand von zwei Ebenen in den Prozess integriert. Probabilistische graphische Modelle in Form von bedingten Zufallsfeldern (Conditional Random Fields, CRF) stellen dafür ein geeignetes Werkzeug bereit. Es wird ein hierarchisches Modell erstellt, das einerseits die Laserpunkte und andererseits Segmente bestehend aus Punktgruppen klassifiziert. Jede dieser Teilaufgaben wird mit einem individuellen CRF gelöst. In einem iterativen Prozess werden die beiden Ebenen nacheinander optimiert und reichen ihre Ergebnisse an die jeweils andere Ebene weiter. Für das segmentbasierte CRF ist vor jeder Durchführung eine Aggregation ähnlicher Punkte zu Segmenten erforderlich. Es wird eine neue Methodik vorgeschlagen, die im Zuge der Iterationen eine adaptive, zunehmend besser werdende Anpassung der Segmente an die tatsächlichen Objektgrenzen erlaubt, wodurch sich die Anzahl der Segmentierungsfehler reduziert. Dies verbessert die Klassifikationsqualität von feineren Strukturen wie etwa Autos in der Szene. Weiterhin lassen sich die durch das alternierende Vorgehen entstehenden Zwischenlösungen für die Extraktion von Kontextmerkmalen verwenden. Zwei Gruppen kontextbasierter Segmentmerkmale für luftgestützte Laserdaten werden im Rahmen dieser Arbeit vorgestellt.

Der Fokus der experimentellen Untersuchungen liegt darauf, das vorgestellte Verfahren zu evaluieren und insbesondere den Beitrag der Kontextinformation in Bezug auf die erzielten Genauigkeiten zu ermitteln. Dafür werden reale Datensätze mit bekannten Referenzklassen für jeden individuellen Laserpunkt verwendet. Eine Auswertung ergibt gute Genauigkeitswerte. Anhand von öffentlich verfügbaren Standarddatensätzen zeigt sich eine Überlegenheit hinsichtlich der Gesamtgenauigkeiten gegenüber anderen aktuellen Verfahren. In den Experimenten führt die Berücksichtigung von Kontextinformation im Vergleich zu einem Random Forest Klassifikator zu Genauigkeitssteigerungen um bis zu 5 Prozentpunkte. Besonders in den Daten unterrepräsentierte Objektklassen wie Autos profitieren von Kontext. Ihre Qualität kann um mehr als 20 Prozentpunkte verbessert werden.

Schlagworte: 3D Punktwolke, Klassifikation, Conditional Random Fields, Segmentierung, Zwei-Ebenenmodell

Abstract

Airborne laser scanners are able to acquire large scale, three dimensional topographic information rapidly. An important task is the classification of the gathered data, which are referred to as point clouds. For that purpose, each point is assigned to one object class. The result is used for several applications such as the generation of three dimensional city models.

This thesis deals with the automatic classification of point clouds by exploiting spatial contextual information. The goal is to develop a new approach which enables the integration of context into two layers corresponding to different scales. Probabilistic graphical models such as Conditional Random Fields (CRF) provide an appropriate tool for that purpose. A two-layer hierarchical framework is set up in order to classify points as well as segments. For each layer an independent CRF is applied. Both layers are optimized in an iterative and alternating manner. After each optimization the results are propagated to the other layer in several ways. The segment-based classification requires a clustering step, which aggregates similar points. The new methodology allows for the iterative adaption of the segments to the actual object boundaries in order to reduce the number of segmentation errors. The aim of this procedure is to classify small details such as cars more reliably. Moreover, two groups of contextual segment features are introduced for airborne laser scanner data.

Experiments are performed on real world data sets with reference labels in order to evaluate the proposed method. One aim is to assess the impact of exploiting contextual information on the classification accuracy. The analysis reveals good results. As shown on several benchmark data sets, the proposed framework is able to outperform state-of-the-art approaches in terms of the overall accuracies. Moreover, it becomes apparent that the consideration of context increases the overall accuracies by up to 5 percentages compared to a standard Random Forest classification. Most importantly, underrepresented classes such as cars benefit significantly from incorporating context. The class specific quality values are improved by more than 20 percentages.

Keywords: 3D point cloud, classification, conditional random fields, segmentation, two-layer model

Nomenklatur

Abkürzungen

2D zweidimensional

3D dreidimensional

ALS airborne laser scanning (luftgestütztes Laserscanning)

AMN Associative Markov Networks

CRF Conditional Random Field (bedingtes Zufallsfeld)

DSM digital surface model (digitales Oberflächenmodell)

DTM digital terrain model (digitales Geländemodell)

DTF Decision Tree Field

HMM Hidden Markov Model

HOP higher order potential (Potential für Cliquen höherer Ordnung)

ISPRS International Society for Photogrammetry and Remote Sensing

LIDAR light detection and ranging (Laserscanning)

LBP Loopy Belief Propagation

MAP Maximum A-Posteriori

MRF Markov Random Field (Markov Zufallsfeld)

MST Minimum Spanning Tree (Minimaler Spannbaum)

OSM Open Street Map

RF Random Forest

RGB Rot-Grün-Blau

RGB-D-Sensor Tiefenkamera mit RGB-Kanälen sowie Tiefeninformation

RTF Regression Tree Field

SVM Support Vector Machine

Liste der Symbole

Allgemein

δ_K	Kronecker-Delta
$\mathbb{R}^n$	n-dimensionaler euklidischer Raum
E	Energiefunktion
w, ω	ein Gewicht
$\mathcal{V}$	Menge der Knoten eines Graphen
$\mathcal{E}$	Menge der Kanten eines Graphen
$\mathcal{G}(\mathcal{V}, \mathcal{E})$	Graph bestehend aus Knoten $\mathcal{V}$ und Kanten $\mathcal{E}$

Methodik

CRF^P	Punktweise Klassifikationsebene
CRF^S	Segmentweise Klassifikationsebene
N_{iter}	Anzahl der Iterationen des Verfahrens

Klassifikation

$\mathcal{L}$	Menge der Klassenlabel
l	eine spezifische Klasse $\in \mathcal{L}$
y	eine Zufallsvariable
$\mathbf{y}$	Vektor aller Zufallsvariablen
$\mathbf{f}$	Merkmalsvektor
$\mathcal{N}$	Nachbarschaftsdefinition eines Primitivs $\in$ [Punkt, Segment]
$\mathbf{x}$	Gesamtheit der Beobachtungen
p	Wahrscheinlichkeit

Random Forest

T	Anzahl der Bäume
N_s	Anzahl der Trainingsbeispiele (gesamt)
$N_{s,l}$	Anzahl der Trainingsbeispiele pro Klasse l
M_{split}	Anzahl der zufälligen Merkmale für einen Aufteilungstest
T_{depth}	Tiefe der Bäume
T_{split}	Mindestanzahl Trainingsbeispiele pro Knoten für Teilung
RF^P	Punktweiser RF Klassifikator (unäres Potential)
RF_u^S	Segmentweiser RF Klassifikator (unäres Potential)
RF_p^S	Segmentweiser RF Klassifikator (paarweises Potential)

Conditional Random Fields

φ^P	Unäres Potential (CRF^P)
ψ^P	Paarweises Interaktionspotential (CRF^P)
χ^P	Potential höherer Ordnung (CRF^P)
ξ^P	Unäres Potential der Segment-Konfidenzen (CRF^P)
φ^S	Unäres Potential (CRF^S)
ψ^S	Paarweises Interaktionspotential (CRF^S)
Z	Partitionsfunktion
$\mathcal{H}$	Menge aller Cliquen höherer Ordnung
h	Index einer Clique höherer Ordnung
θ	Glättungsparameter, kontrast-sensitives Potts Modell
q	Sättigungsparameter, robustes P^n Potts Modell (HOP)
$N^{P^n}_{nVCCS}$	Verwendete Anzahl von Segmentierungsergebnissen (HOP)
d^S_{graph}	maximale Distanz für Nachbarschaft zweier Segmente

Segmentierung

R^{SV}_{vox}	Voxelgröße des Octrees
R^{SV}_{saat}	maximale Größe der Supervoxel
ω^{SV}_s	Gewicht der räumlichen Komponente
ω^{SV}_n	Gewicht der Normalenvektoren
ω^{SV}_{rgb}	Gewicht der Farbkomponente
ω^{SV}_{konf}	Gewicht der Konfidenzkomponente

Merkmale

res_{grid}	Rasterauflösung Binärkarte Straße
$S_{closing}$	Größe des Strukturelements Straße
H^{res}_{dist}	Distanzauflösung im Hough-Raum
H^{res}_{ρ}	Winkelauflösung im Hough-Raum
$H_{minVotes}$	erforderliche Anzahl Hough-Stimmen
$H_{lineGap}$	maximal zulässige Datenlücke
$H_{minLength}$	Mindestlänge Geradensegment

Inhaltsverzeichnis

1 Einleitung **1**
1.1 Motivation . . . 1
1.2 Zielsetzung und wissenschaftlicher Beitrag . . . 3
1.3 Aufbau der Arbeit . . . 4

2 Stand der Forschung **5**
2.1 Kontextbasierte Klassifikation . . . 5
2.1.1 Kontext . . . 5
2.1.2 Klassifikation luftgestützter Laserdaten ohne Kontext . . . 6
2.1.3 Conditional Random Fields . . . 7
2.1.4 Alternative Möglichkeiten zur Kontextmodellierung . . . 13
2.2 Segmentierung von Punktwolken . . . 15
2.3 Diskussion . . . 18

3 Grundlagen **21**
3.1 Klassifikation luftgestützter Laserdaten . . . 21
3.1.1 Definition der Nachbarschaft . . . 21
3.1.2 Extraktion der Merkmale . . . 23
3.2 Random Forest Klassifikator . . . 25
3.3 Conditional Random Fields . . . 28
3.4 Segmentierung von Punktwolken . . . 35

4 Hierarchische kontextbasierte Punktwolkenklassifikation **39**
4.1 Konzept des hierarchischen Ansatzes . . . 39
4.2 Punktbasierte Klassifikation . . . 41
4.2.1 Potentiale . . . 42
4.2.2 Inferenz . . . 46
4.2.3 Training . . . 47
4.3 Segmentierung der Punktwolke . . . 48
4.4 Segmentbasierte Klassifikation . . . 50
4.4.1 Potentiale . . . 50
4.4.2 Neue Kontextmerkmale für Segmente . . . 52
4.4.3 Inferenz . . . 54
4.4.4 Training . . . 55
4.5 Diskussion . . . 55

5 Experimente **59**
5.1 Ziel und Bewertungskriterien . . . 59
5.2 Datensätze . . . 61
5.2.1 Hannover . . . 61
5.2.2 Vaihingen . . . 63
5.2.3 Benchmarks . . . 64
5.3 Voruntersuchung: Merkmalsauswahl und Parameterbestimmung . . . 68
5.3.1 Merkmale . . . 68
5.3.2 Parameter . . . 74
5.4 Klassifikationsergebnisse . . . 78
5.4.1 Nutzen von Kontext für die 3D-Punktwolkenklassifikation . . . 80
5.4.2 Vergleich mit anderen Verfahren . . . 88
5.5 Evaluation ausgewählter Aspekte . . . 91
5.5.1 Analyse der Modellkomponenten . . . 91
5.5.2 Konvergenzuntersuchung . . . 97
5.5.3 Trainingsdaten . . . 98
5.5.4 Straßenmerkmale . . . 100
5.6 Diskussion . . . 102

6 Schlussfolgerungen und Ausblick **105**

1 Einleitung

1.1 Motivation

Mit der zunehmenden technologischen Entwicklung steigt weltweit das Interesse an Geodaten, insbesondere in städtischen Gebieten. Im Zeitalter von virtueller Realität, (teil-)autonom agierenden Fahrzeugen sowie der digitalen Gefahrenanalyse z. B. hinsichtlich Naturkatastrophen wie Überschwemmungen sind detaillierte dreidimensionale (3D) Stadtmodelle als Grundlage vieler Anwendungen erforderlich. Auf der anderen Seite führen verbesserte Sensoren u. a. im Bereich der Fernerkundung dazu, dass innerhalb kürzester Zeit großflächige Gebiete erfasst werden können. Die dabei entstehenden Daten müssen jedoch analysiert und ausgewertet werden. Da dies mit einem enormen Arbeitsaufwand und somit hohen Kosten einher geht, ist eine automatisierte Auswertung wünschenswert. Verfahren aus dem Bereich des maschinellen Lernens eignen sich dafür, Auswertungsstrategien, die an bekannten Daten angelernt wurden, auf unbekannte Datenbestände anzuwenden. Auf diese Weise lässt sich der manuelle und kostenintensive Einsatz durch einen Operateur stark reduzieren.

Auf Flugzeugen montierte Laserscanner lassen sich besonders gut zum Erfassen topographischer Informationen verwenden. Durch die Distanzmessungen wird die Erdoberfläche mitsamt den sich darauf befindlichen Objekten wie Gebäuden, Vegetation, Autos usw. unmittelbar in drei Dimensionen abgetastet. Dabei wird ein Objekt jedoch nicht mit einer einzigen Messung erfasst, sondern vielmehr durch eine großen Anzahl von versetzten Einzelreflexionen, repräsentiert als 3D-Punkte, beschrieben. Auf diese Weise gewonnene Datensätze werden als Punktwolke bezeichnet. Die Datenstruktur unterscheidet sich deutlich von Bildern, da keine regelmäßige Anordnung der Punkte vorliegt. Im Gegensatz zu den 2D-Pixeln eines gerasterten Bildes sind die Punkte einer Laserpunktwolke üblicherweise irregulär im 3D-Raum verteilt. Weitere Schwierigkeiten bei der Auswertung treten durch Messrauschen, stark variierende Punktdichten oder Datenlücken (z. B. in verdeckten Bereichen oder bei nicht reflektierenden Materialien) auf. Eines haben Punktwolken jedoch mit Bildern gemeinsam: Eine explizite Beschreibung der Objektklassen liegt nicht vor. Bei der Betrachtung der Rohdaten lassen sich Objekte selbst für den geübten Nutzer nur anhand spezieller Eigenschaften wie der relativen Anordnung oder der Intensitäten des reflektierten Signals erkennen.

Ein erster Schritt zur Erstellung von 3D-Stadtmodellen besteht daher häufig in der Klassifikation der gewonnenen Daten [Axelsson, 1999; Haala & Brenner, 1999; Weinmann, 2016]. Dieser Prozess umfasst die Zuordnung von semantischem Wissen zu den Objekten innerhalb einer Szene. Die einzelnen Primitive, also beispielsweise die Pixel eines Bildes oder die 3D-Punkte im Fall einer Punktwolke, werden dabei anhand bestimmter Charakteristiken - sogenannten *Merkmalen* - kategorisiert. Typische Klassen im städtischen Gebiet sind u. a. *Gebäude*, *Vegetation*, *Straße*. Aufbauend auf diesen Ergebnissen lassen

sich die semantisch angereicherten Daten für die Weiterverarbeitung nutzen.

Insbesondere in städtischen Gebieten erschweren die Gegebenheiten das korrekte Erkennen von Objekten. Durch die dichte Bebauung treten vermehrt Verdeckungen auf. Um eine vollständige Erfassung zu gewährleisten, wird bei der Befliegung die Anzahl der Flugstreifen erhöht, was wiederum speziell in den Bereichen der doppelten Streifenüberdeckung größere Variationen in den Punktdichten hervorruft. Dies kann sich negativ auf die Merkmalsextraktion auswirken, da lokal abweichende Punktverteilungen beobachtet werden. Weiterhin treten in urbanen Gegenden viele unterschiedlich geformte Objekte verschiedenster Materialien auf einer vergleichsweise geringen Fläche auf, welche sich gegenseitig in Bezug auf die Merkmalsausprägungen der Punkte beeinflussen können. Wenn ein Großteil der emittierten Energie des Lasers bereits an einer Dachkante reflektiert wird, kann beispielsweise die Intensität eines Punktes unterhalb des Daches im Vergleich zu seinen Nachbarpunkten einen deutlich abweichenden Wert aufweisen.

Andererseits handelt es sich bei einer Stadt um vom Menschen entworfene Strukturen. Abgesehen von Vegetation sind die meisten Objekte recht regelmäßig geformt und auch die Anordnung der Elemente in der Szene folgt meist speziellen Regeln. So befindet sich ein Auto typischerweise auf einer Straße und ist nicht - ggf. abgesehen von einem Parkhaus - auf Gebäudedächern platziert. Gebäude wiederum stehen in der Nähe einer Straße oder eines Wegs, über die der Zugang gewährleistet wird. Derartige Eigenschaften und Beziehungen zwischen den einzelnen Objektklassen können als *Kontextwissen* aufgefasst werden. Gelingt es, dieses Wissen in den Prozess der Klassifikation einzubinden, kann sich dies positiv auf das Klassifikationsergebnis auswirken. Aktuelle Verfahren des computergestützten Sehens (engl.: computer vision) machen sich Kontextinformation zu Nutze, um auch mit komplizierten Datensätzen gute Ergebnisse bei der Klassifikation der Bildprimitive (Pixel, Segmente, usw.) zu erreichen. Moderne Verfahren, wie etwa probabilistische graphische Modelle, stellen dafür ein geeignetes und flexibles Werkzeug zu Verfügung.

Für die semantische Analyse von (luftgestützten) 3D-Laserpunktwolken ist die Anzahl an entwickelten Methoden in der Literatur im Vergleich zu Verfahren aus dem Bereich der Bildauswertung signifikant geringer. Die unregelmäßige Verteilung der Punkte, die Größe der Datensätze sowie der um eine Dimension erweiterte Raum erschweren die Anwendung. Die vorliegende Arbeit stellt sich dieser Herausforderung und präsentiert ein neues Verfahren, das graphische Modelle zur kontextbasierten Klassifikation von Laserdaten verwendet.

Probabilistische Ansätze sind aus Gründen der Rechenkomplexität insbesondere bei größeren Datensätzen auf die Berücksichtigung eines räumlich begrenzten Kontextbereichs beschränkt. Da bei der ausschließlichen Betrachtung lokaler Bereiche nur ein sehr begrenzter Teil des Kontextwissens berücksichtigt werden kann, ist die Einbeziehung eines größeren Umgebungsbereichs wünschenswert. Beispielsweise werden Punkte im Bereich eines Dachfirsts aufgrund der lokalen Geometrie häufig mit Baumkronen verwechselt. Zieht man eine größere Umgebung in die Untersuchung mit ein, wird die korrekte Zuordnung erleichtert. Um die Einschränkungen eines begrenzten Kontextbereiches und einer hohen Rechenkomplexität zu umgehen, kann es daher hilfreich sein, auf größere Einheiten (*Entitäten*) überzugehen und statt der 3D-Punkte Segmente zu klassifizieren. Allerdings kann es vorkommen, dass Punkte unterschiedlicher Objekte zu einem Segment gruppiert werden und somit Generalisierungsfeh-

ler in diesem Prozess in Kauf genommen werden müssen. Wenn alle Punkte eines Segments derselben Objektklasse zugeordnet werden, können kleinere Details in der Szene auf diese Weise falsch klassifiziert werden und somit verloren gehen. Sinnvoll ist es daher, die punktweise und die segmentweise Klassifikation in einem Mehrebenenmodell zu kombinieren. Auf diesem Wege gelingt es zum einen, die Vorteile beider Repräsentationsformen zu vereinen und Kontext sowohl in einem lokalen als auch in einem gröberen Skalenbereich zu integrieren. Zum anderen kann die Anzahl der Generalisierungsfehler reduziert werden, indem beide Ebenen gegenseitig Informationen über den aktuell wahrscheinlichsten Klassenzustand jedes Punktes austauschen, welche Einfluss auf den Segmentierungsprozess nehmen können.

1.2 Zielsetzung und wissenschaftlicher Beitrag

Diese Arbeit nimmt sich der Herausforderungen bei der Klassifikation luftgestützter Laserdaten in urbanen Gebieten an. Hierfür wird ein neues kontextbasiertes Verfahren entwickelt, welches in einem Mehrebenenmodell die Berücksichtigung von Kontext zweier Skalenebenen kombiniert und gleichzeitig die Segmentierung verbessert. Eine Evaluation der Methodik anhand mehrerer Datensätze untersucht den Nutzen von räumlicher Kontextinformation im Hinblick auf die zu erzielenden Klassifikationsgenauigkeiten. Der wesentliche wissenschaftliche Beitrag lässt sich wie folgt zusammenfassen:

- Die **Entwicklung eines neuen hierarchischen Klassifikationsansatzes** basierend auf graphischen Modellen ermöglicht die Integration von Kontextinformation auf zwei Skalenebenen. Im Rahmen des vollständig probabilistischen Verfahrens lassen sich typische Beziehungen zwischen Objektklassen mittels Trainingsdaten anlernen und in den Klassifikationsprozess integrieren. Umfangreiche Experimente ermitteln die Stärken und Schwächen des Verfahrens.
- **Entwicklung eines verbesserten Segmentierungsverfahrens:** Durch die Berücksichtigung von Vorklassifikationsergebnissen in den Segmentierungsprozess kann der Umfang an Generalisierungsfehlern reduziert und somit die Klassifikationsgenauigkeit erhöht werden.
- **Entwurf neuer Kontextmerkmale** für die Klassifikation flugzeuggestützter Laserdaten. Auf diese Weise lassen sich zusätzlich zum graphischen Modell räumliche Abhängigkeiten auszunutzen. Dies wird am Beispiel von Straßen aufgezeigt, da diese in städtischen Gebieten von strukturierender Bedeutung sind. Es werden Segmentmerkmale zur Beschreibung der Lage und Orientierung eines Objektes zur nächstgelegenen Straße extrahiert.

Auf den Ergebnissen kann aufgebaut werden, um beispielsweise Objekte zu rekonstruieren und für weiterführende Anwendungen zu nutzen. Dies liegt außerhalb der Zielsetzung der vorliegenden Arbeit, so dass darauf nicht eingegangen wird. Im Fokus steht ausschließlich die Klassifikation einschließlich der Extraktion relevanter Merkmale.

1.3 Aufbau der Arbeit

Diese Arbeit gliedert sich wie folgt: In Kapitel 2 wird der aktuelle Stand der Forschung zur kontextbasierten Klassifikation vorgestellt. Die theoretischen Grundlagen zur semantischen Zuordnung von Punktwolken sowie der verwendeten Klassifikationsverfahren werden in Kapitel 3 zusammengefasst. Kapitel 4 widmet sich der Beschreibung und Diskussion der neu entwickelten Methodik, welche anhand von Experimenten in Kapitel 5 evaluiert wird. Die Arbeit endet mit den Schlussfolgerungen und einem Ausblick auf offen gebliebene Fragestellungen in Kapitel 6.

2 Stand der Forschung

Im Folgenden wird ein Überblick über verwandte Arbeiten aus der Literatur gegeben. Dabei liegt der Schwerpunkt auf der Prozessierung von Punktwolken. Das Kapitel ist dreigeteilt und befasst sich mit der kontextbasierten Klassifikation (Abschnitt 2.1) und der Segmentierung von Punktwolken (Abschnitt 2.2). Abschließend werden die Erkenntnisse in Abschnitt 2.3 diskutiert.

2.1 Kontextbasierte Klassifikation

2.1.1 Kontext

In dieser Arbeit wird ein neues Verfahren zur *kontextbasierten* Klassifikation vorgestellt und der Nutzen von *Kontext* für das Klassifikationsergebnis von Laserdaten ermittelt. Dazu ist es notwendig, den Begriff *Kontext* als solchen zunächst zu klären. Eine gute Definition wird von Wolf & Bileschi [2006] aufgestellt: *„Kontext ist jede Art von nutzbarer Information für eine Klassifikation, welche sich nicht unmittelbar aus dem Erscheinungsbild eines zu klassifizierenden Primitivs ableiten lässt.“* Die physikalische Erscheinung eines Primitivs wird dabei üblicherweise anhand von Merkmalen beschrieben, welche bei Bildern z. B. die Intensitätswerte der einzelnen Kanäle oder im Fall von Laserdaten die Intensitäten, Normalenvektoren usw. der 3D-Punkte sein können. Diese Definition wird der vorliegenden Arbeit zugrunde gelegt.

Eine von Oliva & Torralba [2007] durchgeführte Untersuchung der Literatur in Hinblick auf die Rolle von Kontext für die Erkennung von Objekten in Bilddaten ergibt, dass abgeleitete Statistiken über die Zusammensetzung von Objekten einer Szene eine sehr hilfreiche Informationsquelle zur Verbesserung der Detektionsraten darstellen. Galleguillos & Belongie [2010] zeigen verschiedene Einteilungen von Kontext aus den Bereichen der Psychologie und Bildanalyse. Demnach gibt es drei wesentliche Arten von Kontext:

1. *Semantischer Kontext* gibt die Wahrscheinlichkeit für das Vorhandensein eines Objekts in einer Szene einerseits und das gemeinsame Auftreten mit anderen beobachteten Objekten andererseits an. Beispielsweise befindet sich mit hoher Wahrscheinlichkeit eine Straße in einer Fernerkundungsszene, wenn Autos detektiert wurden. Kühe hingegen geben keinen Hinweis auf die Existenz einer Straße, da in diesem Fall die gemeinsame Auftrittswahrscheinlichkeit gering ist.
2. *Räumlicher Kontext* beschreibt, an welcher Position in einer Szene das Auftreten eines Objektes zu erwarten ist. Autos befinden sich beispielsweise mit großer Wahrscheinlichkeit auf einer Straße, jedoch nicht auf einem Baum. Bei terrestrischen Bildern, wie sie typischerweise im Bereich der computergestützten Bildanalyse untersucht werden, lässt sich die Position von Objekten in einem

Bild dank der ähnlichen Aufnahmeausrichtung leichter vorhersagen als bei Fernerkundungsdaten. Während bei erstgenannten Bildern im oberen Bereich der Himmel zu erwarten ist, lässt sich eine solche Aussage z. B. über die Position von Gebäuden bei der Aufnahme aus der Luft wegen einer fehlenden Referenzrichtung nur schwer ableiten.

3. *Maßstabskontext* umfasst die a priori Information über die Größe von Objekten. Durch das Reduzieren des Suchraums auf die wahrscheinlichen Skalenbereiche verringert sich die Rechenzeit und der Anteil an Fehldetektionen.

Am wichtigsten für Anwendungen in der Fernerkundung ist der räumliche Kontext. Implizit enthält dieser jedoch auch semantischen Kontext, da die Modellierung des gemeinsamen Auftretens von Objekten das Auffinden der Objekte in der Szene bedingt [Galleguillos & Belongie, 2010]. Werden die gegenseitigen Abhängigkeiten von benachbarten Primitiven oder Objekten einer Szene berücksichtigt, wird auch von *lokalem Kontext* gesprochen. Im Gegensatz dazu steht der *globale Kontext*, der die Art der Szene als Ganzes kategorisiert [Hinz & Baumgartner, 2003; Galleguillos & Belongie, 2010]. In dieser Arbeit wird die lokalen Ebene betrachtet.

Liegt das Ziel auf der Detektion eines Objekttyps mittels einer modellbasierten Strategie, kann das Objektmodell durch zusätzliche Kontextobjekte erweitert werden, die Hinweise für das gewünschte Objekt geben. So verwenden Hinz & Baumgartner [2003] etwa Straßenmarkierungen als Indikatoren für das Vorhandensein einer Straße. Bezieht man sich wie in dieser Arbeit auf statistische Ansätze für eine datengetriebene Herangehensweise, so lässt sich Kontext vor allem über zwei Optionen bei einer Klassifikation berücksichtigen [Galleguillos & Belongie, 2010; Rottensteiner, 2017]. Auf recht einfache Weise kann der Merkmalsvektor eines jeden Primitivs um sogenannte *Kontextmerkmale* erweitert werden, indem beispielsweise die an den Nachbarprimitiven beobachteten Daten mit hinzugezogen werden [Wolf & Bileschi, 2006]. Die Zuordnung zu einer Objektklasse kann anschließend über einen herkömmlichen Klassifikator (z. B. Random Forests (RF) [Breiman, 2001]) durchgeführt werden, der jedes Primitiv individuell betrachtet. Nachteilig ist jedoch, dass diese Methode die Dimensionen des Merkmalsraums gegebenenfalls deutlich erhöht. Eine andere Möglichkeit bieten graphische Modelle [Bishop, 2006; Förstner, 2013]. Diese probabilistischen Verfahren erlauben die Berücksichtigung der statistischen Abhängigkeiten zwischen den Nachbarn und ermitteln die beste Konfiguration der Klassenlabels für alle Primitive simultan. Typische Beispiele auf Basis ungerichteter Graphen sind *Markov-Zufallsfelder* (Markov Random Fields, MRF) und *bedingte Zufallsfelder* (Conditional Random Fields, CRF), die im weiteren Verlauf dieses Kapitels beschrieben werden.

2.1.2 Klassifikation luftgestützter Laserdaten ohne Kontext

In den vergangenen Jahren hat sich als Forschungsschwerpunkt in der Fernerkundung die Verwendung von überwachten, statistischen Methoden zur Klassifikation heraus gebildet, da diese verglichen mit modellbasierten Ansätzen flexibler mit Variationen in den Objekterscheinungsformen umgehen können. Dabei werden neben generativen Klassifikatoren, welche die Verbundwahrscheinlichkeit der Daten und der Klassenlabels modellieren [Bishop, 2006], für Punktwolken vor allem moderne diskriminative Methoden verwendet [Lodha et al., 2007; Chehata et al., 2009; Mallet, 2010; Guo et al., 2014; Hackel et al., 2016; Weinmann, 2016]. Diese Verfahren haben den Vorteil, dass sie im Vergleich zu

den generativen Modellen üblicherweise weniger Trainingsdaten benötigen, da nicht die gemeinsame Wahrscheinlichkeitsdichte der Daten und der Klassen modelliert werden muss [Rottensteiner, 2017]. Mallet [2010] nutzt zur punktweisen Klassifikation luftgestützter Laserdaten in mehrere Objektklassen z. B. Support Vector Machines (SVM) [Vapnik, 1998], während Chehata et al. [2009] RF für diese Aufgabe einsetzen. Der Nachteil beider vorgestellter Verfahren ist jedoch, dass jeder Punkt unabhängig von den Klassenlabels seiner Nachbarschaft klassifiziert wird. Vorhandene Korrelationen zwischen den Primitiven lassen sich auf diese Weise nicht ausnutzen, obwohl diese statistischen Informationen zu einer Genauigkeitssteigerung beitragen können. Verrauschte Ergebnisse sind typischerweise die Folge [Weinmann, 2016].

Umfassende Informationen zu einer typischen Prozessierungsreihenfolge von Laserdaten im Zuge einer Klassifikation sind in [Weinmann, 2016] zu finden. Weiterhin geben Yan et al. [2015] einen Überblick über aktuell gebräuchliche Verfahren zur Klassifikation luftgestützter Laserdaten.

Schlussfolgerung: Klassifikatoren, die jedes Primitiv individuell betrachten und einer Klasse zuordnen, sind zwar schnell einsetzbar und in vielen Bibliotheken verfügbar, nutzen jedoch nicht die Abhängigkeiten zwischen den Klassenlabels benachbarter Punkte aus. Eine Verbesserung lässt sich durch die Berücksichtigung von Kontextinformation erzielen, anhand derer Rückschlüsse auf die Objekte (und somit Hinweise für deren korrekte Klassifikation) in komplexen Szenen gezogen werden können.

2.1.3 Conditional Random Fields

Berücksichtigung von lokalem Kontext

Eine geeignete statistische Betrachtungsform von Kontext führt zu ungerichteten, graphischen Modellen [Bishop, 2006], bestehend aus Knoten und Kanten. Wichtige Verfahren dieser Gruppe sind MRF [Geman & Geman, 1984] oder CRF [Lafferty et al., 2001; Kumar & Hebert, 2006]. In einem MRF ist das Klassenlabel einer Variablen statistisch abhängig von denen seiner Nachbarn, während die Daten unterschiedlicher Objekte als unabhängig angenommen werden [Li, 2009]. Bedingte Zufallsfelder (CRF) ermöglichen eine allgemeinere Betrachtungsweise durch die Aufhebung der Unabhängigkeitsannahme der Daten. Bei CRF handelt es sich nicht um einen spezifischen Klassifikator, sondern vielmehr um einen flexibel an die jeweilige Aufgabe anpassbaren Rahmen [Kumar & Hebert, 2006]. Die wesentlichen Stellgrößen stellen dabei die Merkmalsauswahl sowie die Definitionen der Potentiale und der Graphstruktur dar.

Die meisten Anwendungen beschränken sich auf paarweise CRF, bei denen maximal zwei Knoten über eine Kante verbunden werden. Ein solches CRF besteht üblicherweise aus zwei Arten von Termen. Der Datenterm beschreibt das *unäre Potential* und modelliert die Klassenzugehörigkeit für jedes Primitiv auf Grundlage seines Merkmalsvektors unabhängig von den Nachbarn. Zur Bestimmung dieses Potentials werden meistens die probabilistischen Ergebnisse eines diskriminativen Klassifikators verwendet, wie z. B. von linearen Modellen [Anguelov et al., 2005; Kumar & Hebert, 2006] oder RF [Shapovalov et al., 2010; Schindler, 2012; Weinmann, 2016]. Der zweite Term wird als *paarweises Potential* bezeichnet und berücksichtigt die Interaktionen eines Primitivs mit seinen Nachbarn anhand

von Kanten im Graphen. Sowohl die Klassenlabels als auch die Daten der adjazenten Primitive lassen sich zur Modellierung der Interaktionen verwenden. Die Wahl des Klassenlabels eines Primitives ist somit bedingt abhängig von den Nachbarn. Häufig werden einfache assoziative Interaktionsmodelle verwendet, die gleiche Klassenlabels an benachbarten Primitiven bevorzugen, wie z. B. das Potts Modell [Potts, 1952] bzw. seine von Boykov & Jolly [2001] eingeführte kontrast-sensitive Variante. Generell zeigt sich, dass die Berücksichtigung von Interaktionen bereits bei einfachen und in ihrer Aussagekraft beschränkten Modellen den Anteil der Fehlklassifikationen reduziert [Schindler, 2012; Ladický et al., 2014]. Dementsprechend haben sich CRF im Bereich der Bildanalyse und inzwischen auch bei vielen Anwendungen aus Photogrammetrie und Fernerkundung als geeigneter Rahmen für Klassifikationsaufgaben etabliert [Schindler, 2012; Förstner, 2013; Rottensteiner, 2017]. Einige beispielhafte Applikationen, die zum Teil auch komplexere CRF verwenden, sind die multi-temporale Satellitenbildauswertung [Hoberg et al., 2015; Kenduiywo et al., 2016], die simultane Klassifikation von Landbedeckung und -nutzung [Albert et al., 2015], die Straßendetektion [Wegner et al., 2015] oder die Klassifikation von Gebäudefassaden in Bildern [Yang & Förstner, 2011].

Schindler [2012] führt einen Vergleich von glättenden Modellen auf hochauflösenden Luftbildern durch. Obwohl die ursprüngliche und auch die kontrast-sensitive Variante des Potts Modells im Vergleich zu einer unabhängigen Klassifikation der Variablen gut abschneiden, wird bei den einfachen Modellen teilweise ein zu starker Glättungseffekt ersichtlich. Ähnliche Erkenntnisse ergeben sich aus den frühen Arbeiten im Bereich der kontextbasierten Klassifikation terrestrischer Laserpunktwolken bei Robotikanwendungen. Die Entwicklungen in diesem Gebiet haben die Klassifikation von Punktwolken unter Berücksichtigung von Kontext wesentlich beeinflusst. Anguelov et al. [2005] unterscheiden bei ihrem Ansatz vier Objektklassen. Dabei kommen sogenannte Associative Markov Networks (AMN) zum Einsatz, welche eine Untergruppe von Zufallsfeldern darstellen. AMN entsprechen im Wesentlichen CRF mit linearen Modellen zur Bestimmung der Potentiale und nutzen ein generalisiertes Potts Modell mit klassenabhängigem Glättungsgewicht. Wie bei allen Varianten des Potts Modells wird auch dabei die Annahme zu Grunde gelegt, dass benachbarte Punkte mit hoher Wahrscheinlichkeit derselben Objektklasse angehören, so dass die Ergebnisse (teilweise zu stark) geglättet werden. Durch dieses ungünstige Glättungsverhalten kann sich die Fehlklassifikation eines Primitivs zusätzlich negativ auf die Umgebung auswirken und schnell auch die Labels der Nachbarprimitive verfälschen [Shapovalov et al., 2010]. Munoz et al. [2008] verwenden ebenfalls ein punktbasiertes AMN. Während die Ausrichtung der Kanten im Graph bei der Version von Anguelov et al. [2005] nicht von Bedeutung ist, erweitern sie das Modell zu einer anisotropen Variante, welche bestimmte Kantenorientierungen bevorzugt. Diese Richtungsinformation ermöglicht eine genauere Klassifikation von Objekten wie beispielsweise Stromleitungen.

Schlussfolgerung: Obwohl die Berücksichtigung der Beziehungen zwischen Primitiven im Allgemeinen empfehlenswert ist, wird bei den genannten Verfahren das Ergebnis durch das assoziative Verhalten zu stark geglättet. Weiterhin lassen sich mit diesen Modellen konkrete Beziehungen zwischen Objektklassen nicht ausdrücken, obwohl derartige Relationen wie *Autos befinden sich auf Straßen* wertvolle Informationen enthalten können. Dies deutet darauf hin, dass assoziative Modelle wie das kontrast-sensitive oder das generalisierte Potts Modell in Form eines AMN für eine genaue Klassifikation weniger geeignet sind. Nicht-assoziative Modelle können an dieser Stelle Abhilfe schaffen [Shapovalov et al.,

2010; Anand et al., 2012; Najafi et al., 2014; Pham et al., 2015]. Auf diese Weise lassen sich die Glättungseffekte reduzieren und die Genauigkeiten steigern. Nachteilig wirkt sich hingegen der gesteigerte Rechenaufwand sowie der Bedarf einer größeren Menge erforderlicher Trainingsdaten aus.

Integration von Kontext eines größeren Betrachtungsbereiches mittels CRF

Paarweise CRF können nicht die gesamten statistischen Zusammenhänge einer Szene modellieren und sind daher in ihrer Aussagekraft eingeschränkt [Kohli et al., 2009; Ladický et al., 2014]. Dies liegt darin begründet, dass üblicherweise nur die Interaktionen in einer lokal begrenzten Nachbarschaft eines jeden Primitivs über Kanten modelliert werden können. Durch die Hinzunahme von weiteren Kanten steigt die Komplexität des CRF jedoch schnell an. Es gibt verschiedene Ansätze, um aussagekräftigere Beziehungen eines gröberen Skalenbereichs zwischen Primitiven einer Szene zu integrieren.

Der erste Punkt besteht in der **Auswahl der als Primitive genutzten Entitäten**, die den zu berücksichtigen Skalenbereich unmittelbar beeinflussen. Pixel von Bildern stellen nach Ansicht von Ren & Malik [2003] keine natürlichen Entitäten dar und sind vielmehr als Konsequenz der diskreten Repräsentationsform von Bildaufnahmen anzusehen. Der Informationsgehalt eines einzelnen Pixels sei demnach gering und im Vergleich zu aussagekräftigeren Segmenten weniger für eine Klassifikation geeignet. Analog dazu könnte man die einzelnen Laserpunkte einer Punktwolke als diskretes Abtastungsergebnis der Erdoberfläche interpretieren. Der Aussage von Ren & Malik [2003] ist in der Hinsicht zuzustimmen, dass eine Klassifikation im herkömmlichen Sinn (ohne Berücksichtigung von Kontext) der angesprochenen Basisentitäten niedrigster Stufe wenig geeignet ist. Verwendet man jedoch beispielsweise im Rahmen eines CRF die Nachbarschaft, so kann darüber zumindest teilweise auf die entsprechenden Objektstrukturen geschlossen werden. Den Grundprimitiven Pixel oder Punkt kann somit leichter eine korrekte semantische Bedeutung zugeordnet werden.

Sowohl die Verwendung von Punkten als auch die der Segmente als zu klassifizierende Primitive ist jeweils an Vor- und Nachteile geknüpft. Während das Operieren auf der Punktebene das Erkennen feinerer Strukturen in den Daten und somit eine detaillierte (und damit im Prinzip eine genauere) Klassifikation ermöglicht, ist die räumliche Reichweite, welche in typischen CRF mit paarweisen Interaktionen abgedeckt wird, lokal sehr begrenzt. Zudem ist der Rechenaufwand im Vergleich zu Segmenten deutlich höher. Dem gegenüber stehen die Segmente, welche jeweils einen Zusammenschluss mehrerer ähnlicher Einzelpunkte darstellen. Durch die räumliche Ausdehnung eines Segmentes wird, verglichen mit Punkten, unmittelbar ein größerer Bereich in der Szene abgedeckt. Die Interaktionen mit den Nachbarsegmenten ermöglichen dann die Berücksichtigung von Informationen eines größeren Skalenbereichs. Mit der reduzierten Anzahl an Primitiven senkt sich zudem der erforderliche Rechenaufwand. Ein Nachteil ist jedoch der auf eine Untersegmentierung zurückzuführende Generalisierungsfehler. Da nach der Klassifikation jedem Punkt eines Segments dasselbe Label zugeordnet wird, können kleinere Details bei Mischsegmenten nicht erkannt werden.

In Bezug auf die Labelzuordnung bei Punktwolken zeichnet sich die Tendenz ab, dass vermehrt Segmente anstelle der Punkte als Primitive herangezogen werden. Beispiele für eine punktweise Klassifikation sind [Anguelov et al., 2005; Munoz et al., 2008; Lu et al., 2009], während die Arbeiten [Shapovalov et al., 2010; Anand et al., 2012; Kähler & Reid, 2013; Kim et al., 2013; Najafi et al.,

2014] und [Vosselman et al., 2017] auf einer Segmentebene operieren. Eine interessante Untersuchung zu der Art der Entitäten ist in [Shapovalov et al., 2010] zu finden. Die Autoren klassifizieren luftgestützte Laserpunktwolken und unterscheiden fünf Objektklassen. Dabei umgehen sie die Nachteile der zuvor verbreiteten AMN durch die Entwicklung eines nicht-assoziativen Markov Netzwerks. Dieses führt nicht primär eine Glättung durch, sondern modelliert die Verbundwahrscheinlichkeiten der Klassen adjazenter Knoten über ein generisches (sogenanntes nicht-assoziatives) Interaktionsmodell. Dazu besteht der erste Schritt in einer Übersegmentierung der Daten, bevor anschließend eine segmentweise CRF Klassifikation durchgeführt wird. Dies reduziert einerseits die Rechenzeit sowie die Anfälligkeit gegenüber Messrauschen, allerdings ist das Ergebnis stark abhängig von der Segmentierung selbst. Wie bei allen segmentbasierten Ansätzen können dabei wichtige Objektdetails oder kleine Objekte mit Ausdehnungen unterhalb der Segmentgröße verloren gehen. Shapovalov et al. [2010] zeigen in einem Versuch, dass sich der durch die Generalisierung hervorgerufene Fehler negativ auf die Gesamtgenauigkeit auswirkt und diese bei den durchgeführten Experimenten um 1-3 Prozentpunkte verschlechtert. Insbesondere Punkte unterrepräsentierter Klassen wie *Autos* werden mit dem Hintergrund verschmolzen. Wenngleich die Anzahl an Fehlklassifikationen bezüglich der Gesamtgenauigkeit vergleichsweise gering erscheint, kann sie für einzelne Klassen deutlich höher sein.

Um das Problem bei der Auswahl der Entitäten - und damit auch bei der Festlegung der Skalenebene - zu umgehen, sind in der Literatur häufig **hierarchische Ansätze** zu finden, z. B. [Kumar & Hebert, 2005; Ladický et al., 2014; Pham et al., 2015]. Sie versuchen durch die Kombination unterschiedlicher Skalenbereiche die Vorteile der verschiedenen Entitäten auszunutzen und gleichzeitig die jeweiligen Nachteile zu kompensieren. Diese Modelle bestehen aus mindestens zwei Hierarchiestufen, welche oft über Kanten verbunden sind. Typische repräsentierte Primitive der einzelnen Stufen sind bei Bildern Pixel und Superpixel, wie etwa in [Ladický et al., 2014]. Üblicherweise wird die Inferenz für das gesamte Modell in einem Schritt durchgeführt, um die bestmögliche Klassenzuordnung für alle Knoten, unabhängig von deren Hierarchiestufe, simultan zu ermitteln. Das dabei auftretende Problem ist jedoch, dass sich die Aggregationen von Primitiven niedrigerer Ebenen während der Optimierung nicht ändern können. Ein Laserpunkt, der einmal einem Segment i zugeordnet wurde, bleibt dessen Bestandteil und kann während der laufenden Inferenz kein Element eines anderen Segmentes j werden. Dies würde die Merkmale der höheren Entitäten verändern und somit die Klassifikationsaufgabe grundlegend neu definieren. Falschzuordnungen in der initialen Segmentierung lassen sich auf diese Weise nicht mehr korrigieren und äußern sich in der finalen Lösung der Inferenz als Fehler. Abhilfe lässt sich durch eine alternierende Optimierung der einzelnen Ebenen schaffen, die jedoch nicht zur global besten Lösung führen muss [Xiong et al., 2011].

Eine weitere Möglichkeit zur Integration großräumigen Kontextwissens in ein graphisches Modell stellen Kanten zwischen räumlich nicht unmittelbar benachbarten Knoten dar. Diese **weitreichenden Interaktionen** repräsentieren regionale Abhängigkeiten. Werden bei einem CRF im Extremfall alle Knoten miteinander verknüpft, bezeichnet man das graphische Modell als vollständig verbunden.

Shapovalov et al. [2013] führen beispielsweise eine Klassifikation von Innenraumdaten durch und verwenden als Entitäten Segmente anstelle der einzelnen Punkte. Sie berücksichtigen Abhängigkeiten über größere Distanzen mittels sogenannter struktureller Verbindungen. Diese verlaufen in speziellen Richtungen wie der Vertikalen, der Richtung zum Sensor oder zur nächstgelegenen Wand. In Innen-

raumszenarien lassen sich die Wände anhand von Heuristiken leicht finden [Shapovalov et al., 2013]. Bei luftgestützten Aufnahmen ist die Anzahl von Laserpunkten auf Wänden bzw. Gebäudefassaden jedoch meistens relativ gering. Die Identifikation solcher Punkte ist eher als Teil der Klassifikationsaufgabe anzusehen und nicht mittels Heuristik ableitbar. Weiterhin sind die Vertikale und die Richtung zum Sensor bei luftgestützten Aufnahmen nahezu identisch. Folglich lässt sich das Verfahren nicht unmittelbar auf die behandelte Aufgabenstellung übertragen. Dennoch ist diese Arbeit ein Beispiel dafür, wie sich in Punktwolken eine relative Anordnung von Objekten ableiten lässt. Lim & Suter [2009] wenden zunächst eine Segmentierung terrestrischer Laserdaten an und klassifizieren anschließend die resultierenden Supervoxel unter Berücksichtigung einer lokalen und einer regionalen Nachbarschaft. Die Einführung verschiedener Skalen, repräsentiert durch weitreichendere Interaktionen im CRF, führt zu einer Verbesserung der Klassifikationsgenauigkeit um 5-10 Prozentpunkte. Dieses Ergebnis zeigt die Bedeutung gröberskaliger Betrachtungen zusätzlich zu lokal begrenzten Punktnachbarschaften.

Obwohl weitreichende Interaktionen zu einer Verbesserung der Klassifikationsgenauigkeit führen können, sind sie mit Nachteilen verbunden. Einerseits erhöht sich die Anzahl der Kanten und somit auch die Komplexität des Optimierungsproblems stark. Andererseits ist (im Fall eines nicht vollständig verbundenes Modells) schon im Vorfeld bei der Erstellung der Graphstruktur festzulegen, zwischen welchen Knoten nutzbringende Abhängigkeiten eingeführt werden können. Dies erfordert Vorwissen und bringt zusätzlichen Aufwand bei der Prozessierung mit sich.

Wie bereits dargelegt wurde, ist die Ausdruckskraft paarweiser Interaktionen in einem graphischen Modell beschränkt. Vor allem an Objektgrenzen können assoziative paarweise Interaktionen zu einer Überglättung führen, so dass die Detektion von feinen Strukturen (z. B. bei Bäumen oder Büschen) durch die Überglättung der Konturen in Bilddaten nur schwer möglich ist [Kohli et al., 2007; Schindler, 2012]. Potentiale in einem CRF lassen sich jedoch nicht nur für zwei Variablen beschreiben, sondern auch für Gruppen mit einer größeren Anzahl an Variablen. Da sie auf *Cliquen höherer Ordnung* operieren, werden sie entsprechend als **Potentiale höherer Ordnung** (engl.: higher order potentials, HOP) bezeichnet und können das Ergebnis einer Klassifikationsaufgabe deutlich verbessern [Kohli et al., 2009; Najafi et al., 2014]. Die meisten Entwicklungen von HOP entstammen aus dem Bereich der Bildklassifikation. Dabei lassen sich besonders bei hierarchischen CRF mit assoziativem Verhalten gute Ergebnisse erzielen (z. B. [Ladický et al., 2014; Wang & Yung, 2015]). Nachteilig ist der erheblich größere Rechenaufwand bei der Suche der optimalen Labelkonfiguration. Dies stellt die wesentliche Einschränkung für die Verwendung von HOP dar, insbesondere bei der Nutzung vieler Variablen und der Modellierung nicht-assoziativer Interaktionen.

Roig et al. [2011] befassen sich mit der simultanen Klassifikation von Objekten aus Bildern mehrerer Kameras mit verschiedenen Ansichten einer Szene. Dazu verwenden die Autoren CRF mit HOP. Durch die Potentiale höherer Ordnung werden Verdeckungen modelliert und die Konsistenz zwischen den verschiedenen Aufnahmen sichergestellt. Zum Vereinfachen der Komplexität bei der Energieminimierung entwickeln Roig et al. [2011] einen iterativen Inferenzalgorithmus. Sie reduzieren die HOP auf unäre Terme, indem die Klassenlabels für die HOP in jeder Iteration als konstant angesehen werden. Ausgehend von dieser Teillösung werden in der folgenden Iteration die Konstanten erneut berechnet und die HOP entsprechend aktualisiert. Durch diese Vorgehensweise wird die Energiefunktion schrittweise approximiert und kann mit gängigen Inferenzmethoden effizient gelöst werden.

Ein bedeutender Beitrag zur Verbreitung von HOP ist eine von Kohli et al. [2007] vorgestellte Untergruppe, für die eine effektive Inferenz über Graph Cuts möglich ist. Das sogenannt P^n *Potts Modell* ist eine Generalisierung des paarweisen Potts Modells auf beliebig viele Variablen. Dabei handelt es sich um ein glättendes Verfahren, welches Homogenität hinsichtlich der vertretenen Objektklassen innerhalb einer Clique bevorzugt. Die Cliquen lassen sich etwa aus Segmentierungsergebnissen generieren. Dieselben Autoren erweitern den Ansatz zum robusten P^n Potts Modell [Kohli et al., 2009], das Abweichungen zum dominanten Label bis zu einem gewissen Anteil linear ansteigend bestraft, aber doch zulässt. Der starken Glättung der ursprünglichen Variante kann auf diese Weise entgegengewirkt werden, um Zuordnungen zu weniger dominanten Klassen besser zu ermöglichen. Zur Klassifikation terrestrischer Punktwolken werden robuste P^n Potts Modelle von Munoz et al. [2009a] genutzt.

Najafi et al. [2014] setzen eine nicht-assoziative Variante eines CRF höherer Ordnung für die Klassifikation terrestrischer und luftgestützter Laserdaten um. Bei dem segmentbasierten Ansatz ergeben sich die Cliquen höherer Ordnung aus den in die xy-Ebene projizierten, überlappenden Segmenten. Weiterhin modellieren die Autoren Höhendifferenzen über ein Potential, das bestimmte Muster in den Daten bevorzugt. Es zeigt sich, dass ein solches Vorgehen zwar für terrestrische Punktwolken hilfreich ist, im Fall einer luftgestützten Aufnahme jedoch die abgeleiteten Merkmale für die vertikalen Höhenunterschiede aufgrund der recht geringen Punktdichte beispielsweise an Fassaden nicht aussagekräftig genug sind. Obwohl CRF höherer Ordnung immer wichtiger werden, ist die Ausnutzung ihres gesamten Potentials für die Klassifikation von Punktwolken weiterhin stark durch den erforderlichen Rechenaufwand im Zuge der Inferenz eingeschränkt. Aus diesem Grund operieren Najafi et al. [2014] einerseits auf Segmenten anstelle von Punkten und beschränken andererseits die maximale Cliquengröße auf sechs Segmente, um die Inferenz mit gängigen Verfahren (hier: Loopy Belief Propagation) durchführen zu können. Dies ist zwar einer der ersten Ansätze zur Modellierung eines nicht-assoziativen Verhaltens zwischen den Objektklassen in einer Clique höherer Ordnung, aber zur praktischen Umsetzung durch die Einschränkungen für die Punktwolkenklassifikation wenig geeignet.

Interessante Varianten eines CRFs stellen die sogenannten *Decision Tree Fields* (DTF) [Nowozin et al., 2011] und *Regression Tree Fields* (RTF) [Jancsary et al., 2012] dar. Es handelt sich in beiden Fällen um in Zufallsfelder integrierte Entscheidungsbäume, die sich jedoch hinsichtlich der jeweils in den Blättern abgespeicherten Information unterscheiden. Für die Zuordnung diskreter Klassenlabels werden bei DTF die mit der entsprechenden Klasse verknüpften Energiewerte in Form eines Histogramms hinterlegt. RTF hingegen ermitteln die kontinuierlichen Zuordnungswahrscheinlichkeiten eines Primitives zu den Klassen, indem Mittelwerte und Kovarianzinformationen für Gaußverteilungen mit den Blättern der Bäume assoziiert werden. Nach dem Durchlauf des Testbeispiels ergibt sich dadurch schnell die entsprechende Verteilung, mit deren Hilfe die zugehörige Energie berechnet werden kann. Beide Methoden können im Vergleich zu aktuellen Standardverfahren bei vergleichbarer Genauigkeit eine schnellere Inferenz erzielen [Kähler & Reid, 2013; Pham et al., 2015]. Zum Anlernen der Energien und Wahrscheinlichkeiten sind jedoch sehr viele Trainingsdaten erforderlich. Der Umfang entspricht bei DTF üblicherweise mehreren Millionen Elementen [Nowozin et al., 2011]. Für die Klassifikation terrestrischer Punktwolken von RGB-Tiefenbildkameras verwenden Pham et al. [2015] RTF, um mit Hilfe der flexiblen Modelle die Komplexität der Inferenz deutlich zu reduzieren. Auf diese Weise können die Autoren nicht-assoziative Beziehungen selbst bei Cliquen höherer Ordnung berücksichtigen. Das dort

vorgeschlagene graphische Modell ist hierarchisch aufgebaut und nutzt Relationen zwischen Segmenten, sogenannten Supersegmenten und Objekten aus. Auf eine punktweise Vorgehensweise verzichten Pham et al. [2015] wegen des dennoch zu groß erscheinenden Rechenaufwands. Somit kann es durch die Segmentierung zu Generalisierungsfehlern kommen. Weiterhin sind im Vergleich zu nicht-assoziativen Interaktionen auf einer paarweisen Ebene deutlich mehr Trainingsdaten notwendig. Die wesentlichen Merkmale basieren auf den Farbinformationen des RGB-D-Sensors, welche für luftgestützte Laserdaten meist nicht vorhanden sind. Die geometrische Anordnung der Punkte ist somit deutlich entscheidender als bei RGB-D-Daten, was die Klassifikation zusätzlich erschwert.

Schlussfolgerung: Generell lässt sich festhalten, dass die Anwendbarkeit für Modelle mit HOP insbesondere bei zahlreichen Variablen und nicht-assoziativen Interaktionen derzeit noch durch das Fehlen effektiver Inferenzverfahren begrenzt ist. Hierarchische Ansätze erweisen sich als geeignetes Mittel zur Integration unterschiedlicher Skalenbereiche, um die jeweiligen Vor- und Nachteile der als Primitive gewählten Entitäten zu kompensieren. Auf der höchsten Ebene lassen sich regionale Interaktionen auf einfache Weise umsetzen, da größere räumliche Bereiche bereits durch die höherwertigen Primitive abgedeckt werden. Die Einführung von weitreichenden Interaktionen auf Punktebene ist jedoch zu rechenaufwändig und erfordert Vorwissen zur geeigneten Festlegung der Graphstruktur.

2.1.4 Alternative Möglichkeiten zur Kontextmodellierung

Eine alternative Herangehensweise zur Integration von Kontext in eine Klassifikation besteht in der Verwendung von **Kontextmerkmalen**, welche auch unabhängig von graphischen Modellen genutzt werden können. Im einfachsten Fall lassen sich auf verschiedenen Skalenbereichen extrahierte Merkmale nutzen, um etwa bei Punkten mit lokal sehr ähnlichen Merkmalen die genauere Klassifikation zu ermöglichen [Blomley et al., 2016b; Weinmann, 2016]. Besonders in urbanen Gebieten ist die Aussagekraft gröberskaliger Merkmale jedoch durch die Vielzahl der auftretenden Objektkonfigurationen teilweise beschränkt, so dass sich nicht alle Mehrdeutigkeiten lösen lassen [Niemeyer et al., 2014]. Anstelle mehrskalig berechneter Merkmale lässt sich Kontext beispielsweise anhand einer Reihe von einfachen, aufeinanderfolgenden Klassifikatoren mittels *Auto-Kontext* [Tu & Bai, 2010] modellieren. Dabei werden die probabilistischen Zwischenergebnisse einer Iteration als Merkmale für die nachfolgende Klassifikation verwendet. Gadde et al. [2016] haben den ursprünglich für Bilddaten entwickelten Ansatz auf die Klassifikation von 3D-Punktwolken übertragen.

Lucchi et al. [2011] stehen dem positiven Beitrag von CRF-ähnlichen Modellen für die Klassifikation kritisch gegenüber. In ihren Experimenten zeigen sie, dass sich bei Bilddatensätzen über eine Klassifikation von Superpixeln in Kombination mit globalen Merkmalen ähnliche Ergebnisse wie bei graphischen Modellen erzielen lassen. Ihre Diskussion zur Gegenüberstellung probabilistischer und heuristischer Ansätze beschränkt sich jedoch nur auf Bilder und CRF Modelle, welche die relative Anordnung von Objekten innerhalb eines Bildes berücksichtigen. Weiterhin zeigen die Autoren den Einfluss globaler Zwänge basierend auf den gemeinsamen Auftrittswahrscheinlichkeiten der Objekte in einer Szene. Das geometrische paarweise Modell ist in der Lage, Beziehungen wie „Himmel soll sich oberhalb von Gras befinden“ auszudrücken. Dies setzt das Vorhandensein einer absoluten Referenzrichtung in allen Bildern voraus. In terrestrischen Aufnahmen ist dies bei horizontalen Aufnahmeausrichtungen übli-

cherweise mit der Vertikalen gegeben, bei einem luftgestützten (Laser-)Datensatz hingegen fehlt eine solche Referenzrichtung. Folglich lässt sich das Modell nur schwer auf Applikationen der luftgestützten Fernerkundung übertragen.

Im Computer Vision Bereich hat sich das Vorwissen über die relative Position von Objekten als hilfreicher Indikator für die korrekte Klassifikation von Bilddaten erwiesen. Bei der Berechnung des *relative location prior* [Gould et al., 2008] werden Wahrscheinlichkeitskarten für das Auftreten der unterschiedlichen Objektklassen anhand von Trainingsdaten ermittelt und als Merkmal für die Klassifikation herangezogen. Potentielle Aufenthaltspositionen eines Objekts in Bezug auf eine eindeutige Referenzrichtung können dementsprechend vorhergesagt werden. Explizit gegeben ist eine solche Referenz für die Ausrichtung von Bildern jedoch nur bei terrestrischen Daten, bei luftgestützten Punktwolken hingegen fehlt eine eindeutige Definition. Golovinskiy et al. [2009] übertragen die Idee auf hauptsächlich von mobilen Systemen erfasste Laserdaten zur Identifikation von Straßenmobiliar wie z. B. Hydranten oder Briefkästen. Größere Objekte wie Gebäude werden jedoch in einem Vorverarbeitungsschritt durch Filterung entfernt. In ihrer Veröffentlichung definieren die Autoren eine Referenzrichtung durch die Orientierung zur nächstgelegenen Straße, deren Positionen OpenStreetMap (OSM) entnommen werden. Demzufolge ist eine externe GIS-Datenbank mit Straßeninformationen erforderlich.

Kontext lässt sich auch über **weitere Ansätze** ohne Ausnutzung graphischer Modelle in den Prozess der semantischen Zuordnung integrieren. Xiong et al. [2011] präsentieren ein Verfahren, bei dem eine punkt- und eine regionenbasierte Klassifikation miteinander interagieren. Sie stellen eine hierarchische Sequenz von verhältnismäßig einfachen Klassifikatoren angewandt auf Punkte und Segmente vor. Angefangen auf einer beliebigen Ebene werden die Ergebnisse einer Klassifikation herangezogen und als Kontextmerkmale im nächsten Schritt berücksichtigt, um das Gesamtergebnis iterativ zu verbessern. In jedem Schritt werden die Zwischenergebnisse als temporär konstant angesehen. Im Vergleich zu einer simultanen Inferenz mittels eines CRF kann auf diese Weise die Annäherung an das globale Optimum der posteriori Verteilung aller Klassenlabels unter Umständen erschwert werden. Dennoch ist die von den Autoren vorgeschlagene Herangehensweise für das in der vorliegende Arbeit entwickelte Verfahren von Bedeutung, da anhand einer solchen Strategie eine Interaktion zweier Skalenebenen mit variablen Gruppenzusammensetzungen auf der generalisierten Ebene durchgeführt werden kann.

Geometrische Beziehungen zwischen Primitiven sowie häufig benachbarte Objektklassen lassen sich mittels *3D entangled forests* [Wolf et al., 2016] auch ohne die Verwendung von graphischen Modellen bei einer Klassifikation berücksichtigen. Als Erweiterung von RF wird der Weg eines zu klassifizierenden 3D-Primitives durch den Zufallsbaum von dessen Position in der Szene und, in den tieferen Ebenen der Bäume, durch das bis dahin aktuell am wahrscheinlichsten erachtete Label beeinflusst. Die von Wolf et al. [2016] vorgestellten 3D-Merkmale zwischen benachbarten Primitiven basieren jedoch auf engen, manuell festgelegten Schwellwerten, etwa hinsichtlich maximal zulässiger euklidischer Abstände und Änderungen der Normalenvektoren, die sich in dieser Art nur in Innenraumszenen verwenden lassen. In aus der Luft aufgenommenen Szenen ist die Variabilität der zu beobachteten Objektkonfigurationen größer, so dass sich solche Merkmale schwer übertragen lassen.

Neuere Ansätze bei der semantischen Zuordnung luftgestützter Laserdaten verwenden *(tiefe) Neuronale Netze* (engl.: (deep) convolutional neural networks, CNN) für die direkte Klassifikation in unter-

schiedliche Objektklassen [Huang & You, 2016] oder mit dem übergeordneten Ziel, die Bodenpunkte für eine Extraktion des Geländemodells zu identifizieren [Hu & Yuan, 2016]. Einerseits haben CNN den Vorteil, im Zuge des Trainings diskriminative Merkmale automatisch zu identifizieren, so dass diese nicht a priori vorgegeben werden müssen. Andererseits ist dieser Gewinn an einen erheblich größeren Umfang erforderlicher Trainingsdaten gekoppelt. Kontextwissen lässt sich hierbei nicht explizit modellieren. Weiterhin ist die Anwendung der CNN auf unregelmäßig verteilte Punktwolken ohne eindeutig definierte Nachbarschaftsstruktur nicht ohne Weiteres möglich [Hackel et al., 2016; Weinmann, 2016]. Die bisherigen Arbeiten führen für diesen Zweck eine Aufteilung des 3D-Raums in regelmäßige Voxel durch, welche anschließend klassifiziert werden.

Schlussfolgerung: Sowohl die Betrachtung von Auftrittswahrscheinlichkeiten als auch die Berücksichtigung von Kontextmerkmalen können zu verbesserten Klassifikationsgenauigkeiten führen, ohne die Komplexität des Problems signifikant zu beeinflussen. Insbesondere bei luftgestützten Laserdaten findet die Umsetzung und Entwicklung von Kontextmerkmalen in der Literatur bisher kaum Beachtung. Die Herausforderung besteht hierbei vor allem im Fehlen einer Referenzrichtung, welche zur Orientierung herangezogen werden kann.

2.2 Segmentierung von Punktwolken

Bei in der Fernerkundung verwendeten Punktwolken handelt es sich üblicherweise um große Datenmengen. Um den Rechenaufwand bei der Prozessierung zu verringern, erfolgt bei vielen Applikationen als Vorverarbeitungsschritt zunächst eine Gruppierung von Laserpunkten mit ähnlichen Eigenschaften zu homogenen Regionen [Melzer, 2007; Aijazi et al., 2013; Papon et al., 2013; Nguatem & Mayer, 2016]. Dieser Prozess wird als Segmentierung bezeichnet [Nguyen & Le, 2013]. Von Vosselman [2013] wird der Begriff der Punktwolkensegmentierung mit Hinblick auf eine nachfolgende Klassifikation etwas enger gefasst. Demnach handelt es sich dabei um die „sinnvolle Zerlegung in kleinere und verbundene Untermengen von 3D-Punkten, welche Teilen von Objekten oder gesamten Objektinstanzen entsprechen“. Eine optimale Segmentierung liegt dann vor, wenn ausschließlich Punkte einer Objektklasse innerhalb eines Segments auftreten [Vosselman, 2013]. Abweichungen davon lassen sich einer der folgenden beiden Kategorien zuordnen:

- Untersegmentierung: Primitive verschiedener Objektklassen werden zu einem Segment verschmolzen. Dadurch entstehen Generalisierungsfehler, da allen Punkten nach der Klassifikation dasselbe Klassenlabel zugeordnet wird.
- Übersegmentierung: Generalisierungsfehlern wird vorgebeugt. Hingegen reduziert sich die Qualität der Segmentmerkmale, da zum einen über weniger Punkte gemittelt wird und das Ergebnis somit anfälliger gegenüber Rauschen ist. Zum anderen sind von den Segmenten abgeleitete Formparameter weniger aussagekräftig.

Das Ziel für die meisten Anwendungen ist es, eine zufriedenstellende Gruppierung der Punkte vorzunehmen, ohne dass einer der beiden genannten Fälle eintritt. Nur dann sind Segmente mit aussagekräftigen Merkmalen gewährleistet, während gleichzeitig der Anteil der Generalisierungsfehler minimal ist. In der Praxis ist dieser Zustand jedoch nur schwer zu erreichen.

Weiterhin lässt sich eine Segmentierung damit motivieren, dass einzelne Laserpunkte (analog zu Pixeln in Bildern) keine natürlichen Entitäten darstellen, sondern vielmehr das Ergebnis einer Diskretisierung sind [Ren & Malik, 2003]. Folglich kann eine Zusammenfassung in aussagekräftige Komponenten zur Erkennung von Objekten hilfreich sein. Dieser Abschnitt soll eine kurze Einordnung bestehender Segmentierungsansätze mit Fokus auf die Punktwolkenverarbeitung geben. Für eine umfangreiche Übersicht wird auf [Nguyen & Le, 2013] verwiesen. Der dort zusammengestellte Überblick bezieht sich zwar hauptsächlich auf die Segmentierung terrestrischer Laserpunktwolken, die meisten Algorithmen lassen sich jedoch auf den Fall der luftgestützten Datenaufnahme übertragen.

In den meisten Anwendungen werden Punktwolken modellbasiert anhand der lokalen Oberflächenform segmentiert [Schnabel et al., 2007; Vosselman & Klein, 2010; Nguatem & Mayer, 2016]. Da viele Applikationen auf die Extraktion von Gebäudeinformation abzielen, spielt die Detektion von Ebenen eine entscheidende Rolle. Über Methoden wie z. B. RANSAC [Fischler & Bolles, 1981] oder eine 3D Hough Transformation kombiniert mit Flächenwachstumsverfahren wie in [Vosselman & Klein, 2010] können diese gut aus der unregelmäßigen Punktwolke extrahiert werden. Planare Segmente sind jedoch ungeeignet, um Objekte wie Vegetation oder Autos zu detektieren. In diesem Bereich ist die Anzahl vorliegender Arbeiten geringer. Golovinskiy & Funkhouser [2009] und Reitberger et al. [2009] befassen sich mit der Aufspaltung großer Segmente zu einzelnen Objektinstanzen bestimmter Klassen über graphenbasierte Methoden. Beide Ansätze sind jedoch aufgrund des erforderlichen Vorwissens über die Objekte hinsichtlich ihrer Einsatzfähigkeit für eine holistische Szenensegmentierung beschränkt. Vosselman [2013] stellt eine Kombination mehrerer Verfahren vor, um eine vollständige Segmentierung samt nicht-planarer Komponenten zu erhalten. Für eine wie in dieser Dissertation angestrebte vollständige Segmentierung einer Szene mit unterschiedlichsten Objektarten sind solche modellbasierten Strategien aufgrund des fehlenden Modellwissens weniger geeignet als datengetriebene Ansätze.

Zu diesen gehören etwa unüberwachte Clustering-Algorithmen wie k-means [Lloyd, 1982]. Zusätzlich zur Position der Punkte in der Szene lassen sich weitere Merkmale berücksichtigen, da die Cluster im Merkmalsraum detektiert werden. Iterativ erfolgt eine Minimierung der quadratischen euklidischen Distanzen zwischen den Clusterzentren und den Merkmalen der Punkte im Merkmalsraum. Nachteilig wirkt sich aus, dass die Anzahl der zu findenden Segmente durch den Nutzer a priori festgelegt werden muss, was bei unbekannten Szenen schwer ist. Weiterhin erfordert die Bestimmung vieler Cluster einen großen Rechen- und Zeitaufwand. Eine Variante von k-means, welche auf die Segmentierung von Punktwolken optimiert ist, stellt die *Voxel Cloud Connectivity Segmentation* (VCCS) [Papon et al., 2013] dar. Dabei wird der 3D-Raum zunächst über eine Octree-Struktur partitioniert, um anschließend effizient lokal ähnliche Punkte gruppieren zu können. Das Übersegmentierungsverfahren führt zu guten Ergebnissen und wird in vielen Arbeiten verwendet [Stutz, 2015; Wolf et al., 2016].

Felzenszwalb & Huttenlocher [2004] stellen eine häufig genutzte und im Folgenden als FH-Algorithmus bezeichnete Methode vor. Dabei erfolgt die Segmentierung anhand von Merkmalen effizient über einen graphenbasierten Ansatz, der einfache Entscheidungen nach der Greedy-Strategie trifft, aber dennoch globale Kriterien optimiert. Die Laufzeit verhält sich etwa linear zur Anzahl der Kanten des Graphen. Ursprünglich wurde das FH-Verfahren für 2D-Bilddaten entwickelt, bei denen die zu segmentierenden Pixel in einem regelmäßigen Raster angeordnet sind. Brenner [2016] nutzt FH zur Segmentierung von Tiefenbildern aus mobilen Laserdaten. Da der Methode ein Graph zugrunde liegt,

lässt sie sich ebenfalls auf andere Strukturen wie irreguläre 3D-Punktwolken anwenden [Sima & Nüchter, 2013]. Für die in dieser Dissertation angestrebte Zielsetzung sind neben des festzulegenden und von der Szene abhängigen Glättungsparameters besonders die unterschiedlichen Größen der entstehenden Segmente nachteilig. Ohne Weiteres kann keine Punktanzahl für die maximale Segmentgröße angegeben werden, so dass die Segmente hinsichtlich ihrer Kardinalitäten stark heterogen sind.

Die hier vorgestellten Verfahren fügen benachbarte Primitive basierend auf vergleichsweise einfachen Homogenitätsannahmen zu Segmenten zusammen, welche meistens auf der lokalen Geometrie beruhen. Dies birgt jedoch das Risiko, Primitive unterschiedlicher semantischer Gruppen zu einer Komponente zu verschmelzen, was insbesondere für eine anschließende Klassifikation zu unwiderruflichen Untersegmentierungsfehlern führt. Für eine aussagekräftige Beschreibung einer Szene unter Ausnutzung von Segmenten ist der Prozess der Segmentierung eng gekoppelt mit dem der Klassifikation. Nur bei einer zufriedenstellenden Segmentierung können auch bei der Klassifikation gute Ergebnisse erzielt werden. Es ist somit naheliegend, beide Aufgaben gemeinsam zu lösen [Gould et al., 2009].

Yao et al. [2012] befassen sich mit einer holistischen Beschreibung von terrestrischen 2D-Bilddaten. Sie lösen das Problem der zeitgleichen Segmentierung und der Bestimmung von Klasse, Position und räumlicher Ausdehnung eines Objektes, des Vorhandenseins einer Klasse in einem Bild, und ermitteln, um welche Szenenkategorie es sich handelt. Dafür verwenden die Autoren ein CRF, das Vorinformation über Objektarten wie *Auto*, *Vogel* oder *Kuh* integriert. Für eine datengetriebene Vorgehensweise, wie sie bei (3D) Fernerkundungsdaten aufgrund der großen Variabilität von Objekten gewünscht ist, lässt sich ein solches Verfahren schwer übertragen.

Insgesamt liegt der Fokus der Arbeiten zur gemeinsamen geometrischen und semantischen Segmentierung eindeutig auf Bildern (z. B. [Gould et al., 2009; Gonfaus et al., 2010; Yao et al., 2012]). Für Punktwolken sind bisher nur wenige Veröffentlichungen bekannt. Dohan et al. [2015] fassen die Ableitung einer möglichst genauen Segmentierung mobiler Laserdaten als Grundvoraussetzung für eine zufriedenstellende Klassifikation speziell kleinerer Objekte wie Autos auf. In ihrer Arbeit wird eine Methode vorgestellt, die beide Komponenten vereint. Aufbauend auf einer Übersegmentierung der Punktwolke wird zusätzlich zu der semantischen Klassifikation ein weiterer Klassifikator angelernt, um in einem hierarchischem Clusteringansatz eine Wahrscheinlichkeit für die Zusammengehörigkeit zweier benachbarter Segmente zu einem einzigen Objekt abzuleiten. Weisen die Merkmale darauf hin, wird eine Verschmelzung durchgeführt. Allerdings werden ausschließlich die kleinen Klassen berücksichtigt, da über eine Vorklassifikation Objekte wie Gebäude, Straße usw. aus den Daten eliminiert werden. Weiterhin nutzen Dohan et al. [2015] die Kontextbeziehungen zwischen den Klassen nicht unmittelbar in Form eines CRF aus. Stattdessen werden Kontextmerkmale extrahiert, die die relative Anordnung von Objekten modellieren - im konkreten Fall die Entfernungsunterschiede der Objekte zur Bordsteinkante. Angelehnt an die Nutzung dieser relativen Merkmale wird in der vorliegenden Dissertation ein ähnliches Merkmal für den luftgestützten Fall vorgestellt. Zur Generierung semantisch homogener Segmente sollen weiterhin die Ergebnisse einer Vorklassifikation berücksichtigt werden.

Schlussfolgerung: Für eine möglichst verlustfreie semantische Zuordnung von Segmenten zu Objekten sollte der Prozess der Segmentierung eng an den der eigentlichen Klassifikation gekoppelt werden. Für Laserdaten gibt es erst wenige Entwicklungen, die sich mit diesem Aspekt befassen.

2.3 Diskussion

Die Analyse der verwandten Literatur zeigt, dass der Großteil der Arbeiten auf regelmäßigen (terrestrischen) Bilddaten statt auf irregulär angeordneten Punktwolken basiert. Viele Anwendungen lassen sich weiterhin wegen zugrundeliegender Annahmen z. B. bezüglich der Aufnahmerichtung nicht ohne Weiteres auf den luftgestützten Fall übertragen. Hinsichtlich der Integration von Kontext in eine Klassifikation kann festgehalten werden, dass diese im Allgemeinen zur Steigerung der Genauigkeiten empfehlenswert ist. Ein CRF stellt einen geeigneten Rahmen für diesen Zweck zur Verfügung. Da assoziative paarweise CRF jedoch zu Glättungseffekten führen, sollten komplexere Modelle berücksichtigt werden. Den Nutzen generischer Interaktionen auf Punktebene, welche sich insbesondere bei unterrepräsentierten Klassen positiv auswirken, haben dieser Arbeit vorausgehende Untersuchungen gezeigt [Niemeyer et al., 2011, 2014]. Einige dabei beobachtete Fehlzuordnungen sind auf den lokal zu begrenzten Kontextbereich zurückzuführen. Dies ist z. B. bei Gebäuden mit Giebeldächern der Fall, da die lokale Verteilung der Laserpunkte auf Giebeln jener von Vegetation ähnelt. Um Kontextinformationen mehrerer Maßstabsebenen zu integrieren und auf diese Weise die Vorteile einer punkt- und segmentweisen Klassifikation zu kombinieren, empfiehlt sich ein hierarchischer Ansatz. Die angesprochenen Fehler lassen sich durch die Hinzunahme einer zusätzlichen Segmentebene teilweise deutlich reduzieren [Niemeyer et al., 2015, 2016].

Allerdings bringt eine Segmentierung auch nachteilige Effekte mit sich. Durch die Gruppierung von Punkten zu Segmenten entstehen Generalisierungsfehler, welche sich insbesondere auf die Klassifikationsraten kleiner Strukturen und Objekte in den Daten negativ auswirken. Inspiriert durch [Gould et al., 2009] und [Dohan et al., 2015] sollen diese durch die Integration von Vorklassifikationsergebnissen in den Segmentierungsprozess reduziert werden. Dabei ist die Bildung einfacher Zusammenhangskomponenten basierend auf Klassenlabels nicht ausreichend. Wie Niemeyer et al. [2015] zeigen, führt dies zu isolierten Punkten, welche keinem Segment zugeordnet und demzufolge nicht bei der Klassifikation dieser Entitäten berücksichtigt werden können. Die Entwicklung einer verbesserten Segmentierungsmethode zur Integration in ein hierarchisches Verfahren ist somit notwendig und soll in dieser Dissertation vorgestellt werden. Die Verwendung von HOP auf der punktbasierten Ebene in Kombination mit einem weiterentwickelten Segmentierungsverfahren, welches auf den Konfidenzwerten einer Vorklassifikation basiert, sollen diese Probleme lösen. Eine zusätzliche Herausforderung bei der Ableitung geeigneter Segmente besteht in der Definition passender Segmentgrößen, welche wesentlich die Aussagekraft extrahierter Merkmale beeinflusst [Niemeyer et al., 2015, 2016]. Für einen gewinnbringenden Effekt sollten Segmente weder zu klein, noch zu groß sein. Repräsentiert ein Segment ein Objekt (z. B. ein gesamtes Gebäude), verlieren typische Merkmale wie Normalenvektoren aufgrund der unterschiedlichen Ausrichtungen einzelner Gebäudekomponenten ihren diskriminativen Charakter. Auch hierbei gilt es, einen geeigneten Kompromiss hinsichtlich der Segmentgrößen zu finden.

Mittels generischer Interaktionen zwischen den Segmenten lassen sich im Vergleich zur Punktebene aussagekräftige Beziehungen zwischen Objektklassen repräsentieren. Ein glättendes Verhalten ist zur Ausnutzung des gröberskaligen Kontexts nicht gewünscht. Die Kombination eines generischen Modells mit HOP erhöht die Komplexität des Inferenzproblems jedoch signifikant, so dass die Berechnung sehr aufwändig wird. Für diesen Zweck wird in der vorliegenden Arbeit das hierarchische Modell ähnlich

wie bei [Xiong et al., 2011] iterativ optimiert. Dabei werden die probabilistischen Ergebnisse jeweils an die andere Ebene weitergereicht, um a) die Segmentierung als solche zu verbessern und b) um Kontextmerkmale abzuleiten. Die alternierende Vorgehensweise hat weiterhin den Vorteil, dass auf Punktebene die Cliquen höherer Ordnung effizient mittels des robusten P^n Potts Modells gelöst, aber dennoch generische Interaktionen auf der Segmentebene modelliert werden können. Das Ergebnis der segmentbasierten Klassifikation mit einem komplexeren Interaktionsmodell lässt sich ähnlich zu [Roig et al., 2011] als zusätzlicher unärer Term, der temporär als konstant angesehen wird, in das punktweise CRF integrieren. Da allen Punkten innerhalb eines Segmentes dessen Konfidenzen zugeordnet werden, kann man das zusätzliche Unärpotential ebenfalls als HOP interpretieren.

Eine Reihe von Ansätzen leitet für terrestrische Anwendungsfälle Kontextmerkmale ab, welche die Klassifikationsergebnisse verbessern. Diese basieren auf den Resultaten einer Vorklassifikation [Tu & Bai, 2010; Gadde et al., 2016] oder setzen eine Referenzrichtung voraus [Gould et al., 2008; Golovinskiy et al., 2009], welche jedoch bei luftgestützten Laserdaten nicht explizit gegeben sind. In der Literatur wird die Umsetzung von Kontextmerkmalen für diesen Datentyp bisher wenig behandelt, so dass hier Weiterentwicklungen erforderlich sind. Eine weitere Zielsetzung dieser Arbeit ist es daher, für Aufnahmen aus der Luft ebenfalls Kontextmerkmale heranzuziehen und den Einfluss auf die Klassifikationsgenauigkeit zu untersuchen. Die vorläufigen Klassenkonfidenzen bilden dabei die Grundlage für die Extraktion zweier Merkmalsgruppen: Einerseits werden die Konfidenzen selbst als Merkmal integriert, andererseits dienen sie zur Approximation des Straßenverlaufs in der Szene, um Abstand und Orientierung eines Objekts zum nächstgelegenen Straßenabschnitt zu ermitteln. Straßen haben sich bei terrestrischen Applikationen für Punktwolken als geeignete Indikatoren zur Festlegung einer lokalen Referenzrichtung erwiesen [Golovinskiy et al., 2009; Dohan et al., 2015], da sie städtische Gebiete strukturieren und gut identifizierbar sind. Dies soll nun auf luftgestützte Daten übertragen werden.

3 Grundlagen

Dieses Kapitel gibt einen Überblick über die theoretischen Grundlagen der wesentlichen Aspekte dieser Arbeit. Abschnitt 3.1 befasst sich mit der Klassifikation von Laserdaten hinsichtlich Nachbarschaftsdefinition und Merkmalsextraktion. Daran schließt sich die Beschreibung des verwendeten Random Forest Klassifikators in Abschnitt 3.2 an. Für die Methodik sind Conditional Random Fields grundlegend. Diese werden in Abschnitt 3.3 erläutert. Das Kapitel endet mit der Beschreibung des zum Einsatz kommenden Segmentierungsverfahrens *Voxel Cloud Connectivity Segmentation* (Abschnitt 3.4).

3.1 Klassifikation luftgestützter Laserdaten

Bevor auf die eigentliche Klassifikation eingegangen wird, sollen zunächst die grundlegenden Begriffe definiert werden. Ein *Punkt* repräsentiert im Folgenden ein geometrisches Primitiv und beschreibt eine Position im euklidischer Raum $\mathbb{R}^D$. Falls nicht anders erwähnt, ist dieser Raum dreidimensional ($D = 3$) mit den Koordinaten XYZ. Die XY Ebene ist horizontal ausgerichtet und die Z-Achse zeigt senkrecht nach oben. Ein solcher Punkt hat keine geometrischen Dimensionen wie Länge, Fläche oder Volumen. Eine *Punktwolke* beschreibt einen Datensatz bestehend aus 3D-Punkten, welche im Gegensatz zu 2D-Bildern üblicherweise irregulär im Raum verteilt sind und nicht einer festen Rasterstruktur folgen. Für einen Laserpunkt ist im Allgemeinen die punktweise Repräsentation gebräuchlich, obwohl sich dieser im Gegensatz zum geometrischen Primitiv *Punkt* aufgrund des Divergenzwinkels des Laserstrahls tatsächlich aus der reflektierten Energie einer beleuchteten Fläche ergibt [Beraldin et al., 2010].

Üblicherweise werden zur Klassifikation Punktmerkmale nicht nur aus den Beobachtungen eines einzelnen Laserpunkts herangezogen, sondern auch aus einer lokalen Punktnachbarschaft extrahiert. Aus diesem Grund unterteilt Weinmann [2016] den Vorgang einer Klassifikation von Punktwolken in drei wesentliche Komponenten: Der erste Schritt besteht in der Definition einer lokalen Nachbarschaft für jeden 3D-Punkt. Anschließend folgt die Extraktion der Merkmale, teilweise unter Verwendung der vorher festgelegten Nachbarschaft. Diese beiden Schritte sind spezifisch für Laserdaten und werden in diesem Abschnitt behandelt. Die dritte Komponente stellt die eigentliche Klassifikation aller 3D-Punkte basierend auf den Merkmalen dar, um jedem Laserpunkt ein Klassenlabel zuzuweisen. Die Funktionsweise des in dieser Arbeit verwendeten Klassifikators wird in Abschnitt 3.2 vorgestellt.

3.1.1 Definition der Nachbarschaft

Nachbarschaftsbeziehungen zwischen einzelnen zu klassifizierenden Primitiven sind für die Merkmalsextraktion (sowie für den Aufbau eines graphischen Modells, siehe Abschnitt 3.3) von grundlegender

Bedeutung. Dies trifft für unregelmäßige Punktwolken genauso zu wie für Bilddaten. Betrachtet man die Pixel eines Bildes, so sind durch die rasterförmige Anordnung die adjazenten Primitive in Form von 4er- oder 8er-Nachbarschaft unmittelbar gegeben. Die Definition einer Nachbarschaft ist in diesem Fall somit naheliegend. Bei Bildsegmenten geht der Vorteil der Regelmäßigkeit verloren. Noch komplexer wird die Aufgabe für Punktwolken, da die Primitive unregelmäßig im 3D-Raum verteilt sind. Zum Auffinden adjazenter Punkte muss folglich eine aufwändigere Methode herangezogen werden, bei der Annahmen über die Nachbarschaft festzulegen sind. In den meisten Fällen erfordert dies die Einführung mindestens eines weiteren Parameters.

Eine Umgebung wird häufig über eine kugelförmige ($\mathcal{N}_{3D}^{r}$) [Lee & Schenk, 2002] oder eine zylindrische Nachbarschaft ($\mathcal{N}_{2D}^{r}$) [Filin & Pfeifer, 2005] mit einem manuell festgelegten Radius r realisiert (Abbildung 3.1), wie es auch zur Berechnung von Merkmalen in einer lokalen Nachbarschaft üblich ist [Weinmann, 2016]. In beiden Fällen kann die Anzahl der Nachbarn variieren. Eine weitere Möglichkeit besteht darin, für jeden Punkt eine festgelegte Anzahl k nächstgelegener Punkte anhand ihrer Distanz in 2D oder 3D zuzuweisen [Shapovalov et al., 2010]. Auf diese Art definierte Nachbarschaften werden im Folgenden entsprechend als $\mathcal{N}_{2D}^{k}$ und $\mathcal{N}_{3D}^{k}$ bezeichnet. Die Parameter r oder k sind typischerweise für jeden Datensatz individuell zu wählen und werden basierend auf Vorwissen oder Heuristiken festgesetzt [Weinmann, 2016]. Die Auswahl von einem einzigen Wert für k bei $\mathcal{N}_{2D}^{k}$ und $\mathcal{N}_{3D}^{k}$ ist für die Merkmalsextraktion weiterhin ungünstig, da sich die Nachbarschaftsdefinition beispielsweise bei variierenden Punktdichten nicht an die jeweilige Situation anpassen kann. Aus diesem Grund schlägt Weinmann [2016] vor, stattdessen für jeden Punkt individuell einen optimalen Wert für $k \in [k_{min}, k_{max}]$ über die Minimierung der lokalen Eigenentropie abzuleiten, um seine lokale Nachbarschaft $\mathcal{N}_{2D}^{k,opt}$ bzw. $\mathcal{N}_{3D}^{k,opt}$ zu definieren. Gesucht wird dabei derjenige beste Skalenwert für jeden Punkt, bei dem die Energiefunktion $E_\lambda = -e_1 \ln(e_1) - e_2 \ln(e_2) - e_3 \ln(e_3)$ der normierten Eigenwerte (e_1, e_2, e_3) der Kovarianzmatrix, welche sich aus den 3D-Koordinaten aller Punkte in der lokalen Nachbarschaft ableitet, minimal wird. Die auf diese Weise gewählte Nachbarschaft hat den Vorteil, dass die Merkmale robust in Bezug auf variierende Punktdichten in einem Datensatz sind und aussagekräftiger werden [Weinmann, 2016]. Eine vom Nutzer vorgegebene konkrete Stellgröße zur Festlegung einer Nachbarschaft (z. B. durch einen Radius r oder eine feste Anzahl k Nachbarn pro Punkte) entfällt somit und wird durch

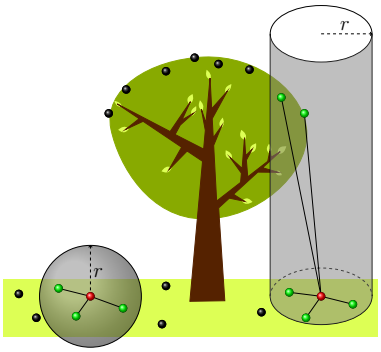


Abbildung 3.1: Bestimmung der Punktnachbarschaft für den jeweils rot markierten Punkt in einer kugelförmigen Suchumgebung $\mathcal{N}_{3D}$ (links) und einer zylindrischen Suchumgebung $\mathcal{N}_{2D}$ (rechts). Laserpunkte innerhalb der Umgebung werden als Nachbarn angesehen und sind grün markiert. In schwarz sind nicht berücksichtigte Laserpunkte dargestellt.

die weniger kritischen Parameter k_{min}, k_{max} zur Beschreibung des Intervalls ersetzt. Dies macht das Verfahren unabhängig vom Vorwissen über einen Datensatz und erleichtert die Bedienbarkeit. Zudem konnte Weinmann [2016] bei einem Vergleich unterschiedlicher Nachbarschaftsdefinitionen deutliche Genauigkeitssteigerungen der Ergebnisse durch die eigenentropiebasierte Skalenauswahl nachweisen.

3.1.2 Extraktion der Merkmale

Die Grundlage einer jeden Klassifikationsaufgabe stellen die berücksichtigten Merkmale der Primitive dar. Unabhängig von der Art der Daten (z. B. Bilder oder Punktwolken) lassen sich die Objektklassen nur dann fehlerfrei voneinander trennen, wenn die Merkmale im Merkmalsraum eindeutig abgrenzbaren Ballungen entsprechen. Klassifikationsfehler resultieren aus den Überlappungen der Ballungen. Die Hinzunahme weiterer Merkmale erhöht die Dimension des Merkmalsraums und kann zu einer sichereren Trennung der Klassen führen, falls auf diese Weise neue Informationen integriert werden. Dennoch sollte man darauf achten, nicht zu viele Merkmale zu verwenden, denn das *Hughes-Phänomen* kann die Klassifikationsgenauigkeit bei Erhöhung der Merkmalsdimension und konstant bleibendem Trainingsdatenumfang ab einem gewissen Grad sinken lassen [Hughes, 1968]. Weiterhin kann eine geeignete Auswahl an Merkmalen die Klassifikationsgenauigkeit steigern und gleichzeitig zu einer signifikanten Reduktion von Laufzeit und Speicherbedarf während der Berechnung führen [Weinmann, 2016].

Punktmerkmale: Für die punktweise Klassifikation von Laserdaten gibt es einen Satz von Merkmalen, welcher sich durch die umfangreiche Untersuchung von Chehata et al. [2009] mittlerweile als Standard herausgebildet hat. Alle oder zumindest einige der Merkmale werden beispielsweise in den Arbeiten [Shapovalov et al., 2010; Hackel et al., 2016; Weinmann, 2016] verwendet. Eine Gruppe umfasst die Merkmale basierend auf der Signalform. Aus dem zeitlichen Verlauf des zurückgeworfenen Signals lässt sich zunächst die *Intensität* ableiten sowie das *Echo-Verhältnis*, welches die Echo-Nummer jedes Punktes mit der Gesamtanzahl der Echos eines Pulses ins Verhältnis setzt. Dies ist hilfreich bei der Detektion von Vegetation, da dann häufig mehrere Echos innerhalb eines Pulses zu verzeichnen sind und der Quotient einen Wert kleiner als eins ergibt. Eine größere Gruppe an Merkmalen beschreibt die lokale Geometrie der 3D-Punkte in der Punktwolke. Die *Höhe eines Punktes über dem Gelände* hat sich insgesamt als besonders wichtig herausgestellt [Chehata et al., 2009; Horvat et al., 2016; Weinmann, 2016], wobei zur Bestimmung dieses Merkmals das Gelände vorab approximiert werden muss. Weitere geometrische Merkmale basieren auf den Eigenwerten $\lambda_1, \lambda_2, \lambda_3$ ($\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$) bzw. auf den normalisierten Eigenwerten e_1, e_2, e_3 mit $e_1 + e_2 + e_3 = 1$. Es handelt sich um *Linearität*, *Planarität*, *Streuung*, *Anisotropie*, *Omnivarianz*, *Eigenentropie* und *Krümmung*. Sie werden auf Grundlage der Matrix der Momente zweiter Ordnung aller Punktkoordinaten in einer lokalen Nachbarschaft $\mathcal{N}_{3D}^{k,opt}$ nach Pauly et al. [2003] folgendermaßen berechnet:

Linearität	$L_\lambda = (\lambda_1 - \lambda_2)/\lambda_1$	(3.1)	Planarität	$P_\lambda = (\lambda_2 - \lambda_3)/\lambda_1$	(3.5)
Streuung	$S_\lambda = \lambda_3/\lambda_1$	(3.2)	Anisotropie	$A_\lambda = (\lambda_1 - \lambda_3)/\lambda_1$	(3.6)
Krümmung	$C_\lambda = \lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$	(3.3)	Eigenentropie	$E_\lambda = -\sum_{i=1}^{3} e_i \ln e_i$	(3.7)
Omnivarianz	$O_\lambda = \sqrt[3]{e_1 \cdot e_2 \cdot e_3}$	(3.4)			

In der lokalen Nachbarschaft $\mathcal{N}_{3D}^{k,opt}$ wird zudem die *Standardabweichung aller z-Koordinaten* betrachtet. Analog zu [Weinmann, 2016] werden die *Radien der Nachbarschaften* von $\mathcal{N}_{3D}^{k,opt}$ und $\mathcal{N}_{2D}^{k,opt}$ herangezogen, die sich ebenfalls in ein *Verhältnis* setzen lassen. Der Radius berechnet sich aus der euklidischen Distanz zwischen dem aktuellen und dem entferntesten Punkt in der jeweiligen Umgebung.

Zusätzlich zur variablen Nachbarschaftsdefinition dient eine lokale Nachbarschaft mit festem Radius zur Beschreibung des *maximalen Höhenunterschieds in einer 2D-Umgebung*. Vergleicht man weiterhin die 3D-Umgebung mit jener in 2D, lassen sich Aussagen über das *Verhältnis der Punktanzahl* sowie der *Eigenwertsummen* und deren *Verhältnis* treffen. Anhand des Normalenvektors und einer approximierenden lokalen Ebene in $\mathcal{N}_{3D}^{k,opt}$ können weiterhin die *Summe der quadrierten Residuen* sowie der *Abstand eines Punktes zur bestanpassendsten Ebene* ermittelt werden. Die *Vertikalität* ergibt sich über den Zusammenhang $V = 1 - n_z$ aus der vertikalen Komponente n_z des Normalenvektors $\mathbf{n} = [n_x\ n_y\ n_z]^T \in \mathbb{R}^3$ [Weinmann, 2016].

Eine weitere Gruppe an Merkmalen stellen die *Fast Point Feature Histograms* (FPFH) [Rusu et al., 2009] dar. Dieser Deskriptor repräsentiert die geometrischen Eigenschaften der lokalen Nachbarschaft (Radius r_{NN}^P) über ein multi-dimensionales Histogramm bestehend aus 33 Komponenten. Dabei werden für jeden Punkt die Winkel zwischen dem eigenen Normalenvektor und denen der Nachbarprimitive berücksichtigt. Diese Signatur ist rotations-invariant sowie wenig anfällig gegenüber variierenden Punktdichten und Messrauschen [Rusu et al., 2009]. Nachdem FPFH zunächst im Bereich der Robotik genutzt wurden [Rusu et al., 2009], finden sie z. B. bei Najafi et al. [2014] inzwischen auch Anwendung für die Klassifikation luftgestützter Laserpunktwolken.

Segmentmerkmale: Für eine segmentierte Punktwolke lassen sich etwa die Mittelwerte (gekennzeichnet als $\emptyset$) und Standardabweichungen der genannten punktweisen Merkmale verwenden, aber auch weitere Eigenschaften z. B. aus der Form des Segments ableiten. Es ergeben sich sechs Gruppen von Merkmalen. Der erste Satz entstammt der Signalform der einzelnen Punkte ($\emptyset$ *Intensität*, *Standardabweichung Intensität*, $\emptyset$ *Echoverhältnis*). Wie bei den Punkten stellt auch die Höhe die Grundlage für einige wichtige Merkmale dar: $\emptyset$ *Höhe über Grund*, *min.* und *max. Höhe*, *max. Höhenunterschied* ΔZ, *Standardabweichung Höhe*. Aus allen Segmentpunkten lassen sich ebenfalls die Eigenwerte bestimmen, um daraus *Ebenheit*, *Linearität*, *Streuung*, *Eigenentropie*, *Omnivarianz* und die *Krümmung* abzuleiten. Auf einer lokalen Ebenenapproximation beruhen *Vertikalität des Segments*, $\emptyset$ *Vertikalität aller Segmentpunkte*, *Standardabweichung der Vertikalität* sowie die *Summe der quadrierten Residuen* zu einer ausgleichenden Ebene durch das Segment. Zusätzliche Informationen über die lokale Geometrie lassen sich zudem über die durchschnittlichen FPFH-Signaturen ableiten ($\emptyset$ *FPFH*). Die letzte Gruppe basiert auf dem minimal umgebenden Quader und beschreibt die Form der Segmente durch *Grundfläche*, *Volumen* sowie dem *Verhältnis aus Länge und Breite* des Quaders. Weiterhin ergibt sich die *Dichte*, indem die Punktanzahl eines Segments durch die *Grundfläche* des Quaders dividiert wird.

Zur Modellierung der Interaktionen zwischen zwei benachbarten Segmenten werden zusätzlich drei weitere Merkmale eingeführt. Der *Winkel zwischen den Segment-Normalenvektoren* kann zur Ähnlichkeitsanalyse der Nachbarn herangezogen werden. Die *kleinste euklidische Distanz zwischen Punkten beider Segmente* beschreibt den 2D-Abstand adjazenter Segmente. Dieser wird um den *geringsten Höhenunterschied am Übergang zwischen den Segmenten* ergänzt, um auch die 3D-Information zu nutzen.

3.2 Random Forest Klassifikator

Eine zentrale Aufgabe im Bereich der Bildverarbeitung und Fernerkundung ist die Klassifikation, also das Zuordnen von semantischer Information zu den Daten, welche beispielsweise Pixel, 3D-Laserpunkte oder Segmente darstellen können. Diese zu klassifizierenden Einheiten werden im Folgenden mit dem Begriff *Primitiv* zusammengefasst, um zunächst keine Einschränkung bezüglich der tatsächlichen Entität und des zugehörigen Skalenbereichs machen zu müssen. Die Menge aller Primitive ist gegeben durch $\mathbb{P} = \{1, \ldots, N_P\}$. Das Ziel der Klassifikation ist nun, jedem Primitiv $i \in \mathbb{P}$ ein Klassenlabel $y_i \in \mathcal{L} = [y^1, y^2, \ldots, y^L]$ der Menge aller Klassen $\mathcal{L}$ mit insgesamt L verfügbaren Kategorien zuzuordnen. Dabei beschreibt jedes Label eine thematischen Klasse. Weiterhin werden die Klassenlabels aller Primitive N_P zu einem Vektor $\mathbf{y} = (y_1, \ldots, y_{N_P})^T$ zusammengefasst.

Zur Bestimmung des optimalen Labels wird für jedes Primitiv ein Merkmalsvektor $\mathbf{f}_i(\mathbf{x})$ extrahiert, wobei $\mathbf{x}$ die gesamten Daten aller Primitive beschreibt. Klassifikationsverfahren lassen sich einer unüberwachten oder einer überwachten Strategie zuordnen. Während im erstgenannten Fall großteils automatisch Ballungen im Merkmalsraum detektiert werden, lernt die überwachte Variante einen Klassifikator an. Dafür sind repräsentative Trainingsdaten mit Referenzlabels erforderlich.

Zufallswälder, im Folgenden als Random Forests (RF) bezeichnet, haben sich zu einem der Standardverfahren zur Klassifikation im Bereich der Fernerkundung entwickelt [Gislason et al., 2006; Chehata et al., 2009; Millard & Richardson, 2015; Belgiu & Drăguţ, 2016]. Bei RF handelt es sich um ein überwachtes, nicht-probabilistisches Verfahren [Bishop, 2006].

Zunächst soll mit der Beschreibung eines einzelnen Entscheidungsbaums begonnen werden. Dabei ist im Allgemeinen zwischen Klassifikations- und Regressionsbäumen (classification and regression trees, CART [Breiman et al., 1984]) zu unterscheiden. Während Erstgenannte ein diskretes Ergebnis liefern, ergeben sich bei den Regressionsbäumen kontinuierliche Ergebnisse [Criminisi & Shotton, 2013]. In dieser Arbeit liegt der Fokus auf den Klassifikationsbäumen. Die Idee eines solchen Baums besteht darin, ein komplexeres Problem auf eine Hierarchie von einfachen Entscheidungen zu reduzieren. Dazu wird eine hierarchische Graphstruktur bestehend aus Knoten und Kanten aufgebaut. Zwei unterschiedliche Typen von Knoten lassen sich unterscheiden: die internen Knoten (Teilungsknoten) sowie die Blätter bzw. -Terminalknoten, welche die finale Entscheidung repräsentieren. Jeder interne Knoten (abgesehen von der Wurzel) hat genau eine ankommende und zwei abgehende Kanten. Es handelt sich somit um einen baumartigen binären Graphen. Wie in Abbildung 3.2 dargestellt, zweigt sich ein Baum ausgehend von einer Wurzel immer weiter auf und teilt dabei den Merkmalsraum. Die Klassifikation wird nun durchgeführt, indem an jedem Teilungsknoten des Baumes (in grün dargestellt) ein Binärtest auf einer Teilmenge der Merkmale des Vektors $\mathbf{f}(\mathbf{x})$ durchgeführt wird. Diese Test entscheidet, ob die Weitergabe an den linken oder rechten Zweig des Baumes erfolgt. Der Merkmalsvektor wird dann an den entsprechend nächsten Teilungsknoten mit einem anderen Test weitergegeben, bis ein Blatt erreicht wird. Die Blätter sind mit den Objektklassen assoziiert. Durch Präsentieren eines unbekannten Merkmalsvektors $\mathbf{f}_i(\mathbf{x})$ von einem Primitiv i kann dieses nach dem Abarbeiten der Einzeltests einer Objektklasse zugeordnet werden [Breiman, 2001; Criminisi & Shotton, 2013].

Der Aufbau des Baums mit der Zuordnung der Tests zu den Teilungsknoten erfolgt in einem Lern-

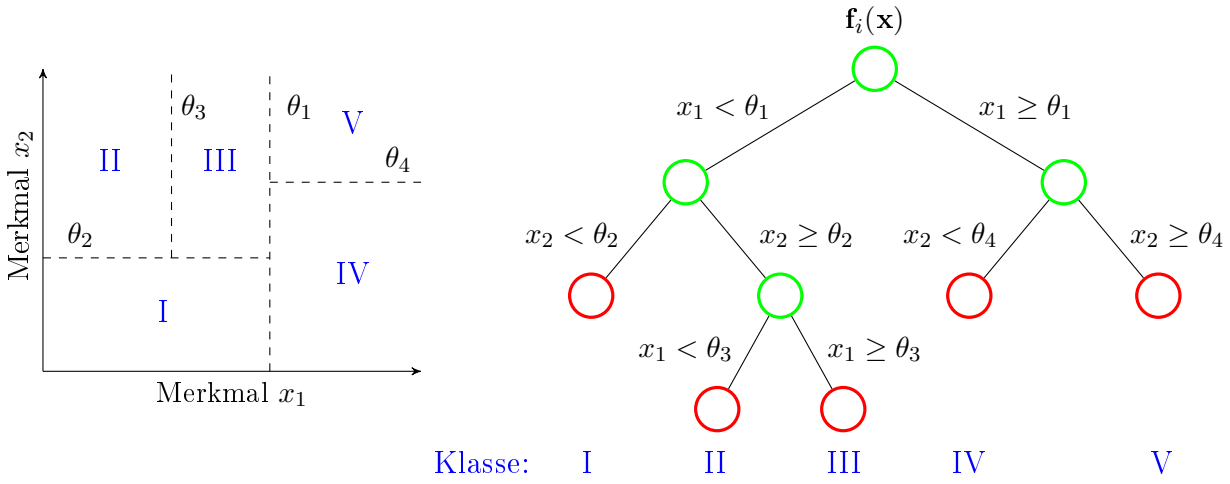


Abbildung 3.2: Prinzip eines Entscheidungsbaums. Basierend auf Tests (Schwellwertabfragen) wird der Merkmalsraum (links) nach und nach eingeteilt. Die Entscheidungsknoten sind in grün und die Blätter in rot dargestellt (rechts). x_i repräsentiert jeweils ein Merkmal, welches mit einem Schwellwert θ_i verglichen wird.

schritt, bei dem Merkmalsvektoren mit bekannten Referenzlabels herangezogen werden. Das Zufallsprinzip spielt hierbei an mehreren Stellen eine entscheidende Rolle. Zunächst wird von der Gesamtmenge der Trainingsdaten nur eine zufällige Teilmenge für den tatsächlichen Baumaufbau verwendet, die im Englischen als „In-bag"-Menge bezeichnet wird. Von mehreren zufällig vorgeschlagen Tests für jeden Knoten wird jeweils derjenige ausgewählt, welcher ein festgelegtes Qualitätskriterium am besten erfüllt. Das Kriterium ist meist verknüpft mit dem Informationszuwachs durch eine neue Trennfläche im Merkmalsraum (beschrieben z. B. durch die Änderung des Gini-Koeffizienten) [Breiman et al., 1984]. Wurde so jeder Knoten mit einem Test assoziiert, wird die zunächst zurückgehaltene Teilmenge der Trainingsdaten („Out-of-bag") verwendet, um die Klassenzuordnung in den Blättern zu ermitteln. Die Merkmalsvektoren dieser Daten werden dafür anhand des aufgebauten Baums bis zu den entsprechenden Blättern durchgereicht. In jedem Blatt wird die Anzahl der auftretenden Merkmalsvektoren pro Objektklasse und daraus ein normalisiertes Histogramm ermittelt. Die am häufigsten vertretene Klasse ist diejenige, für die das Blatt im Zuge der Klassifikation unbekannter Daten stimmen wird. Sie wird mit dem Blatt assoziiert [Breiman, 2001].

Die Verwendung von T individuellen randomisierten Entscheidungsbäume $b_1, \cdots, b_T$ als ein Ensemble führt zu einem RF Klassifikator (Abbildung 3.3). Das Anlernen jedes Baumes erfolgt auf einer zufällig gezogenen Teilmenge der Trainingsdaten, wobei einzelne Trainingsbeispiele mehrfach verwendet werden können (*Bootstrapping* [Efron, 1979]). Die Evaluation von unbekannten Daten erfolgt bei RF, indem jeder Merkmalsvektor $\mathbf{f}_i(\mathbf{x})$ von jedem Baum b_i klassifiziert wird, welcher dann auf Grundlage der jeweiligen Tests für eine Objektklasse l_i stimmt. Als finales Label für ein Primitiv wird die Klasse mit den meisten Stimmen ausgewählt [Breiman, 2001]. Die Methode zum Zusammenfassen mehrerer Einzelprädiktoren zu einem Ergebnis wird in diesem Kontext als *Bagging* (einem Akronym für ***b**ootstrap* und ***agg**regat**ing***) bezeichnet [Breiman, 1996]. Obwohl es sich bei RF um ein nichtprobabilistisches Klassifikationsverfahren handelt, lässt sich ein der Wahrscheinlichkeit ähnliches Maß ableiten. Durch diese Möglichkeit lassen sich RF z. B. gut für die Berechnung von Potentialen in Con-

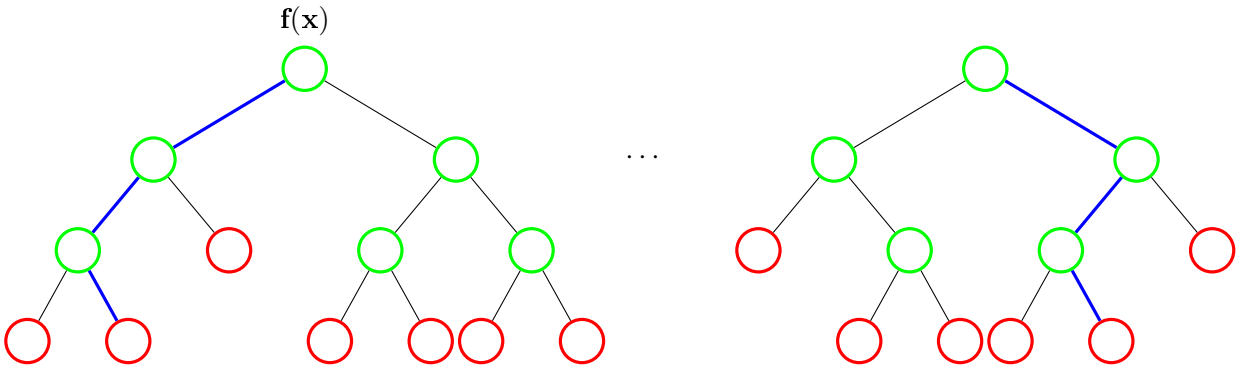


Abbildung 3.3: Prinzip des Random Forest Klassifikators. Ein Merkmalsvektor bekommt von jedem Entscheidungsbaum eine Stimme für eine Klasse. Die finale Zuordnung zu einer Kategorie erfolgt basierend auf einer Mehrheitsentscheidung. Die Entscheidungsknoten sind in grün und die Blätter in rot dargestellt.

ditional Random Fields (Abschnitt 3.3) verwenden. Dazu wird die Anzahl der Stimmen n_l für jede Klasse y^l mit der Anzahl der Bäume T normalisiert:

$$p(y_i = y^l | \mathbf{f}_i(\mathbf{x})) = n_l / T \tag{3.8}$$

Für die Anwendung von RF müssen eine Reihe von Parametern festgelegt werden. Dies ist die Anzahl der Bäume T, die Anzahl der Samples $N_{s,l}$ pro Klasse l im Training, die Tiefe der Bäume T_{depth} sowie die minimale Anzahl von Samples an einem Knoten, um einen neuen Entscheidungstest an dieser Stelle einzuführen (T_{split}). Die Anzahl der zufällig verwendeten Merkmale an einem Knoten (M_{split}) wird üblicherweise als die Quadratwurzel der Dimension des Merkmalsvektors berechnet [Gislason et al., 2006]. An dieser Stelle ist anzumerken, dass für ein gutes Training die gleiche Anzahl $N_{s,l}$ an Samples für jede Objektklasse $l \in \mathcal{L}$ verwendet werden sollte. Der Grund dafür ist, dass beim Aufbau der Bäume die Gesamtgenauigkeit der Klassifikation optimiert wird. Treten in einem Datensatz jedoch einige Klassen nur selten auf, haben diese nur einen sehr geringen Anteil an der Gesamtgenauigkeit. Eine Vielzahl von Untersuchungen empfiehlt daher, für eine unverzerrte Klassifikation beim Training dieselbe Anzahl an Beispielen pro Klasse zu verwenden. Sollten von einer Klasse nicht ausreichend viele Beispiele vorhanden sein, werden diese ggf. mehrmals verwendet [Chen et al., 2004; Millard & Richardson, 2015]. Der Index l von $N_{s,l}$ kann daher vernachlässigt werden, da keine Klassenabhängigkeit mehr gegeben ist. Die Anzahl der gesamten Trainingsbeispiele ergibt sich durch $N_s = N_{s,l} \cdot L$.

Das Verfahren der RF zeichnet sich durch eine gute Handhabung großer Datenmengen in Bezug auf die Anzahl der Primitive und Klassen sowie die Dimension der Merkmalsvektoren aus [Gislason et al., 2006]. RF sind unmittelbar geeignet für die Lösung eines Mehrklassenfalls. Es müssen keine Annahmen über die statistischen Verteilungen der Daten getroffen werden. Die Ergebnisse von RF sind mit denen von Boosting oder SVM vergleichbar bzw. ihnen sogar überlegen [Belgiu & Drăguţ, 2016]. Da es sich um ein diskriminatives Verfahren handelt, welches nur die Grenzen zwischen den Klassen modelliert, sind relativ wenige Trainingsdaten erforderlich. Des Weiteren sind RF verhältnismäßig schnell angelernt; die Trainingszeit t_{train} steigt entsprechend $t_{train} \propto T \cdot \sqrt{M_{split} \cdot N_s \cdot \log(N_s)}$ linear mit der Anzahl der Bäume [Breiman, 2003]. Die ohnehin schon rasche Evaluierung der unbekannten

Daten kann für eine weitere Verbesserung der Laufzeit noch parallelisiert werden. Ein nennenswerter Vorteil gegenüber anderen Verfahren ist zudem die Möglichkeit, dass auf einfache Weise die relative Wichtigkeit der einzelnen Merkmale ermittelt werden kann. Dies erfolgt durch zufälliges Vertauschen der Werte eines Merkmals in den Merkmalsvektoren basierend auf den *Out-of-bag*-Beispielen des Trainingsdatensatzes. Die Information dieses Merkmals wird auf diese Weise verfälscht. Ein Vergleich der resultierenden mit der ursprünglichen Genauigkeit, welche mit den korrekten Werten ermittelt wurde, gibt Auskunft über die Wichtigkeit des Merkmals. Bei Merkmalen mit großem Einfluss wird die Klassifikationsqualität nach Elimination des Merkmals deutlich stärker sinken als bei weniger wichtigen Merkmalen [Breiman, 2001; Belgiu & Drăguţ, 2016]. Nachteilig bei RF hingegen ist die Gefahr der Überanpassung an die Daten, falls eine zu große Baumanzahl verwendet wird. Unter Umständen kann zudem der Speicherbedarf sehr groß werden [Gislason et al., 2006]. Für weitere Details über RF sei an dieser Stelle auf die Arbeiten von Breiman [2001]; Criminisi & Shotton [2013]; Hänsch [2014] sowie Belgiu & Drăguţ [2016] verwiesen.

3.3 Conditional Random Fields

Conditional Random Fields stellen eine flexible Methode dar, um Kontext in eine Klassifikation zu integrieren. Sie gehören zur Familie der ungerichteten graphischen Modelle. Der zugrundeliegende Graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ besteht aus einer Menge von Knoten $\mathcal{V}$ und Kanten $\mathcal{E}$. Die Aufgabe der Klassifikation ist es, jedem Knoten $n_i \in \mathcal{V}$ basierend auf den beobachteten Daten $\mathbf{x}$ ein Objektlabel y_i aus dem Satz der Klassen $\mathcal{L}$ zuzuordnen. Die Knoten repräsentieren somit die zu klassifizierenden Zufallsvariablen.

Die Modellierung der a posteriori Wahrscheinlichkeiten $p(\mathbf{y}|\mathbf{x})$ für die Labels $\mathbf{y}$ bei gegebenen Daten $\mathbf{x}$ erfolgt mit Hilfe einer Gibbs Verteilung [Kumar & Hebert, 2006]

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{i \in \mathcal{V}} \varphi_i(\mathbf{x}, y_i) \prod_{e_{i,j} \in \mathcal{E}} \psi_{ij}(\mathbf{x}, y_i, y_j)^{\omega_p} \prod_{h \in \mathcal{H}} \psi_h(\mathbf{x}_h, \mathbf{y}_h)^{\omega_h}. \tag{3.9}$$

Im einfachsten Fall bestehen CRF aus einem Datenterm und einem paarweisem Interaktionsterm, aus denen sich die beste (gemeinsame) Labelkonfiguration aller Primitive ermitteln lässt. Dabei wird der Datenterm $\varphi_i(\mathbf{x}, y_i)$ als *unäres Potential* bezeichnet. Er beschreibt für jeden Knoten $i \in \mathcal{V}$ die Wahrscheinlichkeit, einer bestimmten Klasse anzugehören. Dies erfolgt über eine individuelle Betrachtung der einzelnen Knoten ohne Berücksichtigung der Nachbarinformationen. Die paarweise Kontextinformation wird durch die Kanten $\mathcal{E}$ in das graphische Modell integriert. Eine Kante e_{ij} verbindet dabei zwei adjazente Knoten $n_i, n_j \in \mathcal{V}$, deren Labels statistisch voneinander abhängig sind. Diese Information wird über das *paarweise Interaktionspotential* $\psi_{ij}(\mathbf{x}, y_i, y_j)$ modelliert, das dabei neben den beiden beteiligten Knotenlabels auch die Daten der Nachbarn berücksichtigt. Der Theorie entsprechend lassen sich für unterschiedliche Knoten bzw. Kanten die korrespondierenden Potentiale mit verschiedenen Modellen berechnen, was über die Indices i bei φ_i bzw. i, j bei ψ_{ij} ausgedrückt wird. In der Anwendung wird davon jedoch nur selten Gebrauch gemacht, so dass in dieser Arbeit im Folgenden auf die Indices aus Gründen der Übersichtlichkeit verzichtet wird. Das paarweise Modell lässt sich entsprechend für Cliquen höherer Ordnung (>2) erweitern, indem das Potential höherer Ordnung

(higher order potential, HOP) $\psi_h(\mathbf{x}_h, \mathbf{y}_h)$ eingeführt wird. Dieses ergibt sich für jede Clique $h \in \mathcal{H}$ aus den Merkmalen $\mathbf{x}_h$ und den Punktlabels $\mathbf{y}_h$ in der Clique. Die Parameter ω_p und ω_h in Gleichung 3.9 gewichten das paarweise Potential und das HOP relativ zum Unärpotential. Sie können entweder manuell festgesetzt oder z. B. über Kreuzvalidierung [Shotton et al., 2009] bzw. die in Abschnitt 3.3 beschriebene Methode anhand von Trainingsdaten angelernt werden. Durch die Normierung mit der Partitionsfunktion $Z(\mathbf{x})$ lassen sich die Potentiale als Wahrscheinlichkeiten interpretieren. Im Zuge der Inferenz werden die Klassenlabels $\mathbf{y}$ aller Primitive simultan durch das Maximieren der a posteriori Wahrscheinlichkeiten (Maximum a Posteriori, MAP) ermittelt, d. h. $\mathbf{y} = \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})$. Äquivalent dazu lässt sich die Energiefunktion E minimieren, welche sich über den Zusammenhang

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp(-E) \tag{3.10}$$

aus Gleichung 3.9 ableiten lässt. Die dem CRF zugrunde liegende Graphstruktur ist primär abhängig von der Art der Daten (etwa Bilddaten oder unregelmäßige Punktwolken) und den gewählten Primitiven (Pixel, 2D-Segmente, 3D-Punkte, 3D-Segmente, etc.). Im Folgenden werden die einzelnen Potentiale näher beschrieben.

Unäres Potential: Das unäre Potential $\varphi(\mathbf{x}, y_i)$ in der Gleichung 3.9 beschreibt die Klassenzugehörigkeit zu den L Klassen für jeden einzelnen Knoten des Graphen basierend auf seinem Merkmalsvektor $\mathbf{f}_i(\mathbf{x})$. Für jeden Knoten n_i wird ein solcher Merkmalsvektor aus den unmittelbar an diesem Knoten beobachteten Daten $\mathbf{x}_i$ sowie zusätzlich aus denen in einer lokalen Nachbarschaft extrahiert. Das Potential lässt sich beispielsweise entsprechend der Wahrscheinlichkeit von y_i bei gegebenen Daten $\mathbf{x}$ definieren, d. h. es gilt $\varphi(\mathbf{x}, y_i) = p(y_i|\mathbf{f}_i(\mathbf{x}))$. Prinzipiell kann jedes diskriminative Klassifikationsverfahren mit einem probabilistischen Ergebnis für die Berechnung des unären Potentials verwendet werden [Kumar & Hebert, 2006]. Häufig werden für diese Aufgabe z. B. lineare Modelle [Kumar & Hebert, 2006; Lim & Suter, 2009; Munoz et al., 2009b; Xiong et al., 2011] oder in aktuelleren Arbeiten RF [Shapovalov et al., 2010; Albert et al., 2015; Wegner et al., 2015; Wolf et al., 2015] eingesetzt. Die Wahrscheinlichkeiten im Fall des RF Klassifikators werden dabei entsprechend Gleichung 3.8 aus den Klassifikationsergebnissen der Entscheidungsbäume abgeleitet.

Paarweises Interaktionspotential: Während das unäre Potential für jedes Primitiv die Klassenlabels anhand des jeweiligen Merkmalsvektors $\mathbf{f}_i(\mathbf{x})$ unabhängig von seiner Umgebung bestimmt, berücksichtigt das Interaktionspotential $\psi(\boldsymbol{x}, y_i, y_j)$ in Gleichung 3.9 die Konfiguration von paarweisen benachbarten Klassenlabels. Die Abhängigkeit eines Primitivs i von seinem Nachbarn j wird im graphischen Modell über die Kanten $e_{i,j} \in \mathcal{E}$ zwischen den Knoten n_i und n_j, repräsentiert. Mit dem Aufbau des Graphen wird über die Kanten $\mathcal{E}$ festgelegt, wie die Nachbarschaft $\mathcal{N}_i$ jedes Primitives i definiert ist. Die beste Klassenkonfiguration ergibt sich aus einer Analyse der beiden adjazenten Knotenlabels sowie der Daten $\mathbf{x}$, welche durch einen Merkmalsvektor $\mathbf{f}_{ij}(\mathbf{x})$ repräsentiert werden. Dieser setzt sich üblicherweise aus den beiden beteiligten Knotenmerkmalsvektoren $\mathbf{f}_i(\mathbf{x})$ und $\mathbf{f}_j(\mathbf{x})$ zusammen, ist aber im Prinzip nicht auf diese Definition beschränkt. Häufig wird entweder die euklidische Distanz $\mathbf{f}_{ij}(\mathbf{x}) = d_{ij} = ||\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_j(\mathbf{x})||$ zwischen den Vektoren berechnet, die elementweise Differenz $\mathbf{f}_{ij}(\mathbf{x}) = [\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_j(\mathbf{x})]$ verwendet oder ein Ansatz gewählt, bei dem beide Vektoren durch Aneinanderhängen gemeinsam als Kantenmerkmalsvektor $\mathbf{f}_{ij}(\mathbf{x}) = [\mathbf{f}_i^T(\mathbf{x}), \mathbf{f}_j^T(\mathbf{x})]^T$ verwendet werden. Auch

Kombinationen oder spezifische Interaktionsmerkmale lassen sich integrieren [Rottensteiner, 2017].

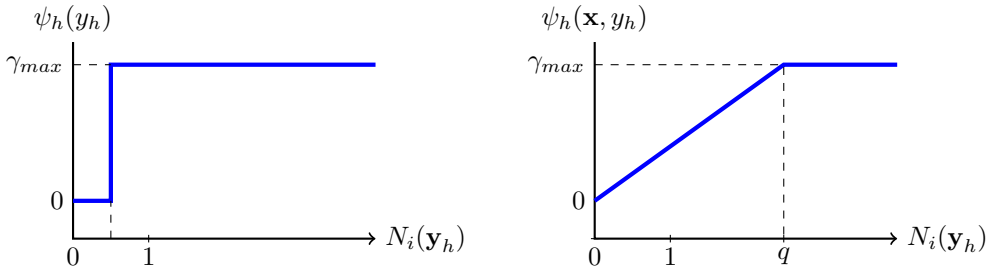
Es gibt verschiedene und unterschiedlich komplexe Möglichkeiten, die Interaktionen der beiden beteiligten Knoten zu modellieren. Viele Anwendungen nutzen eine Form des Potts Modells [Potts, 1952], das eine Verallgemeinerung des Ising Modells auf den Mehrklassenfall darstellt und einen glättenden Effekt auf die Klassenlabels hat [Schindler, 2012]. Die einfachste Variante lässt die Daten $\mathbf{x}$ außer Betracht und favorisiert das Auftreten gleicher Objektklassen an benachbarten Primitiven durch den Zusammenhang $\log \psi(y_i, y_j) = \beta \cdot \delta_K(y_i = y_j)$. Dabei nimmt die Kronecker-Deltafunktion $\delta_K(y_i = y_j)$ nur bei Labelgleichheit den Wert eins an, ansonsten gibt sie den Wert null zurück. Das manuell vorgegebene oder antrainierte Gewicht β beschreibt die Stärke der Glättung. Typischerweise wird dieses Modell bei MRF eingesetzt, da es unabhängig von den Daten operiert. In Untersuchungen hat sich jedoch eine deutliche Tendenz zur Überglättung abgezeichnet. Details in den Strukturen der Szene können dementsprechend verloren gehen [Kumar & Hebert, 2006].

Das für CRF häufig verwendete kontrast-sensitive Potts Modell [Boykov & Jolly, 2001] ist assoziativ und bevorzugt identische Labels für beide durch eine Kante verbundene Knoten. Im Gegensatz zum einfachen Potts Modell hängt der Grad der Glättung dabei von den Daten $\mathbf{x}$ ab. Es basiert auf der Annahme, dass zwei Knoten i und j mit ähnlichen Merkmalsvektoren $\mathbf{f}_i(\mathbf{x})$ und $\mathbf{f}_j(\mathbf{x})$ wahrscheinlich der gleichen Objektklasse angehören. Diese Ähnlichkeit wird beim kontrast-sensitiven Potts Modell anhand der euklidischen Distanz d_{ij} der beiden Merkmalsvektoren bestimmt. In diesem Modell wird somit der Interaktionsmerkmalsvektor $\mathbf{f}_{ij}(\mathbf{x})$ durch ein einziges Merkmal d_{ij} repräsentiert. Zur Berechnung des Potentials gilt dann

$$\log \psi_{ij}(\mathbf{x}, y_i, y_j) = \delta_K(y_i = y_j) \cdot \left(\theta + (1 - \theta) \cdot \exp\left(-d_{ij}^2 / 2\sigma^2\right)\right). \tag{3.11}$$

Die relativen Gewichte zwischen dem datenabhängigen und dem datenunabhängigen Glättungsterm in Gleichung 3.11 wird über den Parameter $\theta \in [0; 1]$ gesteuert. Wenn θ auf den Wert null gesetzt wird, beeinflussen ausschließlich die Daten den Grad der Glättung. Bei ähnlichen Merkmalsvektoren beider Knoten ist die euklidische Distanz d_{ij} klein und die Zugehörigkeitswahrscheinlichkeit beider Knoten i und j zur gleichen Objektklasse groß; folglich wird stark geglättet. Auf der anderen Seite ergibt sich für $\theta = 1$ ein einfaches Potts Modell, welches unabhängig von den Daten glättet. Der Parameter σ repräsentiert die mittlere quadratische Distanz der Knotenmerkmalsvektoren. Er wird anhand von Trainingsdaten beim Aufbau des Graphen bestimmt. Im Gegensatz zu einer Glättung mit einem reinen Potts Modell können durch die Berücksichtigung der Daten feine Strukturen in der Szene detektiert werden, wenn sie verglichen zu den Nachbarprimitiven Unterschiede in ihren Merkmalen aufweisen. Viele auf CRF basierende Anwendungen versuchen einen Wahrscheinlichkeitswert zu ermitteln, mit dem an beiden Endknoten einer Kante identische Klassenlabels vorliegen. Anstelle eines Potts Modells kann entsprechend $\psi(\mathbf{x}, y_i, y_j) \propto p(y_i = y_j | \mathbf{f}_{ij}(\mathbf{x}))$ also auch ein beliebiger diskriminativer Klassifikator zur Bestimmung der a posteriori Wahrscheinlichkeiten für eine Labelgleichheit herangezogen werden [Kumar & Hebert, 2006].

Für die Verwendung eines komplexeren Interaktionsmodells lässt sich die Beschränkung auf identische Klassenlabels aufheben, so dass das Prinzip zur Bestimmung der optimalen Klassenkonfiguration im Allgemeinen übertragen werden kann. Es wird als *generisches Modell* bezeichnet und führt üblicher-



(a) P^n Potts Modell (b) robustes P^n Potts Modell

Abbildung 3.4: Verhalten der Energiefunktion beim a) starren und b) robusten P^n Potts Modell (nach [Kohli et al., 2009]). Während der Clique bei a) bereits bei einem einzigen vom dominanten Label abweichenden Knoten die maximale Energie zugeordnet wird, erfolgt bei b) ein linearer Anstieg bis zu einem gewissen Grad q an Abweichungen.

weise verglichen zu den zuvor beschriebenen assoziativen Modellen zu einer weniger starken Glättung. In diesem Fall wird die Verbundwahrscheinlichkeit $\psi(\boldsymbol{x}, y_i, y_j) = p(y_i, y_j|\mathbf{f}_{ij}(\mathbf{x}))$ für jede Konfiguration der beiden Knotenlabels y_i und y_j bei einem gegebenen Kantenmerkmalsvektor $\mathbf{f}_{ij}(\mathbf{x})$ bestimmt. Diese ergibt sich z. B. aus einem während des Trainings angelernten Klassifikator, der somit für unbekannte Daten die wahrscheinlichste Klassenkonfiguration ermittelt. Folglich lässt sich ausdrücken, dass einige Klassenrelationen mit einer größeren Wahrscheinlichkeit bei gegebenen Daten gemeinsam auftreten als andere. Der Nachteil dieses Verfahrens ist die größere Anzahl der zu ermittelnden Parameter.

Potentiale höherer Ordnung: HOP ($\psi_h(\mathbf{x}_h, y_h)$ in Gleichung 3.9) modellieren Zusammenhänge von mehr als zwei Variablen in einer Clique. Assoziative paarweise Interaktionsmodelle führen insbesondere an Objektgrenzen zu einer Überglättung und erschweren so die Detektion feiner Strukturen [Kohli et al., 2007; Schindler, 2012]. Viele Anwendungen zeigen, dass HOP die Ergebnisse einer Klassifikationsaufgabe verbessern können [Kohli et al., 2009; Najafi et al., 2014]. Der wesentliche Nachteil ist jedoch, dass die Inferenz bei der Verwendung solcher Potentiale sehr rechenaufwändig und somit schwierig wird. Dies ist besonders dann der Fall, wenn einerseits generische Interaktionen zwischen den Cliquen modelliert und andererseits große Cliquen bestehend aus vielen Knoten betrachtet werden sollen, wie es z. B. bei Segmenten einer Laserpunktwolke der Fall wäre. Die Entwicklung geeigneter Inferenzmethoden für komplexe HOP stellt ein aktuelles Forschungsthema dar [Pham et al., 2015].

Das sogenannte P^n Potts Modell [Kohli et al., 2007] repräsentiert eine Untergruppe der HOP, für die sich die Inferenz dank einer speziellen Formulierung der Potentiale effizient durchführen lässt. Es handelt sich um ein assoziatives Verfahren, welches ähnlich dem paarweisen Potts Modell für alle Knoten einer Clique das gleiche Klassenlabel bevorzugt. In Abbildung 3.4a ist das Verhalten als Energiefunktion E gezeigt. Die minimale Energie wird einer Clique zugeordnet, wenn alle enthaltenen Knoten ein und dasselbe Klassenlabel annehmen. Sobald mindestens ein Knoten mit einem abweichenden Label auftritt, wird diese Clique maximal bestraft. Das Potential ergibt sich durch $\psi_h(y_h) \propto \exp(-E(y_h))$. In der Weiterentwicklung [Kohli et al., 2009] wird eine robuste Variante vorgestellt, die Abweichungen vom dominanten Klassenlabel in einer Clique bestraft. Das Modell bevorzugt eine Clique h mit Kardinalität $|h|$ abhängig vom Anteil der beteiligten Knoten, die ein von der dominanten Objektklasse abweichendes Label aufweisen. Dieser Anteil wird als $N_i(\mathbf{y}_h)$ bezeichnet, wobei $N_i(\mathbf{y}_h) = \min_l(|h| - n_l(\mathbf{y}_h))$ gilt.

Die Anzahl der zu einer spezifischen Klasse l zugeordneten Variablen ist gegeben durch $n_l(\mathbf{y}_h)$. Der Sättigungsparameter q des robusten Modells legt fest, wie groß der Anteil der abweichenden Klassenlabels $N_i(\mathbf{y}_h)$ innerhalb einer Clique sein darf, bevor dieser die maximalen Strafenergie γ_{max} zugeordnet wird (Abbildung 3.4b). Ist $N_i(\mathbf{y}_h)$ kleiner als q, wird die Energie entsprechend Gleichung 3.12 linear reduziert:

$$\log \psi_h(\mathbf{x}_h, \mathbf{y}_h) \propto \begin{cases} -N_i(\mathbf{y}_h)\frac{1}{q}\gamma_{max} & \text{falls } N_i(\mathbf{y}_h) \leq q, \\ -\gamma_{max} & \text{sonst} \end{cases} \tag{3.12}$$

Die Festlegung des Modells erfordert die Definition des Sättigungsparameters q sowie einen Wert für die maximale Bestrafung γ_{max}. Letzterer kann beispielsweise anhand einer Bewertungsfunktion zur Beschreibung der Qualität einer einzelnen Clique auf Grundlage der Merkmale $\mathbf{x}_h$ bestimmt werden [Kohli et al., 2009], wodurch sich in $\psi_h(\mathbf{x}_h, \mathbf{y}_h)$ die Abhängigkeit von den Beobachtungen erklärt. Cliquen lassen sich z. B. durch die Anwendung von Segmentierungsalgorithmen generieren, wobei jedes Segment eine Clique repräsentiert.

Die effiziente Lösung der HOP dieser Art basiert auf der Transformation der Cliquen höherer Ordnung in paarweise Interaktionen mit Inferenz durch das Graph Cuts Verfahren [Kohli et al., 2007, 2009]. Dabei muss die Energiefunktion die Submodularitätsbedingung erfüllen, was zu Einschränkungen hinsichtlich der Modellierungsmöglichkeiten von Cliquenpotentialen führt. Auf die Inferenz mittels Graph Cuts wird im nächsten Abschnitt eingegangen. Es kann festgehalten werden, dass das robuste P^n Potts Modell immer ein dominantes Klassenlabel innerhalb der Clique fördert, auch wenn dieses Label aufgrund der Daten nicht unbedingt das korrekte für diese Clique ist. Weiterhin können sowohl Interaktionen zwischen den Cliquen als auch generische Relationen einzelner Knoten innerhalb einer Clique auf diese Weise nicht repräsentiert werden.

Inferenz

Die Inferenz ermittelt für alle Knoten des CRF simultan die wahrscheinlichste Konfiguration der Klassenlabels $\mathbf{y}$. Dies erfolgt durch die Maximierung der Wahrscheinlichkeiten bzw. durch die Minimierung der Energiefunktion. Sobald im Zuge einer Klassifikation mehr als zwei Klassen unterschieden werden sollen, wird das in Gleichung 3.9 formulierte Problem NP-hart und kann nur approximativ gelöst werden [Kolmogorov & Zabih, 2004]. Es gibt eine Reihe von Verfahren zur Durchführung der Inferenz. Hier sollen die beiden am häufigsten verwendeten und auch in dieser Arbeit genutzten Methoden Graph Cuts und Loopy Belief Propagation kurz vorgestellt werden. Alternativen sind z. B. das auf Sampling basierende *Simulated Annealing* [Kirkpatrick et al., 1983] oder *Iterative Conditional Modes* (ICM) [Besag, 1986].

Graph Cuts: Viele Anwendungen im Bereich der Computer Vision lassen sich als Energieminimierungsaufgaben formulieren [Kolmogorov & Zabih, 2004]. Ein weit verbreitetes Verfahren zur Lösung solcher Probleme basiert auf Graph Cuts [Boykov et al., 2001]. Für eine binäre Klassifikation wird der Graph des CRF um zwei weitere, als Quelle s und Senke t bezeichnete Knoten erweitert. Diese beiden Terminal-Knoten repräsentieren jeweils eine Klasse und sind anfangs über Kanten mit allen Primitiven verbunden. Durch eine Invertierung der Interaktionspotentiale ergeben sich *Kosten* für die

Kanten zwischen den Primitiven. Analog werden die Kosten der Kanten zur Quelle bzw. Senke aus den unären Potentialen der Primitive abgeleitet. Graph Cuts ordnet jeden Knoten v_i entweder s oder t zu, um auf diese Weise eine binäre Klassifikation in die beiden disjunkten Teilmengen *Vordergrund* und *Hintergrund* vorzunehmen. Der Zusammenhang ist in Abbildung 3.5 verdeutlicht. Die Zuordnung erfolgt anhand eines Schnitts $\mathcal{S}$, für den die Summe der Kosten aller entfernten Kanten minimal ist. Zur Berechnung kommt meistens der *Max-Flow-Min-Cut*-Algorithmus [Boykov & Kolmogorov, 2004] zum Einsatz. Er basiert auf dem Theorem von Ford & Fulkerson [1962], welches das Ermitteln der minimalen Schnittkosten dem Auffinden des maximalen Flusses in einem Flussnetzwerk gleichstellt. Damit Graph Cuts angewendet werden kann, muss das Interaktionspotential der Submodularitätsbedingung $\log\psi(-1,+1) + \log\psi(+1,-1) \leq \log\psi(-1,-1) + \log\psi(+1,+1)$ genügen. Diese Bedingung stellt eine Einschränkung dar, wenn Graph Cuts für die Inferenz von graphischen Modellen verwendet werden soll. Im binären Fall kann die Einhaltung für das Ising Modell, welches das Pendant zum Potts Modell im Mehrklassenfall darstellt, garantiert werden [Ivănescu, 1965]. Für ein Zweiklassenproblem lässt sich auf diese Weise eine exakte Lösung in Form des globalen Minimums von $p(\mathbf{y}|\mathbf{x})$ finden.

Um das binäre Verfahren auf ein Mehrklassenproblem zu übertragen, muss dieses auf mehrere Zweiklassenfälle abgebildet werden. Dies ist jedoch mit dem Nachteil verbunden, dass sich die Lösung nicht mehr exakt bestimmten lässt [Boykov et al., 2001]. Die Minimierung der Energie erfolgt im Mehrklassenfall meist über eine der beiden von Boykov et al. [2001] vorgeschlagen Varianten α-*Expansion* (welche auf die Arbeiten von Veksler [1999] zurückzuführen sind) und α-β-*Swaps*. Iterativ werden bei beiden Verfahren die Klassenzugehörigkeiten der Knoten v_i variiert, um die minimalen Kosten für einen Schnitt zu identifizieren. Beide Verfahren sind sehr effizient, da sie den gleichzeitigen Wechsel der Labels für eine große Anzahl von Knoten ermöglichen. Bei α-*Expansion* wird in jeder Iteration im Zuge einer binären Klassifikation jeder Knoten dahingehend überprüft, ob er der Klasse α zuzuordnen ist. Falls nicht, behält er sein aktuelles Label. Dies wird für jede Klasse α wiederholt. α-β-*Swaps* hingegen erlaubt es, in einem Iterationsschritt allen Knoten einer Klasse α die Klasse β anzunehmen. Verringert sich die Energie durch den Austausch, wird die neue Konfiguration angenommen. Dieser Vorgang muss für jedes mögliche Paar an Klassen (α, β) wiederholt werden.

Auch im Mehrklassenfall muss die Submodularitätsbedingung erfüllt werden. Das (kontrast-sensitive) Potts Modell entspricht diesen Anforderungen. Für komplexere Interaktionen wie z. B. das generische Modell kann dies jedoch nicht garantiert werden. In einem solchen Fall ist Graph Cuts als Inferenzverfahren ungeeignet. Falls die Anforderungen jedoch erfüllt werden, lassen sich mit Graph Cuts gute Ergebnisse erzielen [Szeliski et al., 2008]. Insbesondere im Zusammenhang mit robusten P^n Potts Modellen haben sie sich durch die von Kohli et al. [2009] vorgestellten Arbeiten als Standard etabliert, da sich über Graph Cuts die auf diese Weise formulierte Energie von Cliquen höherer Ordnung effizient in niedriger polynomischer Komplexität optimieren lässt. Eine Zusammenfassung über mittels Graph Cuts lösbare Energiefunktionen ist z. B. in [Kolmogorov & Zabih, 2004] zu finden.

Loopy Belief Propagation (LBP): LBP gehört zur Gruppe der Message-Passing-Algorithmen und approximiert die optimale Labelkonfiguration in Graphen, welche auch Zyklen enthalten können [Frey & MacKay, 1998]. Die ursprüngliche Version der Belief Propagation war zunächst nur auf planare und gerichtete Graphen beschränkt [Pearl, 1982]. Die Grundidee besteht aus den beiden Komponenten *Senden* und *Empfangen* von Messages, welche iterativ wiederholt werden, bis die beste Labelkon-

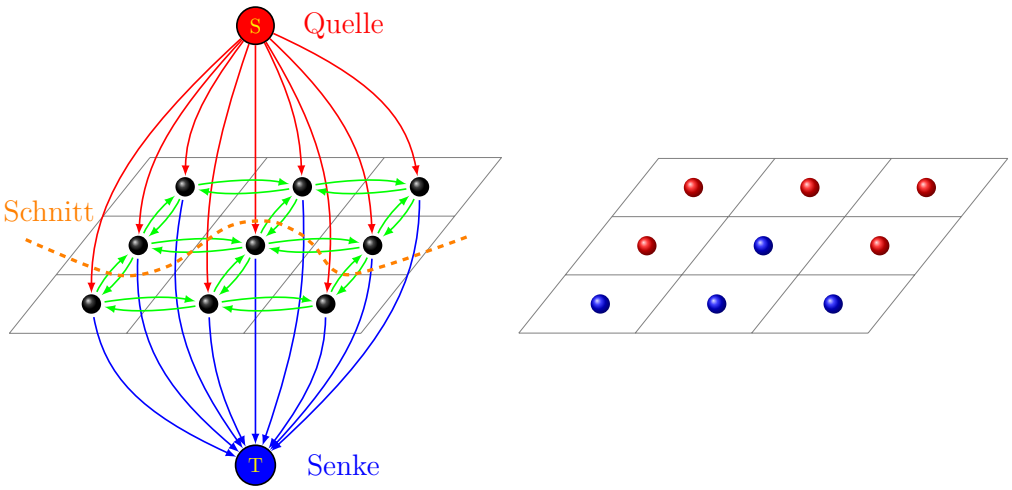


Abbildung 3.5: Graph Cuts: In der Abbildung links wird der Aufbau des Graphen $\mathcal{G}$ veranschaulicht. Jeder Bildknoten (schwarz) ist über Kanten mit seinen Nachbarknoten sowie mit den beiden Terminal-Knoten s und t verbunden. Der Schnitt zur Einteilung in zwei disjunkte Teilgraphen erfolgt entlang des Pfades, bei dem die geringsten Kosten anfallen. Die rechte Abbildung zeigt das resultierende klassifizierte Bild nach Durchführung des Graph Cuts-Verfahrens. Jedes Pixel ist entweder mit der Quelle oder mit der Senke verbunden. Diese Terminal-Knoten repräsentieren die beiden Objektklassen.

figuration gefunden wurde. Dies ist nach einer festgelegten Iterationsanzahl oder bei Erreichen eines Konvergenzkriteriums der Fall. Beim *Senden* schicken die Knoten Nachrichten an alle adjazenten Kanten. Anschließend werden beim *Empfangen* für jeden Knoten die eintreffenden Nachrichten über die Kanten ausgewertet und zur neuen Berechnung ihrer Konfidenzen (engl.: beliefs) verwendet. Es gibt dabei zwei wesentliche Berechnungsvarianten:

1. *Sum-Product*: Diese Methode ermittelt die Randverteilungen der Klassenlabels jedes Knotens innerhalb des graphischen Modells, welche durch die Konfidenzwerte repräsentiert werden [Szeliski et al., 2008]. Die Sum-Product-Variante führt bei baumartigen Graphstrukturen (ohne Zyklen) zur exakten Lösung [Frey & MacKay, 1998; Bishop, 2006].
2. *Max-Product*: Bei der von Kolmogorov [2006] vorgestellten Version wird die a posteriori Wahrscheinlichkeit durch die Konfidenzwerte approximiert. Wird aus numerischen Gründen mit dem Logarithmus der Wahrscheinlichkeiten gearbeitet, so bezeichnet man das Verfahren als *Max-Sum*-Methode [Bishop, 2006].

Bei LBP handelt es sich im Allgemeinen um eine approximative Inferenzmethode ohne Garantie zur Konvergenz bzw. zum Auffinden der optimalen Lösung bei Graphen mit Zyklen. Dennoch hat es sich neben den Graph Cuts als Standardverfahren etabliert und führt in vielen Anwendungen zu guten Ergebnissen. Bei einem von Szeliski et al. [2008] durchgeführten Vergleich verschiedener Inferenzverfahren zur Energieminimierung eines MRFs gehört LBP zu den am besten abschneidenden Methoden bei Computer Vision Anwendungen. Die Methode ist nicht auf submodulare Potentiale beschränkt, was einen wesentlicher Vorteil im Vergleich zu Graph Cuts darstellt.

Training

Bei einem CRF handelt es sich um ein überwachtes Klassifikationsverfahren. Aus diesem Grund ist das Anlernen der Klassifikatoren für die Potentialberechnungen sowie der weiteren Parameter erforderlich. Eine notwendige Bedingung sind dabei repräsentative Trainingsdaten mit Referenzlabels. Da über das graphische Modell zusätzlich zu der knotenweisen Klassifikation auch die Beziehungen zwischen benachbarten Knoten berücksichtigt werden, reichen einzelne gelabelte Objekte in diesem Fall nicht aus. Vielmehr muss ein vollständig mit Referenzlabels zugeordneter Teilausschnitt einer Szene herangezogen werden, um den Kontext entsprechend anlernen zu können.

Die Klassifikatoren lassen sich abhängig von der Formulierung der Potentiale gemeinsam (z.B über die Broyden-Fletcher–Goldfarb-Shanno-Methode (BFGS) [Liu & Nocedal, 1989]) oder getrennt voneinander antrainieren [Shotton et al., 2009]. In dieser Arbeit werden sie zunächst einzeln angelernt, anschließend erfolgt das Optimieren der relativen Potentialgewichte $\mathbf{\Omega} = [\omega^p, \omega^h]^T$. Für diesen Zweck eigenen sich Ansätze wie Kreuzvalidierung oder direkte Suchmethoden [Hooke & Jeeves, 1961; Kramer, 2010], da diese keine Berechnung der Gradienten erfordern. Das letztgenannte Verfahren findet in dieser Arbeit Anwendung.

Eine Optimierung durch *direkte Suche* umfasst die sequentielle Untersuchung verschiedener temporärer Lösungen, welche hinsichtlich eines Qualitätsmaßes jeweils mit der aktuell besten Lösung verglichen werden. Dies beinhaltet auch die Strategie zur Bestimmung der nächsten temporären Lösung als Funktion der vorherigen Ergebnisse [Hooke & Jeeves, 1961]. Die Methode arbeitet gradientenfrei und bietet somit den Vorteil, auch für Zielfunktionen ohne berechenbare Gradienten eine Näherung an die beste Lösung finden zu können. Dabei wird die Anzahl der korrekt klassifizierten Knoten maximiert. Das Auffinden der optimalen Parameter ist jedoch nicht garantiert [Hooke & Jeeves, 1961; Kramer, 2010]. Bei der lokalen und iterativen Suche wird zunächst eine Lösung mittels manuell vorgegebenen Näherungsparametern erzielt. Anschließend erfolgen Wiederholungen der Berechnung, indem jeweils ein Gewicht um eine Schrittweite $\pm\Delta\omega$ variiert wird, während die anderen Parameter konstant bleiben. Sollte sich das Ergebnis in einem der beiden Fällen verbessern, dient dieser Wert als Initialisierung des Parameters für die nächste Iteration. Wenn sich durch Modifikation eines Parameters kein Genauigkeitsgewinn erzielen lässt, ist die Abbruchbedingung erreicht. In diesem Fall wird der nächste Parameter evaluiert. Dies wird solange wiederholt, bis jeder Parameter überprüft wurde und sich während eines Durchlaufs keine Komponente von $\mathbf{\Omega}$ mehr ändert.

3.4 Segmentierung von Punktwolken

In dieser Arbeit wird das Segmentierungsverfahren *Voxel Cloud Connectivity Segmentation* (VCCS) [Papon et al., 2013] zur Einteilung einer Punktwolke in sogenannte Supervoxel verwendet. Häufig wird unter dem Begriff Supervoxel die einfache Erweiterung von 2D-Superpixel-Algorithmen auf 3D-Volumen verstanden. Diese ergeben sich in vielen Fällen aus einer Bildsequenz, so dass die Zeit als dritte Dimension fungiert. Der Datenbestand wird jedoch immer noch durch regelmäßig angeordnete Pixel des Bildes repräsentiert. VCCS hingegen führt die Segmentierung unmittelbar auf den unregelmäßig im $\mathbb{R}^3$ verteilten Daten wie Punktwolken durch. Dabei wird explizit berücksichtigt, dass diese Daten

weder regelmäßig angeordnet noch gleichmäßig in der Szene verteilt sind [Papon et al., 2013].

Es handelt sich um eine Variante der unüberwachten k-means Segmentierung mit zwei wesentlichen Unterschieden: Zum einen wird der gesamte 3D-Raum über einen Octree partitioniert, um in den Daten eine gleichmäßige Verteilung der Saatpunkte für die Supervoxel zu gewährleisten. Der 3D-Raum wird zunächst durch die Voxel des Octrees aufgeteilt, welche im nächsten Schritt über ein Regionenwachstum zu den Supervoxeln zusammengefasst werden. Im Vergleich zu global arbeitenden Verfahren ermöglicht diese Partitionierung eine effiziente Prozessierung auch großer Szenen [Papon et al., 2013]. Zum anderen berücksichtigt VCCS bei der Aggregation die Nachbarschaften. Dadurch ist sichergestellt, dass sich ein Supervoxel nicht über Objektgrenzen hinaus ausdehnen kann, wenn sich die Objekte im $\mathbb{R}^3$ nicht berühren. Eine räumliche Verbundenheit benachbarter und mit Daten belegter Voxel wird somit gefordert. Bei einem standardmäßigen k-means ist ein solches Verhalten nicht gewährleistet. Insbesondere durch den zweiten Unterschied eignet sich VCCS gut für die Anwendung in dieser Arbeit, da ebenfalls eine lokale Nachbarschaft und somit Kontext bei der Segmentierung berücksichtigt wird.

Das von Papon et al. [2013] vorgestellte Prinzip lässt sich mit folgenden Schritten zusammenfassen: Grundlegend für das Verfahren sind die Informationen über benachbarte Voxel. Daher wird zur Modellierung der Adjazenzen zu Beginn ein Octree mit einer 26er-Nachbarschaft und einer Voxelauflösung von R_{vox}^{SV} aufgebaut. Anschließend werden die Saatpunkte der Supervoxel bestimmt. Ihre Anzahl und Positionen ergeben sich aus der Auflösung der Saatpunkte (R_{saat}^{SV}), einem weiteren manuell festgelegten Wert. Über ein iteratives und lokales k-means Verfahren werden die Supervoxel nun erweitert, falls die benachbarten Voxel ähnliche Eigenschaften aufweisen wie das Zentrum des aktuellen Supervoxels. Von allen direkten Nachbarn wird immer nur dasjenige Voxel dem Supervoxel zugeordnet, bei dem ein Ähnlichkeitskriterium (D) maximal ist. Dessen Nachbarn werden für die folgenden Betrachtungen als weitere Ebene eines Suchbaums für dieses Supervoxel hinzugefügt. Anschließend wird - einer Breitensuche im Octree folgend - mit dem nächsten Supervoxel fortgefahren. Auf diese Weise wachsen alle Supervoxel mit der gleichen Geschwindigkeit, bis jedes Voxel einem Supervoxel zugeordnet ist. Die Parameter sind in Abbildung 3.6 veranschaulicht.

VCCS wurde von Papon et al. [2013] ursprünglich für die Segmentierung von Daten einer RGB-D Tiefenbildkamera entwickelt, bei der Punktwolken mit Farbinformation entstehen. Bei der in der vorliegenden Dissertation verwendeten Implementierung[1] erfolgt die Zuordnung zu einem Supervoxel auf der Grundlage dreier Komponenten:

$$D = \sqrt{\frac{\omega_s^{SV} D_s^2}{R_{saat}^{SV}} + \omega_n^{SV} D_n^2 + \omega_{rgb}^{SV} D_{rgb}^2} \tag{3.13}$$

In Gleichung 3.13 setzt sich das Ähnlichkeitskriterium D aus der räumlichen Nähe (D_s) (normalisiert mit der Auflösung der Saatvoxel R_{saat}^{SV}), dem Winkel zwischen Normalenvektoren (D_n) sowie der Übereinstimmung in Bezug auf die RGB-Farbwerte benachbarter Voxel zusammen (D_{rgb}). Der letztgenannte Term ergibt sich als euklidische Distanz im RGB-Farbraum.

Der relative Einfluss der Terme zueinander lässt sich über die manuell vorzugebenden Gewichte

[1] Verwendet wird die Implementierung der Point Cloud Library (PCL), Version 1.8.0, `http://pointclouds.org/`

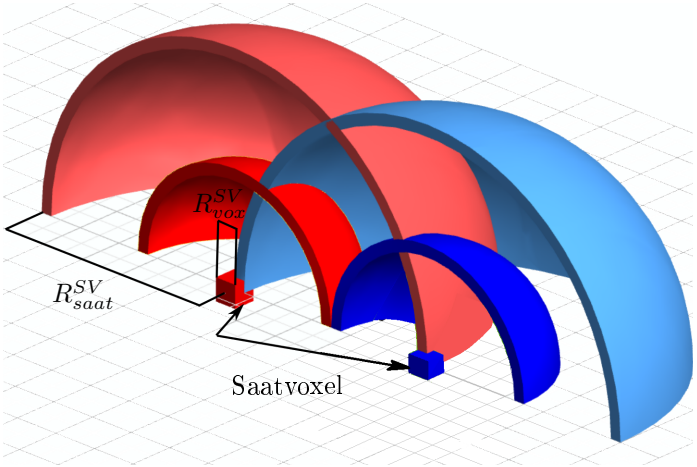


Abbildung 3.6: Parameter zur Erstellung der Supervoxel, entnommen aus [Papon et al., 2013] und bearbeitet. Die Saatvoxel mit der Auflösung R^{SV}_{vox} dehnen sich unter Berücksichtigung des Ähnlichkeitskriteriums D maximal bis zu einer Auflösung R^{SV}_{saat} aus.

ω^{SV}_s, ω^{SV}_n und ω^{SV}_{rgb} steuern. Dabei gilt $\omega^{SV}_s + \omega^{SV}_n + \omega^{SV}_{rgb} = 1$. Ein großes Gewicht korrespondiert mit einem starken Einfluss der jeweiligen Komponente. Hervorzuheben ist insbesondere ω^{SV}_s, über das die Form der resultierenden Supervoxel gesteuert werden kann. Während ein hohes Gewicht zu gleichmäßigen würfelförmigen Supervoxeln führt, passt sich die Form der Voxel bei einem niedrigen Gewicht der lokalen Gegebenheiten entsprechend der anderen beiden Kriterien an. In diesem Fall sind die resultierenden Voxel weniger regelmäßig.

Für das in dieser Arbeit entwickelte Verfahren wird in Abschnitt 4.3 eine neue Variante von VCCS vorgestellt, welche bei der Segmentierung die Berücksichtigung von Konfidenzwerten aus einer zuvor durchgeführten Klassifikation erlaubt.

4 Hierarchische kontextbasierte Punktwolkenklassifikation

Dieses Kapitel stellt den Kern der vorliegenden Arbeit dar und beschreibt das neu entwickelte Verfahren zur Klassifikation der Punktwolken unter Berücksichtigung von Kontext. Dazu wird zu Beginn in Abschnitt 4.1 eine Übersicht über die Methode gegeben, bevor die detaillierte Beschreibung der drei wesentlichen Komponenten des Verfahrens folgt. Diese sind die punktbasierte Klassifikation (Abschnitt 4.2), die Bildung von Segmenten (Abschnitt 4.3) sowie die Klassifikation auf Segmentebene zur Integration von Kontext aus einem anderen Skalenbereich (Abschnitt 4.4). Eine theoretische Diskussion des Ansatzes in Abschnitt 4.5 schließt dieses Kapitel.

4.1 Konzept des hierarchischen Ansatzes

Die in dieser Arbeit vorgestellte Methodik verfolgt zwei wesentliche Zielsetzungen, um die Objektlabels für jeden Laserpunkt bestmöglich zu bestimmen. Zum einen soll Kontextinformation in zwei unterschiedlichen Skalenbereichen bei der Klassifikation berücksichtigt werden. Da im Zuge der Betrachtung eines gröberskaligen Bereichs basierend auf Segmenten mit Generalisierungsfehlern zu rechnen ist (Abschnitt 2.2), soll das zweite Augenmerk auf der Detektion feinerer Strukturen in den Daten liegen. Insbesondere bei den in den Daten weniger stark vertretenen Objektklassen kann es vorkommen, dass diese Details durch die Glättung verloren gehen. Die Absicht ist es also, diese Strukturen trotzdem zuverlässig identifizieren zu können.

Zur Realisierung wird ein hierarchisches und probabilistisches Verfahren vorgestellt, das aus zwei iterativ miteinander interagierenden Ebenen besteht. In jeder Ebene wird eine Klassifikation mittels eines CRF durchgeführt, wobei in beiden Fällen dieselbe Klassenstruktur $\mathcal{L}$ zugrunde liegt. Der Unterschied besteht im Wesentlichen in der Art der zu klassifizierenden Entitäten. In der unteren Ebene (CRF^P) repräsentieren die 3D-Laserpunkte die Knoten des Graphen, in der oberen (CRF^S) sind es aus dem Ergebnis von CRF^P abgeleitete Segmente. Als Notation wird im Folgenden der hochgestellte Index P verwendet, um auf die punktweise Ebene zu verweisen. Der Index S hingegen indiziert Komponenten der Segmentebene. Der Aufbau des Ansatzes ist in Abbildung 4.1 skizziert. Die Aufgabe des punktweisen CRF^P ist es, lokalen Kontext aus der unmittelbaren Nachbarschaft in die Klassifikation einzubeziehen. Hierbei sollten möglichst feine Strukturen erkannt werden, welche im Zuge einer direkten Segmentierung verloren gehen könnten. Als Interaktionsmethodik wird das robuste P^n Potts Modell verwendet. Zusätzlich zu den paarweisen Interaktionen werden über die HOP die Ergebnisse mehrerer Vorsegmentierungen mit verschiedenen Auflösungsstufen berücksichtigt, so dass sich auch hierbei unterschiedliche Skalenbereiche auf einem lokalen Niveau abdecken lassen. Über eine Optimie-

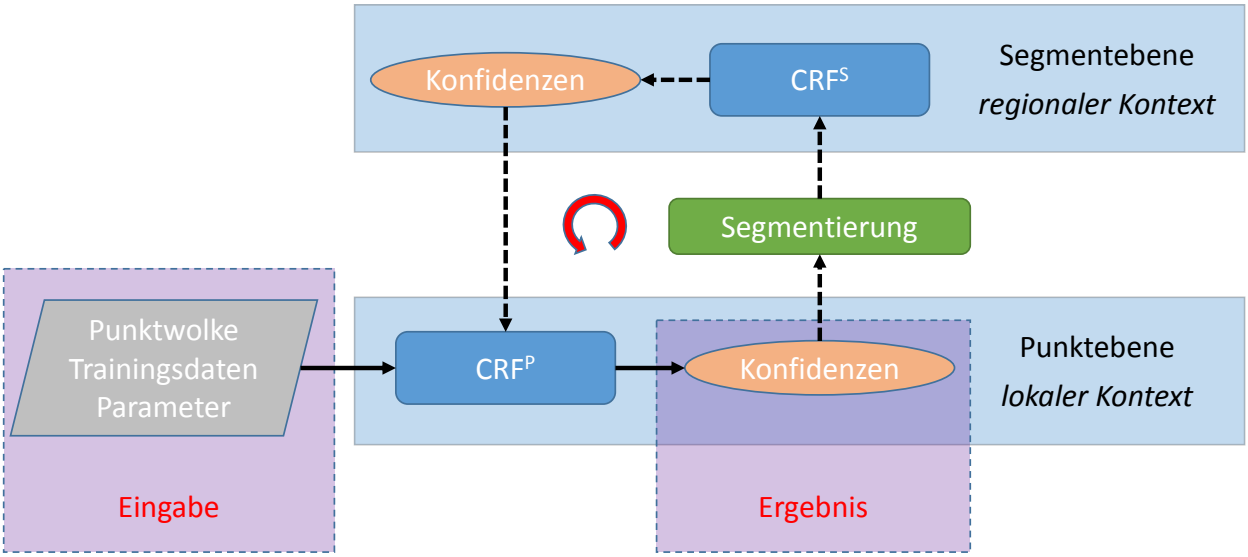


Abbildung 4.1: Ablaufdiagramm des vorgestellten hierarchischen Verfahrens. Die zu klassifizierende Punktwolke, Trainingsdaten in Form von voll ausgelabelten Trainingspunktwolken sowie die Parameter werden zur Durchführung benötigt. Zunächst werden die punktweisen Merkmale extrahiert, und jedem 3D-Punkt wird in CRF^P ein Klassenlabel zugeordnet. Dieser Schritt berücksichtigt Kontext in einer lokalen Umgebung. Die Ergebnisse werden danach an die nächste Stufe weitergegeben. Diese besteht aus der Segmentierung und Klassifikation der Segmente in CRF^S. Die sich ergebenden Konfidenzen, welche auf regionalem Kontext basieren, werden in einem weiteren Iterationsschritt in CRF^P berücksichtigt und sollen die punktweisen Ergebnisse verbessern. Das Alternieren zwischen CRF^P und CRF^S lässt sich im Prinzip beliebig oft wiederholen.

rung werden die wahrscheinlichsten Klassenlabels für die Punkte ermittelt. Dank der robusten Variante des P^n Potts Modells wird ein gewisser Anteil an unterschiedlichen Labeln innerhalb einer Clique toleriert. Insgesamt können auf diese Weise Ergebnisse der punktweisen Klassifikation erzielt werden, die im Vergleich zu einem generischen Ansatz zwar eine leichte, datenabhängige Glättung aufweisen, aber dennoch feine Strukturen besser erkennen sollten als ein nicht robustes P^n Potts Modell oder eine unmittelbar durchgeführte Klassifikation von Segmenten.

Aufbauend auf diesen Ergebnissen erfolgt eine Segmentierung basierend auf VCCS [Papon et al., 2013], wobei benachbarte Punkte mit ähnlichen Normalenvektoren und punktbasierten Konfidenzen zu Segmenten zusammengefasst werden. Die Integration der vorherigen Prädiktionen kann zu einer Reduktion der Generalisierungsfehler bei der Segmentbildung führen, da nur Laserpunkte, die mit einer hohen Wahrscheinlichkeit zur selben Klasse gehören, zu einem Segment aggregiert werden. Weiterhin ergibt sich durch die Berücksichtigung der CRF^P-Ergebnisse für jedes Segment eine Vorinformation über die Wahrscheinlichkeitsverteilung der Klassenzugehörigkeit.

In dem als obere Ebene definierten CRF^S repräsentieren die Segmente die Knoten des graphischen Modells. Die Aufgabe ist es nun, alle Interaktionen zwischen den Segmenten aller Klassen zu modellieren und die wahrscheinlichste Konfiguration zu identifizieren. Durch die Verwendung von Segmenten wird ein größerer räumlicher Kontextbereich einbezogen als auf der Punktebene. Zusätzlich

zu den bereits berücksichtigten lokalen Punktmerkmalen lässt sich nun ein weiterer Satz von Merkmalen auf Basis der Segmente verwenden, der hilfreich für die korrekte Klassifikation sein kann. Das Vorhandensein der Klassenvorinformation für die Segmente ermöglicht die Extraktion von zusätzlichen Merkmalen. Im konkreten Fall werden neue Kontextmerkmale entwickelt, welche Distanz und Orientierung der Segmente zur nächstgelegenen Straße modellieren. Dies lässt sich umsetzen, da die Straße in der Regel bereits auf der Punktebene ausreichend genau detektiert wird. Straßen strukturieren urbane Gebiete und bilden die typischen „Achsen“, an denen sich andere Objekte orientieren. In Anlehnung an den *relative location prior* [Gould et al., 2008], welcher ursprünglich für terrestrische optische Aufnahmen mit festgelegter Kameraausrichtung definiert wurde, kann somit auch für den Luftaufnahmefall eine Referenzrichtung festgelegt werden. Eine weitere Gruppe von Kontextmerkmalen analysiert die punktbasierten Zuordnungswahrscheinlichkeiten der Nachbarsegmente, um die gemeinsamen Auftrittswahrscheinlichkeiten der Klassen zu modellieren. Dieser Merkmalssatz identifiziert somit häufig auftretende Nachbarschaften und repräsentiert die Co-Occurrence Information in einer anderen Form als beispielsweise [Kosov et al., 2013; Ladický et al., 2013].

Der Durchlauf von CRF^P und CRF^S lässt sich iterativ wiederholen. Um die Ergebnisse der Segmentklassifikation bei der folgenden punktweisen Klassifikation zu integrieren, werden deren Konfidenzwerte auf die Punkte projiziert und als weiteres Potential in der nächsten Iteration von CRF^P berücksichtigt. Auf diese Weise können etwaige, in der punktbasierten Klassifikation entstandene Fehler für die Segmentierung korrigiert werden. Da dieser zusätzliche Term einen ähnlichen Effekt wie das P^n Potts Modell aufweist, kann in den Folgeiterationen auf die Modellierung des robusten HOP verzichtet werden.

4.2 Punktbasierte Klassifikation

Zunächst wird CRF^P beschrieben, welches die Punktebene des hierarchischen Verfahrens darstellt. Jeder einzelne 3D-Laserpunkt repräsentiert einen Knoten $v_i^P \in \mathcal{V}^P$ des Graphen $\mathcal{G}^P(\mathcal{V}^P, \mathcal{E}^P)$ und ist über Kanten $e_{ij}^P \in \mathcal{E}^P$ mit seinen Nachbarpunkten verbunden. Jedem Knoten soll ein Label y_i^P der Menge $\mathcal{L}$ zugeordnet werden. Die a posteriori Wahrscheinlichkeiten $p^P(\mathbf{y}^P|\mathbf{x})$ von CRF^P ergeben sich aus der Multiplikation von insgesamt vier Termen:

$$p^P(\mathbf{y}^P|\mathbf{x}) = \frac{1}{Z^P(\mathbf{x})} \prod_{i\in\mathcal{V}^P} \varphi^P(\mathbf{x}, y_i) \cdot \prod_{e_{ij}\in\mathcal{E}^P} \psi^P(\mathbf{x}, y_i, y_j)^{\omega_\psi^P} \cdot \prod_{h\in\mathcal{H}^P} \chi^P(\mathbf{x}_h, \mathbf{y}_h)^{\omega_\chi^P} \cdot \prod_{i\in\mathcal{V}^P} \xi^P(p_i^S)^{\omega_\xi^P}. \quad (4.1)$$

Die beiden Terme $\varphi^P(\mathbf{x}, y_i)$ und $\psi^P(\mathbf{x}, y_i, y_j)$ in Gleichung 4.1 werden als *unäres* und *paarweises Potential* bezeichnet. $\chi^P(\mathbf{x}_h, \mathbf{y}_h)$ modelliert die *Potentiale höherer Ordnung* für jede Clique $h \in \mathcal{H}^P$. Weiterhin wird ein zusätzlicher unärer Term $\xi^P(p^S)$ eingeführt, der ebenfalls implizit Cliquen höherer Ordnung in Betracht zieht. Er berücksichtigt die Konfidenzwerte p^S der segmentbasierten Klassifikation und integriert diese in das punktweise CRF. Über die Parameter ω_ψ^P, ω_χ^P und ω_ξ^P lassen sich die Potentiale relativ zueinander gewichten. Eine Normalisierungskonstante $Z^P(\mathbf{x})$ skaliert das Produkt der Potentiale und macht es so als Wahrscheinlichkeit interpretierbar. Die einzelnen Terme werden in den nachfolgenden Abschnitten näher erläutert.

4.2.1 Potentiale

Unäres Potential

Für jeden betrachteten Knoten i des Graphen wird die Relation zwischen dem Klassenlabel y_i und seinem Merkmalsvektor $\mathbf{f}_i^P(\mathbf{x})$ über das unäre Potential $\varphi^P(\mathbf{x}, y_i)$ bestimmt. $\mathbf{f}_i^P(\mathbf{x})$ setzt sich dabei aus Merkmalen zusammen, welche aus den unmittelbar an diesem Laserpunkt beobachteten Daten sowie zusätzlich aus denen in einer lokalen Nachbarschaft extrahiert wurden. Die im konkreten Fall verwendeten Merkmale sind in Abschnitt 5.3.1 beschrieben. Das Potential entspricht der Wahrscheinlichkeit von y_i bei gegebenen Daten $\mathbf{x}$, d. h. $\varphi^P(\mathbf{x}, y_i) = p(y_i|\mathbf{f}_i^P(\mathbf{x}))$. In dieser Arbeit wird dafür der von Breiman [2001] vorgeschlagene RF Klassifikator verwendet. Das probabilistische Ergebnis erhält man, indem die Stimmenanzahl der Entscheidungsbäume durch die Anzahl der Bäume normiert wird (Gleichung 3.8). Der RF Klassifikator hat sich als eine geeignete Wahl für die Klassifikation von Laserdaten erwiesen, da er zu guten Ergebnissen führt und effizient mit großen Datenmengen umgehen kann [Chehata et al., 2009; Hackel et al., 2016; Weinmann, 2016]. In der Untersuchung von Niemeyer et al. [2013] wird ein Vergleich von zwei Klassifikationen mittels CRF durchgeführt, deren Potentiale einerseits mit RF und andererseits mit linearen Modellen berechnet werden. Mit einer Steigerung der Gesamtgenauigkeit von zwei Prozentpunkten schneiden RF dabei besser ab.

Wie bereits in Abschnitt 3.2 beschrieben wurde, müssen für den im Folgenden als RF^P bezeichneten Klassifikator einige weitere Parameter festgelegt werden. Dies sind insbesondere die Anzahl der Bäume (T), die Anzahl der Trainingsstichproben pro Klasse ($N_{s,l}$) im Training, die Tiefe der Bäume (T_{depth}) sowie die minimale Anzahl von Samples an einem Knoten, um einen neuen Entscheidungstest an dieser Stelle einzuführen (T_{split}). Die Anzahl der zufällig verwendeten Merkmale an einem Knoten (M_{split}) wird entsprechend der üblichen Vorgehensweise als die Quadratwurzel der Dimension des Merkmalsvektors berechnet [Gislason et al., 2006]. Für die Implementierung werden die randomisierten Entscheidungsbäume der Open-Source C++-Bibliothek OpenCV[1] verwendet.

Paarweises Interaktionspotential

Der Term $\psi^P(\boldsymbol{x}, y_i, y_j)$ in Gleichung 4.1 repräsentiert die paarweisen Interaktionen und integriert auf diese Weise Kontextbeziehungen explizit in den Klassifikationsprozess. Er modelliert die gegenseitige Abhängigkeit eines Knotens i und seines adjazenten Knotens j unter Berücksichtigung beider Knotenlabels sowie der beobachteten Daten $\mathbf{x}$. An dieser Stelle wird für CRF^P ein kontrast-sensitives Potts Modell analog zu Gleichung 3.11 verwendet, welches die Ähnlichkeit beider adjazenter Merkmalsvektoren über deren euklidische Distanz $d_{ij} = ||\mathbf{f}_i^P(\mathbf{x}) - \mathbf{f}_j^P(\mathbf{x})||$ darstellt:

$$\log \psi^P(\mathbf{x}, y_i, y_j) = \delta_K(y_i = y_j) \cdot \left(\theta + (1 - \theta) e^{-\frac{d_{ij}^2}{2\sigma^2}} \right) \tag{4.2}$$

Für die vorliegende Anwendung ist eine möglichst geringe Glättung wünschenswert, um auch die kleineren Strukturen korrekt klassifizieren zu können. Daher ist die ausschließliche Betrachtung der

[1] http://opencv.org/, Version 3.0.0

Daten zur Regulierung der Glättungsstärke heranzuziehen, indem die datenunabhängige Glättung in Gleichung 4.2 über $\theta = 0$ deaktiviert wird. Ein deutlicher Glättungseffekt wird somit nur erzielt, sofern die Daten dies durch ähnliche Merkmalsvektoren adjazenter Knoten indizieren. Dem Modell liegt die Annahme zugrunde, dass benachbarte Laserpunkte ähnliche Merkmale aufweisen, falls diese zur gleichen Objektklasse gehören.

Eine Alternative zur Modellierung der Interaktionen stellt ein generischer Ansatz dar, welcher für jede Klassenkombination die Verbundwahrscheinlichkeit ermittelt. Dies hat zur Folge, dass nicht notwendigerweise ein Glättungseffekt eintritt und das Klassifikationsergebnis dementsprechend recht inhomogen sein kann. In [Niemeyer et al., 2015] hat sich gezeigt, dass ein solches Verhalten an dieser Stelle eher ungeeignet ist, da sich bei einem zu inhomogenen Ergebnis die Schwierigkeit ergibt, aussagekräftige Segmente basierend auf den Punktlabels zu erstellen (Abschnitt 4.3). Insbesondere die zahlreichen Einzelpunkte mit einem anderen Klassenlabel als ihre Nachbarpunkte sind schwierig zu handhaben. Ein leicht glättendes Verfahren unter Berücksichtigung der Daten ist somit vorzuziehen.

Zur Konstruktion der zugrunde liegenden Graphstruktur $\mathcal{G}^P(\mathcal{V}^P, \mathcal{E}^P)$ wird jeder 3D-Punkt als Knoten angesehen und mit seinen k räumlich nächsten Nachbarn über ungerichtete Kanten verbunden, was eine übliche Vorgehensweise ist [Munoz et al., 2008; Shapovalov et al., 2010]. Anders als bei der Merkmalsextraktion wird somit keine optimale Nachbarschaft für jeden Punkt identifiziert, sondern eine feste und vergleichsweise geringe Anzahl von k nächsten Nachbarn herangezogen. Dies lässt sich mit der Aufgabe der paarweisen Interaktionen in CRF^P und dem Zusammenspiel mit den Cliquen höherer Ordnung begründen. Beide Potentiale haben einen datenabhängigen Glättungseffekt. Da bereits die Cliquen der HOP eine etwas größere Nachbarschaft um jeden Punkt berücksichtigen, kann zur Laufzeitverbesserung auf zusätzliche paarweise Interaktionen verzichtet werden. Einige wenige paarweise Kanten sind jedoch hilfreich, um durch den unmittelbaren Vergleich zweier Nachbarpunkte feine Strukturen identifizieren zu können. In Bezug auf die Dimension des Suchraums ist für eine Klassifikation luftgestützter Laserdaten entsprechend Abschnitt 3.1.1 eine 2D-Nachbarschaft gegenüber der 3D-Variante zu bevorzugen [Shapovalov et al., 2010]. Während bei $\mathcal{N}_{3D}$ nur ein sehr begrenzter Ausschnitt der lokalen Geometrie modelliert werden kann, repräsentiert $\mathcal{N}_{2D}$ auch Kanten mit einem deutlichen vertikalen Höhenunterschied zwischen Punkten. So können mittels $\mathcal{N}_{2D}$ z. B. Punkte auf Baumkronen oder Dachkanten mit denen am Boden verbunden werden (Abbildung 3.1, Seite 22), was mit $\mathcal{N}_{3D}$ meistens nicht möglich ist. Diese Information hat sich für die Klassifikation als hilfreich herausgestellt [Shapovalov et al., 2010]. Zum Festlegen der Nachbarschaft für jeden Punkt wird daher ein zweidimensionaler Suchraum verwendet, welcher als $\mathcal{N}_{2D}^k$ bezeichnet wird. Die Kanten des graphischen Modells $\mathcal{G}^P(\mathcal{V}^P, \mathcal{E}^P)$ ergeben sich, indem alle zu $\mathcal{N}_{2D}^k$ gehörenden Punkte mit dem aktuellen Punkt i verbunden werden.

Potentiale höherer Ordnung

Abhängig von der aktuellen Iteration werden HOP auf zwei verschiedene Weisen in das neue Klassifikationsverfahren integriert. Dabei unterscheidet sich die erste Iteration von allen folgenden, da zu diesem Zeitpunkt noch keine Ergebnisse von der segmentbasierten Zuordnung vorliegen und somit wesentliche Komponenten fehlen.

Robustes P^n Potts Modell: Um das Problem der bereits angesprochenen Einzelpunkte mit von ihren lokalen Umgebungen isolierten Objektklassen [Niemeyer et al., 2015] zu reduzieren, wird in Gleichung 4.1 mit dem Term χ^P zusätzlich ein Potential für Cliquen höherer Ordnung eingeführt. Die Inferenz für HOP mit großen Cliquen und die Modellierung generischer Interaktionen ist nicht effizient lösbar. Zum weiteren datenabhängigen Glätten unter Berücksichtigung einer größeren räumlichen Umgebung (verglichen mit denen der paarweisen Interaktionen) wird daher für χ^P ein robustes P^n Potts Modell verwendet. Es ermöglicht eine *weiche* Klassifikation [Kohli et al., 2009] der Daten basierend auf den Cliquen höherer Ordnung. *Weich* bedeutet in diesem Fall, dass Punkte durchaus einer vom dominanten Label abweichenden Klasse zugeordnet werden können, solange deren Anteil nicht zu groß wird und die Merkmalsvektoren dies indizieren. Auf diese Weise können feine Strukturen besser identifiziert werden als bei der nicht-robusten Variante des P^n Potts Modells, da der Überglättungseffekt reduziert wird. Für die Berechnung werden zunächst die Cliquen definiert, welche einzelne Primitive zu Gruppen zusammenfassen. Für diese Aufgabe können beispielsweise Segmentierungsverfahren wie *k-means* [MacQueen, 1967], *Mean Shift* [Comaniciu & Meer, 2002] oder VCCS angewendet werden. Das robuste P^n Potts Modell bietet weiterhin die Möglichkeit, überlappende Cliquen einzubeziehen [Kohli et al., 2009]. Folglich können beispielsweise die Ergebnisse von Segmentierungsverfahren mit verschiedenen Parametereinstellungen zur Definition der Cliquen herangezogen werden. Durch die unterschiedliche Form und Größe der Cliquen wird die Wahrscheinlichkeit erhöht, die vorhandenen Strukturen in den Daten gut beschreiben zu können. Im Allgemeinen ist auch die Nutzung verschiedener Segmentierungsverfahren möglich. Als einfaches Verfahren wird in dieser Arbeit das in Abschnitt 3.4 beschriebene VCCS-Clustering verwendet. Durch die lokale und auf irreguläre 3D-Punktwolken zugeschnittene Funktionsweise arbeitet es deutlich schneller als k-means. Es können mehrere Anwendungen ($N^{P^n}_{nVCCS}$) mit verschiedenen Parametersätzen durchgeführt werden, um die Größe und Form der Cliquen zu variieren.

Für alle auf diese Art definierten Cliquen wird nun das Potential $\chi^P(\mathbf{x}_h, \mathbf{y}_h)$ analog zu Gleichung 3.12 berechnet, bei dem die Bestrafung einer Clique h bis zu einem Sättigungsgrad q linear mit der Anzahl abweichender Klassenlabels $N_i(\mathbf{y}_h)$ ansteigt [Kohli et al., 2009]:

$$\log \chi^P(\mathbf{x}_h, \mathbf{y}_h) \propto \begin{cases} -N_i(\mathbf{y}_h)\frac{1}{q}\gamma_{max} & \text{falls } N_i(\mathbf{y}_h) \leq q, \\ -\gamma_{max} & \text{sonst} \end{cases} \tag{4.3}$$

$$\text{mit } \gamma_{max} = \exp\left(-\| \sum_{i=1}^{|h|} (\mathbf{f}_i(\mathbf{x}) - \boldsymbol{\mu})^2 \| / |h|\right). \tag{4.4}$$

Die maximale Bestrafung γ_{max} (Gleichung 4.4) hängt von der Qualität einer Clique h ab. Um dieses Maß zu quantifizieren, wird in Anlehnung an [Kohli et al., 2009] die Merkmalsvarianz der beteiligten 3D-Punkte herangezogen. Dabei repräsentiert $\boldsymbol{\mu}$ den Mittelwertvektor der Merkmale aller Punkte in h. Diesem Qualitätsmaß liegt die Annahme zu Grunde, dass Punkte mit ähnlichen Merkmalen wahrscheinlich der gleichen Objektklasse angehören. Abweichungen vom dominanten Klassenlabel einzelner Variablen werden in diesem Fall stark bestraft. Auf der anderen Seite lässt sich über Punkte mit variierenden Merkmalsausprägungen innerhalb einer Clique keine zuverlässige Aussage über die Gleichheit

aller Labels treffen - solche Konfigurationen werden dementsprechend weniger stark bestraft. Der Zusammenhang wird in Gleichung 4.4 beschrieben.

Die Berechnung von $\chi^P(\mathbf{x}_h, \mathbf{y}_h)$ findet nur in der initialen Iteration statt. Anschließend wird $\omega_\chi^P = 0$ gesetzt, da verbesserte Cliquen aus der segmentweisen Klassifikation herangezogen und über das im folgenden Abschnitt beschriebene Potential integriert werden können.

Unäres Potential aus der segmentbasierten Klassifikation: Die punktbasierte Klassifikation hat den Vorteil, dass im Gegensatz zu einer gröberskaligen Segmentklassifizierung weniger Generalisierungsfehler verursacht werden, da keine feste Segmentierung erforderlich ist. Andererseits ist durch die Repräsentation der einzelnen Laserpunkte als Knoten ein erhöhter Rechenbedarf erforderlich. Dementsprechend können in den meisten vergleichbaren Ansätzen keine Kontextinformationen aus einem größeren Skalenbereich integriert werden (z. B. [Shapovalov et al., 2010; Weinmann, 2016]). Um dieses Problem zu umgehen, wird in dieser Arbeit ein hierarchischer Ansatz mit gegenseitiger Beeinflussung beider Ebenen vorgestellt. Die Segmentebene CRF^S (Abschnitt 4.4.1) gewährleistet eine größere räumliche Abdeckung und berücksichtigt ein komplexeres Interaktionsmodell im Vergleich zu den Einzelpunkten. Zur Integration der auf diese Weise bestimmten höherwertigen Information wird ein weiteres unäres Potential $\xi^P(p^S)$ in Gleichung 4.1 eingeführt. Konkret berücksichtigt es die *Konfidenzwerte* $p^S(\mathbf{y}^S|\mathbf{x})$, welche in CRF^S als Resultat der Inferenz mit dem Max-Sum-Algorithmus entstehen. Für jedes Segment ist dementsprechend ein solcher Wert für jede Klasse verfügbar. Zum Übertragen auf die punktweise Ebene werden die Konfidenzwerte jedes Segments allen korrespondierenden Mitgliedspunkten zugeordnet. Das Potential ergibt sich aus diesen Konfidenzwerten, d. h. $\xi^P(p^S) = p^S(\mathbf{y}^S|\mathbf{x})$. Auf diese Weise lässt sich die aus dem segmentbasierten CRF erlangte Erkenntnis als Vorwissen in die punktweise Klassifikation integrieren. Dies kann z. B. helfen, wenn die lokale Umgebung eines Punktes dafür spricht, dass es sich um einen auf einer Baumkrone befindlichen Laserpunkt handelt. Wird nun der größere Bereich hinzugenommen, wird schnell klar, dass es sich stattdessen um einen Punkt auf einem Giebeldach eines Gebäudes handeln muss (Abbildung 4.2). Dieses Potential hilft somit, die Ergebnisse des generischen Interaktionsmodells der Segmente in der punktweisen Klassifikation zu berücksichtigen. Prinzipiell wird auf diese Weise die Strategie der Reduktion von Cliquen höherer Ordnung auf ein unäres Potential verfolgt, mit der sich HOP über herkömmliche Inferenzmethoden berücksichtigen lassen [Roig et al., 2011].

Bei der ersten initialen Durchführung von CRF^P ist die Information über die Segmente und deren Konfidenzen noch nicht verfügbar. Das Potential $\xi^P(p^S)$ kann dementsprechend erst von der zweiten Iteration an berücksichtigt werden, d. h. es gilt $\omega_\xi^P = 0$ für den ersten Durchlauf. Eine weitere Einschränkung ist bei Punkten zu beachten, die sich für CRF^S keinem Segment zuordnen ließen. Da in diesen Fällen keine Konfidenzwerte für die Segmente vorliegen, wird für jeden Punkt eine Gleichverteilung der Klassenwahrscheinlichkeiten für das Potential angenommen. Somit wird das punktweise Ergebnis nicht negativ beeinflusst.

Die Auswirkungen des Potentials $\xi^P(p^S)$ ähneln jenem des robusten P^n Potts Modells in dem Sinne, dass das Potential sich leicht glättend auf die Klassenlabels der zuvor zu einem Segment zusammengefassten Punkte auswirkt. Aus diesem Grund ersetzt es von der zweiten Iteration an den Term $\chi^P(\mathbf{x}_h, \mathbf{y}_h)$, indem ab diesem Zeitpunkt $\omega_\chi^P = 0$ gesetzt wird. Das P^n Potts Modell ist dennoch hilfreich,

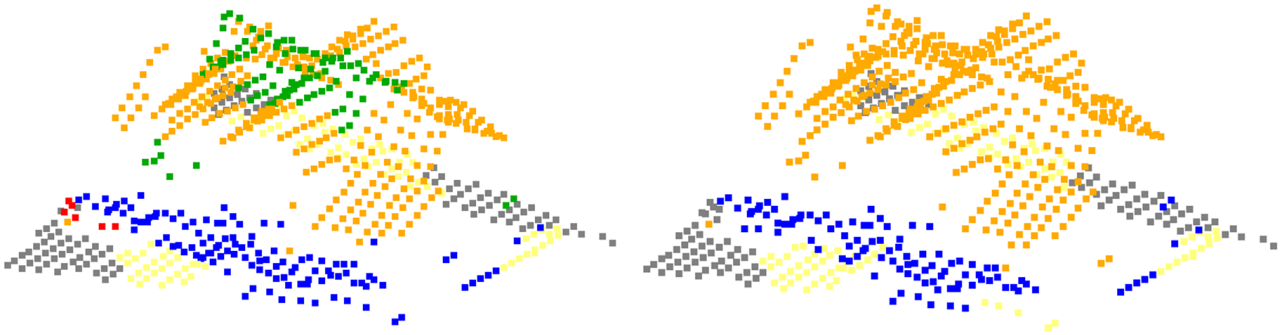


Abbildung 4.2: Laserpunkte auf einem Dachgiebel werden im linken Bild fälschlicherweise als *Baum* (grün) klassifiziert. Im rechten Bild konnten die Fehler durch die Hinzunahme gröberskaliger Information korrigiert werden. Alle Punkte des Daches sind nun der Klasse *Gebäude* (orange) zugeordnet. Weiterhin zu sehen sind die Klassen *Boden* (gelb), *Straße* (grau), *Auto* (rot) sowie *niedrige Vegetation* (blau).

um eine passende initiale Segmentierung erzielen zu können. Zur Reduzierung des Rechenaufwands und ggf. zur Vermeidung einer Überglättung wird auf das zeitgleiche Verwenden von $\chi^P(\mathbf{x}_h, \mathbf{y}_h)$ und $\xi^P(p^S)$ in den Folgeiterationen (> 1) verzichtet.

4.2.2 Inferenz

Inferenz beschreibt die Aufgabe, in einem graphischen Modell die optimale Konfiguration der Objektklassen für alle Knoten zu bestimmen. Da die Durchführung einer Inferenz bei CRF mit Cliquen höherer Ordnung und generischen Interaktionsmodellen eine große Herausforderung darstellt, wird für den vorgeschlagenen Ansatz auf das robuste P^n Potts Modell zurückgegriffen. Es ist zwar einerseits von seiner Aussagekraft etwas eingeschränkt, lässt sich andererseits aber mit dem in [Kohli et al., 2009] dargelegten Ansatz über Graph Cuts (Abschnitt 3.3) sehr effektiv mit einem niedrigen polynomischen Zeitaufwand lösen. Dementsprechend wird dieses Verfahren auch hier zur Inferenz herangezogen. Dabei wird auf die von den Autoren [Kohli et al., 2009] bereitgestellte Implementierung zurückgegriffen, welche zur Minimierung der Energie α-Expansion mit dem Min-Cut-Max-Flow-Verfahren [Boykov & Kolmogorov, 2004] kombiniert. Im Unterschied zur ursprünglichen Anwendung von Kohli et al. [2009] wird mit dem unären Potential $\xi^P(p^S)$ der segmentbasierten Klassifikation ein zusätzlicher Term in Gleichung 4.1 von CRF^P eingeführt. Um die entsprechenden Energiekomponenten in der Inferenz berücksichtigen zu können, werden diese nach der relativen Gewichtung mit dem anderen unären Potential für jeden Knoten kombiniert. Unabhängig von der Iteration (und den berücksichtigten Potentialen) wird somit Graph Cuts als Inferenzverfahren verwendet.

Für die Weiterverarbeitung sind Konfidenzwerte für die Klassenzuordnungen aller Punkte notwendig. Diese lassen sich aus der Energiefunktion des α-Expansion-Ansatzes nach der letzten Iteration ableiten. Sie basieren somit auf den ermittelten Klassenlabels. Grundlegend für die Berechnung ist der aus Gleichung 3.10 abgeleitete Zusammenhang $E = -\log(p)$, über den sich die Potentiale in Energieterme überführen lassen. Folgende Schritte werden durchgeführt:

1. Summiere die Energiewerte, welche sich als negativer Logarithmus der gewichteten unären Po-

tentiale ergeben, für jeden Punkt und jede Klasse.

2. Addiere dazu die Energie jeder paarweisen Interaktion. Dies erfolgt, indem für jeden an einer Kante beteiligten Knoten für alle Klassen ungleich des Knotenlabels die Energiebeträge der Kante summiert werden.

3. Erstelle für jede Clique höherer Ordnung ein Histogramm der enthaltenen Klassenlabels, um entsprechend Gleichung 4.3 die Energie pro Klasse zu berechnen. Initialisiere für jeden Mitgliedsknoten einer Clique die Maximalenergie γ_{max} für alle Klassen außer der derzeit vorliegenden Klasse. Ordne dieser den entsprechende Wert aus dem Histogramm zu.

4. Transformiere die Energie für jeden Knoten zu Konfidenzwerten durch Exponenzieren der negativen Energie mit anschließender Normierung

$$p_i(y_i^l) = \frac{\exp(-E(y_i^l))}{\sum_l \exp\left(-E(y^l)\right)}. \tag{4.5}$$

Auf diese Weise ergeben sich der Wahrscheinlichkeit ähnliche Maße, welche für die weitere Prozessierung der Laserdaten herangezogen werden.

4.2.3 Training

Das vorgestellte Verfahren lässt sich der Gruppe der überwachten Klassifikationsmethoden zuordnen. Ein wesentlicher Schritt ist somit das Anlernen des Klassifikators sowie in diesem Fall der Gewichtsparameter auf der Grundlage eines vollständig annotierten Trainingsdatensatzes. Zum Lernen der paarweisen Interaktionen und der Cliquen höherer Ordnung ist es wichtig, dass ein Ausschnitt der Szene vollständig gelabelt ist.

Für die punktbasierte Ebene müssen der Klassifikator RF^P zur Bestimmung des unären Potentials sowie die relativen Gewichte $\mathbf{\Omega}^P = (\omega_\psi^P, \omega_\chi^P, \omega_\xi^P)^T$ zwischen den einzelnen Potentialen antrainiert werden. Da die zur Modellierung der Cliquen höherer Ordnung verwendeten Potentiale von der Iteration abhängen, erfolgt im ersten Durchgang die Optimierung von ω_ψ^P und ω_χ^P. In der zweiten Iteration werden dann geeignete (und für alle folgenden Iterationen gültige) Werte für ω_χ^P und ω_ξ^P ermittelt. Aufgrund der neuen Zusammensetzung der Potentialterme wird ω_χ^P an dieser Stelle erneut bestimmt.

Zum Training wird der gesamte Trainingsdatensatz zunächst in zwei Teilgebiete $\mathfrak{T}_1$ und $\mathfrak{T}_2$ aufgeteilt, um die Gefahr einer Überanpassung an die Daten zu reduzieren [Bishop, 2006]. Während $\mathfrak{T}_1$ für das Anlernen der Potentiale verwendet wird, dient $\mathfrak{T}_2$ anschließend als Grundlage zum Optimieren von $\mathbf{\Omega}^P$. In einem ersten Schritt werden die T einzelnen Entscheidungsbäume von RF^P unter Zuhilfenahme der Trainingsdaten $\mathfrak{T}_1$ aufgebaut. Für jeden Entscheidungsknoten wird in der Implementierung von OpenCV[2] basierend auf dem Gini-Index das am diskriminativsten erscheinende Merkmal aus einer zufällig gewählten Untermenge zugeordnet.

Zum Ermitteln der besten Parameter wird die direkte Suche [Hooke & Jeeves, 1961] (Abschnitt 3.3) verwendet. Das gradientenfreie Lernverfahren optimiert die Gewichte iterativ und nacheinander

[2]Dokumentation OpenCV, `http://docs.opencv.org/3.0.0/dc/dd6/ml_intro.html` (abgerufen am 25.05.2016)

durch das Maximieren der Anzahl korrekt klassifizierter Knoten. Dazu wird die Klassifikation der Trainingsdaten einige Male für ein aktives Gewicht mit leicht veränderten Werten wiederholt. Jener Wert, welcher zur höchsten Genauigkeit für den Validierungsdatensatz $\mathfrak{T}_2$ führt, wird als neuer Wert verwendet. Anschließend wird der nächste Parameter optimiert. Der Vorgang endet, sobald sich durch die Veränderung aller Parameter kein weiterer Genauigkeitsgewinn an $\mathfrak{T}_2$ erzielen lässt.

4.3 Segmentierung der Punktwolke

Basierend auf den Ergebnissen der punktweisen Klassifikation CRF^P werden die Segmente für das anschließende CRF^S extrahiert. Das Ziel ist es, möglichst nur solche Punkte zu einem Segment zusammenzufassen, welche einem (Teil-)Objekt angehören. Auch soll die Aggregation von Punkten verschiedener Objektklassen zu einem Segment nach Möglichkeit vermieden werden, da dies zu Fehlern bei der nachfolgenden segmentbasierten Klassifikation dieses Iterationsschrittes führt. Dazu ist es ratsam, einerseits die geometrischen Beziehungen benachbarter Punkte sowie andererseits auch die Vorinformation über die Konfidenzen der Klassenzugehörigkeiten aus CRF^P mit einzubeziehen. Folglich handelt es sich um eine Segmentierungsaufgabe, bei der diese Aspekte berücksichtigt werden sollen. Ein für diesen Zweck geeignetes Verfahren stellt die von Papon et al. [2013] vorgeschlagene und in Abschnitt 3.4 beschriebene Methode VCCS dar. Sie ist unmittelbar für die (Über-)Segmentierung von unregelmäßigen Punktwolken mit variierenden Punktdichten ausgelegt, wird häufig verwendet (z. B. in [Wolf et al., 2016]) und zeichnet sich bei einem Vergleich verschiedener Verfahren durch einen sehr geringen Untersegmentierungsfehler aus [Stutz, 2015].

In Erweiterung zur Anwendung von [Papon et al., 2013] sowie zur Implementierung von VCCS in der Point Cloud Library sind zwei wesentliche Änderungen für die hier vorgeschlagene Methodik erforderlich. Einerseits stehen keine RGB-Farbinformationen zur Verfügung, da ausschließlich mit den Laserdaten gearbeitet wird. Demzufolge kann die letzte der drei Komponenten *räumliche Distanz* (D_s), *Winkel zwischen Normalen* (D_n) und *Farbähnlichkeit* (D_{rgb}) in Gleichung 3.13 vernachlässigt werden. An dessen Stelle soll der Untersegmentierungsfehler durch die Integration vorheriger Klassifikationsergebnisse reduziert werden. Zu diesem Zweck wird ein zusätzlicher Term D_{konf} zur Bestimmung des neuen Ähnlichkeitskriteriums D eingeführt:

$$
\begin{aligned}
D &= \sqrt{\frac{\omega_s^{SV} D_s^2}{R_{saat}^{SV}} + \omega_n^{SV} D_n^2 + \omega_{konf}^{SV} D_{konf}^2} \qquad (4.6)\\
\text{mit } D_s &= ||\mathbf{v}_i - \mathbf{v}_j||, \mathbf{v}_i, \mathbf{v}_j \in \mathbb{R}^3\\
D_n &= 1 - \langle \mathbf{n}_i, \mathbf{n}_j \rangle\\
D_{konf} &= \frac{1}{\sqrt{2}} \sqrt{\sum_{l=1}^{|\mathcal{L}|} (\sqrt{p_l} - \sqrt{q_l})^2}
\end{aligned}
$$

Der räumliche Abstand D_s ergibt sich aus den Koordinaten $\mathbf{v}_i = [x_i\ y_i\ z_i]^T, \mathbf{v}_j = [x_j\ y_j\ z_j]^T$ zweier benachbarter Primitive i und j. Analog berechnet sich D_n aus dem Winkel zwischen den entsprechenden Normalenvektoren $\mathbf{n}_i$ und $\mathbf{n}_j$, welche mit dem Verfahren nach [Boulch & Marlet, 2012] bestimmt

werden. Um die Ähnlichkeit zweier diskreter Konfidenzvektoren aus CRF^P zu beschreiben, wird in Gleichung 4.6 zur Bestimmung von D_{konf} die diskrete Hellinger Distanz [Hellinger, 1909] verwendet. Diese Norm ist das Äquivalent zur Distanz im euklidischen Raum. Hierbei ergibt sich ein geringer Wert, wenn beide Primitive eine ähnliche Verteilung der wahrscheinlichkeitsähnlichen Maße aufweisen. Entsprechend gehören sie voraussichtlich derselben Objektklasse an. Diese Ausnutzung der kontinuierlichen Konfidenzwerte erweist sich vorteilhaft gegenüber der Untersuchung von diskreten Klassenlabels, da auch in diesem Fall verfälschende Effekte durch eine Generalisierung vermieden werden. So kann es beispielsweise sein, dass durch die MAP-Methode in CRF^P zwei benachbarte Punkte unterschiedlichen Objektklassen zugeordnet werden, obwohl die Konfidenzen sehr ähnlich sind. Um dies zu verdeutlichen, sei das Beispiel $\mathbf{p}_i = [0\ 0\ 0{,}51\ 0{,}49]^T$ und $\mathbf{q}_j = [0\ 0\ 0{,}48\ 0{,}52]^T$ gegeben. Bei einer ausschließlichen Berücksichtigung der diskreten Label wären beide Punkte unterschiedlich, da sich aus der MAP-Methode $y_i = 3$ und $y_j = 4$ ergäbe. In beiden Fällen ist die Entscheidung zwischen Klasse 3 und 4 jedoch nicht eindeutig, und die Punkte weisen ähnliche Verteilungen auf. Dementsprechend sollen beide Punkte in einem Segment zusammengefasst werden, um durch die Hinzunahme von gröberskaligen Merkmalen eventuell eine Entscheidung herbeiführen zu können. Prinzipiell handelt es sich somit also um eine *weiche Segmentierung*. Das ausschließliche Auffinden von Zusammenhangskomponenten mit benachbarten Punkten gleicher Klassen hat sich in [Niemeyer et al., 2015] als ungeeignet herausgestellt, da viele Segmente aufgrund inhomogener Labels an schwer zu klassifizierenden Objekten nur aus einem einzigen Punkt bestanden.

Analog zu Gleichung 3.13 ist auch in Gleichung 4.6 eine relative Gewichtung der einzelnen Komponenten über die Parameter ω_s^{SV}, ω_n^{SV} und ω_{konf}^{SV} möglich. Nimmt man die beiden Stellgrößen für Voxelauflösung (R_{vox}^{SV}) und Abstand der Saatpunkte (R_{saat}^{SV}) hinzu, sind insgesamt fünf manuell festzulegende und einfach zu interpretierende Parameter für die Segmentierung erforderlich. Die Voxelauflösung hängt im Wesentlichen von der Punktdichte in der Szene ab. Je kleiner die Auflösung, umso mehr Details können in der Szene erfasst werden. Wählt man sie jedoch zu gering, entstehen leere Voxel ohne 3D-Punkte, über die sich die Supervoxel nicht ausbreiten können. Dementsprechend sollte die Auflösung so festgesetzt werden, dass im Schnitt mindestens ein Punkt enthalten ist. Ähnliches gilt für den Abstand der Saatpunkte R_{saat}^{SV}, der die Größe der Supervoxel bestimmt. Bei einem kleinen Wert entstehen viele kleine Segmente, die jedoch nur eine geringe räumliche Abdeckung aufweisen. Andererseits führt ein großer Wert zu einer Verschmelzung von Objekten (etwa zweier benachbarter Autos), was die Aussagekraft einiger Merkmale beeinflussen kann. Auch hierbei muss ein Kompromiss gefunden werden. Ein entsprechendes Experiment des Einflusses der Parameter auf die Segmentierung findet sich in Abschnitt 5.3.2.

Ein weiterer Vorteil der leichten Übersegmentierung gegenüber anderen Segmentierungsverfahren besteht in der recht hohen Anzahl der resultierenden Segmente, da für das Training der RF Klassifikatoren ausreichend viele Trainingsbeispiele vorhanden sein müssen. Aufgrund der komplexeren Klassifikationsaufgabe ist dies insbesondere bei der Modellierung der paarweisen Segmentinteraktionen der Fall. Andere Verfahren, wie das Regionenwachstum oder die Methode von Felzenszwalb & Huttenlocher [2004], neigen zur Generierung von wenigen, aber teilweise sehr großen Segmenten, welche eher gesamte Objekte approximieren. Dies reduziert nicht nur die Anzahl der zur Verfügung stehenden Trainingsbeispiele, sondern kann auch die Aussagekraft der Merkmale negativ beeinflussen, falls bei-

spielsweise Punkte mit stark unterschiedlichen Merkmalsvektoren kombiniert werden. In diesem Fall wäre ein anderer Merkmalssatz erforderlich, der sich z. B. nicht aus den Mittelwerten seiner Komponenten (Abschnitt 3.1.2) ergibt. Aufgrund der unterschiedlichen Objekttypen in der Szene kann sich die Ableitung diskriminativer und objektbezogener Merkmale als schwierig erweisen.

4.4 Segmentbasierte Klassifikation

Nach der Durchführung der punktbasierten Klassifikation in CRF^P und der anschließenden Segmentbildung entsprechend Abschnitt 4.3 besteht der nächste wesentliche Schritt des vorgestellten Verfahrens in der Klassifikation der Segmente. Das CRF dieser Ebene wird im Folgenden als CRF^S bezeichnet. Die Verwendung von Segmenten hat den Vorteil, dass diese einen größeren räumlichen Bereich abdecken und demnach durch diesen Informationsgewinn für das Auffinden der besten Punktlabels beisteuern können. Für Segmente lässt sich ein weiterer Satz von Merkmalen ableiten. Außerdem sollen durch ein flexibleres Interaktionsmodell die Klassenrelationen zwischen den einzelnen Segmenten berücksichtigt werden. Demzufolge liegt der Fokus dieser Ebene nicht auf der datenabhängigen Glättung, sondern vielmehr auf der Ausnutzung möglichst aussagekräftiger Kontextinformation dieses Skalenbereichs. Das Ergebnis dieses Schrittes wird in der folgenden Iteration als Vorwissen über das unäre segmentbasierte Potential $\xi^P(p^S)$ in die punktweise Klassifikation integriert.

In CRF^S stellen die Segmente die wesentlichen Entitäten dar; jedes einzelne Segment wird im graphischen Modell $\mathcal{G}^S(\mathcal{V}^S, \mathcal{E}^S)$ als Knoten der Menge $\mathcal{V}^S$ repräsentiert. Die Menge der Kanten $\mathcal{E}^S$ stellt entsprechend die Verbindungen zwischen jeweils zwei Segmenten $\{i, j\}$ dar. Weiterhin entspricht der Satz der Objektklassen ($\mathcal{L}$) jenem der punktweisen Zuordnung. Das Label eines Segments i wird mit y_i^S angegeben und der Vektor $\mathbf{y}^S$ beinhaltet die Klassen aller Segmente. Die Zuordnungswahrscheinlichkeiten $p^S(\mathbf{y}^S|\mathbf{x})$ ergeben sich aus

$$p^S(\mathbf{y}^S|\mathbf{x}) = \frac{1}{Z^S(\mathbf{x})} \prod_{i \in \mathcal{V}^S} \varphi^S(\mathbf{x}, y_i^S) \cdot \prod_{e_{i,j} \in \mathcal{E}^S} \psi^S(\mathbf{x}, y_i^S, y_j^S)^{\omega_\psi^S}. \tag{4.7}$$

Anders als auf der punktweisen Ebene werden in Gleichung 4.7 nur zwei Terme für die Berechnung verwendet, welche keine Cliquen höherer Ordnung enthalten. Das unäre Potential $\varphi^S(\mathbf{x}, y_i^S)$ wird für jedes Segment berechnet. $\psi^S(\mathbf{x}, y_i^S, y_j^S)$ repräsentiert das paarweise Interaktionsmodell und modelliert die Beziehungen zwischen adjazenten Segmenten. Es wird für jede Kante des Graphen bestimmt. Das Produkt der über ω_ψ^S relativ zueinander gewichteten Potentiale wird über die Partitionsfunktion $Z^S(\mathbf{x})$ normiert.

4.4.1 Potentiale

Unäres Potential

Das unäre Potential $\varphi^S(\mathbf{x}, y_i^S)$ modelliert in diesem Fall für jeden Knoten des Graphen (bzw. für jedes Segment) die Wahrscheinlichkeiten für die Klassenzugehörigkeiten basierend auf seinem Merkmals-

vektor $\mathbf{f}_i^S(\mathbf{x})$. Dabei erfolgt die Bestimmung des Potentials ebenfalls mithilfe eines individuellen RF Klassifikators, der im Folgenden als RF_u^S bezeichnet wird. Der Index u kennzeichnet dabei das unäre Potential. Da für die Segmente ein anderer Satz von Merkmalen $\mathbf{f}_i^S(\mathbf{x})$ verwendet wird, ist dieser Klassifikator nicht identisch mit RF^P des punktweisen CRF^P und muss neu antrainiert werden. Das unäre Potential wird analog zu CRF^P durch die Wahrscheinlichkeit von y_i^S bei gegebenen Daten $\mathbf{x}$ definiert, d. h. $\varphi^S(\mathbf{x}, y_i^S) = p(y_i^S|\mathbf{f}_i^S(\mathbf{x}))$.

Paarweises Interaktionspotential

Der wesentliche Unterschied in der Zielsetzung der beiden interagierenden Ebenen CRF^P und CRF^S im vorgestellten Verfahren liegt bei den Interaktionspotentialen. Während auf punktweiser Ebene durch das kontrast-sensitive Potts Modell eine datenabhängige Glättung stattfindet, wird für die Segmentebene ein komplexeres Modell berücksichtigt. Es wird ein generischer Ansatz zur Repräsentation der Verbundwahrscheinlichkeiten aller Klassenrelationen gewählt. Ein solches generisches Modell kann modellieren, dass bei gegebenen Daten eine konkrete Klassenbeziehung einer anderen Relation vorzuziehen ist. Das paarweise Potential ist proportional zu den Verbundwahrscheinlichkeiten der beiden benachbarten Knotenklassen y_i^S und y_j^S. Für diesen Zweck wird für jede Kante $e_{ij} \in \mathcal{E}^S$ aus den Daten $\mathbf{x}$ ein Merkmalsvektor $\mathbf{f}_{ij}^S(\mathbf{x})$ extrahiert. Ein diskriminativer Klassifikator entscheidet anschließend, welche Konfiguration von Klassen an den adjazenten Segmenten bei den gegebenen Merkmalen $\mathbf{f}_{ij}^S(\mathbf{x})$ am wahrscheinlichsten auftritt.

Wie bereits im Fall von CRF^P muss zum Aufbau des Graphen zunächst festgelegt werden, welche Segmente als benachbart anzusehen sind. Dabei ist die Nachbarschaftsdefinition von Segmenten im $\mathbb{R}^3$ ebenfalls nicht explizit gegeben. Für diesen Zweck wird mittels einer Adjazenzmatrix bestimmt, welche Segmente einen Abstand kleiner als ein manuell vorgegebener Distanzschwellwert d_{graph}^S auf Basis ihrer Mitgliedspunkte aufweisen. Diese Distanz wird über eine auf die xy-Ebene projizierte 2D-Suchumgebung realisiert. Alle adjazenten Knoten (Segmente) werden anschließend mittels Kanten verbunden. Die Wahl einer zylindrischen Suchumgebung ist wie bei CRF^P auf die Begründung zurückzuführen, dass dadurch die wichtigen vertikalen Beziehungen zwischen den Segmenten über Kanten modelliert werden. Formal ist das Potential dann gegeben durch

$$\psi^S(\boldsymbol{x}, y_i^S, y_j^S) = p(y_i^S, y_j^S|\mathbf{f}_{ij}^S(\mathbf{x})) \cdot \frac{1}{N_{k_i}}. \tag{4.8}$$

Ergänzend zu Gleichung 3.11 erfolgt an dieser Stelle eine Normierung des Potentials mit der Anzahl der Nachbarn N_{k_i} für jeden Knoten i, da die Nachbarschaftsgröße variiert. Dies stellt für alle Kanten denselben Einfluss auf die Klassifikation sicher [Wegner, 2011; Weinmann, 2016].

Die konkrete Umsetzung zur Berechnung der Interaktionspotentiale erfolgt analog zum unären Potential. Es wird ein individueller Klassifikator RF_p^S aufgebaut, um die Verbundwahrscheinlichkeiten $p^S(y_i^S, y_j^S|\mathbf{f}_{ij}^S(\mathbf{x}))$ zu ermitteln. Der wesentliche Unterschied besteht in der Anzahl der über den RF zu separierenden Klassen. Anstelle der beim unären Potential notwendigen Einteilung in L Klassen müssen zur Modellierung der Verbundwahrscheinlichkeit maximal L^2 Klassen angelernt und unterschieden werden. Somit wird jede Klassenkombination als eigene Klasse angesehen. Als Nachteil dieses Mo-

dells ist der Bedarf einer größeren Menge an Trainingsdaten zu nennen, da mehr Parameter bestimmt werden müssen. Für das Anlernen des Interaktionen ist im Allgemeinen eine vollständig annotierte Trainingspunktwolke hilfreich, um diskriminative Trainingsbeispiele für die Relationen zu erhalten. Die Referenzlabels können beispielsweise über manuelle Zuordnung der 3D-Punkte zu den Klassen oder - zumindest unterstützend - über ein Verfahren wie [Li et al., 2016] automatisch generiert werden.

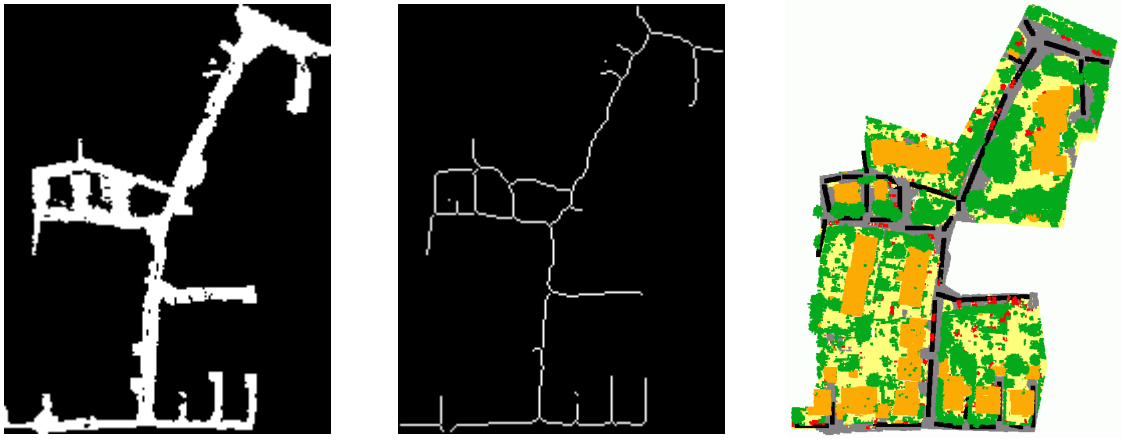
4.4.2 Neue Kontextmerkmale für Segmente

Die aus der punktweisen Klassifikation erhaltene Vorinformation über das wahrscheinliche Klassenlabel eines Segments ermöglicht die Extraktion neuer Segmentmerkmale, die den Kontext modellieren.

Nachbarschaftsklassenverteilung: In Anlehnung an [Xiong et al., 2011] und [Tu & Bai, 2010] wird die erste hier vorgestellte Merkmalsgruppe aus den klassenspezifischen Konfidenzen einer vorherigen Klassifikation berechnet, welche sich aus L Merkmalskomponenten zusammensetzt. Für jedes Segment werden alle benachbarten 3D Punkte innerhalb einer 3D Umgebung mit dem Radius r_{hist} hinsichtlich ihrer Konfidenzen aus CRF^P analysiert. Diese werden in einem Histogramm aufgetragen und anschließend normalisiert, um die prozentualen Klassenverteilungen der Nachbarn um jedes Segment zu ermitteln. Durch die Normalisierung entfällt der Einfluss unterschiedlicher Segmentgrößen. Die resultierenden relativen Häufigkeiten werden direkt als L Merkmale für die Segmente herangezogen. Auf diese Weise lässt sich die typische Nachbarschaft modellieren, wodurch die gemeinsame Auftrittswahrscheinlichkeit berücksichtigt wird. So sind Autosegmente beispielsweise in der Regel hauptsächlich von Straßenpunkten umgeben. Zudem ist die Merkmalsgruppe hilfreich, um typische Fehlklassifikationen von Giebeldächern zu korrigieren, denn ein Segment der Klasse *Baum* ist selten vollständig von Gebäudedächern umgeben. Da diese Form die gesamte Umgebung eines Objektes einbezieht, ist die Information etwas umfassender als beispielsweise die einfache Gewichtung des Interaktionspotentials mittels der Auftrittswahrscheinlichkeiten, wie es etwa von Kosov et al. [2013] vorgeschlagen wird.

Straßenmerkmale: Zusätzlich zu den Nachbarschaftsklassenverteilungen werden weitere Merkmale entwickelt, welche die relative Lage und Ausrichtung von Segmenten in der Szene formulieren. Die Herausforderung bei luftgestützten Daten besteht in der Definition der Referenzrichtung (Kapitel 2), da diese nicht explizit gegeben ist. Inspiriert durch die Idee von Golovinskiy et al. [2009] soll für diesen Zweck der nächstgelegene Straßenabschnitt als Bezug dienen. Urbane Gebiete sind wesentlich durch Straßen strukturiert. Straßen lassen sich weiterhin gut in den Scandaten detektieren, so dass diese als geeignet erscheinen.

Abgeleitet werden die Merkmale *Orientierung eines Segments relativ zum nächstgelegenen Straßensegment*, sowie der *kürzeste* und *durchschnittliche 2D-Abstand* zwischen einer Straße und allen Punkten eines Segmentes. Die im Folgenden beschriebene Vorgehensweise zur Extraktion des Straßenverlaufs beschränkt sich auf 2D. Der Verlauf der Straße wird dabei aus den Punkten aller Segmente, welche der Klasse *Straße* zugeordnet sind, approximiert. Zunächst wird eine Binärkarte der Straße durch Projektion in die xy-Ebene mit vorgegebenen Rasterauflösung res_{grid} erzeugt, bei der jedem Pixel mit mindestens einem enthaltenen Straßenpunkt der Wert 1 zugewiesen wird. Ein beispielhaftes Ergebnis ist in Abbildung 4.3a dargestellt. Da die Straßenränder durch Überdeckungen keine klar definierten Kanten aufweisen, erfolgt eine morphologische Schließung (*Closing*) mit einem kreisapproximieren-



(a) Binäre Straßenmaske (b) Skelett (c) Extrahierte Straßen

Abbildung 4.3: Approximation des Straßenverlaufs

den Strukturelement der Dimension $S_{closing} \times S_{closing}$. Weiterhin werden auf diese Weise Datenlücken geschlossen, die aufgrund von Verdeckungen z. B. durch Autos entstehen. Aus dem resultierenden Straßenverlauf wird das Skelett nach der Methode von Zhang & Suen [1984] gebildet (Abbildung 4.3b).

Die geraden Teilsegmente des Straßennetzes werden nun mittels probabilistischer Hough-Transformation [Matas et al., 2000] detektiert. Dazu müssen einige weitere Parameter festgelegt werden. Konkret sind dies die Distanz- und Winkelauflösung (H^{res}_{dist} und H^{res}_{ρ}) im Hough-Raum, die minimale erforderliche Stimmenanzahl für eine Hypothese ($H_{minVotes}$), die minimale Länge des Geradenstücks ($H_{minLength}$) sowie die maximal erlaubte Länge einer Datenlücke innerhalb eines Liniensegments ($H_{lineGap}$). Die Implementierung in OpenCV liefert unmittelbar die Start- und Endpunkte jedes Teilsegments, die zur Berechnung der Orientierungen herangezogen werden. Eine lineare Interpolation entlang jeder Geraden zwischen diesen beiden Punkt ermöglicht die effiziente Berechnung der Durchschnittsentfernung eines Segments zur Straße mittels eines kd-Baums.

Nachdem der Straßenverlauf auf die beschriebene Weise approximiert ist, lassen sich die entsprechenden Merkmale ableiten. Bei Objekten mit Flächen größer als R^{SV}_{saat} werden Form und Orientierung der einzelnen Supervoxel durch die Übersegmentierung nicht sehr aussagekräftig sein, da die Voxel nicht die Objekte repräsentieren. Aus diesem Grund ist ein weiterer Schritt notwendig, um benachbarte Supervoxel derselben Klasse temporär für die Merkmalsextraktion zu Objekten zu vereinen. Dies erfolgt durch das Bilden der Zusammenhangskomponenten für benachbarte Segmente gleichen Labels basierend auf den Ergebnissen der Vorklassifikation. Jede Komponente dient als Approximation eines Objektes, und kann daher zur Bestimmung von Orientierung sowie minimalem und mittlerem Abstand zur Straße herangezogen werden. Die berechneten Merkmale lassen sich anschließend auf die korrespondierenden Supervoxel übertragen.

Mittels eines kd-Baums wird für jeden Punkt eines *Objekts* der 2D-Abstand zu seinem nächsten Straßensegment bestimmt und anschließend gemittelt. Diese Distanz dient als erstes Merkmal *mittlere Distanz zur Straße*. Zur Steigerung der Robustheit bei langgestreckten und eventuell nicht gerade verlaufenden Objekten wird dabei jeder Punkt analysiert, anstatt nur die Distanz zwischen dem Schwerpunkt und der nächsten Straße zu berechnen. Das auf diese Weise identifizierte nächstgelege-

ne Straßensegment S_i wird ebenfalls verwendet, um die *kürzeste 2D-Distanz des Objektes zur Straße* zu berechnen. Zusätzlich zu den Abständen wird die Ausrichtung der Objekte relativ zur Straße untersucht und als Merkmal herangezogen. Die Orientierung des Segmentes ergibt sich durch eine Hauptachsentransformation der Koordinaten aller beteiligter Punkte und wird durch den Eigenvektor zur ersten Hauptkomponente $\vec{EV}_1$ repräsentiert. Die Ausrichtung $\vec{S}$ des nächstgelegenen Straßensegments S_i kann aus den Koordinaten des Start- und Endpunktes ermittelt werden. Sind beide Größen vorhanden, so ergibt sich der *Winkel zwischen Segment und Straße* aus dem Skalarprodukt, d. h. $\alpha = |\arccos(\vec{EV}_1 \circ \vec{S})/(|\vec{EV}_1| \cdot |\vec{S}|)|$. Eine Untersuchung zum Einfluss dieser Merkmale erfolgt in Abschnitt 5.3.1. Ein Experiment in Abschnitt 5.5.4 untersucht ferner, ob ein Genauigkeitsgewinn durch die Hinzunahme der Straßeninformation erzielt werden kann und wie die Approximation des Straßenverlaufs im Vergleich zu externen GIS-Daten abschneidet.

Die Extraktion des Straßenverlaufs auf diese Weise ist basierend auf den Merkmalen typischerweise nur möglich, wenn die Intensitätsinformation des Lasersignals zu jedem Punkt zur Verfügung steht. Intensitäten sind wesentlich, um zwischen asphaltierter Fläche und bewachsenem Boden zu unterscheiden. Weitere aus der Signalform abgeleitete Merkmale, wie etwa die Echobreite, könnten die Detektionsgenauigkeit weiter verbessern. Als Nachteil ist zu nennen, dass nicht ausschließlich Straßen, sondern vielmehr die versiegelten Flächen einer Szene gefunden werden. Es kann sich somit beispielsweise auch um Plätze o. ä. handeln, wodurch die Position und Orientierung der ermittelten Straßensegmente negativ beeinflusst werden kann.

Sollten Intensitäten nicht verfügbar sein, können alternativ zur Ableitung des genäherten Straßenverlaufs aus einer Vorklassifikation ebenfalls Daten externer GIS-Datenbanken, wie etwa OpenStreetMap (OSM), herangezogen werden. Dies ist jedoch mit dem Nachteil zusätzlicher erforderlicher Daten verbunden, deren Qualität unbekannt ist. Wird die Straße unmittelbar aus den Scandaten geschätzt, handelt es sich um ein aktuelles Abbild der Topographie. Ein vergleichendes Experiment in Abschnitt 5.5.4 stellt die Ergebnisse beider Varianten gegenüber.

4.4.3 Inferenz

Für das generische Interaktionsmodell kann die Einhaltung der Submodularitätsbedingung nicht garantiert werden [Rottensteiner, 2017], wodurch sich das für CRF^P genutzte Inferenzverfahren Graph Cuts nicht anwenden lässt. Als Alternative kommt daher das Message-Passing-Verfahren LBP in der Max-Sum-Variante (Abschnitt 3.3) zum Einsatz. Bei LBP handelt es sich neben Graph Cuts um ein Standardverfahren zur Durchführung der Inferenz. Obwohl bei dieser Methode keine Garantie zum Auffinden der optimalen Lösung gewährleistet ist, führt es in den meisten Fällen zu guten Ergebnissen und gehört bei Vergleichen zu den besten Inferenzverfahren [Bishop, 2006; Szeliski et al., 2008]. Die aus LBP resultierenden Konfidenzwerte werden in der nächsten punktweisen Klassifikation CRF^P der folgenden Iteration durch das Potential $\xi^P(p^S) = p^S(\mathbf{y}^S|\mathbf{x})$ berücksichtigt.

4.4.4 Training

Für die Segmentebene werden zwei voneinander unabhängige RF Klassifikatoren für die beiden Potentiale auf Grundlage der Trainingsdaten $\mathfrak{T}_1$ (Abschnitt 4.2.3) angelernt. Es ist anzumerken, dass ebenfalls vollständig annotierte Punktwolken für den Trainingsschritt hilfreich sind. Da nun die Klassifikation auf Basis von Segmenten erfolgt, welche abhängig von den punktweisen Labels generiert werden, sind zunächst keine Referenzsegmente in den Trainingsdaten vorhanden. Diese werden erzeugt, indem auch die Trainingsdaten punktweise klassifiziert und anschließend nach dem in Abschnitt 4.3 beschriebenen Verfahren zu Segmenten aggregiert werden. Anschließend wird anhand der punktweisen Referenzlabels die dominante Klasse jedes Segments ermittelt und als Referenzlabel angesehen. Bei Mehrdeutigkeiten wird das Segment nicht für das Training herangezogen.

Weiterhin muss in diesem Fall nur ein Parameter ω_ψ^S gelernt werden, welcher das unäre und das paarweise Potential relativ zueinander gewichtet. Die Optimierung erfolgt über direkte Suche (Abschnitt 3.3) anhand der Trainingsdaten $\mathfrak{T}_2$.

4.5 Diskussion

Konzept: Die Realisierung des neu vorgestellten, hierarchischen Ansatzes optimiert die beiden Klassifikationsebenen nacheinander und getrennt voneinander. Zwischen den entsprechenden Inferenzschritten werden die Informationen in die jeweils andere Ebene propagiert, indem einerseits die Segmente für CRF^S auf Grundlage des CRF^P Ergebnisses generiert und andererseits die Konfidenzwerte von CRF^S als zusätzliches Potential $\xi^P(p^S)$ in Gleichung 4.1 in der nächsten Iteration von CRF^P berücksichtigt werden. Der Grund für die alternierende Methode ist, dass sich in jedem Iterationsschritt die Segmente von CRF^S ändern können und somit die Knoten innerhalb des graphischen Modells variieren. Eine simultane Inferenz beider Ebenen wäre daher sehr aufwändig. Sie hätte jedoch den Vorteil, eine Lösung für eine einzige Energiefunktion erzielen und somit die besten Ergebnisse für Punkt- und Segmentebene gemeinsam ermitteln zu können. Im Sinne von [Roig et al., 2011] interpretiert, kann die alternierende Methode hingegen als punktweise Klassifikation mit komplexen HOP durch die segmentbasierte Ebene angesehen werden. Demnach erfolgt eine iterative und approximative Minimierung der Energiefunktionen beider Ebenen. Es gibt allerdings keine Garantie, die optimale Lösung für beide Ebenen zu identifizieren. Durch das alternierende Verfahren können nach einer Optimierung Fehler entstehen, da nicht alle Informationen zeitgleich mit einbezogen werden. Diese Fehlentscheidungen bei der Labelzuordnung können jedoch im Laufe der Iterationen oft wieder korrigiert werden, falls die Merkmale und der entsprechende Kontext dies induzieren. Die Bedingung dafür ist, dass die fehlerhaft klassifizierten Punkte auf Grundlage ihrer vorläufigen Objektklassen im Zuge der Segmentierung nicht mit den Nachbarpunkten zu einem einzigen Segment zusammengefasst werden. Solange die Fehlklassifikationen eine individuelle Einheit darstellen, ist eine anschließende Korrektur prinzipiell möglich. Die Prozedur endet nach einer manuell festgesetzten Anzahl von Iterationsschritten (N_{iter}). Ein Experiment dazu ist in Abschnitt 5.5.2 zu finden.

Kontextmerkmale: Die in dieser Arbeit entwickelten Merkmale zum *Abstand* und zur *Orientierung eines Segments zur nächstgelegenen Straße* basieren auf einem aus der Vorklassifikation abgeleiteten

und approximierten Straßenverlauf. Somit ist eine gute Detektion der entsprechenden Laserpunkte in CRF^P notwendig. Das wesentliche Merkmal zur Unterscheidung von grasbedecktem Boden und versiegelten Flächen stellt die Intensität des Lasersignals dar. Ist diese nicht verfügbar, lassen sich die neuen Merkmale nur ungenau bestimmen. Auch bei vorhandener Intensität kann bei der Einzelpunktklassifikation nicht zwischen einer Straße und versiegelten Flächen im Allgemeinen unterschieden werden. Eine entscheidende Aufgabe besteht somit in der Approximation des Straßenverlaufs ausgehend von allen Punkten der Klasse *versiegelte Fläche*. Dies erfolgt wie in Abschnitt 4.4.2 beschrieben durch eine Skelettbildung einer 2D-Binärkarte aller Kandidatenpunkte. Probleme können dabei in Gebieten größerer versiegelter Flächen auftreten, welche keine Straßen repräsentieren, wie es z. B. bei Parkplätzen der Fall ist. In solchen Bereichen ergeben sich u. U. durch die Skelettbildung Artefakte im approximierten Straßenverlauf, bei denen die Ausrichtung und Position der ermittelten Geradenabschnitte nicht aussagekräftig ist. Dies kann u. U. zu Fehlern in der Folgeklassifikation führen.

Die andere Gruppe von Kontextmerkmalen für die Segmente wird durch die Nachbarschaftsklassenverteilung repräsentiert. Sie dient dazu, die a priori Information der vorgeschalteten CRF^P Klassifikation auf die Segmentebene zu propagieren und dabei die gemeinsamen Auftrittswahrscheinlichkeiten benachbarter Klassen zu modellieren. Anstatt die Nachbarschaft zu berücksichtigen, lassen sich alternativ die Konfidenzwerte aller Punkte jedes Segments selbst mitteln und unmittelbar als entsprechende Merkmalsgruppe heranziehen. In diesem Fall ist der Einfluss der Vorklassifikation jedoch sehr groß, da ein Großteil der Laserpunkte bereits ausschließlich anhand dieser L Merkmale der korrekten Klasse im Training der RF zugeordnet werden kann. Fehler in der punktweisen Klassifikation können sich daher schnell auch auf die Entitäten der höheren Ebene übertragen. Demzufolge ist diese Variante ungeeignet. Eine weitere Möglichkeit besteht darin, die punktweisen Konfidenzen nicht als Merkmale, sondern als zusätzlichen Potentialterm in die Klassifikation der Segmente zu integrieren. Analog zu $\xi^P(p^S)$ von CRF^P können sie in Gleichung 4.7 als unäres Potential berücksichtigt werden, welches für jedes Segment proportional zur klassenweise über alle Segmentpunkte gemittelten Konfidenzen ist. Anhand eines Gewichts lässt sich der relative Einfluss zu den anderen beiden Termen regulieren. Gegen diese Variante spricht das aufwändige generische Modell zur Repräsentation der Segmentinteraktionen. Um den größtmöglichen Nutzen zu erzielen, sind zur zuverlässigen Trennung der maximal L^2 Kategorien ausreichend viele diskriminative Merkmale erforderlich. Die Integration der Auftrittswahrscheinlichkeiten liefert hierbei nutzbringende Informationen. Aus diesem Grund werden die Konfidenzen wie beschrieben als Merkmale anstatt als zusätzliches Potential berücksichtigt.

Trainingsdaten: Für das überwachte Verfahren sind Trainingsdaten zur Optimierung der relativen Potentialgewichte sowie für den Aufbau der drei RF Klassifikatoren erforderlich. Diese sind zeit- und kostenintensiv in ihrer Erstellung. Es ist also wünschenswert, einen möglichst geringen Anteil des Datensatzes manuell annotieren zu müssen. Der Einfluss der Trainingsgebietsgröße wird in Abschnitt 5.5.3 experimentell ermittelt. Für einen operationellen Einsatz des Verfahrens spielt auch die Übertragbarkeit auf andere Untersuchungsgebiete eine wichtige Rolle. So könnte eine repräsentative Szene mit Referenzlabels als Grundlage für die Klassifikation ähnlicher Daten dienen. Im besten Fall lassen sich die Klassifikatoren unmittelbar ohne weitere Anpassungen auf die unbekannten Daten übertragen. Zur Beibehaltung des Genauigkeitsniveaus werden jedoch ähnliche Bedingungen und Parameter (etwa während der Befliegung) vorausgesetzt, um ähnliche Punktdichten, Intensitätswerte usw. zu erhalten,

auf denen die Merkmalsberechnung fußt. Eine hilfreiche Technik zur Adaption antrainierter Klassifikatoren auf neue Szenen kann *Transferlernen* darstellen (z. B. [Paul et al., 2016]). Dies kann zwar u. U. für den neuen Datensatz ebenfalls einige Referenzlabels zur Anpassung eines gelernten Klassifikators an die leicht abweichende Verteilung der Merkmale im Merkmalsraum erfordern, jedoch ist der notwendige Umfang an Trainingsbeispielen geringer. Die Umsetzung einer solchen Herangehensweise steht allerdings nicht im Fokus dieser Arbeit.

Bei der hier vorgestellten Klassifikationsmethodik lässt sich auf Grundlage manuell annotierter Punktwolken üblicherweise eine ausreichend hohe Punktanzahl zum Aufbau des punktweisen RF^P Klassifikators finden. Ungünstiger verhält es sich mit den Zufallsbäumen der Segmentebene, da die Anzahl der zur Verfügung stehenden Primitive im Zuge der Aggregation zu Segmenten reduziert wird. Somit ist die Menge der nutzbaren Trainingsdaten für CRF^S geringer. Diesem Verhalten wird durch die leichte Übersegmentierung mit VCCS entgegengewirkt, das somit im Vergleich zu anderen gängigen Segmentierungsverfahren wie etwa FH [Felzenszwalb & Huttenlocher, 2004] bevorzugt wird. Dies macht sich besonders hinsichtlich des RF_p^S bemerkbar, da zur Modellierung generischer Interaktionen jede mindestens einmal auftretende Beziehung als eigenständige Klasse interpretiert wird, was zu maximal L^2 Kategorien führt. Einige Interaktionen, z. B. weniger dominanter Objektklassen, sind seltener vorhanden. Mit kleiner werdendem Umfang an Trainingsdaten sinkt die Wahrscheinlichkeit, unterschiedliche Merkmalsausprägungen aussagekräftig modellieren zu können. Ist für eine Interaktion zwischen zwei Klassen kein Beispiel in den Trainingsdaten vorhanden, wird die entsprechende Verbundwahrscheinlichkeit auf null gesetzt und die Relation beim Klassifikator RF_p^S nicht in die Gruppe der zu unterscheidende Klassen aufgenommen. Auf diese Weise erklärt es sich, weshalb maximal L^2 anzulernen sind.

Implementierung: Hinsichtlich des Rechenaufwands ist festzuhalten, dass die punktbasierten Merkmale von CRF^P nur einmal vor Beginn der eigentlichen Klassifikationsaufgabe berechnet werden. Entsprechend kann auch der RF^P Klassifikator des unären punktweisen Potentials als Vorverarbeitungsschritt ein einziges Mal antrainiert und für alle Iterationen verwendet werden. Diese Vorgehensweise ermöglicht eine Reduktion des Rechenaufwands. Wenn ausreichend große Mengen an Trainingsdaten zur Verfügung stehen, kann der Klassifikator stattdessen in jedem Schritt erneut angelernt werden. Dies führt einerseits zu einer zusätzlichen Variabilität, da mit großer Wahrscheinlichkeit unterschiedliche Zufallsbeispiele zum Aufbau der Entscheidungsbäume gezogen werden. Andererseits wird jedoch mehr Zeit für die zusätzlichen Trainingsschritte benötigt.

Für die Segmentierung wird das performante VCCS Verfahren verwendet. Dabei führt insbesondere die Einteilung der Punktwolke durch den Octree und die folgende, lokal operierende Segmentierung zu einer im Vergleich zu anderen Methoden guten Laufzeit [Papon et al., 2013].

Die Zusammensetzung der Segmente kann sich in jedem Iterationsschritt ändern. Dementsprechend müssen sowohl die Segmentmerkmale als auch die beiden RF Klassifikatoren des unären bzw. paarweisen Potentials in jeder Wiederholung neu extrahiert bzw. antrainiert werden. Während durch die Segmentierung eine geringere Anzahl an Trainingsbeispielen die Zeit für den Aufbau des unären Klassifikators reduziert, ergibt sich der größte Zeitbedarf aus dem Training des Interaktionspotentials. Durch die Modellierung aller auftretenden Klasseninteraktionen ist eine große Objektklassenzahl erforderlich,

die anhand eines RF zugeordnet werden können muss. Dieser RF mit maximal L^2 Klassen benötigt dementsprechend etwas Zeit. Obwohl sich RF bestens für die Parallelisierung eignen, arbeitet die aktuell verwendete Implementierung seriell. Für das Anlernen der Zufallswälder und die sich anschließende Klassifikation besteht somit Möglichkeit zur weiteren Laufzeitoptimierung.

Neben den Klassifikatoren werden auch die Gewichte $\mathbf{\Omega}^P$ und ω_ψ^S anhand von Trainingsdaten ermittelt. In der initialen Iteration von CRF^P werden zunächst ω_ψ^P und ω_χ^P bestimmt. Die Optimierung des paarweisen Interaktionspotentials auf Segmentebene ω_ψ^S schließt sich im Zuge der initialen Segmentklassifikation an. In der zweiten Iteration ändern sich die berücksichtigten Potentiale von CRF^P, so dass nun ω_ψ^P erneut in Kombination mit ω_ξ^P zur Modellierung des Kontextterms $\xi^P(p^S)$ ermittelt werden muss. Ab diesem Zeitpunkt bleiben alle antrainierten relativen Gewichte für die folgenden Iterationen konstant. Dabei liegt die Annahme zugrunde, dass bereits beim ersten Durchlauf typischerweise Segmente ausreichender Qualität ermittelt werden, um die Relationen zwischen den Objekten möglichst realistisch modellieren zu können.

5 Experimente

Das Kapitel analysiert die neue Methodik auf der Grundlage experimenteller Untersuchungen. In Abschnitt 5.1 werden die Zielsetzungen und Bewertungskriterien zusammengefasst, dann werden die verwendeten Datensätze vorgestellt (Abschnitt 5.2). Eine Voruntersuchung zur Identifikation geeigneter Parameter wird in Abschnitt 5.3 durchgeführt. Danach schließen sich Experimente zur Auswertung der Klassifikationsgenauigkeiten (Abschnitt 5.4) und ausgewählter Aspekte (Abschnitt 5.5) an. Das Kapitel endet mit einer Diskussion in Abschnitt 5.6.

5.1 Ziel und Bewertungskriterien

Die Zielsetzung dieser Arbeit ist die Entwicklung eines Klassifikationsansatzes für luftgestützte Punktwolken unter Berücksichtigung von Kontext. Dabei wird im Folgenden die Eignung des Verfahrens anhand von experimentellen Untersuchungen überprüft, um qualitative und quantitative Aussagen über dessen Potential treffen zu können. Es wird folgenden Fragestellungen nachgegangen:

1) Was sind geeignete Einstellungen für kritische Parameter der RF Klassifikatoren (T, $N_{s,l}$, T_{depth}, T_{split}, M_{split}) und jene der Segmentierung (R_{vox}^{SV}, R_{saat}^{SV}, ω_s^{SV}, ω_n^{SV}, ω_{konf}^{SV})? Welche Merkmale sollten verwendet werden? Ist die Berücksichtigung des Vorwissens aus CRF^P hilfreich für die Qualität der Segmentierung?

Im Rahmen einer Voruntersuchung (Abschnitt 5.3) gilt es, die angesprochenen Aspekte zu untersuchen. Für diese Analysen werden noch nicht die Evaluierungsgebiete herangezogen, auf denen im späteren Verlauf des Kapitels die Genauigkeitsanalysen durchgeführt werden.

2) Welche Genauigkeiten können mit der neuen Methodik erreicht werden? Inwiefern wirkt sich die Integration von Kontext auf die Ergebnisse aus? Wie lassen sich die Resultate im Vergleich zu anderen Verfahren einordnen?

Ein wesentlicher Teil der Experimente befasst sich mit der Qualität der semantischen Zuordnung auf Grundlage realer Daten. Die Ergebnisse sind in Abschnitt 5.4 zusammengestellt. Es wird einerseits aufgezeigt, in welchen Situationen das hierarchische Verfahren Vorteile aufweist. Andererseits werden auftretende Nachteile diskutiert. Der Schwerpunkt der Auswertung liegt auf der Untersuchung des Beitrags von Kontext. Daher stehen nicht ausschließlich die absoluten Klassifikationsgenauigkeiten im Vordergrund, sondern auch die durch Kontext hervorgerufenen Änderungen in Bezug zu einer semantischen Zuordnung ohne Berücksichtigung von Kontext.

Ergebnis / Referenz	Objektart 1	Objektart 2
Objektart 1	t_{11}	t_{12}
Objektart 2	t_{21}	t_{22}

Tabelle 5.1: Aufbau einer Konfusionsmatrix.

3) Wie groß ist der Beitrag einzelner Systemkomponenten der Methodik? Konvergiert das Verfahren? Welchen Einfluss hat die Größe des gewählten Trainingsgebiets? Bietet die Integration externer GIS-Straßendaten einen Vorteil für die Extraktion der neuen Straßenmerkmale?

Abschnitt 5.5 befasst sich mit der Analyse ausgewählter Aspekte, die im Hinblick auf die Methodik und die Anwendbarkeit des Verfahrens als wichtig erachtet werden.

Als wichtige Grundlagen für die Beantwortung der Fragen und die Evaluation der Ergebnisse werden zunächst die herangezogenen Bewertungskriterien erläutert. Eine quantitative Auswertung von Klassifikationsergebnisse basiert auf sogenannten Konfusionsmatrizen. Dazu werden die Objektklassen aller klassifizierten Primitive mit einer Referenzklasse verglichen und die Ergebnisse in einem zweidimensionalen Histogramm aufgetragen. Aus einer solchen Matrix lassen sich Aussagen über die erzielte Genauigkeit in Hinblick auf unterschiedliche Aspekte ableiten. Anhand der beispielhaften Matrix mit zwei Objektklassen (Tabelle 5.1) werden die Maße im Folgenden erläutert. Die Elemente t_{ii} auf der Hauptdiagonalen repräsentierten die korrekten Klassenzuordnungen, während die Fehler auf der Nebendiagonalen (t_{ij} mit $i \neq j$) verzeichnet werden. Um eine Aussage über die allgemeine *Gesamtgenauigkeit* OA (overall accuracy) der Klassifikation anhand eines Wertes treffen zu können, berechnet man häufig den Quotienten aus der Anzahl der korrekten Zuordnungen und der Anzahl aller Primitive $N = \sum_i \sum_j t_{ij}$:

$$\text{Gesamtgenauigkeit OA} = \frac{\sum_i t_{ii}}{N} \cdot 100\,\%. \tag{5.1}$$

Ein ähnliches Maß stellt der *Cohen κ-Koeffizient* dar, der auch als κ-Index bezeichnet wird und die Übereinstimmung zwischen Klassifikationsergebnis und Referenz ausdrückt. Im Vergleich zur Gesamtgenauigkeit wird dieser Wert weniger stark durch die Anzahl der Beispiele pro Klasse beeinflusst, wodurch er das Ergebnis weniger dominanter Klassen besser repräsentiert. Er berechnet sich über

$$\kappa = \frac{p_0 - p_c}{1 - p_c} \cdot 100\% \tag{5.2}$$
$$\text{mit } p_0 = \frac{\sum_i t_{ii}}{N},\ p_c = \frac{\sum_i (\sum_j t_{ji} \cdot \sum_j t_{ij})}{N^2}$$

Dabei ergibt sich die tatsächliche Übereinstimmung p_0 (entsprechend der Gesamtgenauigkeit) aus den Diagonalelementen der Konfusionsmatrix und die erwartete Kongruenz p_c aus den Zeilen- und Spaltensummen.

Mit der Gesamtgenauigkeit und dem κ-Koeffizienten lässt sich zwar die erzielte Güte der Klassifikation einordnen, allerdings gibt der jeweilige Wert keinen Aufschluss über die erreichten Genauigkeiten in Bezug auf einzelne Objektklassen. Es lässt sich also nicht feststellen, ob alle Klassen gleich genau erkannt wurden oder ob bei bestimmten Klassen vergleichsweise viele Fehler zu beobachten sind. Für

diesen Zweck werden daher drei andere Maße herangezogen, welche sich auf einzelne Klassen beziehen:

$$\text{Korrektheit K} = \frac{t_{ii}}{\sum_i t_{ij}} \cdot 100\,\%, \tag{5.3}$$

$$\text{Vollständigkeit V} = \frac{t_{ii}}{\sum_j t_{ij}} \cdot 100\,\%, \tag{5.4}$$

$$\text{Qualität Q} = \frac{\text{V} \cdot \text{K}}{\text{V} + \text{K} - \text{V} \cdot \text{K}} \cdot 100\,\%. \tag{5.5}$$

Die *Korrektheit* K (Gleichung 5.3) beantwortet nach Heipke et al. [1997] die Frage, wie viele Primitive einer bestimmten Klasse im Ergebnis auch in der Referenz zu dieser Klasse gehören. Sie beschreibt damit über den Zusammenhang $100\,\% - K$ die Fehler 1. Art. Analog dazu werden die Fehler 2. Art durch $100\,\% - V$ über die *Vollständigkeit* V (Gleichung 5.4) angegeben. Die Vollständigkeit sagt aus, wie viele Primitive einer Referenzklasse richtig zugeordnet werden [Heipke et al., 1997]. Das dritte klassenbezogene Genauigkeitsmaß stellt die *Qualität* Q (Gleichung 5.5) dar. Sie kombiniert Korrektheit und Vollständigkeit in einem einzigen Wert und ermöglicht somit einen einfachen Vergleich der Ergebnisse [Rutzinger et al., 2009]. Für die Evaluierung im Rahmen eines Benchmarks wird weiterhin der *F1-Score* als harmonisches Mittel aus V und K hinzugezogen:

$$\text{F1} = \frac{2 \cdot V \cdot K}{V + K} \cdot 100\,\%. \tag{5.6}$$

Zur Einschätzung der Rechenzeit sind Informationen über die verwendete Hardware erforderlich. Sofern nicht anders gekennzeichnet, beziehen sich die Werte auf einen Intel Core i7-3820 Prozessor mit 64 GB Arbeitsspeicher. Die entwickelte C++-Implementierung bietet weiteres Potential zur Laufzeitoptimierung, wodurch die angegebenen Werte hinsichtlich der Rechenzeit als Anhaltspunkte zu interpretieren sind. Neben der verfügbaren Hardware ist diese wesentlich von der Komplexität des graphischen Modells (Punktanzahl einer Szene, Definition des Graphen, Anzahl der Merkmale, usw.) abhängig.

5.2 Datensätze

5.2.1 Hannover

Als erster Datensatz zur Evaluierung der neuen Methodik wird ein Ausschnitt einer Laserscanner-Befliegung der Stadt Hannover (Deutschland) verwendet. Es handelt sich um ein Gebiet im Stadtteil Nordstadt, welches im östlichen Bereich eine dichte Bebauung, im südwestlichen Bereich einige Gebäude der Universität und im Norden einen großen mit Vegetation bewachsenen Friedhof umfasst. Der Datensatz wurde im März 2010 bei einer mittleren Flughöhe von 500 m über Grund mit einem Riegl LMS-Q550 Sensor aufgenommen. Das hier zur Untersuchung herangezogene Teilgebiet hat eine Größe von 258.000 m^2 und enthält 2.241.111 Punkte. Dies entspricht einer mittleren Punktdichte von ca. 8,7 Punkte/m^2. Erfasst wurden die Intensitätswerte der Punkte sowie mehrfache Echos eines ausgesandten Pulses.

Klasse	Training $\mathfrak{T}_1$		Validierung $\mathfrak{T}_2$		Test $\mathfrak{E}$		Gesamt	
Boden	214.833	28,6 %	100.443	24,4 %	299.197	27,7 %	614.473	27,4 %
Straße	126.718	16,9 %	69.896	17,0 %	210.946	19,5 %	407.560	18,2 %
Auto	6.486	0,9 %	6.642	1,6 %	15.704	1,5 %	28.832	1,3 %
Vegetation	224.868	30,0 %	77.139	18,8 %	216.700	20,1 %	518.707	23,1 %
Gebäude	177.535	23,6 %	157.024	38,2 %	336.980	31,2 %	671.539	30,0 %
Summe	750.440	100 %	411.144	100 %	1.079.527	100 %	2.241.111	100 %

Tabelle 5.2: Referenzdaten Hannover: Anzahl der Punkte pro Klasse.

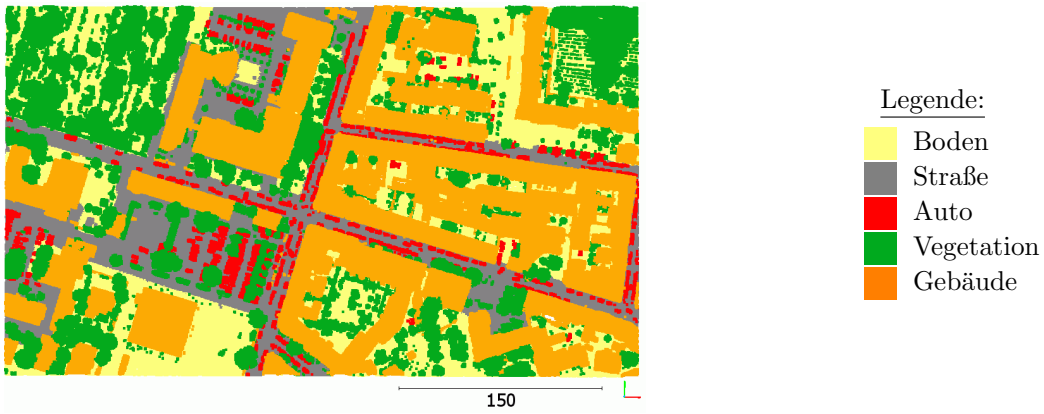


Abbildung 5.1: Testgebiet $\mathfrak{E}$ von Datensatz Hannover. Die Farbgebung entspricht den Referenzklassen.

Der Datensatz wird für die Durchführung der Experimente in drei Gebiete unterteilt. Das Trainingsgebiet $\mathfrak{T}_1$ mit 750.440 Punkten dient zum Anlernen der Klassifikatoren. Auf einem Validierungsbereich $\mathfrak{T}_2$ (411.144 Punkte) werden anschließend die Gewichte optimiert, bevor die eigentliche Klassifikation eines Testgebiets $\mathfrak{E}$ (Abbildung 5.1) mit knapp über einer Million Punkte erfolgt.

Für den Datensatz wurden im Zuge dieser Arbeit alle 3D-Punkte manuell einer der fünf Objektklassen *Boden*, *Straße*, *Auto*, *Vegetation* und *Gebäude* zugeordnet. *Straße* umfasst dabei alle versiegelten Bodenflächen. Eine Übersicht über die Klassenverteilung in den Daten ist in Tabelle 5.2 gegeben. Auffällig ist die geringe Punktanzahl bei der Klasse *Auto*, zu der nur 1,3 % aller Punkte zählen. Sie ist aufgrund der vergleichsweise geringen Auftrittshäufigkeit und der kleinen Objektausdehnungen von Autos unterrepräsentiert in den Daten.

Die Annotationen dienen als Referenzlabels. Sie sind einerseits für das Anlernen der Klassifikatoren ($\mathfrak{T}_1$) sowie der Parameter ($\mathfrak{T}_2$) notwendig und werden andererseits zur Evaluierung von $\mathfrak{E}$ herangezogen. Zum Training des Verfahrens ist eine vollständig gelabelte 3D-Szene hilfreich. Die manuelle Erstellung der Referenzdaten basiert im Wesentlichen auf den Z-Koordinaten und den Intensitäten der Punkte. Optische Luftbilder standen als zusätzliche Informationsquelle nicht zur Verfügung. Demzufolge war die Zuordnung der Laserpunkte zu einer bestimmten Objektklasse nicht in jedem Fall eindeutig möglich und es kann insbesondere an den Objektübergängen vereinzelt zu Fehlern in der Referenz kommen.

Abbildung 5.2: Nördlicher und südlicher Teil des Testgebiets 𝔈 von Datensatz Vaihingen an der Enz. Die Farbgebung entspricht den Referenzklassen.

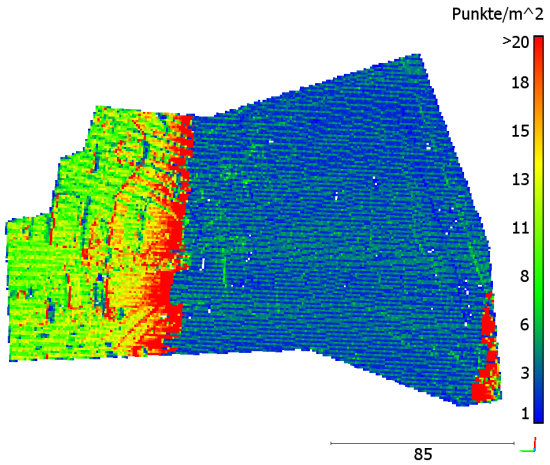


Abbildung 5.3: Variation der Punktdichte im Testgebiet Vaihingen (nördlicher Teil). In rot werden Rasterzellen mit >20 Punkten/m^2 visualisiert. Maximal treten 178 Punkte/m^2 auf.

5.2.2 Vaihingen

Der zweite Datensatz wurde von der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V. (DGPF) bereitgestellt [Cramer, 2010] und umfasst den Innenstadtbereich von Vaihingen an der Enz (Deutschland). Die Laserdaten wurden im Zuge einer Befliegung am 21. August 2008 mit einem Leica ALS50 Sensor aufgenommen. Ein Teil der Szene ist in Abbildung 5.2 dargestellt. Das Sichtfeld betrug 45° und die mittlere Höhe über Grund lag bei 500 m. Charakteristisch für den genutzten Laserscanner ist der oszillierende Spiegel, durch den sich eine linienartige Anordnung der 3D-Punkte auf der Oberfläche ergibt. Insbesondere an den Rändern eines Flugstreifens führt die Richtungsänderung zu einer sehr hohen Punktanzahl. Betrachtet man jeden Flugstreifen für sich, so beträgt die Punktdichte im Schnitt etwa 4 Punkte/m^2. Verursacht durch die doppelte Aufnahme in den Bereichen der 30 % Querüberdeckung wird der Effekt der lokal signifikant variierenden Punktdichten zusätzlich deutlich verstärkt (Abbildung 5.3). Maximal wurden pro m^2 bis zu 178 Punkte registriert. Der Median über das gesamte Stadtgebiet von Vaihingen liegt bei 6,7 Punkten/m^2. Aufgezeichnet wurden sowohl die Intensitätswerte des Lasersignals als auch Mehrfachechos. Die Anzahl von Mehrfachreflexionen ist durch die Befliegung im Sommer und die damit verbundene dichte Belau-

Klasse	Training $\mathfrak{T}_1$		Validierung $\mathfrak{T}_2$		Test $\mathfrak{E}$		Gesamt	
Boden	116.444	22,1 %	59.219	26,0 %	98.690	24,0 %	274.353	23,5 %
Straße	145.816	27,7 %	48.499	21,3 %	101.986	24,8 %	296.301	25,4 %
Auto	4.310	0,8 %	1.202	0,5 %	3.708	0,9 %	9.220	0,8 %
Vegetation	130.212	24,8 %	68.310	30,0 %	86.466	21,0 %	284.988	24,5 %
Gebäude	129.185	24,6 %	50.679	22,2 %	120.872	29,4 %	300.736	25,8 %
Summe	525.967	100 %	227.909	100 %	411.722	100 %	1.165.598	100 %

Tabelle 5.3: Referenzdaten Vaihingen: Anzahl der Punkte pro Klasse.

bung der Vegetation im Vergleich zum Datensatz Hannover gering. Bei Bäumen wird somit die meiste Energie bereits von der Baumkrone reflektiert, so dass dort die meisten (und oft einzigen) Echos eines ausgesandten Laserpulses detektiert werden.

Für die Experimente wird ein Ausschnitt herangezogen, welcher in ein etwa 84.250 m^2 großes Trainingsgebiet $\mathfrak{T}$ bestehend aus 753.876 Punkten und ein 411.722 Punkte umfassendes Testgebiet $\mathfrak{E}$ mit einer Fläche von ca. 46.700 m^2 aufgeteilt ist. Die Punktdichte beträgt somit im Schnitt knapp 8,9 Punkte/m^2. Beide Ausschnitte umfassen im wesentlichen Wohngebiete mit unterschiedlichen Bauweisen, welche von offen (Einfamilienhäuser) über Wohnblöcke bis hin zu einer dichteren Bebauung im Innenstadtbereich variieren. Weiterhin sind Straßen, Rasenflächen, Autos, niedrige Vegetation und Bäume in den Szenen enthalten. In dieser Arbeit wird das Trainingsgebiet nochmals geteilt, um einen Validierungsdatensatz $\mathfrak{T}_2$ für die Parameteroptimierung zu erhalten. Die Klassifikatoren werden demnach auf dem Trainingsdatensatz $\mathfrak{T}_1$ mit 525.967 Punkten angelernt und die Optimierung erfolgt auf dem 227.909 Punkte umfassenden Validierungsdatensatz $\mathfrak{T}_2$. Das Testgebiet $\mathfrak{E}$ besteht aus zwei räumlich von einander getrennten Regionen in Vaihingen, die gemeinsam ausgewertet werden. Zur Visualisierung werden die beiden Teile in Abbildung 5.2 einzeln dargestellt.

Es wird eine Einteilung in dieselben fünf Objektklassen wie bei Hannover (*Boden*, *Straße*, *Auto*, *Vegetation*, *Gebäude*) vorgenommen, wodurch eine Vergleichbarkeit ermöglicht wird. Die Erzeugung der Referenzlabels für alle Gebiete erfolgte durch das manuelle Annotieren aller Laserpunkte beruhend auf deren Z-Koordinaten und Intensitäten. Zusätzlich wurde ein Falschfarben-Orthophoto mit einer Bodenauflösung von 8 cm zu Hilfe genommen. Die Aufnahme erfolgte jedoch nicht zeitgleich mit der Laserbefliegung und bildet die Szene nur zweidimensional ab, so dass die optischen Daten nur als Unterstützung herangezogen wurden. Die Verteilung der Objektklassen kann Tabelle 5.3 entnommen werden. Auch in diesem Fall ist die Klasse *Auto* mit nur 0,8 % aller Punkte stark unterrepräsentiert. Die anderen vier Klassen sind in etwa gleichverteilt.

5.2.3 Benchmarks

Neben den beiden Datensätzen Hannover und Vaihingen, auf denen die meisten Untersuchungen beruhen, werden für das Experiment in Abschnitt 5.4.2 Benchmarkdaten zum Vergleich der erzielten Ergebnisse mit anderen Verfahren herangezogen.

a) GML: Dieser Vergleichsdatensatz besteht aus zwei Gebieten GML A (Abbildung 5.4) und GML B (Abbildung 5.5), deren Referenzdaten von Shapovalov et al. [2010] über das Graphics & Media Lab

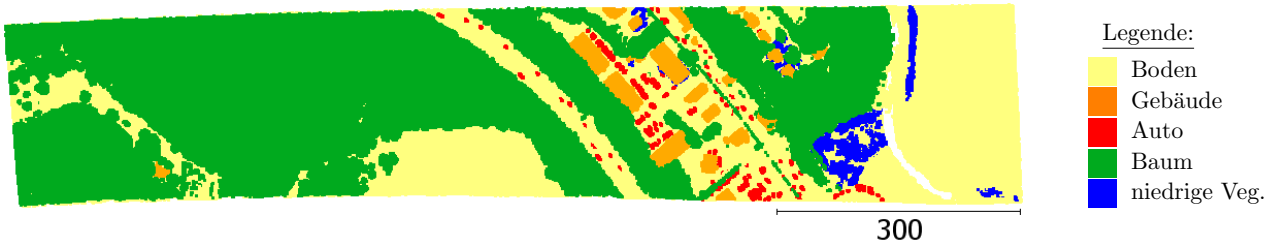


Abbildung 5.4: Testgebiet $\mathfrak{E}$ von Datensatz GML A. Die Farbgebung entspricht den Referenzklassen.

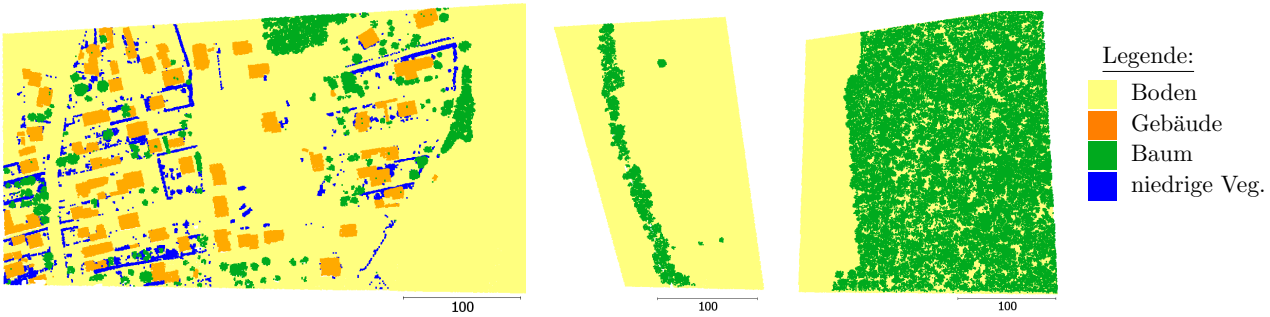


Abbildung 5.5: Testgebiet $\mathfrak{E}$ von Datensatz GML B. Die Farbgebung entspricht den Referenzklassen.

(GML) der Moscow State University zur Verfügung gestellt werden[1]. Die Szene weist durch die freien Bodenflächen und Wälder einen eher ländlichen Charakter mit einigen einzeln stehenden Gebäuden eines Dorfes auf. Die Punktdichte beträgt etwa 2,2 (GML A) bzw. 6,1 Punkte pro m^2 (GML B) bei einer Fläche von ca. 937.419 m^2 bzw. 442.600 m^2. Die Daten erfasste ein ALTM 2050 von Optech Inc. Es handelt sich ausschließlich um die 3D-Koordinaten der Punkte. Informationen bezüglich Intensität, Echo usw. sowie Bilddaten sind nicht vorhanden. Jeder der beiden Bereiche ist zu etwa gleichen Teilen in ein Trainings- und ein Evaluierungsgebiet aufgeteilt. Zum Anlernen der Parameter wird ein Teilbereich des Trainingsgebiets als Validierungssatz $\mathfrak{T}_2$ zurückgehalten. Die Referenzlabels stehen für alle Punkte zur Verfügung und die Auswertung erfolgt eigenverantwortlich. Bei der Interpretation der Ergebnisse sollte das bedacht werden, da die Teilnehmer u. U. leicht unterschiedliche Berechnungsweisen (z. B. bezüglich Rundung) verwendet haben könnten. Beide Teildatensätze GML A und B bestehen aus den Klassen *Boden*, *Gebäude*, *Baum* und *niedrige Vegetation*. Zusätzlich beinhaltet GML A Punkte, die der Klasse *Auto* zugeordnet sind, siehe Tabelle 5.4.

b) Vaihingen ISPRS 9 Klassen: Die Punktwolken von Vaihingen sind ebenfalls Teil des *ISPRS Test Project on Urban Classification, 3D Building Reconstruction and Semantic Labeling* [Rottensteiner et al., 2014]. Dabei handelt es sich um einen von der ISPRS Arbeitsgruppe *3D Szenenanalyse* organisierten Vergleich aktueller Methoden zur Klassifikation und Objektrekonstruktion in Fernerkundungsdaten. Die *3D Labelling challenge* ermöglicht eine einheitliche und vergleichbare Evaluierung der klassifizierten 3D-Punktwolken von Vaihingen, indem die Ergebnisse von den Organisatoren ausgewertet und veröffentlicht[2] werden.

Der Benchmark basiert auf den Trainings- und Evaluierungsdaten, wie sie bereits in Abschnitt 5.2.2 beschrieben wurden. In diesem Fall erfolgt die Zuordnung der Punkte zu den neun Klassen *Stromlei-*

[1] öffentlich verfügbar unter `http://graphics.cs.msu.ru/en/science/research/3dpoint/classification`
[2] `http://www2.isprs.org/vaihingen-3d-semantic-labeling.html`

Klasse	Training $\mathfrak{T}_1$		Validierung $\mathfrak{T}_2$		Test $\mathfrak{E}$		Gesamt	
Boden	512.654	51,2 %	44.488	61,2 %	439.989	43,9 %	997.131	48,0 %
Gebäude	86.968	8,7 %	11276	15,5 %	19.592	2,0 %	117.836	5,7 %
Autos	1.704	0,2 %	129	0,2 %	3.235	0,3 %	5.068	0,2 %
Baum	368.127	36.7 %	14.130	19,4 %	532.094	53,0 %	914.351	44.0 %
niedrige Veg.	32.419	3,2 %	2.674	3,7 %	7.758	0,8 %	42.851	2,1 %
Summe	1.001.872	100 %	72.697	100 %	1.002.668	100 %	2.077.237	100 %

(a) Datensatz GML A

Klasse	Training $\mathfrak{T}_1$		Validierung $\mathfrak{T}_2$		Test $\mathfrak{E}$		Gesamt	
Boden	1.099.364	80,7 %	141.896	78,0 %	977.787	84,2 %	2.219.047	82,0 %
Gebäude	120.897	8,9 %	26.646	14.7 %	54.532	4,7 %	202.075	7,5 %
Baum	104.204	7,6 %	4.770	2,6 %	111.075	9,6 %	220.049	8,1 %
niedrige Veg.	38.429	2,8 %	8.454	4,7 %	17.323	1,5 %	64.206	2,4 %
Summe	1.362.894	100 %	181.766	100 %	1.160.717	100 %	2.705.377	100 %

(b) Datensatz GML B

Tabelle 5.4: Referenzdaten GML: Anzahl der Punkte pro Klasse.

tung, *niedrige Vegetation*, *versiegelte Fläche*, *Auto*, *Zaun/Hecke*, *Dach*, *Fassade*, *Strauch* und *Baum*. Dabei entspricht *niedrige Vegetation* bei der in dieser Dissertation hauptsächlich verwendeten Klassendefinition der Kategorie *Boden*. Die Referenzlabels des Evaluierungsgebiets $\mathfrak{E}_{9K}$ sind unbekannt, so dass diese in der Übersicht über die Auftrittshäufigkeiten der einzelnen Klassen in Tabelle 5.5 nicht angegeben werden können. Der im Folgenden als *Vaihingen ISPRS 9 Klassen* gekennzeichnete Datensatz wird ausschließlich in Abschnitt 5.4.2 zum Vergleich mit anderen Methoden herangezogen. Alle weiteren Experimente basieren auf dem in fünf Klassen unterteilten Datensatz Vaihingen.

Die wesentlichen Eigenschaften der verwendeten Datensätze sind in Tabelle 5.6 zusammengefasst.

Klasse	Training $\mathfrak{T}_1$		Validierung $\mathfrak{T}_2$		Gesamt	
Stromleitung	594	0,1 %	390	0,2 %	984	0,1 %
niedrige Vegetation	116.444	22,1 %	59.219	26,0 %	175.663	23,3 %
versiegelte Fläche	145.816	27,7 %	48.499	21,3 %	194.315	25,8 %
Auto	4.310	0,8 %	1.202	0,5 %	5.512	0,7 %
Zaun	9.291	1,8 %	3.547	1,6 %	12.838	1,7 %
Dach	107.465	20,4 %	43.763	19,2 %	151.228	20,1 %
Fassade	21.126	4,0 %	6.526	2,9 %	27.652	3,7 %
Strauch	29.318	5,6 %	22.273	9,8 %	51.591	6,8 %
Baum	91.603	17,4 %	42.490	18,6 %	134.093	17,8 %
Summe	525.967	100 %	227.909	100 %	753.876	100 %

Tabelle 5.5: Vaihingen ISPRS 9 Klassen: Anzahl der Punkte pro Klasse. Bekannt sind nur die Klassen der Trainingsdaten.

	Hauptdatensätze		**Benchmark Datensätze**		
Bezeichnung	**Hannover**	**Vaihingen**	**GML A**	**GML B**	**Vaihingen ISPRS 9 Klassen**
Szene	städtisch	städtisch	ländlich	ländlich	städtisch
Sensor	Riegl LMS-Q550	Leica ALS50	Optech ALTM 2050	Optech ALTM 2050	Leica ALS50
Flughöhe	500 m	500 m	unbekannt	unbekannt	500 m
Aufnahme	März 2010	Aug. 2008	unbekannt	unbekannt	Aug. 2008
Intensität	x	x	-	-	x
Echonummer	x	x	-	-	x
Anzahl Echos	x	x	-	-	x
Fläche	258.000 m^2	130.950 m^2	937.419 m^2	442.600 m^2	130.950 m^2
Punktanzahl	2.241.111	1.165.598	2.077.237	2.705.377	1.165.598
davon $\mathfrak{T}$	1.161.584	753.876	1.074.569	1.544.660	753.876
davon $\mathfrak{E}$	1.079.527	411.722	1.002.668	1.160.717	411.722
mittlere Punktdichte	8,7 Pkt/m^2	8,9 Pkt/m^2	2,2 Pkt/m^2	6,1 Pkt/m^2	8,9 Pkt/m^2
Referenzlabels	manuell erstellt		von Shapovalov et al. [2010]		von ISPRS
bekannt für $\mathfrak{T}$	x	x	x	x	x
bekannt für $\mathfrak{E}$	x	x	x	x	unveröffentlicht
Klassenanzahl	5	5	5	4	9
Klassen	*Boden* *Straße* *Auto* *Vegetation* *Gebäude*	*Boden* *Straße* *Auto* *Vegetation* *Gebäude*	*Boden* *Gebäude* *Auto* *Baum* *niedrige Veg.*	*Boden* *Gebäude* *Baum* *niedrige Veg.*	*Stromleitung* *niedrige Veg.* *versiegelte Fl.* *Auto* *Zaun/Hecke* *Dach* *Fassade* *Strauch* *Baum*
Eigenschaften		stark variierende Punktdichten; wenige Mehrfachechos	nur Koordinaten aufgezeichnet; absolute Georeferenzierung nicht bekannt		Referenz von $\mathfrak{E}$ nicht öffentlich, Auswertung durch ISPRS

Tabelle 5.6: Übersicht über alle verwendeten Datensätze.

5.3 Voruntersuchung: Merkmalsauswahl und Parameterbestimmung

Im Fokus des folgenden Abschnitts stehen die Auswahl der wichtigsten Merkmale für die Klassifikation sowie die Bestimmung geeigneter Werte für die manuell festzulegenden Parameter des Verfahrens. Zwischen beiden Gruppen liegt eine gegenseitige Abhängigkeit vor, da beispielsweise die Segmentierungsparameter die Segmentformen beeinflussen, welche wiederum zu unterschiedlichen Einflüssen der jeweiligen Segmentmerkmale führen können. Zur Approximation des Suchraums werden die Experimente daher in der Reihenfolge 1) *Bestimmung der punktweisen Merkmale*, 2) *Bestimmung der Parameter des unären RF Klassifikators*, 3) *Bestimmung der Parameter für die VCCS Segmentierung*, 4) *Bestimmung der segmentweisen Merkmale (unär und paarweise)* und 5) *Bestimmung der Parameter beider segmentweisen RF Klassifikatoren (unär und paarweise)* durchgeführt und ausgewertet. Die Beschreibung der Ergebnisse erfolgt hingegen gruppiert nach den Merkmalen (Abschnitt 5.3.1) und den Parametern (Abschnitt 5.3.2). In Abschnitt 5.3.2 werden weiterhin alle zugrundeliegenden Parametereinstellungen zusammengefasst. Für die Experimente werden die RF Klassifikatoren (für die fünf Objektklassen *Boden*, *Straße*, *Auto*, *Vegetation* und *Gebäude*) auf einem Trainingsgebiet $\mathfrak{T}_1$ angelernt. Die Auswertung erfolgt anschließend auf dem zweiten, zunächst zurückgehaltenen Validierungsgebiet $\mathfrak{T}_2$. Das eigentliche Testgebiet ($\mathfrak{E}$) wird für dieses Experiment noch nicht herangezogen, um einer Überanpassung bei der später folgenden Performance-Analyse des Verfahrens entgegenzuwirken.

5.3.1 Merkmale

Das erste vorgestellte Experiment befasst sich mit der Ermittlung der besten Merkmale aus der zur Verfügung stehenden und in Abschnitt 3.1.2 beschriebenen Gesamtmenge. Die Auswahl basiert auf der Wichtigkeit der Merkmale, wie sie sich aus dem RF Klassifikator ergibt. Eine Beschränkung auf wenige, aber besonders diskriminative Merkmale hat zwei wesentliche Vorteile. Einerseits wird dem Problem des in Abschnitt 3.1.2 geschilderten *Hughes-Phänomens* entgegengewirkt. Dies ist sinnvoll, da einige der vorselektierten Merkmale miteinander korreliert sein können. Die Genauigkeit des Ergebnisses soll dadurch nicht negativ beeinflusst werden. Andererseits reduzieren sich durch die Beschränkung auf eine Untermenge an Merkmalen sowohl der erforderliche Speicher- als auch der Rechenzeitbedarf. Letztgenannter Aspekt bezieht sich insbesondere auf das Training des Klassifikators.

Zusätzlich zu den guten Klassifikationsergebnissen besteht ein weiterer Vorteil des RF Klassifikators in der Möglichkeit, auf einfache Weise die Wichtigkeit der beteiligten Merkmale zu ermitteln (Abschnitt 3.2). Einschränkend ist festzuhalten, dass sich die Merkmalswichtigkeiten auf den jeweils verwendeten RF Klassifikator beziehen. Für einen anderen Klassifikatortyp können entsprechend andere Merkmale bedeutender sein. Da in dieser Arbeit jedoch der RF Verwendung findet, kann diese einfache Methode zur Identifikation der Wichtigkeit genutzt werden.

Zur Evaluation der Merkmalswichtigkeiten wird zunächst eine initiale Klassifikation vorgenommen. Hinsichtlich der resultierenden Wichtigkeitsangaben erfolgt eine Sortierung der Merkmale in absteigender Reihenfolge. Nun werden aufeinanderfolgende Klassifikationen mit Teilmengen der Merkmale durchgeführt, wobei die Teilmenge in jeder Klassifikation um das nächst wichtigere Merkmal ergänzt wird. Zur Reduktion von Ausreißern wird die Klassifikation mit demselben Merkmalssatz fünf Mal

wiederholt. Die nachfolgenden Abbildungen zu dieser Untersuchung repräsentieren jeweils den Mittelwert. Da der κ-Koeffizient sensibler auf Änderungen bei weniger dominanten Klassen reagiert als die Gesamtgenauigkeit, wird dieser als Qualitätsmaß herangezogen. Insgesamt werden in den nächsten Unterabschnitten die wesentlichen Merkmale der drei verwendeten RF Klassifikatoren (unäre Potentiale von CRF^P und CRF^S sowie paarweises Interaktionsmodell CRF^S) ermittelt.

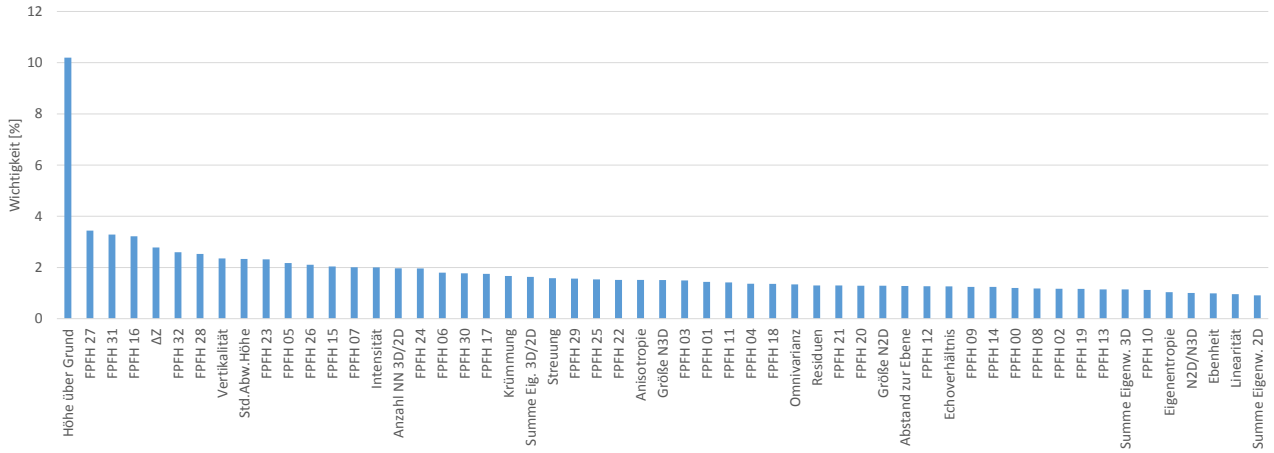
Merkmale der 3D-Punkte

Für die Berechnung der meisten Merkmale ist zunächst die Definition einer Nachbarschaft erforderlich. Sie ergibt sich bei beiden Testgebieten als $\mathcal{N}_{3D}^k$ mit $k \in [10; 400]$ nach dem von Weinmann [2016] vorgeschlagenen Verfahren. Um die hohe Variabilität der Punktdichte (speziell in Vaihingen) zu berücksichtigen, wird ein großes Intervall mit den manuell festgelegten Grenzwerten $k_{min} = 10$ und $k_{max} = 400$ zugelassen. Für die Merkmale beruhend auf einem festen Radius wird $r = 1,25\,\text{m}$ als typischer Wert für luftgestützte Laserdaten mit einer ähnlichen durchschnittlichen Punktdichte [Mallet, 2010] festgelegt. Die robuste Berechnung der Punktnormalen nach [Boulch & Marlet, 2012] betrachtet jeweils die nächstgelegenen $k_{NN}^{Normal} = 25$ Punkte.

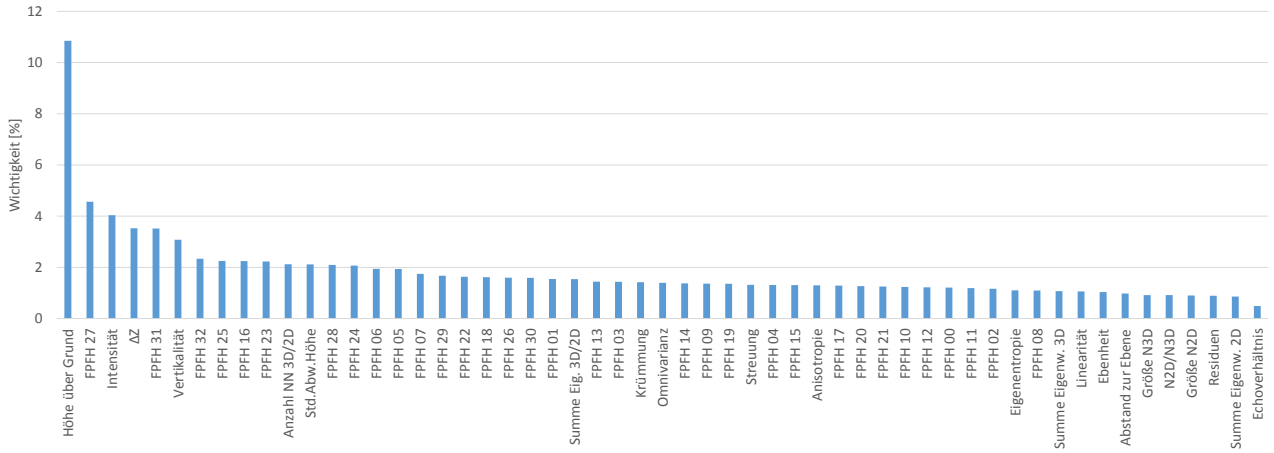
Eine Analyse der Wichtigkeit aller 55 Punktmerkmale zeigt bei beiden Testgebieten ein ähnliches Verhalten. In Abbildung 5.6 sind die Relevanzwerte in absteigender Reihenfolge sortiert für beide Datensätze dargestellt. Für die beiden Gebiete Hannover und Vaihingen wird jeweils der Satz an optimalen Merkmalen zugrunde gelegt, bei der ein sehr gutes Genauigkeitsmaß erlangt wird. Die Hinzunahme weiterer Merkmale führt nicht zu einer deutlichen Verbesserung der Genauigkeit. Die Größe dieser Merkmalssätze wird anhand der Ergebnisse in Abbildung 5.7 ermittelt. Die schwarze Kurve in Abbildung 5.7a zeigt die Sättigung der Genauigkeit bezüglich des κ-Index am Beispiel Hannover. Ab 28 Merkmalen bleibt das Genauigkeitsniveau etwa konstant. Für Vaihingen tritt dieser Effekt ab etwa 23 Merkmalen ein, wie anhand der grauen Kurve in Abbildung 5.7b zu erkennen ist. Eine Schnittmenge von 21 zeigt auf, dass die meisten der auf diese Weise identifizierten Merkmale in beiden Datensätzen von Bedeutung sind (Abbildung 5.6). Um zu gewährleisten, dass alle folgenden Tests auf dem gleichen Merkmalssatz basieren, wird die ausgewählte Menge um die jeweils im anderen Test ebenfalls als wichtig identifizierten Merkmale ergänzt.

Folgende, nach absteigender (gemittelter) Wichtigkeit sortierte Merkmale ergeben sich für den RF des unären Potentials in CRF^P: *Höhe über Grund*, $FPFH_{27}$, $FPFH_{31}$, *max. Höhenunterschied* (ΔZ), *Intensität*, $FPFH_{16}$, *Vertikalität*, $FPFH_{32}$, $FPFH_{28}$, $FPFH_{23}$, *Standardabweichung Höhe*, $FPFH_{05}$, *Verhältnis Punktanzahl 3D/2D*, $FPFH_{24}$, $FPFH_{25}$, $FPFH_{07}$, $FPFH_{06}$, $FPFH_{26}$, $FPFH_{30}$, $FPFH_{15}$, $FPFH_{29}$, *Verhältnis Summe Eigenwerte* 3D/2D, $FPFH_{22}$, *Krümmung*, $FPFH_{17}$, $FPFH_{01}$, $FPFH_{18}$, *Streuung*, *Anisotropie* sowie *Radius Nachbarschaft 3D*. Insgesamt ergibt sich auf Grundlage der Analyse ein Merkmalsvektor $\mathbf{f}_i^P(\mathbf{x})$ bestehend aus 30 Komponenten.

Es wird ersichtlich, dass die *Höhe über Grund* mit 10,2 % (Hannover) bzw. 10,8 % (Vaihingen) das mit Abstand wichtigste Merkmale darstellt. Weiterhin zeigt sich die große Bedeutung vieler FPFH-Komponenten, welche die lokalen Richtungen der Normalenvektoren als Histogramm repräsentieren. Das vergleichsweise schlechte Abschneiden der auf den Eigenwerten basierenden Merkmalsgruppe (*Summe der Eigenwerte in 2D bzw. 3D*, *Eigenentropie*, *Ebenheit*, *Linearität*) ist vermutlich darauf

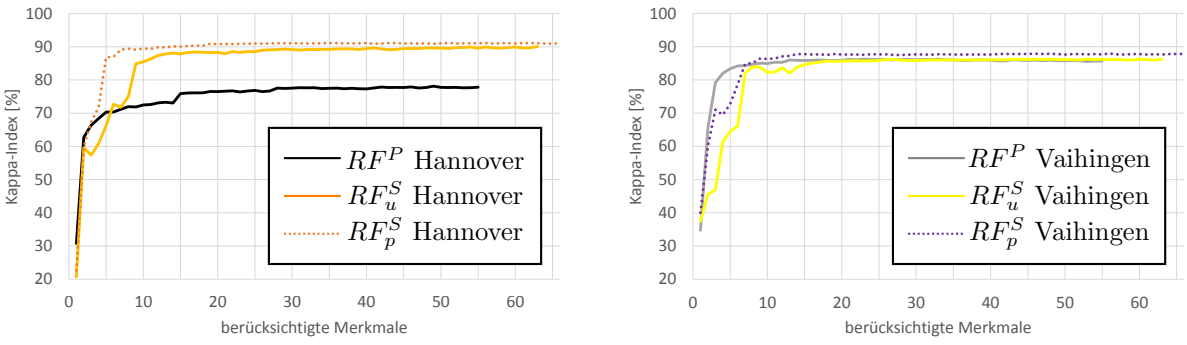


(a) Hannover



(b) Vaihingen

Abbildung 5.6: Relevanz der Merkmale des punktweisen RF (unäres Potential). Die Balkenhöhe entspricht den Wichtigkeitswerten und dient als Grundlage zur Sortierung in abnehmender Reihenfolge von links nach rechts.



(a) Hannover

(b) Vaihingen

Abbildung 5.7: Klassifikationsgenauigkeiten (κ-Index) durch iterative Hinzunahme der nach Wichtigkeit sortierten Merkmale. Gezeigt werden das unäre Potential von CRF^P (repräsentiert durch RF^P) sowie das unäre (RF_u^S) und paarweise Potential (RF_p^S) von CRF^S. Die Segmente für CRF^S wurden mit den in Abschnitt 5.3.2 ermittelten Parametern generiert.

zurückzuführen, dass die entsprechenden geometrischen Beziehungen zwischen den Laserpunkten in einer lokalen Nachbarschaft bereits teilweise durch die FPFH-Merkmale mit beschrieben werden.

Da sowohl die paarweisen Interaktionen als auch die Cliquen höherer Ordnung die Merkmale von Punkten in einer begrenzten lokalen Nachbarschaft miteinander vergleichen, werden für diesen Zweck jeweils nur die beiden punktspezifischen Merkmale *Höhe über Grund* und *Intensität* herangezogen. Wie die vorhergehende Untersuchung gezeigt hat, ist das Echoverhältnis als weiteres Merkmal individueller Punkte nicht sehr diskriminativ. Für die Berechnung aller anderen Merkmale wird bereits eine lokale Nachbarschaft berücksichtigt. Die euklidische Distanz der Merkmale zweier unmittelbar benachbarten Punkte ist in diesem Fall demzufolge gering, was entsprechend des kontrast-sensitiven und P^n Potts Modells tendenziell zu einer stärkeren Glättung der Klassenlabels führen kann. Die beiden für die Bestimmung der paarweisen Interaktionen verwendeten Merkmale werden auch zur Ermittlung der Qualität einer Clique herangezogen (γ_{max} in Gleichung 4.4).

Merkmale der 3D-Segmente

Die meisten Punktmerkmale werden aus einer lokal stark begrenzten Nachbarschaft extrahiert, aus der u. U. nur unzureichende Informationen für eine korrekte Klassifikation bereitstellt werden können. Die Verwendung von Segmenten hingegen ermöglicht die Nutzung von Merkmalen, welche einen größeren geometrischen Bereich beschreiben und dadurch ggf. die Probleme der punktweisen Merkmale kompensieren können. Bei diesem Experiment werden die Segmente basierend auf den in Abschnitt 5.3.2 ermittelten Parametern generiert. Für CRF^S müssen nicht nur die Knotenmerkmale, sondern auch Interaktionsmerkmale zwischen adjazenten Segmenten extrahiert werden. Zum Aufbau des Graphen gelten zwei Primitive als benachbart, wenn es mindestens zwei Punkte beider Segmente gibt, deren euklidischer 2D-Abstand kleiner als d^S_{graph}=1 m ist. Die Anzahl der Nachbarn kann dabei im Gegensatz zu CRF^P variieren. Durch den geringen Wert für d^S_{graph} werden nur unmittelbar benachbarte Primitive als adjazent angesehen. Dennoch wird im Vergleich zu CRF^P eine größere räumliche Ausdehnung durch die Segmente abgedeckt.

Für die Bestimmung der Nachbarschaftsklassenverteilungsmerkmale wird ebenfalls ein kleiner Wert herangezogen ($r_{hist} = 1\,m$), damit die Auftrittswahrscheinlichkeiten insbesondere in städtischen Gebieten mit einer hohen Objektdichte nicht durch eine zu große Nachbarschaft verfälscht werden. Zur Extraktion der kontextbasierten Straßenmerkmale ist ein weiterer Parametersatz erforderlich (Abschnitt 4.4.2). Auf Grundlage einer empirischen Analyse werden folgende Werte gesetzt: $res_{grid} = 1\,\mathrm{m}$, $S_{closing} = 3 \times 3$ Pixel, sowie für die Hough-Transformation $H^{res}_{dist} = 1\,\mathrm{m}$, $H^{res}_{\rho} = 5°$, $H_{lineGap} = 2\,\mathrm{m}$, $H_{minVotes} = 30$ und $H_{minLength} = 15\,\mathrm{m}$.

Für CRF^S werden im generischen Interaktionsmodell alle Klassenrelationen angelernt. Aufgrund der dadurch entstehenden höheren Klassenanzahl ist ein ausreichend diskriminativer Merkmalsvektor $\mathbf{f}^S_{ij}(\mathbf{x})$ für jede Kante notwendig, wodurch sich diese Klassifikation deutlich von den paarweisen Interaktionen in CRF^P unterscheidet. Der häufig für einfache Interaktionsmodelle verwendete Ansatz der Differenzbildung beider Merkmalsvektoren, also $\mathbf{f}^S_{ij}(\mathbf{x}) = [\mathbf{f}^S_i(\mathbf{x}) - \mathbf{f}^S_i(\mathbf{x})]$, ist in diesem Fall nicht ausreichend. In den meisten Fällen weisen benachbarte Segmente ähnliche Merkmale auf, so dass die Differenz Werte nahe null ergibt. Ein solcher (Null-)Vektor ist für die Klassifikation von Kanten nur

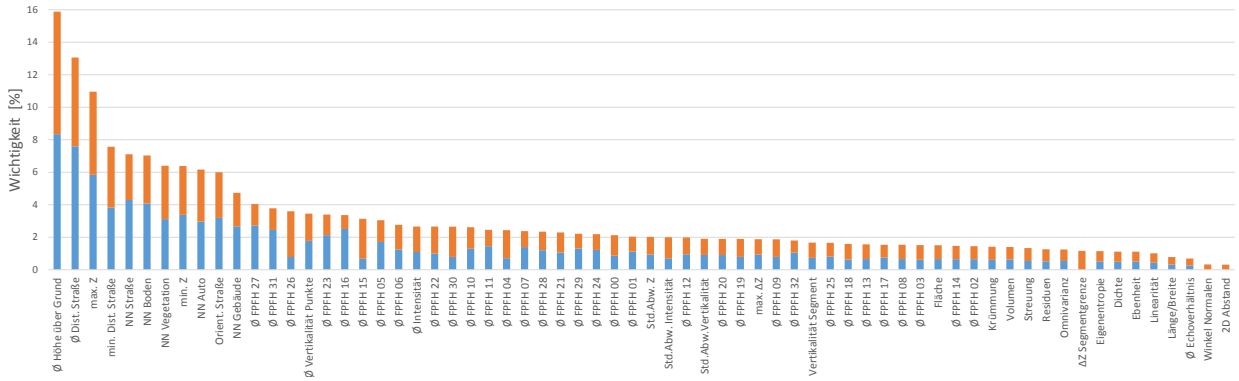
wenig aussagekräftig [Niemeyer et al., 2014]. Hilfreich sind beispielsweise nicht nur die Höhenunterschiede, sondern auch die absoluten Höhenwerte. Nur mit dieser Information kann z. B. unterschieden werden, ob zwei benachbarte Segmente den Boden oder das Flachdach eines Gebäudes beschreiben. Folglich ist ein umfangreicherer Merkmalsvektor erforderlich. In dieser Arbeit wird $\mathbf{f}_{ij}^{S}(\mathbf{x})$ durch das Verketten beider adjazenter Knotenmerkmalsvektoren $\mathbf{f}_{ij}^{S}(\mathbf{x}) = [\mathbf{f}_{i}^{S}(\mathbf{x}); \mathbf{f}_{i}^{S}(\mathbf{x})]^{T}$ gebildet. Weiterhin wird der resultierende Vektor um drei weitere, ausschließliche Interaktionsmerkmale ergänzt. Dabei handelt es sich um den *Winkel zwischen den Segmentnormalen*, den *minimalen 2D-Abstand der Punkte beider Segmente* sowie den *Höhenunterschied an den Segmentgrenzen.*

Auf der segmentbasierten Ebene werden zur Modellierung des unären und paarweisen Interaktionspotentials zwei unabhängige RF Klassifikatoren verwendet. Dementsprechend soll die im Rahmen dieser Untersuchung erfolgte Merkmalsauswahl für beide Fälle möglichst diskriminativ sein. Zunächst werden dazu für beide Datensätze die Wichtigkeiten der Merkmale für den unären und paarweisen RF in fünffacher Versuchswiederholung getrennt bestimmt. Um eine Selektion vorzunehmen, werden anschließend die Summen der Wichtigkeiten gebildet; es ergeben sich die in Abbildung 5.8 dargestellten Diagramme für Hannover und Vaihingen. Die Balkenhöhe setzt sich aus dem Anteil des unären und paarweisen Potentials zusammen. Zur Auswertung der Segmentinteraktionsmerkmale werden die Wichtigkeiten doppelt einfließender Merkmale, wie sie durch das Aneinanderfügen der Merkmalsvektoren entstehen, zuvor aufsummiert und bei der folgenden Analyse als ein einziges Merkmal behandelt.

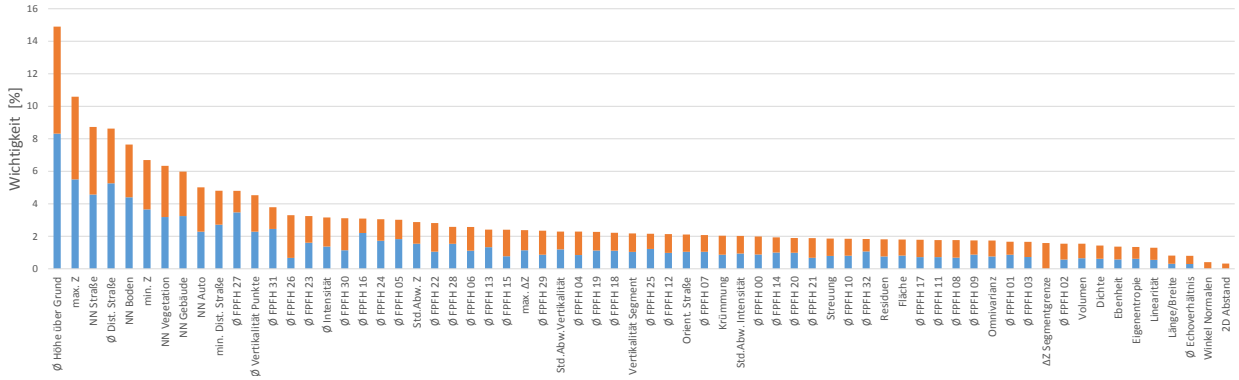
Es zeigt sich, dass für beide Datensätze vor allem Merkmale bezüglich der Höhe (*mittlere, max.* und *min. Höhe*) sowie die neu vorgestellten Kontextmerkmale (Nachbarschaftsklassenverteilungen (gekennzeichnet durch NN in Abbildung 5.8) und Straßenmerkmale) die höchsten Wichtigkeitswerte aufweisen. Dies lässt den positiven Einfluss der Kontextmerkmale erkennen. Von geringer Bedeutung ist die Gruppe der Eigenwertmerkmale. Vermutlich sind die Supervoxel in den meisten Fällen zu symmetrisch aufgebaut, so dass die Punktverteilung innerhalb eines Supervoxels nur wenig aussagekräftig ist. Aufgrund der zu geringen Anzahl an Mehrfachechos liefert auch das durchschnittliche Echoverhältnis nur einen minimalen Beitrag zur Trennung der Objektklassen.

Zusätzlich werden für beide Potentiale mit den nach Relevanz sortierten und sukzessive hinzugefügten Merkmalen die Klassifikationsgenauigkeiten in Form des segmentbezogenen κ-Koeffizienten ausgewertet. Sie sind in Abbildung 5.7a als orangefarbene und braun punktierte Kurven für Hannover bzw. als gelbe und violett punktierte Kurven in Abbildung 5.7b für Vaihingen dargestellt. Auf der x-Achse wird die Anzahl der berücksichtigten Merkmale angegeben. Wichtig ist, dass die Merkmale des paarweisen RF bis auf wenige Ausnahmen doppelt hinzugefügt werden, wodurch der Merkmalssatz deutlich größer ist als bei seinem unären Pendant. Bei beiden Testgebieten tritt sowohl für den unären als auch den paarweisen RF bereits nach wenigen verwendeten Merkmalen ein Sättigungseffekt ein.

Der maximal erzielbare κ-Index dient bei beiden Datensätzen als Indikator für die Identifikation geeigneter Segmentmerkmale. Der gesamte Merkmalssatz für die folgenden Experimente ergibt sich aus der Vereinigung beider Teilmengen, wobei bereits ein großer Anteil identisch ist. Dies sind, in absteigender Wichtigkeit: $\emptyset$ *Höhe über Grund*, $\emptyset$ *Distanz zur Straße*, *max. Höhe*, *Anteil Nachbarklasse Straße*, *Anteil Nachbarklasse Boden*, *min. Höhe*, *min. Distanz zur Straße*, *Anteil Nachbarklasse Vegetation*, *Anteil Nachbarklasse Gebäude*, *Anteil Nachbarklasse Auto*, *Orientierung zur Straße*, $\emptyset$ *Vertikalität*



(a) Hannover



(b) Vaihingen

Abbildung 5.8: Relevanz der Segmentmerkmale für das unäre (blau) und paarweise Potential (orange). Die Balkenhöhe ergibt sich aus der Summe beider Wichtigkeitswerte. Dieser Wert dient als Grundlage zur Sortierung in abnehmender Reihenfolge von links nach rechts. Die durch *NN* gekennzeichneten Merkmale repräsentieren die Nachbarschaftsklassenverteilung für die jeweils angegebene Klasse.

Punkte, *Standardabweichung Höhe*, *∅ Intensität*, *Standardabweichung Vertikalität* und *max.* ΔZ. Sie unterscheiden sich im Wesentlichen in der Reihenfolge der FPFH Merkmale. Da die Trainingszeit beim unären Potential von CRF^S kein kritischer Faktor ist, wird die gesamte Gruppe der FPFH Merkmale bestehend aus 33 Komponenten hinzugefügt, um alle geometrischen Eigenschaften abbilden zu können. Insgesamt setzt sich somit ein Merkmalsvektor $\mathbf{f}_i^S(\mathbf{x})$ für Segmente aus 49 Komponenten zusammen.

Für den Kantenmerkmalsvektor $\mathbf{f}_{ij}^S(\mathbf{x})$ werden die Knotenmerkmale beider beteiligter Segmente $\mathbf{f}_i^S(\mathbf{x})$ und $\mathbf{f}_i^S(\mathbf{x})$ zusammengefügt. Dadurch erhöht sich der Rechenzeit- und Speicherbedarf signifikant. Somit wird die Auswahl geeigneter Merkmale noch bedeutender. Die Analyse ergibt, dass folgende Merkmale - ebenfalls nach absteigender Relevanz sortiert - besonders diskriminativ zur Modellierung der Interaktionen sind: *∅ Höhe über Grund*, *max. Höhe*, *∅ Distanz zur Straße*, *Anteil Nachbarklasse Straße*, *Anteil Nachbarklasse Vegetation*, *Anteil Nachbarklasse Boden*, *min. Höhe*, *Anteil Nachbarklasse Auto*, *min. Distanz zur Straße*, $FPFH_{26}$, *Anteil Nachbarklasse Gebäude*, $FPFH_{15}$, *∅ Vertikalität Punkte*, *Orientierung zur Straße*, $FPFH_{30}$, $FPFH_{22}$, *∅ Intensität*, $FPFH_{04}$, $FPFH_{23}$ *Höhenunterschied an Segmentgrenze* und $FPFH_{29}$. Abgesehen vom reinen Interaktionsmerkmal *Höhenunterschied an Segmentgrenze* werden alle Merkmale durch das Aneinanderhängen zweifach verwendet. Die Dimension von $\mathbf{f}_{ij}^S(\mathbf{x})$ beläuft sich somit auf 41.

Überprüfung der Merkmalsauswahl

Für die Klassifikation mittels RF (unäres Potential CRF^P, unäres und paarweises Potential von CRF^S) ist es sinnvoll, die Merkmale auf den Wertebereich [0;1] zu normieren, um die durch verschiedene Skalenbereiche hervorgerufenen Effekte bei der Klassifikation zu reduzieren. Dazu werden die Ausprägungen jedes Merkmals aus dem Intervall zwischen dem unteren und oberen 2,5 %-Quantil, welche sich aus den Trainingsdaten ergeben, linear auf das Intervall [0;1] skaliert. Zur Reduktion von Ausreißern in den Merkmalen werden weiterhin Werte innerhalb der unteren bzw. oberen 2,5 % auf die entsprechenden Grenzwerte des neuen Intervalls abgebildet.

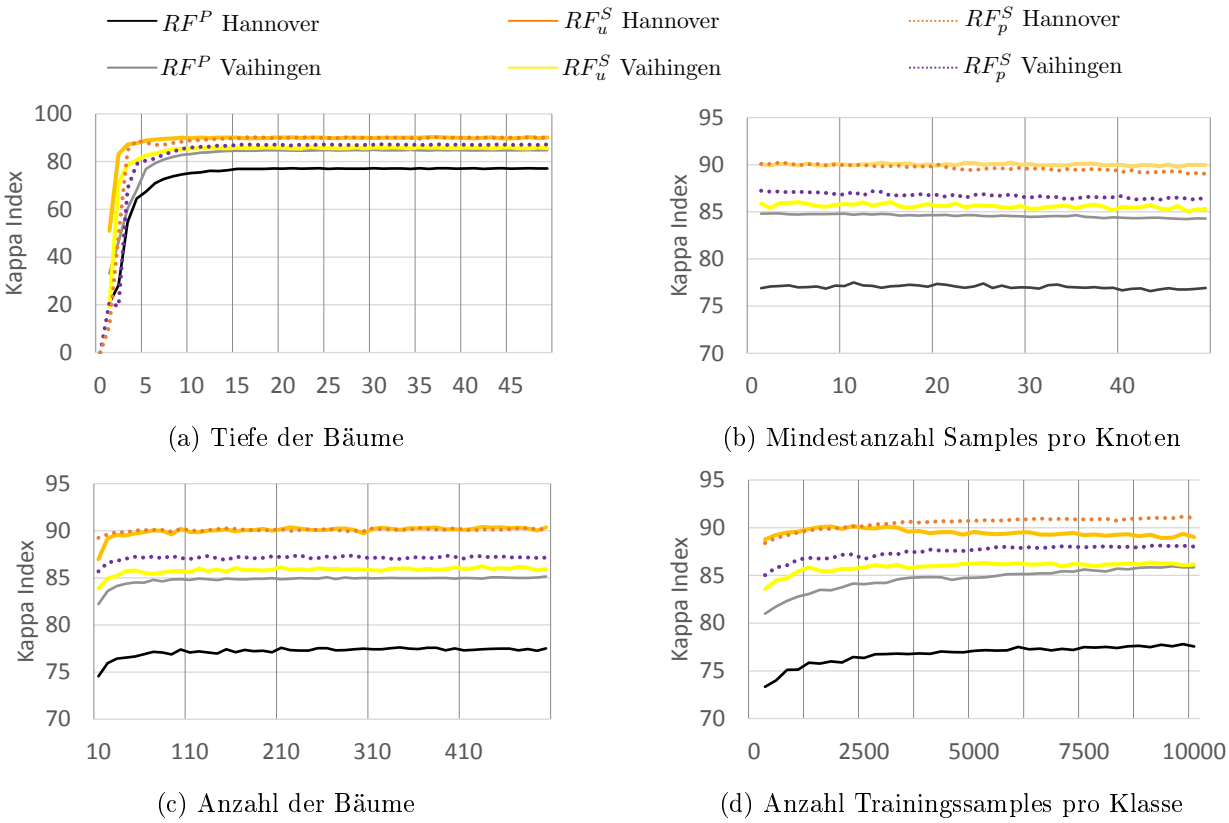
Abschließend soll die Auswirkung der beschriebenen Merkmalsauswahl auf die Ergebnisse im Vergleich zum gesamten Merkmalssatz untersucht werden. Sowohl für Hannover als auch für Vaihingen ergibt sich jeweils bei allen drei RF ein leichter Genauigkeitsgewinn von bis zu 0,6 Prozentpunkten hinsichtlich des κ-Koeffizienten bei Verwendung der reduzierten Merkmalsmenge. Weiterhin verringert sich die benötigte Trainingszeit insbesondere bei den RF des segmentweisen Interaktionspotentials signifikant. Für Hannover reduziert sich die Zeit von 21:32 min auf 4:59 min, was einer Ersparnis von 75 % entspricht. Analog verringert sich die Trainingszeit für Vaihingen um 74 % von 19:34 min auf 5:09 min. Da die RF in jeder Iteration neu trainiert werden, führt die Reduktion zu einer deutlichen Verkürzung der Gesamtlaufzeit. Zusammenfassend zeigt sich, dass die Auswahl einen positiven Effekt sowohl auf Genauigkeit als auch auf die Laufzeit hat. Der reduzierte Merkmalssatz kann somit für die Experimente in Abschnitt 5.4 verwendet werden.

5.3.2 Parameter

Dieser Abschnitt befasst sich mit der Festlegung der für das Verfahren erforderlichen Parameter. In den Unterabschnitten werden auf der Basis von Experimenten geeignete Einstellungen für die RF Klassifikatoren sowie für die Segmentierung ermittelt. Danach folgt eine Zusammenstellung aller Parameter.

Random Forest Parameter

Für eine Klassifikation mittels RF müssen zunächst einige Parameter festgelegt werden. Dabei handelt es sich entsprechend Abschnitt 3.2 um 1) die Anzahl der Bäume T, 2) die Anzahl der Samples $N_{s,l}$ pro Klasse l im Training, 3) die Tiefe der Bäume T_{depth}, 4) die minimale Anzahl von Samples an einem Knoten, um einen neuen Entscheidungstest an dieser Stelle einzuführen (T_{split}) sowie 5) die Anzahl der zufällig verwendeten Merkmale an einem Knoten (M_{split}). Entsprechend der üblichen Vorgehensweise in der Literatur ergibt sich M_{split} aus der Quadratwurzel der Dimension des Merkmalsvektors [Gislason et al., 2006; Belgiu & Drăguţ, 2016]. Für die verbleibenden vier Parameter sollen in diesem Abschnitt die optimalen Werte bestimmt werden. Im Rahmen der experimentellen Untersuchung wird der Einfluss eines jeden Parameters auf den κ-Index des Validierungsgebietes bestimmt. Iterativ wird jeweils ein Parameterwert variiert und die RF Klassifikation durchgeführt. Dabei liegt die Annahme zugrunde, dass die voneinander unabhängig bestimmten optimalen Parameter auch in Kombination zu einem guten Ergebnis führen. Die empirisch ermittelten Standardeinstellung der Parameter werden für das Experiment auf $T = 100$, $N_{s,l} = 5000$, $T_{depth} = 25$ und $T_{split} = 5$ gesetzt. Abbildung 5.9 zeigt das



(a) Tiefe der Bäume

(b) Mindestanzahl Samples pro Knoten

(c) Anzahl der Bäume

(d) Anzahl Trainingssamples pro Klasse

Abbildung 5.9: Variation der wesentlichen Parametereinstellungen der RF Klassifikatoren. Gezeigt werden jeweils die aus fünf Wiederholungen gemittelten Genauigkeiten der drei voneinander unabhängigen RF (unäres Potential CRF^P, unäres Potential CRF^S sowie das paarweise Interaktionspotential CRF^S) für die Datensätze Hannover und Vaihingen.

Ergebnis der Testreihen aller drei RF für die beiden Datensätze Hannover und Vaihingen in Bezug auf die erzielten κ-Indices. Zur Steigerung der Aussagekraft wurde jeder Test fünf Mal wiederholt - entsprechend handelt es sich in Abbildung 5.9 um die jeweiligen Mittelwerte der κ-Indices. Ebenfalls untersucht wurde die jeweils benötigte Trainingszeit. Aus Gründen der Übersichtlichkeit wird diese jedoch nicht dargestellt, sondern ggf. bei der Auswertung berücksichtigt.

Insgesamt stellen sich die gewählten Standardparameter bereits als eine gute Wahl heraus. Bei der Verwendung von gebräuchlichen Einstellungen sind die Auswirkungen auf die Genauigkeiten recht gering, so dass die Wahl nicht kritisch ist. Im Folgenden wird die genaue Bestimmung geeigneter Parameterwerte näher beschrieben. Wählt man für die **Tiefe der Bäume** T_{depth} einen zu geringen Wert, so reduziert sich die Genauigkeit bei allen drei RF (Abbildung 5.9a). Sie wird jedoch recht schnell ab einer Tiefe von ca. 10 konstant, so dass keine wesentliche Änderung mehr festzustellen sind. Da sich dieser Parameter nur marginal auf die Rechenzeit auswirkt, wird der Wert $T_{depth} = 20$ ausgewählt, um eine stabile Genauigkeit zu gewährleisten. In Abbildung 5.9b zeigt sich, dass die **Anzahl der Trainingsbeispiele pro Knoten** im gewählten Intervall zwischen 1 und 50 nahezu keinen Einfluss auf die Genauigkeit (und auch auf die Rechenzeit) hat. Folglich kann an dieser Stelle der Standardwert $T_{split} = 5$ beibehalten werden. Bei den nächsten beiden Parametern hat die Laufzeit eine wesentlich größere Bedeutung. Dies trifft insbesondere für den RF zum Modellieren der Interaktionen in CRF^S

aufgrund der hohen Klassenanzahl zu. Die **Anzahl der Bäume** (T, Abbildung 5.9c) führt in allen Fällen ab einem Wert von ca. 130 zu einer stabilen Genauigkeit bei einer moderaten Trainingszeit, welche entsprechend Abschnitt 3.2 linear mit der Anzahl der Bäume ansteigt. Beim paarweisen RF von CRF^S ist beispielsweise ein Anstieg von $3\,s$ ($T = 10$) auf $140\,s$ ($T = 500$) zu beobachten. Bezüglich der **Anzahl an Trainingsbeispielen pro Klasse** $N_{s,l}$ kann man in Abbildung 5.9d eine Genauigkeitssteigerung bei Erhöhung der Anzahl erkennen. Der Sättigungseffekt ist im Vergleich zu den anderen Parametern etwas weniger stark ausgeprägt. Folglich muss ein Kompromiss zwischen der Genauigkeit und benötigten Trainingszeit gefunden werden. Ein geeigneter Parameterwert für den unären Term auf Punktebene scheint $N_{s,l} = 10.000$ zu sein. Dies ist möglich, da viele Trainingsbeispiele zu Verfügung stehen. Auf Segmentebene hingegen werden für beide RF jeweils $N_{s,l} = 5.000$ verwendet, da in den Testdatensätzen die Anzahl der Trainingsbeispiele pro Kombination geringer ist und die benötigte Trainingszeit ohne großen Genauigkeitsverlust reduziert werden kann. Für den bereits beim vorherigen Parameter genannten RF steigt die Trainingszeit von $3\,s$ ($N_{s,l} = 250$) auf $206\,s$ ($N_{s,l} = 10.000$) an.

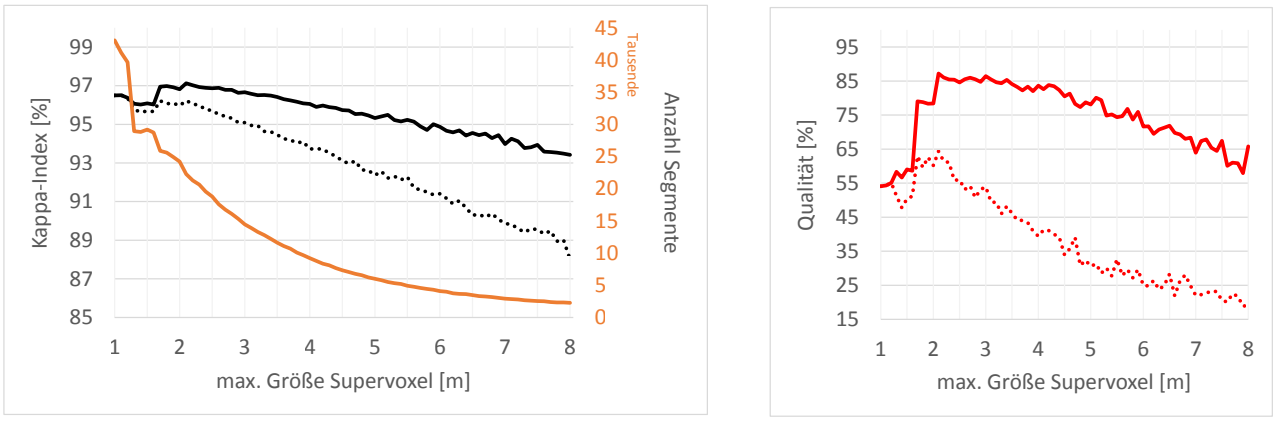
Zusammenfassend werden für alle drei verwendeten RF in den nachfolgenden Experimenten die Parameter $T_{depth} = 20$, $T_{split} = 5$ und $T = 130$ genutzt. Die Anzahl der Trainingssamples pro Klasse beträgt für CRF^P $N_{s,l} = 10.000$ sowie $N_{s,l} = 5.000$ für die beiden RF von CRF^S.

Parameter der Segmentierung

Dieser Abschnitt befasst sich mit der Festlegung geeigneter Parametereinstellungen für die Aggregation der einzelnen Punkte über VCCS. Festzulegen sind die Voxelauflösung des Octrees (R^{SV}_{vox}), die maximale Größe der Supervoxel (R^{SV}_{saat}) sowie die drei relativen Gewichte ω^{SV}_s, ω^{SV}_n, und ω^{SV}_{konf}. Das Gewicht ω^{SV}_s steuert die Form der Supervoxel und wird auf den Wert 0 gesetzt, um eine möglichst genaue Adaption der Segmentgrenzen an die anderen Komponenten zu gewährleisten. Dies wirkt Untersegmentierungsfehlern entgegen. Die beiden verbleibenden Gewichte zur Berücksichtigung der Normalen und Konfidenzen sollen einen gleich starken Einfluss haben und werden auf $\omega^{SV}_n = \omega^{SV}_{konf} = 0,5$ gesetzt.

Da R^{SV}_{vox} die kleinste darstellbare Einheit nach der Segmentierung repräsentiert, empfiehlt sich ein möglichst kleiner Wert. Andererseits sollte sich in den meisten Voxeln noch mindestens ein Laserpunkt befinden, um eine korrekte Ausdehnung der Supervoxel zu gewährleisten. Durch den Zusammenhang mit der Punktdichte des Datensatzes kann er somit nicht beliebig klein gewählt werden. Dies würde zu Punkten ohne Segmentzugehörigkeit führen. Als Kompromiss wird $R^{SV}_{vox} = 0,75\,\mathrm{m}$ gesetzt.

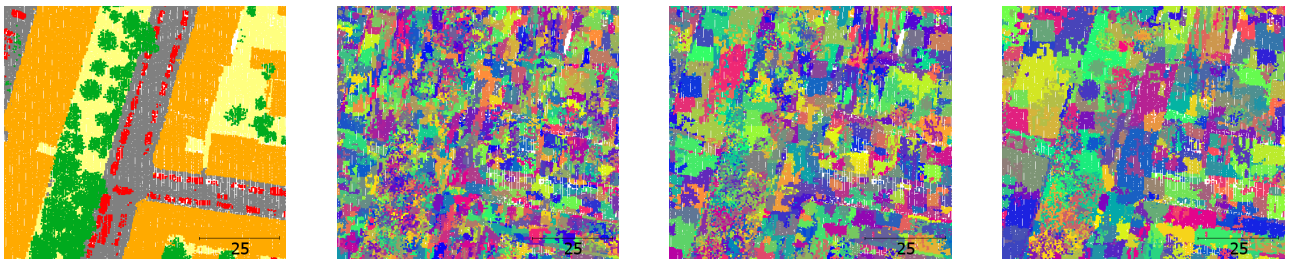
Die Auflösung der Saatpunkte R^{SV}_{saat} entspricht der max. zu erwartenden Größe der Supervoxel. Der Einfluss auf den κ-Index und die Anzahl der resultierenden Segmente wurde für beide Datensätze experimentell ermittelt. Aufbauend auf den Klassifikationsergebnissen von CRF^P wird für das Validierungsgebiet R^{SV}_{saat} in 0,1 m-Schritten von 1 m bis 8 m erhöht. Jedem Supervoxel wird dann auf Grundlage einer Mehrheitsentscheidung ein Klassenlabel zugeordnet und mit einem Referenzlabel (ebenfalls aus einer Mehrheitsentscheidung der punktweisen Referenzklassen) verglichen, um den κ-Index sowie die weiteren Genauigkeitsmaße ableiten zu können. Eine mehrfache Wiederholung des Versuchs ist aufgrund der deterministischen Funktionsweise von VCCS nicht erforderlich. In den Abbildungen 5.10a und 5.10b wird das Ergebnis für Hannover dargestellt, die Ergebnisse für Vaihingen sehen sehr ähnlich



(a) κ-Index und Segmentanzahl

(b) Qualität Klasse *Auto*

Abbildung 5.10: Einfluss der maximalen Größe eines Supervoxels auf das Ergebnis. In a) wird der κ-Index für das Validierungsgebiet Hannover in schwarz dargestellt. Der durchgezogene Linie ergibt sich unter Berücksichtigung der Konfidenz bei der Segmentierung, die punktierte Linie ist das Ergebnis ohne Konfidenzwerte. Auf der zweiten Achse (in orange) wird die Anzahl der Segmente [in Tausend] veranschaulicht. Diagramm b) zeigt den Einfluss auf die Qualität der Klasse *Auto* mit und ohne Beachtung der Konfidenz über den durchgängigen bzw. punktierten Verlauf.



(a) Referenz (90×90 m)

(b) $R^{SV}_{saat} = 3$ m

(c) $R^{SV}_{saat} = 5$ m

(d) $R^{SV}_{saat} = 7$ m

Abbildung 5.11: Einfluss der variierenden Supervoxelauflösung R^{SV}_{saat} mit 3, 5 und 7 m auf die (zufällig einfärbten) Segmente des in a) gezeigten Ausschnitts der Szene Hannover.

aus. Der visuelle Unterschied der variierenden Auflösungen wird für $R^{SV}_{saat} = 3, 5$ und 7 m in Abbildung 5.11 gezeigt. Je kleiner R^{SV}_{saat} gewählt wird (x-Achse in Abbildung 5.10a), umso mehr Segmente entstehen (orange). Der Generalisierungsfehler verhält sich invers dazu. Durch das Zusammenfassen einer größeren Punktanzahl zu einem Supervoxel werden mehr Entitäten verschiedener semantischer Gruppen aggregiert, wodurch der κ-Index um 3,1 Prozentpunkte (schwarze, durchgängige Kurve) und in Vaihingen um 3,7 Prozentpunkte sinkt. Ein Sonderfall ergibt sich für den Bereich <2,1 m. Papon et al. [2013] entsprechend muss R^{SV}_{saat} signifikant größer als R^{SV}_{vox} sein, um aussagekräftige Supervoxel zu erhalten. Dieser Zustand scheint bis zum dreifachen Wert von R^{SV}_{vox} noch nicht erreicht zu sein. Besonders die Qualität unterrepräsentierter Klassen mit kleinen Objekten wie Autos wird durch die Wahl eines zu großen Wertes für R^{SV}_{saat} negativ beeinflusst, wie das Beispiel in Abbildung 5.10b (durchgezogene Kurve) zeigt. Die Qualität nimmt um 33 Prozentpunkte (von 87 % auf 54 %) ab (in Vaihingen sogar um 45,5 Prozentpunkte). Als Kompromiss wird $R^{SV}_{saat} = 3$ m gesetzt. Dies ermöglicht einerseits die korrekte Modellierung auch kleinerer Objekte, und kann andererseits abhängig von der Position

des Voxels in der Szene auch zur Extraktion aussagekräftiger (Form-)Merkmale herangezogen werden. Weiterhin ist eine vergleichsweise große Anzahl an Segmenten verfügbar, die für das Training der beiden RF in CRF^S genutzt werden können.

In einem weiteren Experiment soll der Nutzen der Konfidenzwerte bei der Segmentierung analysiert werden. Dazu werden die bereits beschriebenen Testreihen mit geänderten Gewichten wiederholt. Die Deaktivierung von ω_{konf}^{SV} führt dazu, dass die Segmentierung ausschließlich auf den Normalenvektoren der Punkte beruhen. Entsprechende Ergebnisse für den κ-Index und die Qualität der Klasse *Auto* am Beispiel Hannover sind in den Abbildungen 5.10a bzw. 5.10b als punktierte Kurven dargestellt. Man kann erkennen, dass die Genauigkeiten in diesem Fall stärker abfallen. Die Berücksichtigung der Konfidenzen führt somit zu einer deutlich geringeren Abhängigkeit von R_{saat}^{SV}. Auch für kleinere Werte von $R_{saat}^{SV} < 2,1$ sind die Genauigkeiten durch die Hinzunahme der Vorklassifikation besser. Die maximale Qualität der Klasse *Auto* beträgt beispielsweise 87,2 % mit Konfidenzen verglichen zu 64,4 % ohne Berücksichtigung der a priori Information. Folglich ist die Integration zur Reduktion der Segmentierungsfehler empfehlenswert. Insbesondere Klassen mit kleineren Objekten profitieren von der Nutzung der Konfidenzen.

Übersicht aller Parameter

Eine Übersicht über alle Parameter ist in Tabelle 5.7 gegeben. Die bisher noch nicht beschriebenen Parameter sollen nun kurz erläutert werden. In der ersten Iteration von CRF^P erfolgt die Cliquengenerierung aus zwei Anwendungen von VCCS (ohne Konfidenzinformationen). Die Auflösungen der Voxel variieren dabei leicht und werden auf $R_{vox}^{SV} = 0{,}75\,\text{m}$, bzw. $R_{vox}^{SV} = 0{,}8\,\text{m}$ gesetzt. Die Variation dient dazu, in einer Auflösungsstufe eventuell nicht optimal zusammengefasste Objekte an Voxelrändern mit der anderen Segmentierung zu kompensieren. Gleiches gilt für die maximale Größe der Supervoxels R_{saat}^{SV}, welche zu 2 bzw. 2,5 m festgelegt werden. Demnach wird insgesamt eine etwas feinere und eine gröber aufgelöste Segmentierung durchgeführt, um zwei Typen von Cliquen zu erhalten. Auf die Hinzunahme weiterer, größerer Segmente wird verzichtet, um einer Überglättung kleinerer Objekte mit Ausdehnungen $< R_{saat}^{SV}$ entgegenzuwirken. Der Sättigungsparameter q muss per Definition kleiner als 0,5 sein [Kohli et al., 2009]. Er wird auf 0,4 gesetzt, um einen möglichst großen Anteil abweichender Labels innerhalb einer Clique zuzulassen, welche feine Strukturen in der Szene darstellen könnten. Der Glättungseffekt wird dadurch geringfügig reduziert. Weiterhin werden für den Aufbau des Graphen in CRF^P für jeden Punkt die $k = 3$ nächsten Nachbarn in der $\mathcal{N}_{2D}^k$ Suchumgebung herangezogen. Drei bis fünf Kanten haben sich für assoziative Modelle bewährt [Munoz et al., 2008].

5.4 Klassifikationsergebnisse

Die folgenden Experimente untersuchen die Leistungsfähigkeit des neu entwickelten Verfahrens auf Grundlage der Evaluationsgebiete $\mathfrak{E}$. Abschnitt 5.4.1 analysiert die Ergebnisse für die Datensätze von Hannover und Vaihingen ausführlich. In Abschnitt 5.4.2 schließt sich ein Vergleich mit anderen aktuellen Verfahren anhand von Benchmark Datensätzen an.

Allgemein	
maximale Anzahl Iterationen	N_{iter}
Relative Gewichte der Potentiale	$\omega_{\psi}^{P}, \omega_{\chi}^{P}, \omega_{\xi}^{P}, \omega_{\psi}^{S}$ (werden gelernt)
Merkmale	
Nachbarschaft Merkmalsberechnung	optimale Nachbarschaft nach Weinmann [2016]
Nachbarschaft Normalenberechnung	$k_{NN}^{Normal} = 25$
Nachbarschaftsklassenverteilung	$r_{hist} = 1\,m$
Parameter für Straßenextraktion	Rasterauflösung $res_{grid} = 1\,\text{m}$
	Größe morphologischer Filter $S_{closing} = 3 \times 3$ Pixel
	Hough - räumliche Auflösung $H_{dist}^{res} = 1\,\text{m}$
	Hough - Winkelauflösung $H_{\rho}^{res} = 5°$
	Hough - Größe Datenlücke $H_{lineGap} = 2\,\text{m}$
	Hough - min. Anzahl Votes $H_{minVotes} = 30$
	Hough - min. Straßenlänge $H_{minLength} = 15\,\text{m}$
Random Forest	
Anzahl Bäume	$T = 130$
maximale Baumtiefe	$T_{depth} = 20$
Mindestanzahl Samples pro Knoten	$T_{split} = 5$
Anzahl Trainingsbeispiele CRF^{P}	$N_{s,l} = 10.000$
Anzahl Trainingsbeispiele CRF^{S}	$N_{s,l} = 5.000$
***CRF*^{P}**	
Anzahl Nachbarpunkte Graph	$k = 3$
Glättungsparameter kontrast-sensitives Potts Modell	$\theta = 0$ (ausschließlich datenabhängige Glättung)
HOP: Anzahl Segmentierungen	2
HOP: 1. Parametersatz VCCS	$R_{vox}^{SV} = 0{,}75\,\text{m}$, $R_{saat}^{SV} = 2\,\text{m}$
HOP: 2. Parametersatz VCCS	$R_{vox}^{SV} = 0{,}8\,\text{m}$, $R_{saat}^{SV} = 2.5\,\text{m}$
HOP: Gewichte	$\omega_{n}^{SV} = 1$, $\omega_{rgb}^{SV} = \omega_{s}^{SV} = 0$
Sättigungsparameter	$q = 0,4$
Segmentierung	
Voxelauflösung	$R_{vox}^{SV} = 0{,}75\,\text{m}$
Auflösung Saatpunkte	$R_{saat}^{SV} = 3\,\text{m}$
Gewichte	$\omega_{s}^{SV} = 0{,}0$, $\omega_{n}^{SV} = 0{,}5$, $\omega_{konf}^{SV} = 0{,}5$
***CRF*^{S}**	
Nachbarschaft Segmente Graph	Punktabstand zweier Segmente kleiner $d_{graph}^{S} = 1\,\text{m}$

Tabelle 5.7: Für die vorgestellte Methodik erforderliche Parameter. Sofern nicht explizit darauf hingewiesen wird, werden die hier angegebenen Werte für die folgenden Experimente verwendet.

		Boden		Straße		Auto		Vegetation		Gebäude	
V	RF^P	69,1		56,8		87,6		91,5		95,5	
	CRF_1^P	76,5	+7,3	59,2	+2,3	82,4	-5,2	92,5	+0,9	97,3	+1,8
	CRF_3^P	81,5	+12,4	61,6	+4,7	80,8	-6,8	91,9	+0,3	97,6	+2,1
K	RF^P	68,4		59,2		53,3		91,8		96,7	
	CRF_1^P	70,4	+1,9	66,7	+7,5	71,0	+17,8	94,6	+2,8	97,3	+0,6
	CRF_3^P	72,4	+4,0	72,3	+13,1	85,8	+32,5	95,7	+3,8	96,8	+0,1
Q	RF^P	52,4		40,8		49,5		84,7		92,5	
	CRF_1^P	57,8	+5,4	45,7	+4,8	61,7	+12,1	87,8	+3,2	94,7	+2,3
	CRF_3^P	62,2	+9,8	49,8	+9,0	71,2	+21,7	88,2	+3,5	94,5	+2,1
OA	RF^P	79,7				κ-Index	RF^P	72,9			
	CRF_1^P	82,9	+3,2				CRF_1^P	77,0	+4,1		
	CRF_3^P	84,7	+5,0				CRF_3^P	79,5	+6,5		

Tabelle 5.8: Erzielte Genauigkeiten in [%] (Vollständigkeit V, Korrektheit K, Qualität Q, Gesamtgenauigkeit OA, κ-Index) für das Testgebiet Hannover. Verglichen werden die Ergebnisse einer kontextfreien Klassifikation RF^P, einer lokalen Kontextberücksichtigung nach CRF_1^P und nach dreimaligem Durchlauf zur Integration gröberskaligen Kontexts (CRF_3^P). Hinter den absoluten Werten werden die erzielten Unterschiede in Bezug auf das Ergebnis von RF^P in Prozentpunkten dargestellt. Verbesserungen sind grün, geringere Genauigkeiten rot gekennzeichnet.

5.4.1 Nutzen von Kontext für die 3D-Punktwolkenklassifikation

Experiment: Zur Evaluation der neuen Methodik werden im Folgenden die Klassifikationsergebnisse der beiden Datensätze Hannover und Vaihingen vorgestellt. Es finden drei Iterationen über CRF^P Anwendung, wodurch der gröbere Maßstab durch CRF^S zweimal integriert wird. Der Fokus der Analyse liegt auf dem Nutzen der Kontextinformation. Daher werden speziell die finalen Ergebnisse mit denen der ersten Zwischenlösung verglichen, da nach dem erstmaligen Durchlauf von CRF^P (im Folgenden CRF_1^P) noch keine gröberskalige Information berücksichtigt wurde. Weiterhin soll zum Vergleich das Ergebnis des punktweisen Klassifikators der ersten Iteration (RF_1^P) herangezogen werden, da RF_1^P (abgesehen von der lokalen Punktnachbarschaft zur Merkmalsberechnung) ohne Kontext klassifiziert. Die Auswertung erfolgt punktweise, d. h. die Segmentkonfidenzen und -labels werden auf die jeweiligen Komponenten projiziert. Als finales Ergebnis wird das Resultat des dritten CRF^P angesehen und als CRF_3^P gekennzeichnet. Zur Klassifikation werden die in Abschnitt 5.3.1 beschriebenen Merkmalssätze verwendet. Die Parametereinstellungen entsprechen Tabelle 5.7, wobei die relativen Gewichte im Rahmen des Trainings optimiert werden.

Ergebnis Hannover: Für das Testgebiet Hannover ergibt sich nach drei Iterationen (CRF_3^P) eine Gesamtgenauigkeit von 84,7 % und ein κ-Index von 79,5 % (Tabelle 5.8). Für die Unterscheidung von fünf Objektklassen stellt dies ein zufriedenstellendes Ergebnis dar. Betrachtet man die Genauigkeiten der einzelnen Klassen, so zeigen sich die besten Qualitätswerte für *Gebäude*. Auf der anderen Seite führen Verwechslungen zwischen *Boden* und *Straße* zu vergleichsweisen geringen Qualitäten von 62,2 % bzw. 49,8 %. Dies ist vor allem auf die größeren Parkplatzflächen in der Szene zurückzuführen, welche aus Schotter bestehen und basierend auf den Intensitätswerten selbst für den menschlichen Operateur

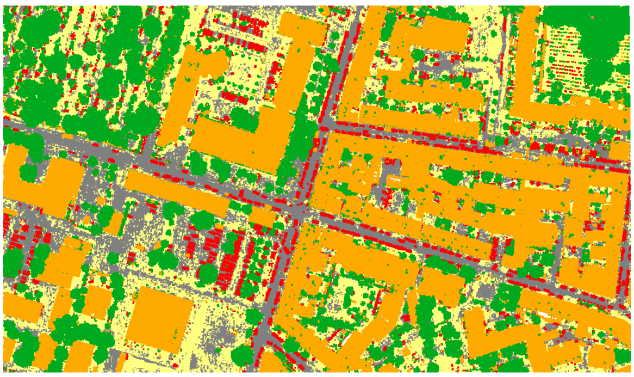
R/E	Boden	Str.	Auto	Veg.	Geb.
Boden	22,6	4,4	0,1	0,4	0,3
Str.	7,2	12,0	0,1	0.1	0,1
Auto	0,1	0,1	1,2	0,0	0,1
Veg.	1,0	0,1	0,0	18,4	0,5
Geb.	0,3	0,0	0,0	0,4	30,5

Tabelle 5.9: Konfusionsmatrix Hannover CRF_3^P [%].

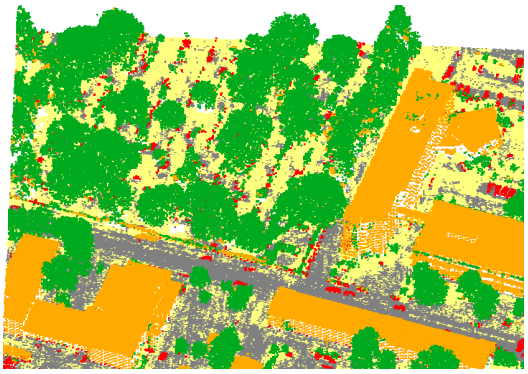
schwer einer der beiden Klassen zuzuordnen sind, so dass eventuell auch die Referenz an diesen Stellen nicht ganz korrekt ist. Wie die visuelle Analyse von Abbildung 5.12g zeigt, wurden die reinen Straßenverläufe hingegen gut detektiert. Dies ist besonders für die kontextbasierten Merkmale wichtig, da sie auf dem Straßennetz aufbauen. Im Bereich der Parkplätze ist ihr diskriminativer Beitrag geringer. Weiterhin zeigt sich, dass für die unterrepräsentierte Klasse *Auto* mit 71,2 % ein guter Wert erzielt wird. Vollständigkeit und Korrektheit betragen 80,8 % bzw. 85,8 %. Nach den drei Iterationen ergibt sich die in Tabelle 5.9 dargestellte Konfusionsmatrix. Verwechslungen sind hauptsächlich zwischen den beiden Klassen *Boden* und *Straße* zu beobachten. Darüber hinaus werden sowohl der natürliche Boden als auch Gebäude zum Teil fälschlicherweise als *Vegetation* klassifiziert. Der quantitative Einfluss von Kontext lässt sich ebenfalls Tabelle 5.8 entnehmen. Dort werden die Ergebnisse einer kontextfreien Klassifikation (RF_1^P), einer lokalen Kontextberücksichtigung auf der Punktebene (CRF_1^P) jenen nach drei Durchläufen des zweistufigen Verfahrens (CRF_3^P) gegenübergestellt. Die relativen Abweichungen in Bezug auf das Referenzergebnis von RF^P sind farbig dargestellt. Grün bedeutet einen Genauigkeitsgewinn, rot entspricht einer Verschlechterung gegenüber RF_1^P. Berücksichtigt werden Vollständigkeit V, Korrektheit K, Qualität sowie Gesamtgenauigkeit OA und der κ-Index. Insgesamt kann die Gesamtgenauigkeit von 79,7 % (RF_1^P) über 82,9 % (CRF_1^P) auf 84,7 % (CRF_3^P) gesteigert werden. Analog dazu verhält sich der κ-Index mit einer Zunahme von 6,5 Prozentpunkten. Nahezu alle Genauigkeitsmaße der einzelnen Klassen profitieren von Kontextinformation. Meistens sind dabei die Ergebnisse von CRF_3^P denen von CRF_1^P überlegen. Hinsichtlich der Vollständigkeit verbessert sich vor allem *Boden*, gefolgt von *Straße*. Ein negativer Effekt des vorgeschlagenen Verfahrens ist in Bezug auf die Vollständigkeit von *Auto* sowohl für CRF_1^P als auch noch deutlicher für CRF_3^P zu beobachten. Dies ist teilweise auf die Glättungseffekte zurückzuführen, wenn sich Punkte auf Autos aufgrund der Merkmale schlecht von ihrer Umgebung trennen lassen. Entsprechend der Konfusionsmatrix (Tabelle 5.9) werden Autos öfter mit einer der beiden Bodenklassen verwechselt. Die gute Vollständigkeit von 87,6 % wird durch die Interaktionen bis auf 80,8 % reduziert. Genau gegenteilig verhält es sich jedoch bei Betrachtung der Korrektheit. Hierbei ist der Gewinn der Klasse *Auto* mit 32,5 Prozentpunkten (CRF_3^P) mit Abstand der größte. Damit kann selbst das gute Ergebnis mit einem Zuwachs von 17,8 Prozentpunkten von CRF_1^P nahezu verdoppelt werden. Auch bei den weiteren Klassen ist der Gewinn im Schnitt leicht oberhalb des Niveaus von V. Insgesamt führt dies bei allen Klassen zu teils signifikanten Genauigkeitssteigerungen bei den Qualitätswerten. Auch bei Q ist für die Klasse *Auto* der größte Effekt von Kontextinformation zu beobachten. Verglichen mit RF_1^P wird die Qualität um 12,1 (CRF_1^P) bzw. 21,7 (CRF_3^P) Prozentpunkte erhöht. Einzig bei *Gebäude* führen bei CRF_3^P die zusätzlichen Iterationen - und damit die Berücksichtigung gröberskaligen Kontexts - zu einem leichten Rückgang von 0,2 Prozentpunkten im Vergleich zu CRF_1^P.

(a) Referenz

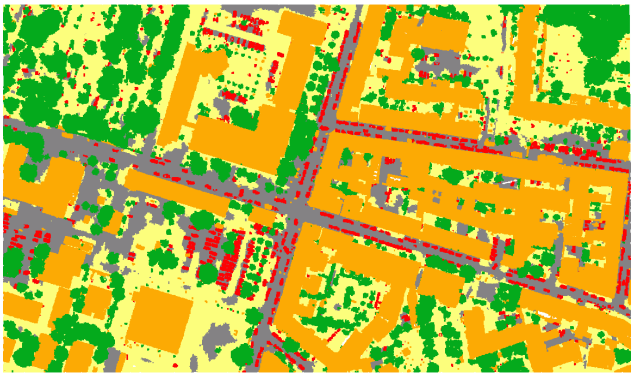
(b) Referenz



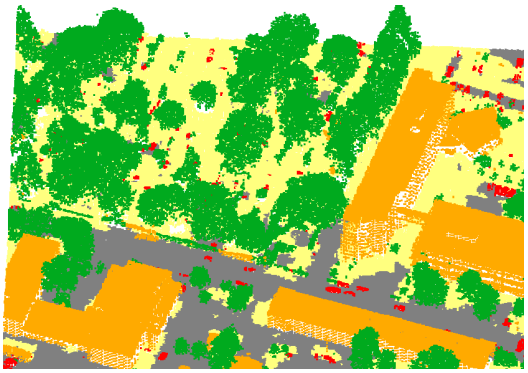
(c) Random Forest RF_1^P



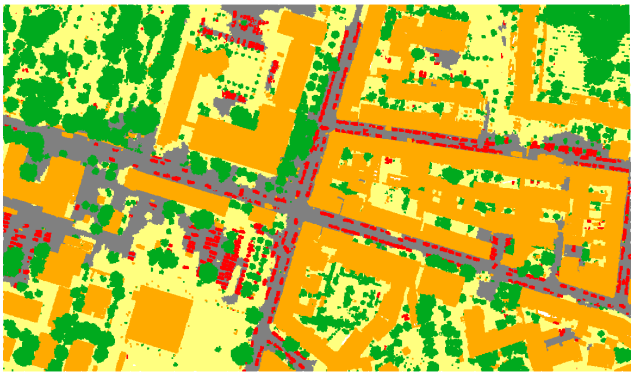
(d) Random Forest RF_1^P



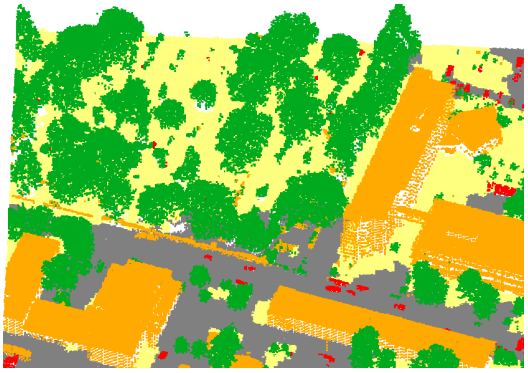
(e) CRF_1^P Iteration 1



(f) CRF_1^P Iteration 1



(g) CRF_3^P Iteration 3



(h) CRF_3^P Iteration 3

Abbildung 5.12: Klassifikationsergebnis Hannover. Die linke Spalte zeigt das gesamte Testgebiet; in der rechten Spalte ist der nord-westliche Ausschnitt detaillierter visualisiert. Unterschieden werden die fünf Klassen *Boden* (gelb), *Straße* (grau), *Auto* (rot), *Vegetation* (grün) und *Gebäude* (orange).

Die Abbildungen 5.12 visualisieren die auf unterschiedliche Weise erlangten Klassifikationsergebnisse für Hannover. Die linke Spalte (a, c, e, g) stellt das gesamte Testgebiet dar, während in den rechten Abbildungen (b, d, f, h) der nordwestliche Bereich der Szene rund um den Friedhof vergrößert zu sehen ist. Den Abbildungen kann der Unterschied zwischen der Referenz (a, b), RF_1^P (c, d), CRF_1^P (e, f) sowie CRF_3^P (g, h) entnommen werden. Die reine RF Klassifikation führt zu einem sehr heterogenen Ergebnis, bei dem insbesondere die Boden- und Straßenpunkte dem Verhalten des sogenannten „Salz- und Pfefferrauschens“ ähneln. Im Bereich des Friedhofs am linken Rand werden viele Hecken fälschlicherweise der Klasse *Auto* zugeordnet (rot). Auch auf Gebäuden sind einige Verwechslungen mit Baumpunkten (grün) zu beobachten. Die Hinzunahme von lokalem Kontext durch die erstmalige Anwendung von CRF^P glättet das Ergebnis deutlich, so dass sich Objekte leichter identifizieren lassen. Die Anzahl der Fehler an den Gebäuden sowie in der Grünanlage sinkt merklich. Die vierte Zeile in Abbildung 5.12 zeigt das endgültige Ergebnis nach drei Iterationen (CRF_3^P); ein Großteil der Verwechslungen zwischen *Vegetation* und *Auto* im Bereich des Friedhofs konnte eliminiert werden.

Für die relativen Potentialgewichte, welche sich invers zum tatsächlichen Einfluss der Terme verhalten, ergeben sich im Training die Werte $\omega_\psi^P = 0,56, \omega_\chi^P = 10,96$ (Iteration 1) bzw. $\omega_\psi^P = 1,1, \omega_\xi^P = 0,5$ (Iteration >1) sowie $\omega_\psi^S = 1$. Demnach sind in Iteration 1 für CRF^P die paarweisen Interaktionen vergleichsweise wichtig, während für die Folgeiterationen das neu eingeführte Potential aus der segmentbasierten Klassifikation von besonderer Bedeutung ist. Die Berechnung dauert inklusive Training ca. 51 min (davon Merkmalsextraktion ca. 2 min, RF Klassifikation ca. 2 min, CRF_1^P ca. 20 min). Aufgrund der Parameteroptimierung nehmen die ersten beiden Iterationen mehr Zeit in Anspruch als die Folgedurchläufe. Für jedes Segment werden durchschnittlich 9,2 Nachbarsegmente gefunden.

Ergebnis Vaihingen: Für das Testgebiet Vaihingen zeigt sich insgesamt ein ähnliches Bild, jedoch ist das nach drei Iterationen erlangte Genauigkeitsniveau mit einer Gesamtgenauigkeit von 87,9 % und einem κ-Koeffizienten von 83,9 % etwas höher (Tabelle 5.10). Auch für diesen Datensatz lässt sich die höchste Qualität mit 89,3 % für die Klasse *Gebäude* erzielen. Mit 83,2 % folgt *Straße* mit sehr guten V und K-Werten von 89,4 bzw. 92,3 %. Dies bestätigt, dass sich die Straße bei Vorhandensein der Intensitätswerte gut identifizieren lässt, solange sich keine versiegelten Plätze in der Szene befinden. Die größten Herausforderungen dieses Datensatzes stellen *Autos* (Q=64,2 %) und *Boden* (Q=67,0 %) dar, was bei Autos auf eine vergleichsweise geringe V von 72,8 % zurückzuführen ist. Wie die Konfusionsmatrix in Tabelle 5.11 zeigt, kommt es bei *Boden* zu Verwechslungen mit *Straße* sowie teilweise mit *Vegetation* (insbesondere kleineren Sträuchern). *Autos* hingegen werden hauptsächlich mit *Vegetation* verwechselt, da sie ähnliche geometrische Eigenschaften wie Sträucher aufweisen. Einige verbleibende Verwechslungen sind weiterhin zwischen *Gebäude* und *Vegetation* zu erkennen. Dies ist vor allem auf Anbauten an Gebäuden wie Schornstein, Balkon usw. zurückzuführen, die teilweise als *Vegetation* identifiziert wurden.

Generell äußert sich die Kontextintegration mit ähnlichen Auswirkungen wie bei Hannover. Die Gesamtgenauigkeit der RF Klassifikation wird um 0,9 (CRF_1^P) bzw. 2,6 (CRF_3^P) Prozentpunkte auf 87,9 % gesteigert. Der Gewinn ist somit verglichen mit Hannover etwas geringer, da die Änderungen die dominanten, punktreichen Klassen wie *Boden* und *Straße* etwas weniger stark beeinflussen, was auf die zuverlässigere Initialklassifikation mittels RF_1^P zurückzuführen ist. Im Gegensatz zu Hannover liegen in der Szene Vaihingen nahezu keinerlei Parkplatzflächen, die sich nur schwer einer der beiden Klassen

		Boden		Straße		Auto		Vegetation		Gebäude	
V	RF^P	75,9		87,5		73,3		87,7		89,8	
	CRF_1^P	76,4	+0,5	89,1	+1,6	74,2	+1,0	88,0	+0,3	90,9	+1,1
	CRF_3^P	79,7	+3,8	89,4	+1,9	72,8	-0,5	88,5	+0,8	93,4	+3,6
K	RF^P	79,1		88,7		44,3		78,6		95,4	
	CRF_1^P	80,3	+1,2	90,1	+1,3	54,7	+10,4	78,4	-0,2	95,7	+0,3
	CRF_3^P	80,8	+1,6	92,3	+3,6	84,5	+40,3	81,4	+2,9	95,3	-0,1
Q	RF^P	63,3		78,7		38,1		70,8		86,1	
	CRF_1^P	64,3	+1,1	81,1	+2,4	46,0	+7,9	70,8	+0,1	87,4	+1,3
	CRF_3^P	67,0	+3,7	83,2	+4,5	64,2	+26,1	73,6	+2,8	89,3	+2,0
OA	RF^P	85,3				κ-Index	RF^P	80,5			
	CRF_1^P	86,2	+0,9				CRF_1^P	81,7	+2,6		
	CRF_3^P	87,9	+2,6				CRF_3^P	83,9	+3,4		

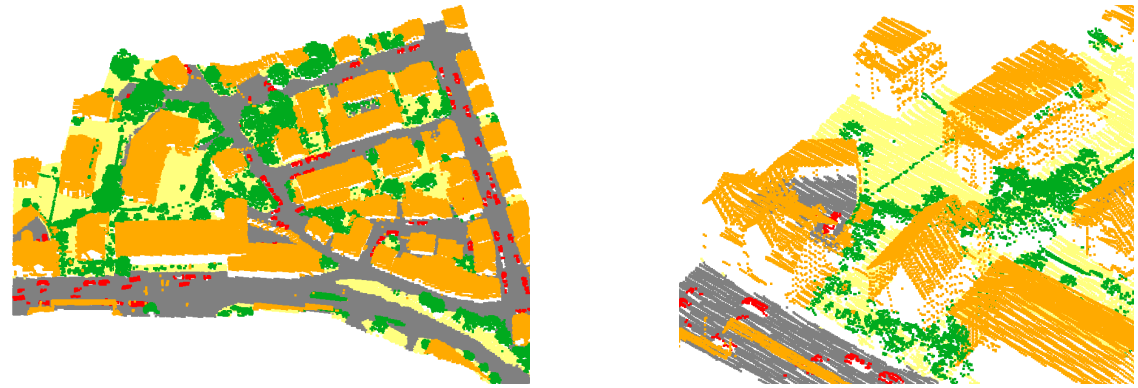
Tabelle 5.10: Erzielte Genauigkeiten in [%] (Vollständigkeit V, Korrektheit K, Qualität Q, Gesamtgenauigkeit OA, κ-Index) für das Testgebiet Vaihingen. Verglichen werden die Ergebnisse einer kontextfreien Klassifikation RF^P, einer lokalen Kontextberücksichtigung nach CRF_1^P und nach dreimaligem Durchlauf zur Integration gröberskaligen Kontexts (CRF_3^P). Hinter den absoluten Werten werden die erzielten Unterschiede in Bezug auf das Ergebnis von RF^P in Prozentpunkten dargestellt. Verbesserungen sind grün, geringere Genauigkeiten rot gekennzeichnet.

R/E	Boden	Str.	Auto	Veg.	Geb.
Boden	19,1	1,8	0,0	2,6	0,5
Str.	2,5	22,1	0,0	0,1	0,1
Auto	0,1	0,0	0,7	0,1	0,0
Veg.	1,5	0,0	0,1	18,6	0,8
Geb.	0,5	0,0	0,0	1,4	27,4

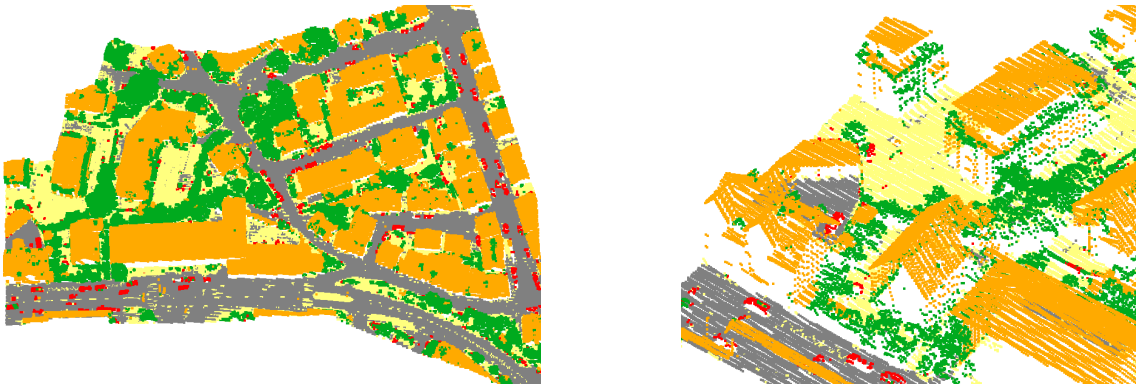
Tabelle 5.11: Konfusionsmatrix Vaihingen CRF_3^P [%].

zuordnen lassen und somit Verwechslungen hervorrufen können. Die klassenweise Auswertung von V, K und Q von Vaihingen lässt abgesehen von zwei Ausnahmen die höchsten Genauigkeiten bei CRF_3^P erkennen. Erneut verschlechtert sich die Vollständigkeit von *Auto* geringfügig durch die Segmentinteraktionen. Analog zu Hannover reduziert diese auch die Korrektheit leicht, und zwar von *Gebäude* um 0,4 Prozentpunkte verglichen mit CRF_1^P. Der deutlichste Gewinn von 40,3 Prozentpunkten wird erneut bei der Korrektheit von *Auto* erzielt. In diesem Fall ist der Nutzen durch die gröberskaligen Interaktionen ca. viermal größer als bei der Berücksichtigung ausschließlich lokaler Beziehungen bei CRF_1^P.

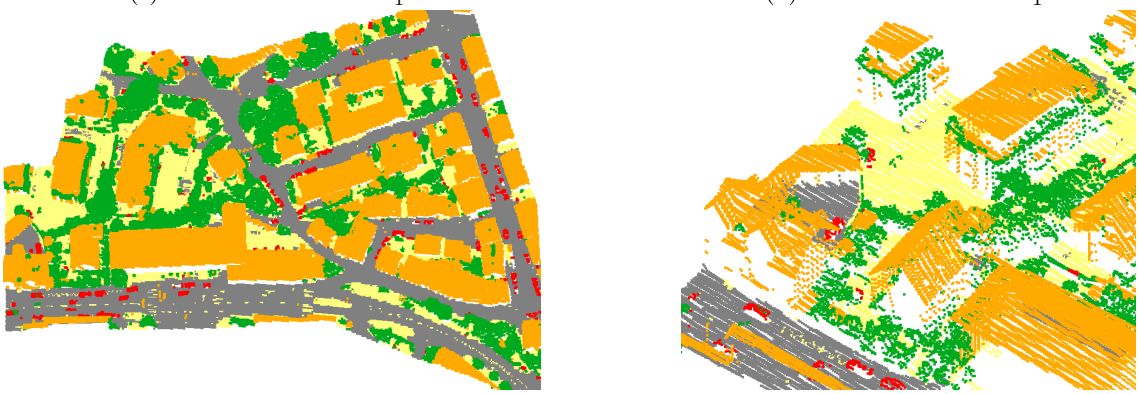
In den Abbildungen 5.13 werden die drei verschiedenen Klassifikationsergebnisse der Referenz des nördlichen Testgebiets visuell gegenüber gestellt. Für RF^P (c, d) zeigt sich erneut eine inhomogene Zuordnung der Klassenlabels. Die CRF-Resultate sind homogener und weisen weniger Fehlzuordnungen auf. Dies ist z. B. gut an den Gebäuden in den rechten Abbildungen (d, f, h) zu erkennen. Durch den Kontext können zahlreiche Verwechslungen erkannt und korrigiert werden, bei denen Fassaden fälschlicherweise als *Vegetation* klassifiziert wurden. Weiterhin verfügt das Gebäude im linken vorde-



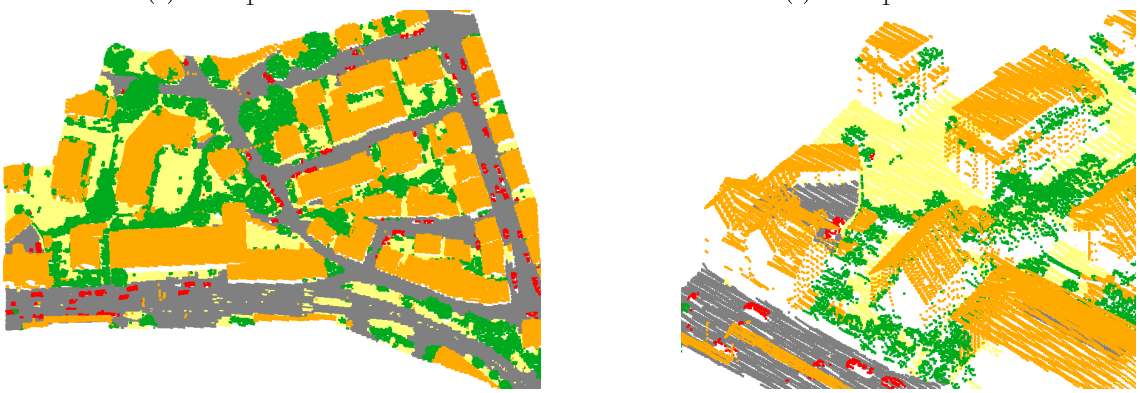
(a) Referenz (b) Referenz



(c) Random Forest RF_1^P (d) Random Forest RF_1^P



(e) CRF_1^P Iteration 1 (f) CRF_1^P Iteration 1



(g) CRF_3^P Iteration 3 (h) CRF_3^P Iteration 3

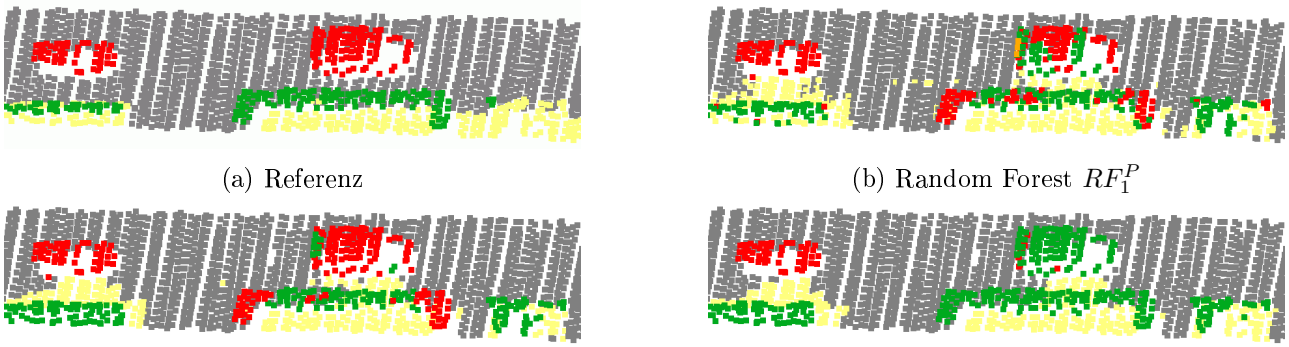
Abbildung 5.13: Klassifikationsergebnis Vaihingen. Die Abbildungen links zeigen das nördliche Testgebiet; rechts ist ein Ausschnitt aus etwas anderer Perspektive detaillierter visualisiert. Unterschieden werden die fünf Klassen *Boden* (gelb), *Straße* (grau), *Auto* (rot), *Vegetation* (grün) und *Gebäude* (orange).

ren Bereich des Ausschnitts über eine komplexe Dachstruktur mit vielen Erkern. Die lokalen Merkmale induzieren bei RF^P (Abbildung 5.13d) teilweise eine Zugehörigkeit zu *Vegetation*. Einige der Fehler lassen sich bereits durch die Hinzunahme lokaler Abhängigkeiten zwischen den Punkten durch CRF_1^P korrigieren (Abbildung 5.13f). Die Anwendung des gesamten Modells (CRF_3^P) führt in Abbildung 5.13h zur Beseitigung aller Verwechslungen, so dass die Gebäudepunkte vollständig und korrekt klassifiziert werden können.

Auch für Vaihingen sollen die antrainierten Potentialgewichte angegeben werden. Sie wurden optimiert auf die Werte $\omega_\psi^P = 0,2$, $\omega_\chi^P = 5,94$ (Iteration 1) und $\omega_\psi^P = 0,56$, $\omega_\xi^P = 0,29$ (Iteration >1) sowie $\omega_\psi^S = 1,1$. Die Berechnung dauert inklusive Training ca. 30 min (davon Extraktion Punktmerkmale ca. 3 min, reine RF Klassifikation ca. 2 min, CRF_1^P ca. 7 min). Die verkürzte Laufzeit im Vergleich zu Hannover ist auf die geringere Punktanzahl zurückzuführen. Dies macht sich besonders bei der Optimierung der punktbasierten Gewichte bemerkbar, da in jedem Iterationsschritt der direkten Suche die Inferenz für das Validierungsgebiet durchzuführen ist. Im Durchschnitt ergeben sich für jedes Segment sechs Kanten. Somit ist für Hannover das Inferenzproblem durch eine größere mittlere Kantenanzahl rechenintensiver.

Diskussion: Insgesamt sind die erzielten Ergebnisse sehr positiv zu beurteilen. Starke gewinnbringende Änderungen sind hauptsächlich bei den weniger dominanten Klassen zu beobachten, daher ist die Auswirkung auf die Gesamtgenauigkeit vergleichsweise gering. Die Verbesserungen lassen sich vor allem anhand der Qualitätsmaße und der visuellen Interpretation erkennen. Wie in den Abbildungen 5.12 gezeigt, ergibt sich dabei meistens ein deutlich homogeneres Ergebnis als bei einer kontextfreien Klassifikation. Dieses kann für nachfolgende Applikationen, wie etwa der Objektrekonstruktion, besser genutzt werden, da sich mehr Informationen über die tatsächliche Objektgeometrie ableiten lassen. Teilweise wird jedoch auch ein geringer Glättungseffekt beobachtet, durch den vereinzelte, durch wenige Punkte beschriebene Objekte nicht mehr korrekt klassifiziert werden können. Es kann zusammengefasst werden, dass sich im Mittel über alle Klassen die Vollständigkeit nach drei Iterationen im Vergleich zu RF_1^P um 2,1 (Hannover) bzw. 1,9 Prozentpunkte (Vaihingen) und die Korrektheit durchschnittlich um 8,9 (Hannover) bzw. 9,7 Prozentpunkte (Vaihingen) erhöht. Somit profitiert insbesondere die Korrektheit von dem auf diese Weise integrierten Kontext. Des Weiteren verbessern sich nahezu alle Qualitätswerte durch CRF_3^P. Für beide Testgebiete ist der größte Gewinn mit 21,7 (Hannover) bzw. 26,1 Prozentpunkten (Vaihingen) für die Klasse *Auto* festzustellen, welche sich durch eine vergleichsweise geringe Punktanzahl auszeichnet. Die Mehrdeutigkeiten im Merkmalsraum bei Punkten unterrepräsentierter Klassen scheinen somit durch die Betrachtung der Umgebung zum Teil gelöst zu werden, so dass sich das korrekte Label zuordnen lässt. Ferner sind deutlichere positive Veränderungen hinsichtlich der Qualität bei schwächeren Initialklassifikationen (RF_1^P von Hannover) zu beobachten. In diesem Fall trägt die Kontextinformation wesentlich zur Verbesserung der Zuordnungen bei.

Wie Abbildung 5.14 zeigt, kann die Berücksichtigung der zusätzlichen Kontextinformation in wenigen Fällen auch Fehlklassifikationen hervorrufen. In dem Ausschnitt ist eine Straße mit zwei Autos (Mitte) zu sehen. Am unteren Rand befinden sich parallel dazu Hecken, welche Grundstücke von der Straße abtrennen. Der Fokus der Untersuchung soll auf dem rechten Auto liegen, da das linke Fahrzeug fehlerfrei identifiziert wird. Bei einer kontextfreien RF_1^P Klassifikation (Abbildung 5.14b) werden viele



(a) Referenz

(b) Random Forest RF_1^P

(c) CRF^P Iteration 1

(d) CRF^P Iteration 3

Abbildung 5.14: Fehlerfortpflanzung durch Kontext bei einem Auto. Unterschieden werden *Boden* (gelb), *Straße* (grau), *Auto* (rot), *Vegetation* (grün) und *Gebäude* (orange).

Punkte auf dem rechten Auto richtig erkannt, es sind jedoch auch Verwechslungen mit *Vegetation* und sogar einigen *Gebäude*punkten zu beobachten. Des Weiteren werden viele Punkte der unmittelbar benachbarten Hecke ebenfalls fälschlicherweise der Klasse *Auto* zugeordnet. Die Verwechslungen sind auf die volumenhafte Punktverteilung zurückzuführen, die bei der Trennung der beiden Klassen auf Grundlage der genutzten Merkmale zu Mehrdeutigkeiten führt. Nach der Integration von lokalem Kontext durch die initiale semantische Zuordnung von CRF_1^P (Abbildung 5.14c) wird die Fehleranzahl auf dem Auto signifikant reduziert. Nur noch vereinzelte Punkte werden der Klasse *Vegetation* zugeordnet. Die Auswirkungen auf die Hecke sind hingegen gering. Das Propagieren der Informationen über zwei Maßstäbe (CRF_3^P, Abbildung 5.14d) korrigiert die dort verbliebenen Fehler bei der Hecke, die nun korrekt und vollständig als *Vegetation* identifiziert ist. Allerdings zeigt sich, dass auch das sich daneben befindliche Auto zu einem Großteil *Vegetation* zugeordnet wurde. An dieser Stelle wirkt sich der Kontext somit nachteilig auf das Ergebnis aus. Eine Ursache für den Fehler könnte in den daneben befindlichen und als *natürlicher Boden* klassifizierten Punkten zu finden sein. Der Anteil der Bodenpunkte an der umgebenen Fläche ist vergleichsweise groß, so dass vermutlich die kontextbasierten Merkmale zur Beschreibung der Nachbarschaftsklassen ein weiterer Auslöser zusätzlich zur Geometrie der Punktverteilung sein können. Da Autos meistens vollständig von Straßenpunkten umgeben sind, kann die Nachbarschaft eines Segments zu *Boden* auf ein Gebüsch hindeuten. Sollten die anderen Segmentmerkmale für das Autosegment eher wenig diskriminativ sein, kann dies den Ausschlag für eine Klassifikation als *Vegetation* geben. Beim linken Auto ist der *Boden*anteil etwas geringer und die ebenere Form des Fahrzeuges lässt sich eindeutiger zuordnen. Durch die nahezu vollständige Fehlklassifikation in eine einheitliche Objektklasse kann bei einer eventuellen Folgeiteration ggf. eine Aggregation zu einem einzigen Segment stattfinden, aus dem sich eindeutige Merkmale bezüglich Form und Aufenthaltsort des Segments zur korrekten Klassifikation extrahieren lassen. Diese Art der Korrektur ist nur möglich, so lange Autopunkte nicht mit der benachbarten Hecke zusammengefasst werden. Die Segmentbildung mittels VCCS, welche ebenfalls Nachbarschaftsbeziehungen berücksichtigt, sollte diesem Problem durch den sich dazwischen befindlichen Straßenstreifen vorbeugen. Folglich könnte dieser Fehler ggf. durch eine höhere Iterationsanzahl eliminiert werden.

5.4.2 Vergleich mit anderen Verfahren

Experiment: In dieser Untersuchung werden die Ergebnisse denen anderer Verfahren gegenübergestellt. Für einen aussagekräftigen Vergleich der Methoden liegen öffentlich zugängliche Benchmark Datensätze mit gleichen semantischen Objektklassen zu Grunde. Auf diese Weise lassen sich die Vor- und Nachteile der jeweiligen Ansätze identifizieren.

a) GML:

Beschreibung: Für die beiden GML Datensätze (GML A und GML B) werden alle dem Autor bekannten Veröffentlichungen herangezogen, die entsprechende Ergebnisse publiziert haben. Zusätzlich zu den eigenen Ergebnissen ergeben sich dadurch vier Vergleichswerte. Shapovalov et al. [2010] führen den Benchmark Datensatz ein und verwenden nicht-assoziative Markov Felder zur Segmentklassifikation. Xiong et al. [2011] propagieren Informationen auf Punkt- und Segmentebene zur Integration von Kontext in einem hierarchischen Ansatz. Najafi et al. [2014] arbeiten ebenfalls mit Segmenten. Sie nutzen ein Verfahren zur Berücksichtigung nicht-assoziativer Beziehungen mit Cliquen höherer Ordnung im Rahmen eines CRF. Bei [Blomley et al., 2016b] steht nicht die Entwicklung eines starken Klassifikationsverfahrens im Mittelpunkt der Untersuchung. Der Fokus liegt stattdessen auf der Ermittlung geeigneter Merkmale und Nachbarschaftsgrößen für deren Berechnung. Die Ergebnisse sollten dementsprechend nur als Ergänzung zu den anderen Verfahren angesehen werden. Da bei Blomley et al. [2016b] viele Experimente durchgeführt wurden, wird repräsentativ das beste Ergebnis einer RF Klassifikation zum Vergleich ausgewählt. Es wurde mit zwei Merkmalsgruppen und einer mehrskaligen Nachbarschaft erzielt, welche sich sowohl aus einer zylindrischen als auch einer kugelförmigen Umgebung ergibt.

Für die Datensätze ist weder Intensitätsinformation vorhanden, noch ist der genaue Ort bekannt. Aus dem Grund kann die Straße nicht extrahiert bzw. durch externe Datenquellen ergänzt werden, so dass sich dementsprechend die Merkmalsgruppe bestehend aus *Orientierung* sowie *minimalem* und *durchschnittlichen Abstand zur nächsten Straße* nicht nutzen lässt. Echoinformationen sind ebenfalls nicht vorhanden. Davon abgesehen werden die in Abschnitt 5.3.1 bestimmten Merkmale extrahiert. Die Berechnung erfolgt mit den in Abschnitt 5.3.2 definierten Parametern mit $N_{iter} = 3$ Iterationen.

Ergebnis: Insgesamt kann die neue Methodik gut die Schwierigkeiten des anspruchsvollen Geländes von GML A handhaben. Die Gesamtgenauigkeit bei GML A wird durch die Iterationen von 94,5 % (CRF_1^P) auf 96,4 % (CRF_3^P) erhöht. Verglichen mit einer RF Klassifikation (unäres Potential in CRF^P) ergibt sich eine Steigerung um 4,3 Prozentpunkte. Datensatz GML B hingegen wird bereits durch den initialen RF mit einer Gesamtgenauigkeit von 97,7 % nahezu fehlerfrei zugeordnet. Aus diesem Grund kann sie durch die Hinzunahme von Kontext nach drei Iterationen nur geringfügig um 0,5 Prozentpunkte gesteigert werden. Das Testgebiet ist somit einfacher und bietet u. a. durch die geringere Klassenanzahl und bessere Separierbarkeit weniger Potential für Verbesserungen durch die Verwendung von Kontext.

Die Tabellen 5.12a und 5.12b stellen die Ergebnisse der unterschiedlichen Verfahren vor. Gezeigt werden die Qualitätswerte der einzelnen Klassen sowie deren Mittelwert. Fett markierte Einträge stellen jeweils das genaueste Ergebnis dar. Abgesehen von zwei Ausnahmen erzielt die neu vorgeschlagene Methodik die besten Ergebnisse, teilweise mit deutlichem Abstand zum Zweitplatzierten. Insbeson-

	Boden	Gebäude	Auto	Bäume	niedrige Veg.	Mittel
Shapovalov et al. [2010]	86,8	53,7	12,6	92,1	8,6	50,7
Xiong et al. [2011]	93,2	71,6	9,2	97,0	**20,0**	58,2
Najafi et al. [2014]	91,3	61,6	**24,9**	95,1	15,9	57,8
Blomley et al. [2016b]	60,4	10,9	12,2	83,1	6,7	34,7
neue Methode (CRF_3^P)	**93,7**	**88,3**	20,0	**97,9**	18,8	**63,7**

(a) Datensatz GML A

	Boden	Gebäude	Bäume	niedrige Veg.	Mittel
Shapovalov et al. [2010]	97,0	72,9	85,7	20,8	69,1
Xiong et al. [2011]	98,0	77,4	94,2	35,6	76,3
Najafi et al. [2014]	98,0	85,2	94,2	38,9	79,1
Blomley et al. [2016b]	85,9	35,9	72,3	20,1	53,5
neue Methode (CRF_3^P)	**98,7**	**92,6**	**94,8**	**42,4**	**82,1**

(b) Datensatz GML B

Tabelle 5.12: Vergleich der erzielten Qualitätswerte [%] für Benchmark Datensatz GML.

dere die Qualität der Gebäudepunkte wird signifikant von 71,6 % [Xiong et al., 2011] auf nun 88,3 % gesteigert. Im Mittel ergibt sich eine Verbesserung von 5,5 Prozentpunkten für GML A bzw. von 3 Prozentpunkten bei GML B.

Eine Herausforderung bei GML A stellt die korrekte Klassifikation der Autos dar. Durch die eng nebeneinander parkenden Fahrzeuge ist die lokale Geometrie der Punkte ähnlich zu niedriger Vegetation, was häufige Verwechslungen mit sich führt. Alle fünf Verfahren haben Schwierigkeiten bei der korrekten Klassifikation von Autos. Die neue Methodik erreicht mit 20,0 % das zweitbeste Ergebnis, liegt damit aber fast 5 Prozentpunkte hinter [Najafi et al., 2014]. Im vorliegenden Fall ist das Problem bereits auf das unäre Potential von CRF^P in der ersten Iteration zurückzuführen. Durch die initiale Klassifikation mittels RF ist ein großer Teil der Autopunkte fälschlicherweise *niedriger Vegetation* zugeordnet. Es fehlt an weiteren diskriminativen Merkmalen wie etwa Intensität, um eine korrekte Zuordnung vorzunehmen. Durch die ähnliche lokale Geometrie zu den Nachbarpunkten findet durch das kontrast-sensitive bzw. das robuste P^n Potts Modell eine weitere Glättung statt. Insgesamt reduziert sich die Qualität aufgrund der niedrigen Vollständigkeit und Korrektheit von 27,1 % auf 20,0 %. Wie bereits anhand von Abbildung 5.14 diskutiert wurde, kann sich der hervorgerufene Glättungseffekt durch mehrdeutige Merkmale negativ auswirken, falls die Daten keinen ausreichenden Hinweis auf unterschiedliche Objektarten geben. Eine Anpassung der Interaktionsmerkmale, der Cliquengrößen während der initialen Iteration sowie der Anzahl der Durchläufe könnte ggf. die Erkennungsrate für Autos verbessern. Das bessere Abschneiden von [Najafi et al., 2014] kann in der mehrskaligen Merkmalsextraktion begründet sein. Die Autoren haben typische Merkmale mit zwei unterschiedlichen Nachbarschaftsgrößen extrahiert, welche eventuell einige Mehrdeutigkeiten bei kleineren Objekten wie Autos auflösen können. Insgesamt muss allerdings bedacht werden, dass die Klasse *Auto* nur durch wenige Punkte vertreten ist. Änderungen einzelner Punktlabels haben daher größeren Einfluss auf die Qualitätswerte.

b) Vaihingen 3D Labelling challenge (ISPRS)

Beschreibung: Ein weiterer Vergleich von Verfahren wird von der ISPRS Arbeitsgruppe II/4 organisiert. Bei der *3D Labelling challenge* steht der Datensatz *Vaihingen ISPRS 9 Klassen* zu Verfügung, auf dem verschiedene Methoden getestet werden können. Die Referenzdaten des Testgebiets sind unbekannt, da die Evaluierung von den Organisatoren vorgenommen wird. Als Vergleichsmaß liegt der Bewertung der F1 Wert (Gleichung 5.6) zugrunde. Bisher wurden mehrere Ergebnisse von fünf weiteren Teilnehmern eingereicht[3]. Für den Vergleich werden die jeweils besten Resultate herangezogen, wobei in einem Fall zwei Ergebnisse eines Teilnehmers genutzt werden. Insgesamt erfolgt somit eine Gegenüberstellung zu sechs Verfahren. Die meisten Ansätze sind bisher unveröffentlicht, so dass nur wenige Details zur konkreten Methodik bekannt sind.

Von Ramiya et al. werden zwei eingereichte Varianten (IIS_5 und IIS_7) basierend auf der Methodik von [Ramiya et al., 2016] als Vergleich verwendet. Beide extrahieren zunächst Segmente aus der Punktwolke, welche anschließend klassifiziert werden. Der Unterschied in den Verfahren besteht in den verwendeten Merkmalen. IIS_5 nutzt zusätzlich zu üblichen geometrischen Merkmalen die Intensität der Laserdaten, bei IIS_7 werden hingegen auf die Punkte projizierte Spektralinformation eines Falschfarben-Orthophotos verwendet. Im Vergleich zu den anderen Verfahren vereinen die Autoren die Klassen *Fassade* und *Dach* zu einer Kategorie, wodurch sich die Klassifikationsaufgabe vereinfacht. Cvirn et al. (UM) nutzen Signalformmerkmale, Texturanalysen sowie geometrische Merkmale für die punktweise Einordnung in Objektkategorien anhand eines überwachten RF Klassifikators. Yang (WhuY) verwendet ein CNN zur semantischen Zuordnung der Punkte. Das Verfahren HM_1 von Steinsiek et al. [2017] ist CRF^P der hier vorgestellten Methodik recht ähnlich. Die Autoren nutzen ein CRF mit RF und kontrast-sensitiver Glättung für die punktweise Klassifikation der Daten. Es besteht aus einer Ebene und kann nur lokalen Kontext modellieren. Blomley et al. [2016a] analysieren die Merkmalsextraktion und reichen dazu verschiedene Ergebnisse beim Benchmark ein. Analog zu GML liegt dabei der Fokus nicht auf der direkten Entwicklung eines Klassifikationsverfahrens, so dass die Ergebnisse nur ergänzend betrachtet werden sollen. Hier wird nur die beste Variante K_LDA berücksichtigt. Sie basiert auf einer punktweisen Klassifikation mittels *linear discriminant analysis* (LDA) und approximiert die Trainingsdaten jeder Klasse durch mehrdimensionale Gaußverteilungen. Das eigene Ergebnis wird mit den in Tabelle 5.7 festgesetzten Parameter und den ausgewählten Merkmalen aus Abschnitt 5.3.1 erzielt. Die Anzahl der Iterationen N_{iter} beträgt drei.

Ergebnis: Eine Übersicht über alle Ergebnisse hinsichtlich der F1-Werte und Gesamtgenauigkeiten (OA) findet sich in Tabelle 5.13. Auch bei diesem Benchmark erzielt die neu vorgeschlagene Klassifikationsmethodik die höchste Gesamtgenauigkeit. Mit 81,6 % ist diese um 0,8 Prozentpunkte besser als das zweitplatzierte Verfahren UM. Die CRF^P ähnliche Methodik von HM_1 erzielt aktuell die drittbeste OA mit 80,5 %. Die Genauigkeiten der übrigen Verfahren befinden sich zum Teil deutlich darunter. Auch bei Betrachtung der klassenweisen F1-Werte spiegelt sich dieser Unterschied wider. Mit der eigenen Methodik können für fünf der neun Klassen (*Auto*, *Zaun*, *Dach*, *Fassade* und *Baum*) die höchsten und für die verbleibenden Kategorien die zweithöchsten F1-Werte erzielt werden. Bei *Strauch* und *versiegelter Fläche* wird das beste Ergebnis von HM_1 fast erreicht. Somit zeigt sich die Stärke des Verfahrens erneut bei den unterrepräsentierten Klassen (*Auto*, *Zaun*, *Fassade*, *Strauch*).

[3]Stand: 17. April 2017; Ergebnisse abrufbar unter `http://www2.isprs.org/vaihingen-3d-semantic-labeling.html`

Methode	Stroml.	Boden	Vers.	Auto	Zaun	Dach	Fas.	Strauch	Baum	OA
IIS_5	54,4	59,7	83,0	33,5	23,5	83,4	–	38,7	56,1	68,0
IIS_7	54,4	65,2	85,0	57,9	28,9	90,9	–	39,5	75,6	76,2
UM	46,1	**79,0**	89,1	47,7	5,2	92,0	52,7	40,9	77,9	80,8
HM_1	**69,8**	73,8	**91,5**	58,2	29,9	91,6	54,7	**47,8**	80,2	80,5
WhuY	24,5	59,2	62,7	41,5	28,0	81,9	53,1	32,0	68,7	64,4
K_LDA	5,9	20,1	61,0	30,1	16,0	60,7	42,8	32,5	64,2	50,2
eigene (CRF_3^P)	59,6	77,5	91,1	**73,1**	**34,0**	**94,2**	**56,3**	46,6	**83,1**	**81,6**

Tabelle 5.13: Benchmarkergebnisse (F1 und OA in [%]) für *Vaihingen ISPRS 9 Klassen.*

Punkte der Klasse *Stromleitung* oberhalb der Gebäudedächer sind ebenfalls nur spärlich in den Gebieten vertreten. Sie stellen in den meisten Fällen Einzelpunkte ohne direkte Nachbarschaftspunkte in 3D dar. Aus diesem Grund lassen sie sich nach der Methodik in Abschnitt 4.3 nur selten zu sinnvollen Segmenten zusammenfassen. CRF^S führt bei dieser Klasse demnach nur zu geringen Verbesserungen. Dennoch wird im Vergleich das zweitbeste Ergebnis erreicht. Die Klassen *Boden* und *versiegelte Fläche* schneiden bei den drei besten Verfahren in etwa ähnlich ab. Auffällig ist, dass insgesamt abgesehen von einem einzigen Wert (F1 von *Boden*) alle höchsten Genauigkeiten von Verfahren basierend auf CRF erzielt werden.

Diskussion: Sowohl bei den beiden Datensätzen von GML als auch bei der ISPRS 3D Labelling challenge ist das vorgestellte Verfahren vergleichbaren Arbeiten überlegen. Die Untersuchungen stellen erneut die Wichtigkeit von Kontext für die Gesamtgenauigkeit und insbesondere für weniger dominante Klassen heraus. Die Auswertung der Klasse *Stromleitungen* zeigt in Bezug auf das eigene Verfahren, dass sich für den Nutzen gröberskaliger Informationen ausreichend viele Punkte einer Klasse in einer lokalen räumlichen Nachbarschaft befinden müssen. Andernfalls lassen sie sich nicht zu Segmenten zusammenfassen und können nicht von weitreichenderen Interaktionen profitieren. Die Zuordnung eines Klassenlabels für solche 3D-Punkte erfolgt dabei ausschließlich auf der Grundlage von CRF^P. Basierend auf einer visuellen Analyse der Zwischenergebnisse können dennoch einige Gruppen von Stromleitungspunkten ermittelt werden, welche zu Segmenten zusammengefasst und klassifiziert werden. Im Laufe der Iterationen verbessert sich auf diese Art die Vollständigkeit der vergleichsweise wenigen Punkte.

5.5 Evaluation ausgewählter Aspekte

5.5.1 Analyse der Modellkomponenten

Experiment: Diese Reihe an Experimenten soll den Beitrag einzelner Komponenten des Verfahrens zum Gesamtsystem aufzeigen. Für die Untersuchungen werden die Laserdaten zunächst mittels RF punktweise klassifiziert. Dieses Ergebnis dient als Bezug, da es einem herkömmlichen Klassifikator ohne Berücksichtigung von Kontextinformationen entspricht. Iterativ werden nun weitere Module des Verfahrens ergänzt, um deren Nutzen zu ermitteln. Auch alternative Ansätze, welche von der vorgestellten Methodik abweichen, werden berücksichtigt, wenn diese als naheliegend erscheinen. Tabelle

5.14a zeigt eine Übersicht über die Untersuchungen.

Die Experimente I-IV analysieren zunächst die punktweise Klassifikation. Ausgehend von der bereits angesprochenen reinen RF_1^P Klassifikation in Exp. I, welche das unäre Potential repräsentiert, werden einzeln das kontrast-sensitive paarweise Potts Modell (Exp. II), ein generisches paarweises Interaktionsmodell (Exp. III) sowie das robuste P^n Potts Modell für Cliquen höherer Ordnung in Exp. IV zusätzlich verwendet. Anstelle von Graph Cuts erfolgt in Exp. III die Inferenz aufgrund der generischen Klassenrelationen mit LBP. Experiment V stellt die in der Methodik verwendete Kombination von CRF^P bestehend aus drei Potentialen dar. Darauf aufbauend ergänzen die Experimente VI-IX eine segmentbasierte Klassifikation. Diese besteht in VI ausschließlich aus einer RF Klassifikation, welche im Folgenden das unäre Potential modelliert. Experiment VII analysiert den Effekt einer kontrast-sensitiven Glättung der Segmente, während die Interaktionen bei Exp. VIII alle Klassenbeziehungen über die Verbundwahrscheinlichkeiten in einem generischen Ansatz berücksichtigt. Untersuchung IX entspricht der Vorherigen mit der Erweiterung über drei Iterationen. Es handelt sich somit um die vorgestellte Methodik (Kapitel 4), bei der die Ergebnisse der verschiedenen Skalenbereiche propagiert werden. Die Potentiale der initialen Iteration werden mit denselben antrainierten RF Klassifikatoren von Exp. VIII berechnet. Das Ziel von Exp. X ist es, die Wichtigkeit der vorher durchgeführten punktweisen Klassifikation zu ermitteln, durch die sich die Kontextmerkmale verwenden lassen. Dabei handelt es sich um eine ausschließliche Segmentklassifikation ohne die Straßen- und Nachbarschaftsklassenverteilungsmerkmale. Da im Modell auch die Segmentierung von den vorherigen Ergebnissen abhängt, wird in diesem Fall eine Standardsegmentierung mit VCCS ebenfalls ohne die Konfidenzen aus CRF^P durchgeführt, d. h. $\omega_{konf}^{SV} = 0$. Das Exp. XI kompensiert die fehlende Vorklassifikation durch eine einfache RF Klassifikation ohne die Berücksichtigung von Interaktionen auf der Punktebene. Somit lassen sich für CRF^S in diesem Fall wieder die Kontextmerkmale verwenden. Es soll analysiert werden, ob die Verwendung eines CRF auf Punktebene einen Vorteil gegenüber der nachbarschaftsunabhängigen Klassifikation mit RF bietet. Das letzte Experiment (XII) ist mit der vorgestellten Methodik (Exp. IX) vergleichbar, da ebenfalls eine dreimalige Iteration über beide Ebenen erfolgt. Im Unterschied zu Exp. IX wird jedoch bei der initialen Durchführung von CRF^P auf das P^n Potts Modell verzichtet, um den Effekt dieses Terms auf das Gesamtergebnis zu ermitteln.

Die Testreihen werden sowohl für Vaihingen als auch für Hannover fünfmal wiederholt. Dabei wird für jedes Experiment einer Testreihe der selbe antrainierte RF^P verwendet. Auch die zugrundeliegenden Graphstrukturen (Kanten und Cliquen) der anderen punktweisen Potentiale sind identisch, falls diese Anwendung finden. Auf diese Weise werden Genauigkeitsschwankungen aufgrund von unterschiedlichen Trainingsbeispielen bzw. Graphen weitgehend vermieden, um den tatsächlichen Beitrag der jeweiligen Komponente genauer bestimmen zu können. Die relativen Gewichte der Potentiale werden hingegen für die einzelnen Konstellationen jeweils neu anhand der Validierungsdaten $\mathfrak{T}_2$ optimiert, um das bestmögliche Ergebnis für die Evaluierungsdaten $\mathfrak{E}$ zu erzielen. Zur Genauigkeitsevaluation von Exp. X und Exp. XI erfolgt die Übertragung der Segmentlabels auf die Punkte.

Ergebnis: In den Tabellen 5.14b bzw. 5.14c sind die über fünf Testläufe gemittelten Ergebnisse aller zwölf Experimente zusammengefasst. Angegeben werden Gesamtgenauigkeit, κ-Index sowie die fünf klassenbezogenen Qualitätswerte. Für das Bezugsexperiment I sind dabei die absoluten Maße dargestellt, alle folgenden Untersuchungen zeigen die Veränderungen auf. Negative relative Werte

repräsentieren eine geringere Genauigkeit, positive hingegen eine Verbesserung verglichen mit Exp. I.

Für beide Datensätze ist die Gesamtgenauigkeit bei einer punktweisen RF Klassifikation ohne die Berücksichtigung von Kontext vergleichsweise niedrig. Abgesehen von einem Experiment (Exp. X bei Vaihingen) erzielen alle weiteren Untersuchungen höhere Gesamtgenauigkeiten. Es werden 79,9 % (Hannover) bzw. 85,2 % (Vaihingen) erreicht. Der Datensatz Hannover scheint somit eine etwas größere Herausforderung für den Klassifikator darzustellen. Wie die folgenden Experimente zeigen, bietet sich hier ein größeres Potential für Verbesserungen durch Kontextinformation. Die Verwendung eines paarweisen Standard-CRF mit datenabhängiger Glättung durch das kontrast-sensitive Potts Modell (Exp. II) führt zu einer Steigerung von 1,3 Prozentpunkten (bzw. 0,4 bei Vaihingen). Generische Interaktionen zwischen den einzelnen Laserpunkten erzielen ein etwas schlechteres Ergebnis. Bemerkenswert ist der Qualitätsgewinn bei der Klasse *Auto* um 9,4 bzw. 7,3 Prozentpunkte. Dies entspricht den Beobachtungen in [Niemeyer et al., 2014], denen zufolge insbesondere unterrepräsentierte Klassen von dieser Art der Kantenmodellierung profitieren. Noch besser werden die Gesamtgenauigkeiten bei der Verwendung des robusten P^n Potts Modells ohne paarweise Interaktionen durch die Glättung in Cliquen höherer Ordnung, wie Exp. IV zeigt. Sie haben vor allem auf die Unterscheidung von *natürlichem Boden* und *Straße* einen nutzbringenden Effekt, unterstützten im Vergleich zu Exp. I aber auch die Identifikation von *Autos*. Die Kombination von RF, paarweisen Kanten und HOP in Exp. V repräsentiert die in Kapitel 4 vorgeschlagene Variante CRF^P der initialen Iteration. Die Genauigkeiten sind in etwa vergleichbar mit Exp. IV, jedoch unterscheiden sich die Ergebnisse für die beiden Datensätze geringfügig voneinander. In Hannover führt die Hinzunahme des kontrast-sensitiven Potts Modells zu einem weiteren Anstieg der OA von 3,0 auf 3,3 Prozentpunkte. Während *Boden* und *Straße* konstant bleiben, wirken sich die paarweisen Beziehungen vor allem auf *Auto*, *Vegetation* und *Gebäude* positiv aus. In Vaihingen hingegen liegt Exp. V um 0,1 Prozentpunkte unterhalb des Genauigkeitsniveaus von Exp. IV. Dies ist im Wesentlichen auf die etwas geringere Qualität bei den Punkten der Klasse *Straße* zurückzuführen. Insgesamt gesehen scheint die Kombination aus RF, kontrast-sensitiven und P^n Potts Modell eine geeignete Wahl zu sein. Generische Punktbeziehungen hingegen helfen durch den lokal zu stark eingeschränkten Kontext nur wenig bei der korrekten Klassifikation.

Die mit dieser Kombination erzielten Konfidenzen dienen als Grundlage für die Segmentierung und anschließende Klassifikation der Segmente mittels RF in Exp. VI. Dabei werden auch die neu vorgestellten Kontextmerkmale herangezogen. Insgesamt steigt durch den Übergang zu Segmenten die Genauigkeit sprunghaft an. Für Hannover ergibt sich eine Verbesserung von 4,9 Prozentpunkten verglichen zu Exp. I bzw. um 1,6 Prozentpunkte zum bisher besten Ergebnis von Exp. V. Besonders profitieren die Klassen *Boden*, *Straße* sowie *Auto*. Analog ist bei Vaihingen ein Anstieg um 2,0 Prozentpunkte im Vergleich zu Exp. I zu beobachten, was 1,1 Prozentpunkte höher ist als das beste punktweise Ergebnis. Bei diesem Datensatz ist der Qualitätsanstieg nicht auf einzelne Klassen zurückzuführen, da sich die Genauigkeitsmaße aller Objektkategorien deutlich verbessern. Experiment VII zufolge führt ein segmentbasiertes CRF mit kontrast-sensitiver Glättung zwischen benachbarten Primitiven bei Hannover zu einem Rückgang und in Vaihingen zu einer leichten Verbesserung der Genauigkeiten verglichen zur RF Klassifikation (Exp. VI).

Anstelle assoziativer Interaktionen werden in den beiden Experimenten VIII und IX generische Beziehungen zwischen den Segmenten mittels Verbundwahrscheinlichkeiten repräsentiert. Die Kom-

Experiment	*punktweise Klassifikation*				*segmentweise Klassifikation*			Durchläufe
	unär	paarweise		HOP	unär	paarweise		
	RF	Potts	generisch	P^n	RF	Potts	generisch	
I	x							1
II	x	x						1
III	x		x					1
IV	x			x				1
V	x	x		x				1
VI	x	x		x	x			1
VII	x	x		x	x	x		1
VIII	x	x		x	x		x	1
IX	x	x		x	x		x	3
X					x		x	1
XI	x				x		x	1
XII	x	x			x		x	3

(a) Übersicht über die Experimente. Dabei entspricht Version IX dem Vorgehen aus Kapitel 4.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
OA	79,9	1,3	0,3	3,0	3,3	4,9	4,3	4,6	**5,2**	2,3	3,2	3,5
Kappa	73,1	1,7	0,3	3,9	4,3	6,3	5,6	6,0	**6,7**	3,0	4,3	4,6
Boden	53,0	1,3	0,4	5,8	5,7	9,9	9,0	9,5	**10,1**	5,2	4,2	4,4
Straße	41,6	2,1	0,6	5,0	4,9	7,4	5,6	8,6	**9,1**	7,0	9,8	7,8
Auto	49,6	5,2	8,0	9,4	12,0	19,6	13,8	20,4	**22,3**	-9,1	19,4	21,3
Vegetation	84,3	1,9	-0,2	2,0	3,3	3,3	3,5	2,8	**3,7**	-0,2	2,0	3,7
Gebäude	92,0	1,1	-0,4	1,7	2,5	**2,8**	2,6	1,4	2,5	-0,1	0,1	2,1
Rechenzeit	5	+1	+34	+5	+6	+11	+13	+25	+70	+13	+23	+65

(b) Datensatz Hannover

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
OA	85,2	0,4	0,3	0,9	0,8	2,0	2,1	2,1	2,4	-0,8	2,1	**2,5**
Kappa	80,4	0,5	0,3	1,2	1,1	2,5	2,7	2,7	3,1	-1,2	2,7	**3,2**
Boden	63,4	0,3	0,1	1,2	1,0	2,6	2,1	2,7	3,1	-2,6	2,6	**3,4**
Straße	78,7	1,1	0,0	2,6	2,3	3,4	3,9	4,0	**4,1**	0,8	3,6	4,1
Auto	37,3	3,5	8,9	7,3	6,7	15,3	12,1	19,9	**23,1**	-11,5	10,9	21,8
Vegetation	70,5	-0,1	0,5	-0,2	-0,0	1,4	1,4	1,3	2,2	-4,9	1,8	**2,7**
Gebäude	85,8	0,5	0,1	1,2	1,1	3,1	3,9	3,0	3,3	0,1	3,2	**3,3**
Rechenzeit	4	+1	+29	+2	+2	+6	+7	+11	+26	+3	+10	+25

(c) Datensatz Vaihingen

Tabelle 5.14: Übersicht über die Experimente und Ergebnisse der Komponentenanalyse, gemittelt über jeweils fünf Wiederholungen. Gezeigt werden OA, κ-Index sowie die Qualität jeder Klasse. Für Exp. I (reine RF Klassifikation) sind die absoluten Werte in % angegeben. Alle weiteren Genauigkeitsmaße innerhalb einer Zeile stellen die Unterschiede zu Exp. I in Prozentpunkten dar. Das beste Ergebnis pro Zeile ist fett markiert. Zusätzlich werden jeweils die Rechenzeiten in Minuten relativ zu Exp. I bereitgestellt. Die Versuchsreihen für Hannover wurden auf einem anderen Rechner als jene für Vaihingen durchgeführt.

ponenten entsprechen somit dem in Kapitel 4 eingeführtem Modell. Der Unterschied zwischen Exp. VIII und Exp. IX besteht in der Anzahl der Durchläufe. Während in Exp. VIII die Abfolge für die punkt- und segmentweise Berechnung nur einmalig erfolgt, wird die Prozedur in Exp. IX dreimal iteriert, um auf diese Weise die Ergebnisse zwischen den Entitätsstufen zu propagieren. Bereits bei Exp. VIII werden für Hannover bessere Resultate erzielt als bei der kontrast-sensitiven Glättung. Die Gesamtgenauigkeit steigt aufgrund einer geringeren Verwechslung zwischen *Boden* und *Straße* um 0,3 Prozentpunkte. Für Vaihingen ergibt sich die gleiche OA wie bei Exp. VII. Es zeigt sich jedoch, dass die generischen Interaktionen die Qualität von *Auto* in beiden Datensätzen deutlich steigert, während sie bei *Gebäude* leichte Verluste hervorruft. Das dreimalige Iterieren in Exp. IX führt bei allen Genauigkeitsmaßen zu weiteren Verbesserungen. Abgesehen von einer Ausnahme (Qualität von Gebäudepunkten) erreicht diese Konfiguration die höchsten Genauigkeiten aller Experimente der Testreihe von Hannover. Im Fall von Vaihingen sind sie mit den besten Ergebnissen aus Exp. XII vergleichbar. Die Gesamtgenauigkeiten können von 4,6 Prozentpunkten nach zwei zusätzlichen Iterationen auf 5,2 Prozentpunkte (Hannover) bzw. von 2,1 auf 2,4 Prozentpunkte (Vaihingen) erhöht werden.

Experiment X zeigt die Wichtigkeit der Vorklassifikation. Ohne die Konfidenzen entstehen einerseits bereits mehr Fehler bei der Segmentierung, welche nun im Wesentlichen auf den Normalenvektoren der 3D-Punkte basiert. Andererseits führt auch der reduzierte Merkmalssatz ohne die neuen Kontextmerkmale zu einer geringeren Separierbarkeit der Klassen. Dies äußert sich durch schlechte Genauigkeitswerte, welche in Vaihingen sogar unterhalb des Niveaus einer RF Klassifikation liegen. Kleinere Strukturen wie Autos werden signifikant schlechter zugeordnet. Wie Exp. XI zeigt, kann bereits eine RF Vorklassifikation die segmentbasierte Zuordnung deutlich verbessern, da die notwendigen Konfidenzen für Segmentierung und Merkmalsextraktion zur Verfügung stehen. Die Genauigkeiten des vollständigen CRF^P aus Exp. VIII werden jedoch insbesondere bei *Auto* nicht erreicht. Zur Analyse des Nutzens des P^n Potts Modells wird als letzte Untersuchung dieser Testreihe Exp. XII durchgeführt. Der Versuchsaufbau entspricht jenem von Exp. IX, allerdings wird bei der ersten der insgesamt drei Iterationen auf die HOP verzichtet. Für den Datensatz Vaihingen ergeben sich in beiden Fällen vergleichbare Ergebnisse, wobei die Genauigkeiten von Exp. XII in den meisten Fällen geringfügig höher x als jene von Exp. IX sind. Die Gesamtgenauigkeit unterscheidet sich um 0,1 Prozentpunkte. Auffällig ist, dass für die Klasse *Straße* ein um 2,1 Prozentpunkte besseres Ergebnis unter Verwendung des P^n Potts Modells in Exp. IX erlangt wird. Beim Datensatz Hannover ist der Nutzen der HOP in der initialen Iteration deutlich stärker ausgeprägt. In Bezug auf die Gesamtgenauigkeit schneidet Exp. IX im Vergleich zu Exp. XII mit 1,8 Prozentpunkten Unterschied merklich besser ab. Das Fehlen des P^n Potts Modells äußert sich in niedrigeren Genauigkeitsmaßen bei allen Klassen.

In den Tabellen 5.14b und 5.14c sind zusätzlich die jeweils über fünf Wiederholungen gemittelten und auf Minuten gerundeten Rechenzeiten angegeben. Analog zu den Genauigkeiten handelt es sich bei Experiment I um die Gesamtdauer, bei allen weiteren Experimenten um den Unterschied hinsichtlich der Rechenzeit verglichen mit der RF Klassifikation. Dabei ist anzumerken, dass die Versuchsreihen für den Datensatz Hannover auf einem anderen Rechner (Intel Xeon CPU E5-2667 v2 @ 3,30 GHz mit 39 GB Arbeitsspeicher) durchgeführt wurden, wodurch die absoluten Prozessierungszeiten nicht mit den Angaben in Abschnitt 5.4.1 (Hannover) überein stimmen. Dennoch lassen sich die relativen Auswirkungen unterschiedlicher Komponentenkonfigurationen auf die Prozessierungszeit aus den Wer-

ten ableiten. Aus den Angaben wird ersichtlich, dass die Berücksichtigung von Kanteninformationen in Form eines kontrast-sensitiven Potts Modells (Exp. II) bei beiden Datensätzen die Laufzeit der einfachen RF Klassifikation nur geringfügig um ca. 1 min erhöht. Das Anlernen generischer Punktinteraktionen in Experiment III ist hingegen zeitaufwändig und verlängert die Laufzeit von RF um etwa 30 min. Insbesondere beim Datensatz Hannover führt das Parametertraining der relativen Potentialgewichte zu einer etwas längeren Prozessierungszeit von 5 bzw. 6 min für die Experimente IV und V, da bei den gegebenen Startwerten eine größere Anzahl an Iterationsschritten bis zur Konvergenz erforderlich sind. Für Vaihingen ist für beide Experimente nur ein minimaler Unterschied von 2 min zu beobachten. In beiden Datensätzen ist der Aufwand für die zusätzlichen paarweisen Interaktionen in Experiment IV verglichen zu Variante IV vernachlässigbar gering. In Exp. XI findet erstmalig der Übergang auf die Segmentebene statt, wodurch sich bei beiden Datensätzen der Sprung der Bearbeitungszeit im Vergleich zu den vorhergehenden Untersuchungen erklären lässt. Auch bei CRF^S erfordert die zusätzliche kontrast-sensitive Glättung über paarweise Kanten (Exp. XII) mit nur wenigen Sekunden Unterschied keine signifikant längere Rechenzeit als die RF Klassifikation der Primitive. Die 2 min bei Hannover bzw. 1 min bei Vaihingen ergibt sich aus der weiteren Parameteroptimierung zur Bestimmung des relativen Potentialgewichts in CRF^S. Verwendet man anstelle des Potts Modells ein generisches Interaktionsmodell für die Segmente, muss ein weiterer RF Klassifikator angelernt werden. Dies resultiert in einer um 12 min (Hannover) und 4 min (Vaihingen) erhöhten Rechenzeit von Experiment VIII. Für die in Kapitel 4 vorgeschlagene Konfiguration (Exp. IX) wird durch die alternierende Vorgehensweise CRF^P dreimal und CRF^S zweimal berechnet, um auf diese Weise die Kontextinformationen zu propagieren. Dadurch wird erneut ein größerer Zeitbedarf erforderlich. Für Hannover ergibt sich im Durchschnitt die absolute Rechenzeit von 75 min, während diese für den Datensatz Vaihingen etwa 30 min beträgt. Experiment X wendet ausschließlich die segmentbasierte Klassifikation an, um die Bedeutung von der vorhergehenden punktweisen Klassifikation (CRF^P) zu ermitteln. Dadurch kann die Berechnungszeit verkürzt werden. Für eine homogene Darstellung in den Tabellen 5.14b und 5.14c wurden die tatsächlichen Laufzeiten von Exp. X ebenfalls relativ auf die punktbasierte Klassifikation aus Exp. I bezogen, obwohl der dort verwendete RF Klassifikator an dieser Stelle nicht genutzt wird. Absolut benötigten die Experimente X durchschnittlich 18 min für Hannover und ca. 7 min für Vaihingen. Experiment XI nutzt einen RF Klassifikator vor der Berechnung von CRF^S für eine einfache punktweise Klassifikation. Wie zu erwarten ist, ergibt sich eine etwas kürzere Laufzeit als bei Exp. VIII, bei dem das vollständige CRF^P angewendet wird. Im letzten Experiment XII wird durch das Deaktivieren des P^n Potts Modells in der initialen CRF^P Klassifikation der Einfluss dieses Potentials ermittelt. Nach drei Iterationen ergibt sich eine nur geringfügig kürzere Laufzeit als bei dem Vergleichsexperiment IX, da im Wesentlichen CRF_1^P von der Änderung betroffen ist. Der Unterschied beträgt für Hannover 5 min und für Vaihingen 1 min.

Diskussion: Beide Datensätze weisen ein leicht unterschiedliches Verhalten auf. In Hannover haben die Verwechslungen zwischen beiden Bodenklassen (*Boden* und *Straße*) großen Einfluss auf die Genauigkeitswerte. Dies liegt vor allem an den Parkflächen, welche mit Schotter bedeckt sind und sich somit ausschließlich aufgrund der Intensitätswerte keiner der beiden Klassen eindeutig zuordnen lassen. Im Allgemeinen werden die Ergebnisse der ausführlichen Untersuchung in Abschnitt 5.4.1 auch durch die zusätzlichen Testreihen bestätigt. Die Verbesserungen im Vergleich zu RF^P erreichen dem-

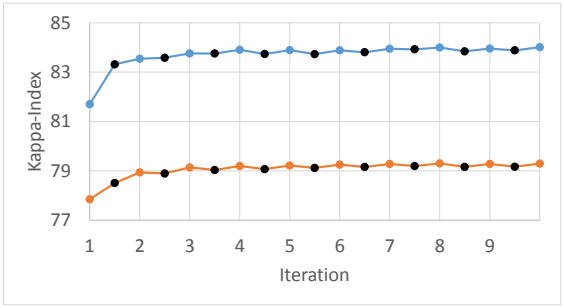
nach für Hannover und Vaihingen jeweils ein ähnliches Niveau zu den aus fünf Durchläufen gemittelten Resultaten in Tabelle 5.14c. Ein hierarchisches Verfahren führt im Vergleich zu einer ausschließlichen Klassifikation auf einer Punkt- oder Segmentebene zu einem deutlichen Genauigkeitsanstieg. Der erforderliche Mehraufwand für die Berechnung des umfangreicheren Verfahrens bestehend aus zwei Ebenen ist dem Ergebnis dienlich und ermöglicht eine signifikant höhere Qualität bei der Klassifikation unterrepräsentierter Objektklassen. Weiterhin wird aus den Experimenten in Tabelle 5.14 die Bedeutung des aufwändigen CRF^P anstelle einer Vorklassifikation mittels RF ersichtlich, da die Genauigkeiten vieler Klassen erst durch das volle Modell der Punktebene profitieren. Anhand der Ergebnisse kann gezeigt werden, dass die zusätzliche Verwendung der Potentiale höherer Ordnung durch das P^n Potts Modell in der ersten von drei Iterationen zu vergleichbaren (Vaihingen) oder deutlich besseren Ergebnissen (Hannover) im Vergleich zu einer Vernachlässigung dieses Terms führen, weshalb der geringfügig höhere Rechenaufwand in Kauf genommen werden kann.

5.5.2 Konvergenzuntersuchung

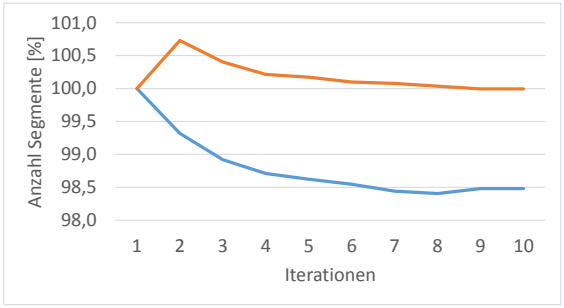
Experiment: Im Folgenden soll das Konvergenzverhalten der vorgestellten Methodik evaluiert werden. Dazu wird die Berechnung mit einer hohen Iterationszahl durchgeführt. Die Untersuchung soll zeigen, ob und wie sich die Genauigkeitsmaße von Schritt zu Schritt unterscheiden. Der Unterschied zwischen punktweiser und segmentbasierter Klassifikation wird dabei ebenfalls betrachtet. Weiterhin gilt es, eine geeignete Einstellung des Parameters N_{iter} für die Anzahl an Durchläufen zu ermitteln. Für das Experiment werden fünf Testreihen mit jeweils zehn Iterationen berechnet. Die Analyse basiert auf den gemittelten Werten jeder Iteration.

Ergebnis: Die Diagramme in Abbildung 5.15 zeigen die Ergebnisse der Konvergenzuntersuchung für die beiden Datensätze Hannover (orange) und Vaihingen (blau). Dabei visualisiert Abbildung 5.15a die pro Iteration erzielte Genauigkeit repräsentiert durch den κ-Index. Die Zwischenergebnisse der segmentbasierten Klassifikation CRF^S werden durch schwarze Markierungen illustriert. Generell lässt sich für beide Datensätze eine leichte, kontinuierliche Steigerung der Genauigkeiten feststellen. Während anfangs größere Qualitätssteigerungen zu beobachten sind, konvergiert die Genauigkeit etwa ab der dritten Iteration. Besonders die initiale Hinzunahme des gröberskaligen Kontexts durch CRF^S führt zu deutlichen Verbesserungen von Iteration 1 auf 2. Weiterhin lässt sich ein Konvergenzverhalten zwischen der punkt- und segmentbasierten Ebene feststellen, wobei (abgesehen vom ersten Durchlauf) die Genauigkeit von CRF^S etwas unterhalb von CRF^P liegt. Dies ist vermutlich auf eine geringe Anzahl verbleibender Generalisierungsfehler im Zuge der Segmentierung zurückzuführen.

Die rechte Abbildung 5.15b veranschaulicht die relative Anzahl der generierten Segmente pro Iterationsschritt. Im Wesentlichen bedingt durch die verschiedenen Gebietsgrößen variiert die absolute Zahl der Segmente in beiden Datensätzen deutlich. So ergeben sich während der ersten Iteration im Durchschnitt etwa 32.400 Segmente für Hannover und 9.115 Segmente für das Testgebiet von Vaihingen. Diese Werte entsprechen 100 % im Diagramm 5.15b; um einen Vergleich zwischen den Datenätzen zu ermöglichen, wird eine relative Repräsentationsform gewählt. Die Segmentanzahl jeder Folgeiterationen bezieht sich dabei jeweils auf die erstmalig erhaltene Anzahl. Analog zur linken Abbildung wird das Ergebnis von Hannover in orange und jenes von Vaihingen in blau dargestellt.



(a) κ-Index



(b) Segmentanzahl

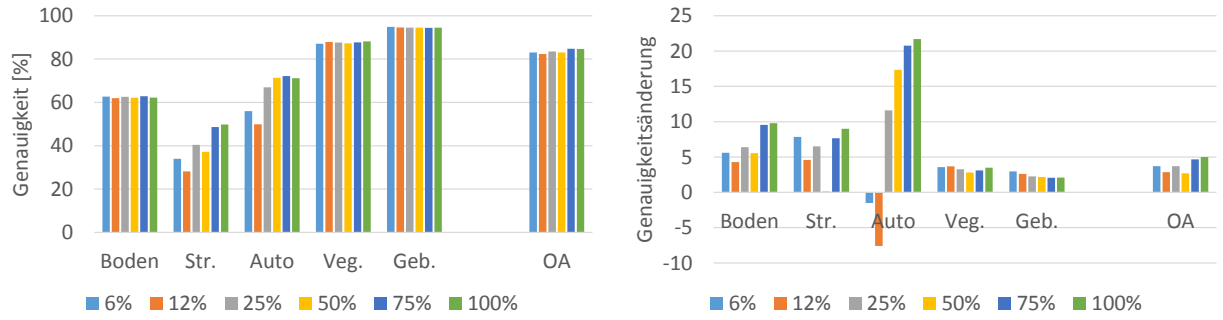
Abbildung 5.15: Konvergenzanalyse a) des κ-Index und b) der Segmentanzahl. Die orangefarbene Kurve stellt das Ergebnis von Hannover dar, blau jene von Vaihingen. In a) sind die Zwischenergebnisse von CRF^S schwarz markiert.

Beide Kurven weisen im Verlauf der Iterationen eine leichte Veränderung der Segmentanzahl auf, welche sich etwa ab der siebten Iteration zu stabilisieren scheint. Bei Vaihingen liegt dieser Wert ca. 1,5 Prozentpunkte unterhalb des Startwerts. In Bezug auf Hannover (orangefarbene Kurve) wird zu Beginn ein geringfügiger Anstieg der Segmentanzahl um 0,7 Prozentpunkte ersichtlich. Dies kann auf eine vergleichsweise ungeeignete Anfangssegmentierung hinweisen, bei der im Zuge von einigen Untersegmentierungsfehlern zunächst unterschiedliche Objekte zu Segmenten zusammengefasst wurden. Die inhomogenen Konfidenzwerte innerhalb dieser Segmente führen zu deren Aufspaltung. Die auf diese Weise neu entstehenden, kleineren Einheiten können in den Folgeiterationen die feinen Strukturen in der Szene besser beschreiben und werden entsprechend wieder neu kombiniert, wodurch sich die Anzahl der Segmente wieder etwas verringert. Fehler der initialen Klassifikation lassen sich so korrigieren. Für den Datensatz Hannover entspricht die stabile Segmentanzahl nach sieben Iterationen und gemittelt über die fünf Testreihen in etwa wieder der anfänglichen Anzahl. Dabei sollte beachtet werden, dass sich dennoch die Zusammensetzung aller Segmente während der Iterationsschritte ändern kann.

Diskussion: Die geringfügige Veränderung der Segmentanzahl im Laufe der Berechnung ist auf die adaptive Segmentierungsverbesserung durch die aktuell vorliegenden Konfidenzen zurückzuführen. Viele Punkte scheinen ihre Zugehörigkeit zu den Segmenten im Laufe der Iterationen zu wechseln, so dass sich auch deren Zusammensetzung jeweils neu gestaltet. Demzufolge erscheint es sinnvoll, die beiden segmentbasierten RF ebenfalls in jeder Iteration neu anzulernen, anstatt auf einmalig aufgebaute Klassifikatoren zurückzugreifen. Unter Berücksichtigung der Ergebnisse aus den Abbildungen 5.15a und 5.15b erweist sich beispielsweise die Anwendung dreier Iterationen als geeigneter Wert für N_{iter} als Kompromiss zwischen Genauigkeit und Rechenaufwand.

5.5.3 Trainingsdaten

Experiment: Für die Funktionsweise des überwachten Klassifikationsverfahrens sind die Trainingsdaten grundlegend. Sie werden herangezogen, um einerseits die drei RF Klassifikatoren anzulernen und um andererseits auf den zunächst zurückgehaltenen Validierungsdaten eine Optimierung der relativen Potentialgewichte vorzunehmen. Dieses Experiment analysiert den Einfluss des zugrunde liegenden Trainingsdatensatzes hinsichtlich seiner Größe. Dafür werden die Gebiete $\mathfrak{T}_1$ und $\mathfrak{T}_2$ des Datensatzes



(a) Ergebnis CRF_3^P (b) Gewinn von CRF_3^P verglichen mit RF_1^P

Abbildung 5.16: Einfluss des reduzierten Trainingsdatensatzes auf die Klassenqualität und OA [%].

Hannover durch das Auswählen zusammenhängender Teilbereiche auf ca. 75 %, 50 %, 25 %, 12 % und 6 % ihrer ursprünglichen Größe reduziert. Es wurde dabei auf die näherungsweise Beibehaltung der Klassenproportionen geachtet. Jedes Paar bestehend aus $\mathfrak{T}_1$ und $\mathfrak{T}_2$ dient als Trainingsdatensatz für unabhängig voneinander durchgeführte Vergleichsberechnungen über drei Iterationen auf dem gesamten Testgebiet $\mathfrak{E}$. Zum Vergleich wird das Ergebnis aus Abschnitt 5.4.1 herangezogen, bei dem 100 % der Trainingsdaten zur Verfügung standen. Die in Abschnitt 5.3.2 identifizierten RF Parameter bleiben bei diesem Versuch unverändert, so dass einzelne Trainingsbeispiele in den reduzierten Datensätzen häufiger verwendet werden. Die nachfolgenden Ergebnisse beziehen sich auf Mittelwerte zweier Testreihen, welche sich beide ähnlich verhalten. Die Untersuchung soll evaluieren, wie viele Trainingsdaten für ein zufriedenstellendes Klassifikationsergebnis notwendig sind, da die Erstellung von Referenzlabels durch das manuelle Annotieren mit einem hohen Zeit- und Kostenaufwand verbunden ist. Die Menge der zur Verfügung stehenden Trainingsdaten wird sich vor allem auf die Klassifikatoren RF_u^S und RF_p^S von CRF^S auswirken, da auf dieser Ebene durch die zu Segmenten aggregierten Punkte im Vergleich zu CRF^P weniger Trainingsbeispiele vorhanden sind. Insbesondere RF_p^S erfordert für jede Beziehung ausreichende Trainingsmengen, um die Daten einer der maximal L^2 Klassen zuzuordnen.

Ergebnis: Die Ergebnisse sind in Abbildung 5.16a zusammengefasst. Gruppiert nach der Klasse bzw. OA werden die erzielten Qualitätswerte bzw. Genauigkeiten durch die Balkenhöhen repräsentiert. Eine Auswertung der Gesamtgenauigkeiten lässt keine wesentliche Abhängigkeit zur Größe des Trainingsgebietes erkennen. Maximal unterscheiden sie sich um 2,5 Prozentpunkte. Ähnlich verhält es sich mit den klassenbezogenen Qualitäten von *Boden*, *Vegetation* und *Gebäude*, bei denen sich der Einfluss des Trainingsdatenumfangs um höchstens 1,1 Prozentpunkte auswirkt. Hingegen sind bei den beiden Klassen *Straße* und *Auto* deutliche Auswirkungen zu beobachten. Im maximalen Fall wird die Qualität um etwa 22 Prozentpunkte gesteigert. Bei diesen Klassen zeichnet sich die Tendenz ab, dass mehr Trainingsbeispiele zu besseren Genauigkeiten führen.

Diskussion: Für alle sechs untersuchten Genauigkeitsmaße ist keine strikte lineare Abhängigkeit zur Anzahl der verfügbaren Trainingsdaten erkennbar. So werden die durchschnittlich schlechtesten Ergebnisse für den auf 12 % reduzierten Datensatz erzielt. Dies kann auf eine gestiegene Sensitivität hinsichtlich der Qualität einzelner Trainingsbeispiele hinweisen. Bei wenigen Trainingsdaten wird demnach die diskriminative Aussagekraft der Merkmale bedeutender, da einzelne Primitive mehrfach für das Anlernen der RF verwendet werden. Potentielle Fehler in den Referenzlabels der Trainingsdaten

oder schwer trennbare Merkmale eines Primitivs können demnach die Genauigkeit in einem größeren Maße negativ beeinflussen als bei einer ausreichend großen Menge individueller Trainingsbeispiele. Aus dem Vergleich der CRF_3^P Ergebnisse mit denen von RF_1^P (Abbildung 5.16b) wird weiterhin ersichtlich, dass der über die Iterationen integrierte Kontext bei den unsichereren Klassen *Boden*, *Straße* und *Auto* mit zunehmender Anzahl an Trainingsdaten zu einer Genauigkeitssteigerung beiträgt. Dies kann vor allem an einem robuster angelerntem RF_p^S liegen. Eine nicht ausreichende Menge an Trainingsdaten führt bei *Auto* sogar zu einer Qualitätsminderung. Zusammenfassend lässt sich ein leichter Einfluss des Trainingsdatenumfangs auf die Gesamtgenauigkeit festhalten. Stehen mehr Trainingsdaten zur Verfügung, profitieren insbesondere die weniger dominanten Klassen und solche mit mehrdeutigen Merkmalen von der Kontextinformation.

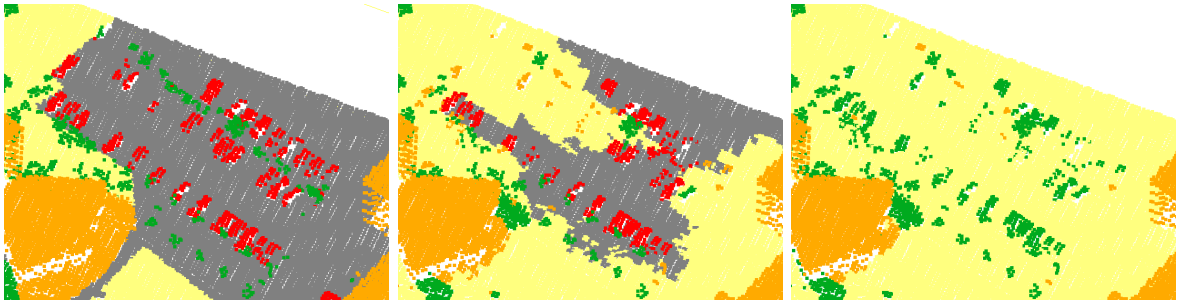
5.5.4 Straßenmerkmale

Experiment: Diese Arbeit stellt zwei Gruppen für segment- bzw. objektbasierte Kontextmerkmale vor. Im Rahmen dieses Experiments sollen die Merkmale *Orientierung*, *mittlerer* sowie *kürzester Abstand zur nächstgelegenen Straße* im Hinblick auf die zugrundeliegende Datenquelle evaluiert werden. Der Nutzen der neuen Kontextmerkmale wurde bereits in Abschnitt 5.3.1 anhand hoher Wichtigkeitswerte nachgewiesen. Vergleichbare Ansätze basierend auf terrestrischen Applikationen sind bei der Extraktion relativer Merkmale in Bezug zur Straße auf externe GIS-Daten wie etwa OSM angewiesen [Golovinskiy et al., 2009]. Die Realisierung des hierarchischen Modells in der vorliegenden Arbeit kann durch die alternierende Vorgehensweise auf die Zwischenlösungen einer vorherigen punktweisen Klassifikation zurückgreifen, um den Straßenverlauf zu approximieren. Es gilt zu untersuchen, ob bei der Berücksichtigung von GIS-Daten signifikante Unterschiede erzielt werden.

Aufbauend auf den Ergebnissen von CRF^P vergleicht diese Untersuchung Klassifikationsvarianten mit unterschiedlichen Merkmalssätzen bei einer einmaligen Durchführung von CRF^S. Es werden alle in Abschnitt 5.3.1 ausgewählten Segmentmerkmale mit Ausnahme der Nachbarschaftsklassenverteilung verwendet, um den Einfluss anderer Kontextmerkmale zu vermeiden. Variante 1, im Folgenden als R_{OSM} bezeichnet, verwendet zur Extraktion der Straßenmerkmale einen vorgegebenen Straßenverlauf von OSM. Zur Diskretisierung der Vektordaten werden zwischen den Start- und Endpunkten eines Liniensegments mit einem Abstand von 1 m Punkte interpoliert und als eigenständige Punktwolke gespeichert. Das Auffinden der nächstgelegenen Segmentpunkte erfolgt dann mittels kd-Baum. In Variante 2 wird der Straßenverlauf wie in Abschnitt 4.4.2 beschrieben über eine Binärkarte mit anschließender Hough-Transformation auf Grundlage der CRF^P-Vorklassifikationsergebnisse approximiert. Dementsprechend wird sie als $R_{appr.}$ gekennzeichnet. Die Gegenüberstellung zwischen $R_{appr.}$ und R_{OSM} soll ggf. Rückschlüsse auf die Sensitivität der Merkmale ermöglichen, da die kontinuierlichen OSM-Daten den tatsächlichen Straßenverlauf ohne Parkplätze beschreiben. Durch die Approximation kann ein Straßenabschnitt hingegen in seiner Ausrichtung vom tatsächlichen Verlauf u. U. etwas abweichen. Beiden Tests liegt dasselbe Segmentierungsergebnis zugrunde. Durch die unterschiedlichen Merkmalssätze müssen die Klassifikatoren RF_u^S und RF_p^S für jede Variante neu antrainiert werden. Auf das Optimieren des relativen Potentialgewichtes ω_ψ^S wird verzichtet, um den effektiven Unterschied zwischen den Varianten bestimmen und somit den Einfluss der Straßenmerkmale ermitteln zu können.

		OA	κ-Index	Boden	Straße	Auto	Vegetation	Gebäude
Hannover	R_{OSM}	83,8	78,3	61,1	53,7	64,4	83,4	91,4
	$R_{appr.}$	+0,6	+0,7	+3,4	-2,9	+2,4	+0,5	-0,1
Vaihingen	R_{OSM}	85,1	80,2	65,8	81,8	48,5	66,1	82,8
	$R_{appr.}$	+0,7	+0,9	-0,4	-0,8	-1,2	+2,4	+2,7

Tabelle 5.15: Ergebnisvergleich von R_{OSM} (Absolutwerte in [%]) und $R_{appr.}$ (relativ zu R_{OSM}).



(a) Referenz (b) Ergebnis $R_{appr.}$ (c) Ergebnis R_{OSM}

Abbildung 5.17: Vergleich der Referenz mit $R_{appr.}$ und R_{OSM} im Bereich eines Parkplatz. In den GIS-Informationen sind nur die tatsächlichen Straßenverläufe enthalten, Parkplätze werden im zugrundeliegenden Datensatz nicht berücksichtigt. Aufgrund der großen Distanzen zur nächstgelegenen Straße kommt es daher bei R_{OSM} in diesem Beispiel zu Fehlklassifikationen.

Es wird auf $\omega_{\psi}^{S} = 1$ gesetzt. Für Hannover und Vaihingen werden jeweils drei Testreihen erstellt.

Ergebnis: Eine Übersicht über die gemittelten Genauigkeitswerte findet sich in Tabelle 5.15. Die Ergebnisse von $R_{appr.}$ stellen die Abweichung zu den absoluten Genauigkeitsmaßen von R_{OSM} dar. Mit 0,6 bzw. 0,7 Prozentpunkten Unterschied wird sowohl für Hannover als auch für Vaihingen eine geringfügig höhere OA mit dem approximierten Straßenverlauf auf Grundlage der Vorklassifikation erzielt. In Bezug auf die einzelnen Klassenqualitäten lässt sich festhalten, dass der approximierte Verlauf von $R_{appr.}$ zu einer etwas geringeren Qualität der Klasse *Straßen* führt. Im Gegensatz dazu verbessert sich *Vegetation*. Die verbleibenden Klassen beider Datensätze unterscheiden sich hingegen. Bei Hannover profitiert *Auto* von der approximierten Straßeninformation. Dies kann zum Teil auf die parkenden Fahrzeuge zurückgeführt werden, welche abseits der Straße (und somit entfernt vom nächstgelegenen Straßenpolygon in OSM) auf den Parkplätzen abgestellt sind. Ein Beispiel dafür ist in Abbildung 5.17 gezeigt. Anhand der Intensitäten lassen sich in diesen Gebieten versiegelte Flächen mit CRF^{P} partiell identifizieren, wodurch auch dort ein Straßensegment mittels Hough-Transformation angenähert werden kann und sich Punkte der Klasse *Auto* besser klassifizieren lassen (Abbildung 5.17b). Bei R_{OSM} (Abbildung 5.17c) spricht die Entfernung und Orientierung dieser Auto- und Bodensegmente gegen deren korrekte Zuordnung. Beim Datensatz Vaihingen ermöglicht die Nutzung des externen Straßenverlaufs von OSM eine bessere Identifikation von Punkten dieser Klasse.

Diskussion: Insgesamt sind die Ergebnisse ähnlich, so dass sich beide Quellen zur Extraktion der Merkmale nutzen lassen. Einerseits kann R_{OSM} zu Verwechslungen z. B. in Bereichen ohne eindeutige Fahrbahn führen. Andererseits profitiert die Klassifikationsgenauigkeit in Regionen mit geradlinigem

Straßenverlauf von der Verwendung externer Daten. Die Merkmale beschreiben die tatsächliche Situation in diesem Fall etwas präziser und gewinnen an Aussagekraft. Die neue Methode bietet den Vorteil, diese Merkmalsgruppe unabhängig von externen Datenquellen verwenden zu können. Bei fehlenden Intensitätswerten kann für diesen Zweck auf GIS-Daten zurückgegriffen werden, falls deren Qualität und Vollständigkeit ausreichend ist. Die Wichtigkeitsanalyse und dieses Experiment zeigen einen positiven Beitrag der untersuchten Merkmalsgruppe. Mit der derzeitigen Umsetzung reagieren die Merkmale allerdings vergleichsweise sensitiv auf Abweichungen des Straßenverlaufs. Voraussetzung für diskriminative Merkmale sind weiterhin Segmente, welche die einzelnen Objektinstanzen hinsichtlich Orientierung und Position gut beschreiben. Das Verschmelzen mehrerer Instanzen zu einem Objekt, etwa bei dicht nebeneinander angeordneten Bäumen, Gebäuden oder Autos, kann fehlerhafte Merkmale verursachen. Zur Steigerung der Robustheit des Verfahrens sollten diese Aspekte verbessert werden.

5.6 Diskussion

Auf Grundlage der Ergebnisse kann zusammenfassend gesagt werden, dass die in Abschnitt 1.2 aufgestellten Forschungsziele zu weiten Teilen erreicht werden konnten. Anhand der experimentellen Untersuchungen lässt sich die gute Genauigkeit der mit dem hier vorgestellten neuen Verfahren erzielten Ergebnisse nachweisen. Es zeigt sich gegenüber aktuellen Methoden hinsichtlich der erreichten Klassifikationsqualität bei Benchmark Datensätzen überlegen. Für alle drei Testgebiete können - teilweise mit deutlichem Vorsprung - die besten Gesamtgenauigkeiten erreicht werden (Abschnitt 5.4.2). Im Fokus dieser Arbeit steht die Integration von Kontextwissen, dessen Nutzen anhand der Experimente u. a. in Abschnitt 5.4.1 belegt wird. Die Modellierung statistischer Abhängigkeiten zwischen den Primitiven trägt zur Verbesserung der Ergebnisse bei. Verglichen mit einer herkömmlichen RF Klassifikation ohne Betrachtung des Kontexts (RF_1^P) kann die Gesamtgenauigkeit um bis zu fünf Prozentpunkte gesteigert werden. In herausfordernden Szenen mit einer unsicheren Initialklassifikation lassen sich dabei stärkere Verbesserungen erzielen. Im besonderen Maße profitieren die in einem Datensatz unterrepräsentierten Objektklassen. Die Kontextinformation kann dabei helfen, die Anzahl falscher Zuordnungen signifikant zu reduzieren, um auf diese Weise die Qualitätswerte dieser Klassen zu steigern. Die Stärke der Verbesserung scheint dabei mit dem Umfang der zur Verfügung stehenden Trainingsdaten zu korrelieren (Abschnitt 5.5.3). Diese Abhängigkeit ist für die übrigen Klassen weniger ausgeprägt.

Für das hierarchische Modell ist eine Segmentierung der Punkte erforderlich. Im Rahmen einer Aggregation besteht allgemein die Gefahr der Untersegmentierung, wodurch u. U. kleinere Strukturen in den Daten mit der Umgebung gruppiert werden und sich anschließend nicht mehr korrekt klassifizieren lassen. Das zweite Forschungsziel bestand daher in der Reduktion dieser Fehler, um dadurch zusätzlich die Zuordnungsgenauigkeiten von Details und Objektkategorien weniger Punkte zu erhöhen. Zur Reduktion dieser Fehler werden Konfidenzen einer a priori Klassifikation als zusätzliches Kriterium in VCCS integriert. In Abschnitt 5.5.2 wurde gezeigt, dass sich auf diese Weise die Segmentierung in jeder Iteration bis zu einer Konvergenz an die Daten adaptiert. Diese fortschreitende Anpassung der Segmentierung stellt den Grund für die alternierende Optimierung beider Hierarchieebenen dar, um auf diesem Wege Fehler zu vermeiden. Zudem belegen die Experimente durch die Berücksichtigung der Konfidenzen eine geringere Sensitivität des manuell vorzugebenden Parameters hinsichtlich der

Segmentgröße (Abschnitt 5.3.2). Auch dieses Ziel kann somit als erreicht betrachtet werden.

Das dritte Ziel bestand im Entwurf neuer Kontextmerkmale für luftgestützte Laserdaten. Es werden zwei Merkmalsgruppen entwickelt, die einerseits die Auftrittswahrscheinlichkeiten von Objektklassen benachbarter Segmente und andererseits die relative Orientierung und Anordnung von Objekten in der Szene modellieren. Eine Wichtigkeitsanalyse in Abschnitt 5.3.1 bestätigt den nutzbringenden Beitrag aller Kontextmerkmale. Insbesondere die zweite Merkmalsgruppe stellt einen Lösungsansatz dar, um anhand der Straßen eine relative Referenzrichtung für Objekte auch bei luftgestützten Laserpunktwolken festzulegen. Dies erfordert entweder eine gute Approximation des Straßenverlaufs auf Grundlage der Vorklassifikationsergebnisse oder die Integration externer GIS-Daten. Das Experiment in Abschnitt 5.5.4 hat eine vergleichsweise hohe Abhängigkeit von der Qualität der Datengrundlage aufgezeigt. An dieser Stelle besteht Potential für weitere Verbesserungen.

6 Schlussfolgerungen und Ausblick

In dieser Arbeit wurde ein neuer Ansatz zur Integration von mehrskaliger Kontextinformation für die Klassifikation luftgestützter Laserdaten vorgestellt. Er beruht auf einer Kombination zweier Conditional Random Fields, welche jeweils die statistischen Abhängigkeiten zu den räumlichen Nachbarschaften eines Primitives in den Prozess der semantischen Zuordnung integrieren. Dabei wird einerseits die lokale Umgebung individueller 3D-Laserpunkte zur datenabhängigen Glättung der Klassenlabels herangezogen. Andererseits ermöglicht die Berücksichtigung komplexer generischer Interaktionen zwischen benachbarten Segmenten die Modellierung der gemeinsamen Auftrittswahrscheinlichkeit zweier Klassen bei entsprechenden Daten. Ein Vorteil der beschriebenen Vorgehensweise besteht in der adaptiven Verbesserung der Segmentierung zur Reduktion von Untersegmentierungsfehlern. Die alternierende Optimierung beider Ebenen erlaubt es dabei, in einem vorherigen Schritt ermittelte Klassenkonfidenzen in den Segmentierungsprozess zu integrieren. Dazu wurde die *Voxel Cloud Connectivity Segmentation* um einen zusätzlichen Term erweitert. Weiterhin lassen sich im Rahmen der vorgestellten Methodik neue kontextbasierte Segmentmerkmale für luftgestützte Laserdaten ableiten.

Bei den experimentellen Untersuchungen werden für verschiedene Datensätze gute Klassifikationsgenauigkeiten erreicht. Vergleiche anhand von Benchmark Datensätzen zeigen eine Überlegenheit des Verfahrens gegenüber anderen Methoden hinsichtlich der Gesamtgenauigkeiten. Der wesentliche Fokus der Arbeit liegt auf der Analyse, inwiefern Kontextinformation die Ergebnisse beeinflusst. Im Allgemeinen profitieren nahezu alle Qualitätsmaße von der Modellierung der statistischen Abhängigkeiten. So werden beispielsweise die Gesamtgenauigkeiten eines herkömmlichen Klassifikators ohne Betrachtung von Kontext um bis zu 5 Prozentpunkte gesteigert. Insbesondere bei Klassen mit einer geringeren Punktanzahl in den Daten ist eine signifikante Verbesserung zu erkennen. Zum Beispiel ergibt sich für die Klasse *Auto* eine Steigerung der Qualität um über 20 Prozentpunkte. Dies ist vor allem auf eine Reduktion falsch positiver Zuordnungen zu diesen Klassen zurückzuführen. Folglich ist die Integration von Kontext in den Klassifikationsprozess von Punktwolken mit Hinblick auf die Genauigkeiten empfehlenswert, obwohl dies mit einem höheren Rechen- und Speicherbedarf einhergeht.

Im Rahmen dieser Dissertation konnten die in Kapitel 1 definierten Forschungsziele erreicht werden. Mit dem vorgestellten Verfahren lassen sich schon jetzt sehr gute Ergebnisse erzielen. Aufbauend auf den durch die Experimente gewonnenen Erfahrungen ergeben sich Anregungen für mögliche Folgeentwicklungen, welche zu weiteren Verbesserungen führen können.

Ein Aspekt für weiterführende Untersuchungen besteht in der Identifikation zusätzlicher aussagekräftiger *Segmentmerkmale* für luftgestützte Laserpunktwolken. Bei den meisten der genutzten Merkmale handelt es sich um Mittelwerte punktbasierter Eigenschaften. Charakteristiken der Segmentform und -größe hingegen sind entsprechend der Wichtigkeitsanalyse von geringerer Bedeutung, was auf die

vergleichsweise regelmäßigen Supervoxel zurückzuführen ist. Die genutzten Beispieldatensätze waren hinsichtlich der zur Verfügung stehenden Attribute im umfangreichsten Fall auf Intensität und die Echonummer beschränkt. Die Hinzunahme weiterer Informationen wie beispielsweise aus der Signalform abgeleitete Merkmale (Echobreite, Schräge) [Mallet, 2010] oder Spektralinformationen optischer Sensoren können ggf. zu einer weiteren Genauigkeitssteigerung beitragen.

Wie die Wichtigkeitsanalyse ergeben hat, kann die neu eingeführte Gruppe der kontextbasierten *Straßenmerkmale* bei der Steigerung der Klassifikationsgenauigkeiten helfen. Das weiterführende Experiment zeigt auf, dass die Merkmale noch vergleichsweise sensitiv auf zwei wesentliche Einflüsse reagieren: Sie werden einerseits durch den zugrundeliegenden Straßenverlauf und andererseits von der Qualität der approximierten Objekte beeinflusst, deren Orientierung und Abstand zur Straße bestimmt werden soll. In Bezug auf den ersten Punkt kann über eine Weiterentwicklung des Verfahrens zur Ableitung der Straßensegmente aus den Vorklassifikationsergebnissen nachgedacht werden. Aktuell basiert es auf Heuristiken und ist von mehreren manuell festzulegenden Parametern abhängig. Die Qualität der erstellten Binärkarte ist von Bedeutung, da lokale Verdeckungen oder beispielsweise Hofeinfahrten mit dem derzeit verwendeten Verfahren den tatsächlichen Straßenverlauf beeinflussen können. Eine Option zur Steigerung der Robustheit besteht darin, als zusätzliches Merkmal die Straßenbreite bei der Extraktion mit zu schätzen. Aufbauend auf den Vorklassifikationsergebnissen lässt sich alternativ etwa eine auf die Extraktion von Straßennetzwerken aus luftgestützten Laserdaten optimierte Methode wie beispielsweise von Hui et al. [2016] integrieren, um den Straßenverlauf robust bestimmen zu können. Der zweite Aspekt besteht in der Generierung aussagekräftiger Segmente, welche einzelne Objektinstanzen modellieren. Dies ist vor allem zur Orientierungsbestimmung mittels Hauptachsentransformation notwendig. Die momentane Umsetzung basiert auf der Bestimmung der Zusammenhangskomponente, indem benachbarte Punkte der gleichen Klasse zu einem Segment vereint werden. Die resultierenden Primitive umfassen in einigen Fällen mehrere Instanzen derselben Objektart. Ein ähnliches Ergebnis hat sich bei einem hier nicht dargestellten Experiment mittels FH-Algorithmus [Felzenszwalb & Huttenlocher, 2004] ergeben. Eine Verbesserung der Segmentierung durch Aufspaltung dieser Segmente (z. B. wie [Golovinskiy & Funkhouser, 2009] oder [Reitberger et al., 2009]) kann somit die Aussagekraft der Merkmale noch weiter erhöhen.

Die alternierende Methodik ist sowohl auf der Punktebene als auch auf der Segmentebene mit der mehrmaligen Durchführung einer Inferenz verbunden. Weiterhin müssen in jedem Iterationsschritt die Merkmale der Segmente neu berechnet werden, was insbesondere aufgrund der hochdimensionalen Interaktionsmerkmalsvektoren in Verbindung mit der Kantenanzahl rechen- und speicheraufwändig ist. Eine Weiterentwicklung kann demnach auf die *Reduktion der erforderlichen Ressourcen* abzielen. Dazu gibt es verschiedene Ansätze. Viele Primitive lassen sich vergleichsweise schnell und mit hohen Konfidenzen einer Objektklasse zuordnen, so dass ihr Label während der folgenden Iterationen konstant bleibt. Nach der aktuellen Vorgehensweise werden solche Primitive weiterhin betrachtet. Die Komplexität des Klassifikationsproblems lässt sich reduzieren, indem sicher zugeordnete Primitive von der weiteren Klassifikation ausgeschlossen werden. Unsichere Primitive, die etwa häufig an Objektgrenzen zu beobachten sind, werden hingegen mit dem Ziel der Korrektur weiterhin variabel im Klassifikationsvorgang berücksichtigt. Stabile Primitive können beispielsweise (ähnlich wie von Vosselman et al. [2017] vorgeschlagen) als konstanter Knoten mit ausgehenden gerichteten Kanten in das graphische

Modell integriert werden. Entsprechende Informationen über die eigene Klasse lassen sich weiterhin an die Nachbarn propagieren, ohne eine neue Klassifikation dieses Knotens vornehmen zu müssen. Auf diese Weise kann auch der Umfang der zu extrahierenden Merkmale reduziert werden. Bei dieser Vorgehensweise besteht jedoch die Gefahr, Verwechslungen mit hohen Zuordnungswahrscheinlichkeiten als konstant anzusehen, welche im ungünstigsten Fall den Fehler auf ihre Umgebung propagieren. Weiterhin ist ein Schwellwert zur Bestimmung der als zuverlässig angesehenen Knoten erforderlich.

Die Berücksichtigung von Kontextinformation wird zu einem wesentlichen Teil über die Segmentebene in CRF^S realisiert. Der zugrunde liegende Graph berücksichtigt die unmittelbaren Nachbarn. Eine Einführung von weiterreichenden Kanten zur Modellierung von Interaktionen über größere räumliche Distanzen kann den Nutzen von Kontext ggf. weiter erhöhen. In einer aufbauenden Untersuchung kann ermittelt werden, ob derartige Interaktionen hilfreich sind. Die verfügbare Vorklassifikation lässt sich dabei verwenden, um vor allem nutzbringende Kanten in den Graph zu integrieren. So kann es beispielsweise hilfreich sein, weitreichende Verbindungen zwischen den Autos entlang einer Straße aufzubauen, um daraus sicherer auf die korrekte Zuordnung der Primitive schließen zu können.

Wie bei allen überwachten Klassifikationsverfahren haben die *Trainingsdaten* einen maßgeblichen Einfluss auf die erzielbaren Genauigkeiten. Sie werden für das Optimieren der Parameter und das Anlernen der Klassifikatoren benötigt. Kritisch zeigt sich hierbei in erster Linie der verwendete RF Klassifikator RF_p^S zum Modellieren der Verbundwahrscheinlichkeiten von Segmentinteraktionen. In diesem Fall müssen maximal L^2 Klassen angelernt werden, da derzeit jede Interaktion als individuelle Klasse repräsentiert wird. Eine ausreichende Anzahl an Trainingsbeispielen ist für eine zuverlässige Trennung der Klassen im Merkmalsraum erforderlich. Im Rahmen eines Experiments hat sich die Tendenz gezeigt, dass umfangreichere Trainingsdatensätze zu einem höheren Nutzen von Kontext führen können. An dieser Stelle sind weitere Untersuchungen empfehlenswert, um die tatsächlich benötigte Menge an Trainingsbeispielen zu identifizieren, aber auch die Auswirkungen von Fehlern in den Referenzlabels zu ermitteln. Die manuellen Erstellung von Trainingsdaten stellt einen zeit- und kostenintensiven Prozess dar. Arbeiten wie beispielsweise von Li et al. [2016] versuchen, Referenzdaten möglichst automatisch zu generieren. Sie können genutzt werden, um Trainingsdaten in einem ausreichendem Umfang bereitzustellen. In diesem Zusammenhang sind auch das Transfer-Lernen (z. B. [Paul et al., 2016]) und die Klassifikation unter *label noise*, d. h. mit teilweise fehlerhaften Referenzlabels (z. B. [Maas et al., 2016]) geeignet, um bereits angelernte Klassifikatoren an den Merkmalsraum einer neuen Szene zu adaptieren bzw. um vorhandene GIS-Daten, die veraltet sein können, als Trainingsdaten zu nutzen.

Literaturverzeichnis

Aijazi AK, Checchin P, Trassoudaine L (2013) Segmentation Based Classification of 3D Urban Point Clouds: A Super-Voxel Based Approach with Evaluation. *Remote Sensing*, 5 (4): 1624–1650.

Albert L, Rottensteiner F, Heipke C (2015) An iterative Inference Procedure applying Conditional Random Fields for simultaneous Classification of Land Cover and Land Use. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. II-3/W5: 1–8.

Anand A, Koppula HS, Joachims T, Saxena A (2012) Contextually guided Semantic Labeling and Search for three-dimensional Point Clouds. *The International Journal of Robotics Research*, 32 (1): 19–34.

Anguelov D, Taskar B, Chatalbashev V, Koller D, Gupta D, Heitz G, Ng A (2005) Discriminative Learning of Markov Random Fields for Segmentation of 3D Scan Data. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 169–176.

Axelsson P (1999) Processing of Laser Scanner Data - Algorithms and Applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54 (2): 138–147.

Belgiu M, Drăguţ L (2016) Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114: 24–31.

Beraldin JA, Blais F, Lohr U (2010) Laser Scanning Technology. In: Vosselman G, Maas HG (eds) *Airborne and Terrestrial Laser Scanning* (pp. 1–42), 1. Auflage. Whittles Publishing. Dunbeath, Vereinigtes Königreich.

Besag J (1986) On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society, Series B (Methodological)*, 48 (3): 259–302.

Bishop CM (2006) *Pattern Recognition and Machine Learning*, 1. Auflage. Springer. New York, USA.

Blomley R, Jutzi B, Weinmann M (2016a) 3D Semantic Labeling of Als Point Clouds By Exploiting Multi-Scale , Multi-Type Neighborhoods for Feature Extraction. In: *International Conference on Geographic Object-Based Image Analysis (GEOBIA)*: 1–8.

Blomley R, Jutzi B, Weinmann M (2016b) Multi-Scale Features and Different Neighbourhood Types. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. III-3: 169–176.

Boulch A, Marlet R (2012) Fast and Robust Normal Estimation for Point Clouds with Sharp Features. *Computer Graphics Forum*, 31 (5): 1765–1774.

Boykov Y, Kolmogorov V (2004) An experimental Comparison of Min-Cut/Max- Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (9): 1124–1137.

Boykov Y, Veksler O, Zabih R (2001) Fast approximate Energy Minimization via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (11): 1222–1239.

Boykov YY, Jolly MP (2001) Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in ND Images. In: *IEEE International Conference on Computer Vision (ICCV)*, 1: 105–112.

Breiman L (1996) Bagging Predictors. *Machine Learning*, 24 (2): 123–140.

Breiman L (2001) Random Forests. *Machine learning*, 45 (1): 5–32.

Breiman L (2003) RF / tools - A Class of Two-Eyed Algorithms. In: *SIAM International Conference on Data Mining*: 1–56.

Breiman L, Friedman J, Stone CJ, Olshen R (1984) *Classification and Regression Trees*. Wadsworth International Group. Belmont, Kalifornien, USA.

Brenner C (2016) Scalable Estimation of Precision Maps in a MapReduce Framework. In: *International Conference on Advances in Geographic Information Systems*: 27:1–27:10.

Chehata N, Guo L, Mallet C (2009) Airborne Lidar Feature Selection for Urban Classification using Random Forests. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXXVIII-3/W8: 207–212.

Chen C, Liaw A, Breiman L (2004) *Using Random Forest to learn imbalanced Data*. University of California, Berkeley, USA. Technical report.

Comaniciu D, Meer P (2002) Mean Shift: A robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (5): 603–619.

Cramer M (2010) The DGPF-Test on Digital Airborne Camera Evaluation Overview and Test Design. *Photogrammetrie-Fernerkundung-Geoinformation*, 2010 (2): 73–82.

Criminisi A, Shotton J (2013) *Decision Forests for Computer Vision and Medical Image Analysis*. Springer. London, Vereinigtes Königreich.

Dohan D, Matejek B, Funkhouser T (2015) Learning Hierarchical Semantic Segmentations of LIDAR Data. In: *International Conference on 3D Vision*: 273–281.

Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7 (1): 1–26.

Felzenszwalb PF, Huttenlocher DP (2004) Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59 (2): 1–26.

Filin S, Pfeifer N (2005) Neighborhood Systems for Airborne Laser Data. *Photogrammetric Engineering & Remote Sensing*, 71 (6): 743–755.

Fischler Ma, Bolles RC (1981) Random Sample Consensus: A Paradigm for Model Fitting with Applicatlons to Image Analysis and Automated Cartography. *Communications of the ACM*, 24 (6): 381 – 395.

Ford L, Fulkerson D (1962) *Flows in Networks*. Princeton University Press. Princeton, New Jersey, USA.

Förstner W (2013) Graphical Models in Geodesy and Photogrammetry. *Photogrammetrie - Fernerkundung - Geoinformation*, 2013 (4): 255–267.

Frey B, MacKay DJC (1998) A Revolution: Belief Propagation in Graphs with Cycles. In: *Advances in Neural Information Processing Systems*. The MIT Press, 10: 479–485.

Gadde R, Jampani V, Marlet R, Gehler PV (2016) Efficient 2D and 3D Facade Segmentation using Auto-Context. In: *arXiv*. 1606.06437: 1–8.

Galleguillos C, Belongie S (2010) Context based Object Categorization: A Critical Survey. *Computer Vision and Image Understanding*, 114 (6): 712–722.

Geman S, Geman D (1984) Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 (6): 721–741.

Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random Forests for Land Cover Classification. *Pattern Recognition Letters*, 27 (4): 294–300.

Golovinskiy A, Funkhouser T (2009) Min-Cut based Segmentation of Point Clouds. In: *IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2009*: 39–46.

Golovinskiy A, Kim VG, Funkhouser T (2009) Shape-based Recognition of 3D Point Clouds in Urban Environments. In: *IEEE International Conference on Computer Vision (ICCV)*: 2154–2161.

Gonfaus JM, Boix X, Van De Weijer J, Bagdanov AD, Serrat J, Gonzàlez J (2010) Harmony Potentials for joint Classification and Segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 3280–3287.

Gould S, Fulton R, Koller D (2009) Decomposing a Scene into Geometric and Semantically consistent Regions. In: *IEEE International Conference on Computer Vision (ICCV)*: 1–8.

Gould S, Rodgers J, Cohen D, Elidan G, Koller D (2008) Multi-Class Segmentation with Relative Location Prior. *International Journal of Computer Vision*, 80 (3): 300–316.

Guo B, Huang X, Zhang F, Sohn G (2014) Classification of Airborne Laser Scanning Data using JointBoost. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92: 124–136.

Haala N, Brenner C (1999) Extraction of Buildings and Trees in Urban Environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54 (2-3): 130–137.

Hackel T, Wegner JD, Schindler K (2016) Fast Semantic Segmentation of 3D Point Clouds With Strongly Varying Density. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. III-3: 177–184.

Hänsch R (2014) *Generic Object Categorization in PolSAR Images - and beyond*. Dissertation, Technische Universität Berlin.

Heipke C, Mayer H, Wiedemann C, Jamet O (1997) Evaluation of Automatic Road Extraction. In: *International Archives of Photogrammetry and Remote Sensing*. XXXII-3/4W: 151–156.

Hellinger E (1909) Neue Begründung der Theorie quadratischer Formen von unendlich vielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909 (136): 210–271.

Hinz S, Baumgartner A (2003) Automatic Extraction of Urban Road Networks from Multi-View Aerial Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (1-2): 83–98.

Hoberg T, Rottensteiner F, Feitosa RQ, Heipke C (2015) Conditional Random Fields for Multitemporal and Multiscale Classification of Optical Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53 (2): 659–673.

Hooke R, Jeeves TA (1961) Direct Search Solution of Numerical and Statistical Problems. *Journal of the ACM*, 8 (2): 212–229.

Horvat D, Žalik B, Mongus D (2016) Context-dependent Detection of non-linearly distributed Points for Vegetation Classification in Airborne LiDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116: 1–14.

Hu X, Yuan Y (2016) Deep-Learning-Based Classification for DTM Extraction from ALS Point Cloud. *Remote Sensing*, 8 (9), Nr. 730: 1–16.

Huang J, You S (2016) Point Cloud Labeling using 3D Convolutional Neural Network. In: *International Conference on Pattern Recognition (ICPR)*: 1–6.

Hughes GF (1968) On the Mean Accuracy of Statistical Pattern Recognizers. *IEEE Transactions on Information Theory*, 14 (1): 55–63.

Hui Z, Hu Y, Jin S, Yevenyo YZ (2016) Road Centerline Extraction from Airborne LiDAR Point Cloud based on Hierarchical Fusion and Optimization. *ISPRS Journal of Photogrammetry and Remote Sensing*, 118: 22–36.

Ivănescu P (1965) Some Network Flow Problems Solved with Pseudo-Boolean Programming. *Operations Research*, 13 (3): 388–399.

Jancsary J, Nowozin S, Sharp T, Rother C (2012) Regression Tree Fields An efficient, non-parametric Approach to Image Labeling Problems. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2376–2383.

Kähler O, Reid I (2013) Efficient 3D Scene Labeling using Fields of Trees. In: *IEEE International Conference on Computer Vision (ICCV)*: 3064–3071.

Kenduiywo BK, Bargiel D, Sörgel U (2016) Crop Type Mapping From a Sequence of TerraSAR-X Images With Dynamic Conditional Random Fields. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. III-7: 59–66.

Kim BS, Kohli P, Savarese S (2013) 3D Scene Understanding by Voxel-CRF. In: *IEEE International Conference on Computer Vision (ICCV)*: 1425–1432.

Kirkpatrick S, Gelatt CD, P. VM, Vecchi M (1983) Optimization by Simulated Annealing. *Science*, 220, Nr. 4598: 671–680.

Kohli P, Kumar MP, Torr PHS (2007) P3 & Beyond: Solving Energies with Higher Order Cliques. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1–8.

Kohli P, Ladický LL, Torr PHS (2009) Robust Higher Order Potentials for enforcing Label Consistency. *International Journal of Computer Vision*, 82 (3): 302–324.

Kolmogorov V (2006) Convergent tree-reweighted Message Passing for Energy Minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (10): 1568–1583.

Kolmogorov V, Zabih R (2004) What Energy Functions can be Minimized via Graph Cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (2): 147–159.

Kosov S, Rottensteiner F, Heipke C (2013) Sequential Gaussian Mixture Models for Two-Level Conditional Random Fields. In: *German Conference on Pattern Recognition (GCPR), LNCS 8142*. Springer. Berlin-Heidelberg, Deutschland: 153–163.

Kramer O (2010) Iterated Local Search with Powell's Method: A Memetic Algorithm for Continuous Global Optimization. *Memetic Computing*, 2 (1): 69–83.

Kumar S, Hebert M (2005) A Hierarchical Field Framework for Unified Context-Based Classification. In: *International Conference on Computer Vision (ICCV)*: 1284–1291.

Kumar S, Hebert M (2006) Discriminative Random Fields. *International Journal of Computer Vision*, 68 (2): 179–201.

Ladický L, Russell C, Kohli P, Torr PHS (2013) Inference Methods for CRFs with Co-Occurrence Statistics. *International Journal of Computer Vision*, 103 (2): 213–225.

Ladický L, Russell C, Kohli P, Torr PHS (2014) Associative Hierarchical Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36 (6): 1056–1077.

Lafferty J, McCallum A, Pereira F (2001) Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: *International Conference on Machine Learning*: 282–289.

Lee I, Schenk T (2002) Perceptual Organization of 3D Surface Points. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXXIV-3A: 193–198.

Li SZ (2009) *Markov Random Field Modeling in Image Analysis*. Springer. London, Vereinigtes Königreich.

Li Z, Zhang L, Zhong R, Fang T, Zhang L, Zhang Z (2016) Classification of Urban Point Clouds: A Robust Supervised Approach with automatically generating Training Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10 (3): 1207 – 1220.

Lim EH, Suter D (2009) 3D terrestrial LIDAR Classifications with Super-Voxels and multi-scale Conditional Random Fields. *Computer Aided Design*, 41 (10): 701–710.

Liu DC, Nocedal J (1989) On the Limited Memory BFGS Method for Large Scale Optimization. *Mathematical programming*, 45 (1): 503–528.

Lloyd SP (1982) Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2): 129–137.

Lodha SK, Fitzpatrick DM, Helmbold DP (2007) Aerial Lidar Data Classification using Adaboost. In: *International Conference on 3D Digital Imaging and Modeling*: 435–442.

Lu WL, Murphy KP, Little JJ, Sheffer A, Fu H (2009) A Hybrid Conditional Random Field for Estimating the Underlying Ground Surface from Airborne LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 47 (8): 2913–2922.

Lucchi A, Li Y, Boix X, Smith K, Fua P (2011) Are spatial and global Constraints really necessary for Segmentation? In: *IEEE International Conference on Computer Vision (ICCV)*: 9–16.

Maas A, Rottensteiner F, Heipke C (2016) Using Label Noise Robust Logistic Regression for Automated Updating of Topographic Geospatial Databases. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. III-7: 133–140.

MacQueen JB (1967) Some Methods for classification and Analysis of Multivariate Observations. In: *Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 1: 281–297.

Mallet C (2010) Analysis of Full-Waveform lidar data for urban area mapping. Dissertation, Télécom ParisTech, Paris, Frankreich.

Matas J, Galambos C, Kittler J (2000) Robust Detection of Lines Using the Progressive Probabilistic Hough Transform. *Computer Vision and Image Understanding*, 78 (1): 119–137.

Melzer T (2007) Non-parametric Segmentation of ALS Point Clouds using Mean Shift. *Journal of Applied Geodesy*, 1 (3): 159–170.

Millard K, Richardson M (2015) On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, 7 (7): 8489–8515.

Munoz D, Bagnell JAD, Vandapel N, Hebert M (2009a) Contextual Classification with Functional Max-Margin Markov Networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1–8.

Munoz D, Vandapel N, Hebert M (2008) Directional Associative Markov Network for 3-D Point Cloud Classification. In: *International Symposium on 3D Data Processing, Visualization and Transmission*: 1–8.

Munoz D, Vandapel N, Hebert M (2009b) Onboard Contextual Classification of 3D Point Clouds with Learned High-Order Markov Random Fields. In: *International Conference on Robotics and Automation*: 1–8.

Najafi M, Namin ST, Salzmann M, Petersson L (2014) Non-associative Higher-Order Markov Networks for Point Cloud Classification. In: *European Conference on Computer Vision (ECCV)*: 500–515.

Nguatem W, Mayer H (2016) Contiguous Patch Segmentation in Pointclouds. In: *German Conference on Pattern Recognition (GCPR), LNCS 9796*. Springer International Publishing. Cham, Schweiz: 131–142.

Nguyen A, Le B (2013) 3D Point Cloud Segmentation: A survey. In: *IEEE Conference on Robotics, Automation and Mechatronics (RAM)*: 225–230.

Niemeyer J, Rottensteiner F, Sörgel U (2013) Classification of Urban LiDAR Data using Conditional Random Field and Random Forests. In: *Joint Urban Remote Sensing Event 2013*: 139–142.

Niemeyer J, Rottensteiner F, Sörgel U (2014) Contextual Classification of LiDAR Data and Building Object Detection in Urban Areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 152–165.

Niemeyer J, Rottensteiner F, Sörgel U, Heipke C (2015) Contextual classification of Point Clouds using a Two-Stage CRF. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XL-3/W2: 141–148.

Niemeyer J, Rottensteiner F, Sörgel U, Heipke C (2016) Hierarchical Higher Order CRF for the Classification of Airborne LiDAR Point Clouds in Urban Areas. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XLI-B3: 655–662.

Niemeyer J, Wegner JD, Mallet C, Rottensteiner F, Sörgel U (2011) Conditional Random Fields for Urban Scene Classification with Full Waveform LiDAR Data. In: *Photogrammetric Image Analysis, LNCS 6952*. Springer. Berlin-Heidelberg, Deutschland: 233–244.

Nowozin S, Rother C, Bagon S, Sharp T, Yao B, Kohli P (2011) Decision Tree Fields. In: *IEEE International Conference on Computer Vision (ICCV)*: 1668–1675.

Oliva A, Torralba A (2007) The Role of Context in Object Recognition. *Trends in Cognitive Sciences*, 11 (12): 520–527.

Papon J, Abramov A, Schoeler M, Worgotter F (2013) Voxel Cloud Connectivity Segmentation - Supervoxels for Point Clouds. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2027–2034.

Paul A, Rottensteiner F, Heipke C (2016) Iterative Re-Weighted Instance Transfer for Domain Adaptation. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. III-3: 339–346.

Pauly M, Keiser R, Gross M (2003) Multi-scale Feature Extraction on Point-Sampled Surfaces. *Computer Graphics Forum*, 22 (3): 81–89.

Pearl J (1982) Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach. *National Conference on AI*, : 133–136.

Pham TT, Reid I, Latif Y, Gould S (2015) Hierarchical Higher-order Regression Forest Fields: An Application to 3D Indoor Scene Labelling. In: *International Conference on Computer Vision (ICCV)*: 2246–2254.

Potts R (1952) Some Generalized Order-Disorder Transformations. In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge University Press, 48 (1): 106–109.

Ramiya AM, Nidamanuri RR, Ramakrishnan K (2016) A Supervoxel-based spectro-spatial Approach for 3D Urban Point Cloud Labelling. *International Journal of Remote Sensing*, 37 (17): 4172–4200.

Reitberger J, Schnörr C, Krzystek P, Stilla U (2009) 3D Segmentation of Single Trees exploiting Full Waveform LIDAR Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64 (6): 561–574.

Ren X, Malik J (2003) Learning a Classification Model for Segmentation. *International Conference on Computer Vision (ICCV)*, 1: 10–17.

Roig G, Boix X, Shitrit HB, Fua P (2011) Conditional Random Fields for multi-camera Object Detection. In: *IEEE International Conference on Computer Vision (ICCV)*: 563–570.

Rottensteiner F (2017) Kontextbasierte Ansätze in der Bildanalyse. In: Heipke C (ed) *Handbuch der Geodäsie, Band Photogrammetrie und Fernerkundung* (pp. 555–602). Springer. Berlin-Heidelberg, Deutschland.

Rottensteiner F, Sohn G, Gerke M, Wegner JD, Breitkopf U, Jung J (2014) Results of the ISPRS Benchmark on Urban Object Detection and 3D Building Reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93: 256–271.

Rusu RB, Holzbach A, Blodow N, Beetz M (2009) Fast Geometric Point Labeling using Conditional Random Fields. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*: 7–12.

Rutzinger M, Rottensteiner F, Pfeifer N (2009) A Comparison of Evaluation Techniques for Building Extraction from Airborne Laser Scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2 (1): 11–20.

Schindler K (2012) An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50 (11): 4534–4545.

Schnabel R, Wahl R, Klein R (2007) Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum*, 26 (2): 214–226.

Shapovalov R, Velizhev A, Barinova O (2010) Non-Associative Markov Networks for 3D Point Cloud Classification. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXXVIII-3A: 103–108.

Shapovalov R, Vetrov D, Kohli P (2013) Spatial Inference Machines. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2985–2992.

Shotton J, Winn J, Rother C, Criminisi A (2009) 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.

Sima MC, Nüchter A (2013) An Extension of the Felzenszwalb-Huttenlocher Segmentation to 3D Point Clouds. In: *SPIE International Conference on Machine Vision (ICMV)*. 8783: 6.

Steinsiek M, Polewski P, Yao W, Krzystek P (2017) Semantische Analyse von ALS- und MLS-Daten in urbanen Gebieten mittels Conditional Random Fields. In: *37. Wissenschaftlich-Technische Jahrestagung der DGPF in Würzburg*, Publikationen der DGPF, Band 26: 521–531.

Stutz D (2015) Superpixel Segmentation: An Evaluation. In: *German Conference on Pattern Recognition (GCPR), LNCS 9358*. Springer International Publishing. Cham, Schweiz: 555–562.

Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C (2008) A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (6): 1068–1080.

Tu Z, Bai X (2010) Auto-Context and its Application to High-Level Vision Tasks and 3D Brain Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 (10): 1744–1757.

Vapnik VN (1998) *Statistical Learning Theory*. Wiley. New York, USA.

Veksler O (1999) *Efficient Graph-Based Energy Minimization Methods in Computer Vision*. Dissertation, Cornell University, Ithaca, New York, USA.

Vosselman G (2013) Point Cloud Segmentation for Urban Scene Classification. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXXX-7/W2: 257–262.

Vosselman G, Coenen M, Rottensteiner F (2017) Contextual segment-based Classification of Airborne Laser Scanner Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128: 354–371.

Vosselman G, Klein R (2010) Visualisation and Structuring of Point Clouds. In: Vosselman G, Maas HG (eds) *Airborne and Terrestrial Laser Scanning*, 1. Auflage (pp. 43–79). Whittles Publishing. Dunbeath, Vereinigtes Königreich.

Wang LL, Yung NHC (2015) Improved Hierarchical Conditional Random Field Model for Object Segmentation. *Machine Vision and Applications*, 26 (7): 1027–1043.

Wegner J (2011) Detection and Height Estimation of Buildings based on Fusion of one SAR Acquisition and an Optical Image. *Deutsche Geodätische Kommission, Reihe C*, Heft Nr. 669.

Wegner JD, Montoya-Zegarra JA, Schindler K (2015) Road Networks as Collections of Minimum Cost Paths. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108: 128–137.

Weinmann M (2016) *Reconstruction and Analysis of 3D Scenes.* Springer International Publishing. Cham, Schweiz.

Wolf D, Prankl J, Vincze M (2015) Fast Semantic Segmentation of 3D Point Clouds using a Dense CRF with Learned Parameters. In: *IEEE International Conference on Robotics and Automation (ICRA)*: 4867–4873.

Wolf D, Prankl J, Vincze M (2016) Enhancing Semantic Segmentation for Robotics: The Power of 3-D Entangled Forests. *IEEE Robotics and Automation Letters*, 1 (1): 49–56.

Wolf L, Bileschi S (2006) A Critical View of Context. *International Journal of Computer Vision*, 69 (2): 251–261.

Xiong X, Munoz D, Bagnell JA, Hebert M (2011) 3-D Scene Analysis via Sequenced Predictions over Points and Regions. In: *IEEE International Conference on Robotics and Automation*: 2609–2616.

Yan WY, Shaker A, El-Ashmawy N (2015) Urban Land Cover Classification using Airborne LiDAR Data: A Review. *Remote Sensing of Environment*, 158: 295–310.

Yang MY, Förstner W (2011) A Hierarchical Conditional Random Field Model for Labeling and Classifying Images of Man-made Scenes. In: *IEEE International Conference on Computer Vision/ISPRS Workshop on Computer Vision for Remote Sensing of the Environment*: 196–203.

Yao J, Fidler S, Urtasun R (2012) Describing theScene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 702–709.

Zhang T, Suen C (1984) A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the ACM*, 27 (3): 236–239.

Danksagung

An dieser Stelle möchte ich einigen Personen danken, die mich in den letzten Jahren bei der Erstellung dieser Arbeit direkt und indirekt unterstützt und an mich geglaubt haben:

Mein besonderer Dank gilt meinem Doktorvater Prof. Dr.-Ing. Christian Heipke für die intensive fachliche Betreuung. Er unterstützte mich in jeder Hinsicht, schenkte mir Vertrauen und ermöglichte mir auch Freiräume, eigenen Forschungsinteressen nachzugehen. Darüber hinaus möchte ich apl. Prof. Dr.-Ing. Claus Brenner, Prof. Dr.-Ing. Uwe Sörgel und Prof. Dr. techn. Franz Rottensteiner dafür danken, dass sie das Korreferat für diese Dissertation übernommen und mich mit zahlreichen Gesprächen inspiriert und unterstützt haben. Auch sie haben im wesentlichem Maße zum Erfolg dieser Dissertation beigetragen. Weiterhin danke ich Prof. Dr.-Ing. Monika Sester für die Übernahme des Vorsitzes der Prüfungskommission.

Eine wichtige Komponente zur erfolgreichen Fertigstellung dieser Arbeit bildete die tolle Arbeitsatmosphäre am Institut für Photogrammetrie und GeoInformation (IPI) der Leibniz Universität Hannover. Allen Kollegen und Freunden möchte ich für die schöne gemeinsame Zeit danken. Insbesondere in meiner Bürogemeinschaft mit Alena und Lena gab es interessante fachliche Diskussionen, aufbauende Worte, aber auch zahlreiche Heißgetränke bei lustigen Rätselrunden in den Pausen. Jederzeit konnte ich mich auf ihre Unterstützung verlassen. Darüber hinaus danke ich auch Tobias für die jahrelange Freundschaft und den fachlichen Austausch in allen Belangen.

Ganz besonders möchte ich mich bei meinen Eltern Gabi und Joachim sowie meiner Schwester Anne für ihre uneingeschränkte Unterstützung und Förderung bedanken. Mein abschließender Dank geht an meine Freundin Jessica, die auch in den schwierigen Zeiten immer verständnisvoll war und mir jederzeit Rückhalt und Kraft gegeben hat.

Lebenslauf

Persönliches

Joachim-Christoph Niemeyer

geboren am 26. September 1984 in Heidelberg, Deutschland

Beruflicher Werdegang

seit 05/2017	R&D Software Engineer, Hexagon Technology Center GmbH, Heerbrugg, Schweiz
01/2010 - 12/2016	Wissenschaftlicher Mitarbeiter am Institut für Photogrammetrie und GeoInformation, Leibniz Universität Hannover, Deutschland
10/2009 - 12/2009	Wissenschaftliche Hilfskraft am Institut für Photogrammetrie und GeoInformation, Leibniz Universität Hannover, Deutschland

Auslandsaufenthalt

08/2014 - 11/2014	Gastwissenschaftler am Computer Vision Laboratory, Päpstliche Katholische Universität Rio de Janeiro, Brasilien

Ausbildung

10/2004 - 09/2009	Studium der Geodäsie und Geoinformatik an der Leibniz Universität Hannover, Deutschland. Abschluss: Diplom
08/1997 - 07/2004	Gymnasium Anna-Sophianeum Schöningen, Deutschland. Abschluss: Abitur