Geometric Features for Supervised Classification in Airborne Laser Scanning Data

Zur Erlangung des akademischen Grades einer

DOKTORIN DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften des

Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Phys. Rosmarie Blomley
aus Carlton, Australien

Tag der mündlichen Prüfung:	22. April 2025
Referent:	Prof. Dr.-Ing. Stefan Hinz
Korreferent:	Prof. Dr.-Ing. Markus Gerke

Abstract

This work is about utilizing the unique true-to-scale 3D sampling properties of airborne laser scanning as an active remote sensing technology for automated scene and data understanding by features extracted from the geometric relations among the data. This is particularly challenging given the medium pulse density of 5-20 $^{\mathrm{pulses}}/_{\mathrm{m}^2}$. Despite recent advances in the application of deep learning technology on unstructured point cloud data, the practical part of this research focuses on the traditional domain of machine learning using hand-crafted features, as it was conducted between 2013 and 2018. This is still highly relevant as a baseline for comparing and evaluating the use of different deep learning strategies. To this end, we provide a detailed review and comparison regarding the addressed tasks.

We analysed geometric features both on the more universal level of point-wise semantic labelling and on a more application-specific level of single tree species classification. For both of these applications, we implemented a new type of geometric features, each inspired by surface point descriptors from object recognition. In point-wise semantic labelling (performed on two publicly available benchmark data sets), we found it beneficial to use an adapted version of shape distributions to describe local neighbourhoods, and to use multi-scale, multi-type neighbourhoods as the basis for feature extraction. In single tree species classification (based on 3630 individual tree segments attributed to Scots pine (*Pinus sylvestris L.*), Norway spruce (*Picea abies (L.) Karst.*), and Birch (*Betula pendula Roth.* and *Betula pubescens Ehrh.*)), we found it beneficial to capture the geometric distribution of waveform attributes throughout the tree crown by an adapted version of spin images. This performed well compared to other geometric features and improved classification results when combined with statistical incidence metrics of waveform attributes that do not describe their geometric distribution. In both application cases, our work marked a significant contribution to the field.

In the case of point-wise semantic labelling, we concluded our review of current literature by noting the importance of context. Contextual classification can be achieved by structured prediction, such as conditional random fields in the domain of traditional machine learning. But context is also modelled by deep learning strategies such as convolutional neural networks or transformers. In the case of tree species classification, we could not find evidence in the field for clear advances on the task given the type of data we used. This is likely due to the limited number of classes, the limited amount of training data that can be collected, and the excellent integration of expert knowledge via application-specific hand-crafted features. In this field, modern advances come rather from high-resolution point densities enabled by refined sensor technology and unmanned-aerial-vehicle recording, from multi-spectral laser scanning, or from the combination with other data sources. All of these are more likely to profit from deep learning data analysis. Future research is likely going to profit from the collaborative collection of very large databases, as they are currently being initiated for tree species classification.

Zusammenfassung

In dieser Arbeit werden die charakteristischen maßstabsgetreuen 3D-Abtasteigenschaften von Airborne-Laserscanning als aktive Fernerkundungstechnologie für ein automatisiertes Szenen- und Datenverständnis genutzt, indem Merkmale aus dem geometrischen Zusammenhang der Daten extrahiert werden. Eine besondere Herausforderung ist dabei die mittlere Pulsdichte von 5-20 $^{\text{Pulsen}}/_{\text{m}^2}$. Trotz jüngster Fortschritte in der Anwendung von Deep-Learning-Technologien auf unstrukturierten Punktwolkendaten konzentriert sich der praktische Teil dieser Arbeit, wie er zwischen 2013 und 2018 durchgeführt wurde, auf traditionelles Machine-Learning unter Verwendung manuell erstellter Merkmale. Dies ist als Grundlage für den Vergleich und die Bewertung verschiedener Deep-Learning-Strategien nach wie vor von großer Bedeutung. Für die bearbeiteten Anwendungsbereiche erfolgt hierzu jeweils ein detaillierter Überblick und Vergleich zu aktuellen Studien.

Geometrische Merkmale wurden einerseits im universelleren Kontext der semantischen Einzelpunktklassifikation, andererseits im anwendungsspezifischen Kontext der Einzelbaum-basierten Baumartenklassifikation untersucht. Für beide Anwendungsbereiche wurde jeweils ein neuer Merkmalstyp, inspiriert durch Punktdeskriptoren aus der Objekterkennung, implementiert. In der semantischen Einzelpunktklassifikation zweier öffentlich zugänglicher Benchmarkdatensätze konnten Verbesserungen durch eine angepasste Implementierung von Shape Distributions zur Beschreibung lokaler Nachbarschaften, sowie durch die Kombination von Nachbarschaften unterschiedlicher Größe und Form als Basis der Merkmalsextraktion erzielt werden. Zur Baumartenklassifikation (basierend auf 3630 Einzelbaumsegmenten der Arten Kiefer (*Pinus sylvestris L.*), Fichte (*Picea abies (L.) Karst.*) und Birke (*betula pendula Roth.* sowie *Betula pubescens Ehrh.*)) konnten Verbesserungen durch die Beschreibung der geometrischen Verteilung von Waveform-Attributen innerhalb der Baumkrone, basierend auf einer angepassten Implementierung von Spin Images, erzielt werden. Diese konnten sich im Vergleich zu anderen geometrischen Merkmalen behaupten und verbesserten das Klassifikationsergebnis in Kombination mit statistischen Verteilungsmerkmalen von Waveform Attributen ohne die Berücksichtigung ihrer geometrischen Verteilung. Auf beiden Anwendungsgebieten stellt diese Arbeit einen signifikanten wissenschaftlichen Beitrag dar.

In der Einzelpunktklassifikation stellte sich unter Betrachtung aktueller Literatur die besondere Bedeutung von Kontextinformation heraus. Im Bereich des traditionellen Machine-Learnings kann diese durch kontextbasierte Klassifikation, wie z.B. durch Conditional Random Fields integriert werden. Darüber hinaus wird Kontextinformation aber auch von einigen Deep-Learning-Ansätzen wie z.B. Convolutional Neural Networks oder Transformerarchitekturen modelliert. In der Baumartenklassifikation konnten in aktueller Literatur keine Hinweise auf klare Fortschritte auf Basis des gegebenen Datentyps festgestellt werden. Dies erklärt sich durch die begrenzte Anzahl zu unterscheidender Klassen, durch die begrenzte Menge praktisch erhebbarer Trainingsdaten und durch die hervorragende Integra-

tion von Expertenwissen in anwendungsspezifischen, manuell erstellten Merkmalen. Fortschritte in diesem Anwendungsbereich können heutzutage durch besonders hochauflösende Punktdichten, ermöglicht durch Weiterentwicklungen in der Sensorik und der Dronen-gestützten Aufzeichnung, verzeichnet werden, sowie durch multispektrale Laserscanning-Aufnahmen oder die Kombination mit anderen Datenquellen. Diese Herangehensweisen erlauben überdies einen gewinnbringenderen Einsatz von Deep-Learning in der Datenverarbeitung. Besonders zukunftsweisend sind hierbei aktuelle Bemühungen, z.B. in der Baumartenklassifikation, sehr große Datenbanken in kollaborativer Anstrengung zusammenzutragen.

Contents

Personal Framing 1

List of Publications 3

List of Abbreviations 5

1 Introduction 7
- 1.1 Goals 7
- 1.2 Challenges 7
- 1.3 Objectives 7

2 Fundamentals 9
- 2.1 Airborne Laser Scanning Data 9
  - 2.1.1 System Characteristics 10
  - 2.1.2 Geo-localization of ALS Returns 10
  - 2.1.3 Range Normalization 11
  - 2.1.4 Receiver Effects 12
  - 2.1.5 Waveform Analysis 13
  - 2.1.6 Summary of Data Characteristics 16
- 2.2 Supervised Classification and Semantic Labelling 17
- 2.3 Traditional Machine Learning Methods for Classification 17
  - 2.3.1 Generative Classifiers 17
  - 2.3.2 Discriminative Classifiers 18
  - 2.3.3 Structured Prediction 18
- 2.4 Deep Learning 19
  - 2.4.1 Neural Networks 19
  - 2.4.2 Learning the Weights and Biases 21
  - 2.4.3 Towards Deep Neural Networks 21
  - 2.4.4 3D Deep Learning 22
  - 2.4.5 Characteristics of Deep Learning 23
- 2.5 Evaluation 23
  - 2.5.1 Class-Wise Evaluation Metrics 23
  - 2.5.2 Overall Evaluation Metrics 25

3 Point-Wise Semantic Labelling 27
- 3.1 Introduction 27
  - 3.1.1 Goals 27
  - 3.1.2 Challenges 27
  - 3.1.3 Objectives 28
- 3.2 Related Work 29
  - 3.2.1 Neighbourhood Definition 29
  - 3.2.2 Geometric Feature Types 30
  - 3.2.3 Classification 32
- 3.3 Methodology 32
  - 3.3.1 Neighbourhoods for Feature Extraction 33

3.3.2 Established Features for Comparison . . . 34
3.3.3 Shape Distributions . . . 35
3.3.4 Normalized Height . . . 37
3.3.5 Feature Normalization . . . 37
3.3.6 Classification . . . 38
3.4 Material . . . 39
3.4.1 ISPRS Vaihingen Data Set . . . 39
3.4.2 GML Data Set A . . . 41
3.5 Results . . . 42
3.5.1 Comparison of Shape Distribution and Covariance Features in Binary Classification of Four Classes on Varying Scales . . 42
3.5.2 Analysis of Class Separability by Simple Classifiers . . . 52
3.5.3 Comparison and Combination of Different Feature Types, Neighbourhood Types, and Scales . . . 56
3.6 Discussion . . . 69
3.6.1 Own Work . . . 69
3.6.2 Comparison to Deep Learning . . . 71
3.7 Conclusion . . . 73

4 Object-Wise Classification of Tree Species for Individual Tree Segments . . . 77
4.1 Introduction . . . 77
4.1.1 Goals . . . 77
4.1.2 Challenges . . . 78
4.1.3 Objective . . . 78
4.2 Related Work . . . 79
4.2.1 Single-Tree Segmentation . . . 79
4.2.2 Tree Species Classification . . . 79
4.2.3 Feature Design . . . 80
4.2.4 Large-Scale Structures . . . 81
4.2.5 Medium-Scale Structures . . . 81
4.2.6 Small-Scale Structures . . . 82
4.2.7 Multi-Scale Approaches . . . 84
4.3 Methodology . . . 84
4.3.1 Waveform Processing . . . 84
4.3.2 Established Features for Comparison . . . 85
4.3.3 Spin Image Features . . . 85
4.3.4 Classification . . . 88
4.3.5 Evaluation . . . 89
4.4 Material . . . 89
4.4.1 Study Area . . . 89
4.4.2 Field Data . . . 89
4.4.3 ALS Data . . . 89
4.4.4 Radiometric Correction . . . 90
4.4.5 Tree Segments . . . 91
4.5 Results . . . 92
4.5.1 Choice of Symmetry Axis Compared to Manual Tree Top Measurement . . . 92
4.5.2 Effect of Parameter Choices . . . 94
4.5.3 Geometric Descriptiveness of Spin Images Compared to α-Shape Features . . . 95
4.5.4 Species Classification Improvement . . . 97
4.5.5 Failure Cases . . . 99
4.6 Discussion . . . 100
4.6.1 Own Work . . . 103
4.6.2 Comparison to Deep Learning and High-Resolution Data . . . 105
4.7 Conclusion . . . 109

5 Summary . . . 111

6 Conclusions and Outlook 113

References 115

List of Figures 131

List of Tables 134

Acknowledgments 135

Personal Framing

Having studied physics and completing my diploma in biophysics in 2013, I applied, since I was interested in changing my scientific working area, at the Institute of Photogrammetry and Remote Sensing at the Karlsruhe Institute of Technology (KIT) with Prof. Dr. Stefan Hinz.

I started work there in September 2013, familiarizing myself with the field, and worked on geometric features for airborne laser scanning with publicly available benchmark data, from which I produced my first publication. I also took up a research cooperation with Aarne Hovi and Ilkka Korpela from the University of Helsinki (UoH), Finland, which formed the foundations of my second focus of research during this thesis. In October 2014 I received a doctoral scholarship from the Carl Zeiss foundation and in November/December 2014 I spent a month abroad at the Department of Forest Sciences (UoH).

After intermittent health issues in 2015, I went on to publish two papers in 2016. Following the birth of my first child in June 2016, I went on parental leave and returned to 50 % part-time in January 2017. After two more publications in 2017, my second child was born in June 2018, due to which I went on parental leave until July 2019 and afterwards resumed again to 50 % part-time. During the pandemics I had to leave work for health reasons again, and stayed home giving birth to my third child in November 2020. Supporting a family of five, I was only able to return to 40 % part-time in March 2024.

Due to these personal circumstances, I have been writing up in 2024, while the main scientific working period of this thesis was from 2014 to 2018, as can be seen from the list of supporting publications (page 3 to 4). Upon resuming work in 2024, the field of research had naturally evolved and has undergone quite substantial changes due to the spread of deep learning and evolved sensor technology. The reported work should therefore be seen in light of the scientific circumstances at their time of publication. For both main application scenarios however, a comparison to the current state of the art can be found in the discussion towards the end of each chapter.

List of Publications

This thesis is based on a number of original research articles published during the scientific working period from 2013 to 2018.

Parts of this thesis are taken *verbatim* from these articles. Those quotations are highlighted by corresponding colour bars beside the text, even though minor editorial changes have been done to make both language and the use of specific terms consistent throughout this thesis. If a change altered the sentence, it was marked in square brackets. For the sake of clarity, occlusions are marked by [] instead of [...]. The colour encoding scheme is as follows:

R. Blomley, Ma. Weinmann, J. Leitloff, B. Jutzi (2014) Shape distribution features for point cloud analysis - a geometrical histogram approach on multiple scales. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-3, p. 9-16 (peer-reviewed)

R. Blomley, B. Jutzi, Ma. Weinmann (2016) 3D semantic labeling of ALS point clouds by exploiting multi-scale, multi-type neighborhoods for feature extraction. Proceedings of the International Conference on Geographic Object-Based Image Analysis (GEOBIA), p. 1-8, Editor: N. Kerle

R. Blomley, B. Jutzi, Ma. Weinmann (2016) Classification of airborne laser scanning data using geometric multi-scale features and different neighbourhood types. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. III-3, p. 169-176 (peer-reviewed)

R. Blomley, Ma. Weinmann (2017) Using multi-scale features for the 3D semantic labelling of airborne laser scanning data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. IV-2/W4, p. 43-50 (peer-reviewed)

R. Blomley, A. Hovi, Ma. Weinmann, S. Hinz, I. Korpela, B. Jutzi (2017) Tree species classification using within crown localisation of waveform LiDAR attributes. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 133, p. 142-156 (peer-reviewed)

Ma. Weinmann, **R. Blomley**, Mi. Weinmann, B. Jutzi (2018) Investigations on the potential of binary and multi-class classification for object extraction from airborne laser scanning point clouds. 38. Wissenschaftlich-Technische Jahrestagung der DGPF und PFGK18 Tagung, Publikationen der DGPF, Vol. 27, p. 408-421, Editor: T. P. Kersten

As I do not follow a commercial interest in the publication of this thesis, I retain the right to use and quote the works I published earlier. This is in agreement with the publisher's guidelines.

List of Abbreviations

AGC	automatic/active gain control	90
ALS	airborne laser scanning	78
ALS-HD	very high density ALS	10
CBH	crown base height	81
CHM	canopy height model	78
κ	Cohen's κ coefficient	97
CNN	convolutional neural network	80
CRF	conditional random field	113
DBH	diameter at breast height	77
DR	discrete return	29
DTM	digital terrain model	32
FN	false negatives	24
FP	false positives	24
FW	full waveform	12
FWHM	full width half maximum	90
GML	Graphics and Media Lab	41
GNSS	global navigation satellite system	10
GPS	global positioning system	107
IMU	inertial measuring unit	107
ISPRS	International Society for Photogrammetry and Remote Sensing	39
k**-NN**	k-nearest neighbour	38
LDA	linear discriminant analysis	130
LiDAR	light detection and ranging	77
MCP	mean class precision	133
MCR	mean class recall	133
MLS	mobile laser scanning	29
NEAS	noise-exceeding amplitude sequence	84
NN	nearest neighbour	130
OA	overall accuracy	97
PCA	principal component analysis	86
QDA	quadratic discriminant analysis	33
QSM	quantitative structural model	107
ReLU	rectified linear unit	22
RF	random forest	88
RTK	real-time kinematic positioning	89
RQ	research question	78
SIFT	scale-invariant feature transform	31
SVM	support vector machine	105
SWF	system waveform	90
TLS	terrestrial laser scanning	107
TN	true negatives	24
TP	true positives	24
UAVLS	unmanned aerial vehicle laser scanning	107

Chapter 1
Introduction

Airborne laser scanning (ALS) for landscape survey has unique strengths. Accurate range measurement leads to an isometric, sampled representation of landscape structures, independent of external lighting conditions. As an active sensor technology, it even allows for partial penetration of scattered target structures such as vegetation, so depth and landscape topography can be inferred from the data along with land cover structures.

1.1 Goals

The goal of this thesis is to design hand-crafted features to utilize the geometric strength of ALS data. Discriminative features provide a robust basis for supervised classification tasks and semantic per-point labelling, which may then foster automated scene understanding or robust parameter estimation about the scanned land cover. At the time when the scientific work of this thesis was taken up in 2014, the application of deep learning methods had not yet reached point cloud analysis on a broad level. Today, adaptions of deep learning to the unstructured nature of point clouds (as opposed to the regular grid structure of image data) have become generally available. However, for application scenarios with limited amounts of practically obtainable training data, the use of potent hand-crafted features is still compelling, as the use of expert-knowledge circumvents unnecessary parameter estimation. Therefore, the original goals of this work are still relevant and the results may be used as a strategic template in other applications.

1.2 Challenges

When working with ALS data, it is crucial to understand the strengths and weaknesses of this data type, contingent upon its acquisition process (cf. the 'Fundamentals' in Section 2.1). In order to deduce scientifically sound results, it is also important to follow a clear task, such as clear research questions (RQs) or a clear definition of classes, and to understand possible difficulties inherent in this definition. Furthermore, there has to be a clear reference (ground truth) for a given class definition, which often involves huge manual annotation or mapping efforts.

1.3 Objectives

Therefore, we follow a two-step layout. After an initial outline of the technical fundamentals, the first applicational part of this thesis will focus on a basic, per-point semantic labelling task to enhance the usability of the geometric information, unique to point cloud data. Meanwhile in the second applicational part, we will work with a more complex application scenario, where different qualities of ALS data have to be merged to improve object-wise classification.

In the field of point-wise semantic labelling on ALS point clouds, urban benchmark data sets form a remarkably good basis for comparable research. This data typically offers per-point annotations in accordance to a class definition of urban scene objects like *Roof* or *Building*, *Vegetated Ground*, *Sealed Ground*, *Car*, or *Fence/Hedge*, and depending on the level of annotation detail, maybe *Façade* or *Powerline*.

We therefore formulate the following RQs for this part of our thesis:

- **RQ1:** How can geometric properties be used for point-wise semantic labelling within ALS point clouds? Can we design a novel geometric feature type, which enables advances compared to existing approaches?

- **RQ2:** What is the influence of different neighbourhood types and scales on the descriptiveness of geometric features with respect to different classes?

For our later part on object-wise scene understanding, the in-depth analysis of vegetation offers particular challenges due to its structural variability and partial permeability. Seeking for a clear ground truth, the tree species classification of individual tree segments offers a suitable application case.

Therefore we pose the following RQs for this part our thesis:

- **RQ3:** Is it possible to design a feature type which can be used to improve tree species classification of individual tree segments by capturing the geometric distribution of waveform properties (generated by below-footprint-scale structures) within tree crowns?

- **RQ4:** Given the baseline accuracies in tree species classification by detailed waveform analysis (Hovi et al., 2016), can the accuracy be improved even further by considering the localization of the waveform attributes within the tree crown? If so, how big is the gain?

- **RQ5:** How are the failure cases distributed among tree sizes? Trends are indicative of practical relevance.

Chapter 2

Fundamentals

This chapter aims to summarize the conceptual foundations and fundamental properties of the technologies applied throughout this thesis. The main components here are the data type used, which is specified as airborne laser scanning (ALS) data, and the classification strategies applied for semantic labelling. It is crucial to understand both the characteristics and limitations of these concepts to deduce scientifically founded conclusions. Finally, well-recognised evaluation metrics are introduced here, which enable a differentiated assessment of supervised classification results.

2.1 Airborne Laser Scanning Data

Light detection and ranging (LiDAR) is an active remote sensing technology, designed to measure ranges between a sensor unit and backscatter targets. This is done by emission of a laser pulse or beam and measurement of the time of flight until the reflection from the target is picked up by the LiDAR receiver. Thus the distance to the target can be inferred. There are pulsed or (frequency modulated) continuous wave LiDAR systems, which measure the sensor-target-distance either by time-of-flight or phase shift measurements. Continuous wave systems may also measure target velocities by analysing frequency shifts due to the Doppler effect.[1] Use-cases of LiDAR include static upward-facing setups for atmospheric measurement of aerosols and clouds, or scanning applications like terrestrial laser scanning (TLS), mobile laser scanning (MLS), airborne laser scanning (ALS), or satellite laser scanning, usually employed either for mapping tasks or for autonomous driving and navigation purposes.

In this thesis, we will focus on ALS data only. Application scenarios that utilize the strength of this data type are landscape survey tasks where accurate height information is relevant, such as urban land cover classification (Yan et al., 2015) or ecology-oriented surveying, e.g. for estimation of biomass (Zachary and Wynne, 2005) or carbon storage (Stephens et al., 2007) or for biodiversity monitoring (Fuhr et al., 2022), as well as forest inventories (Latifi et al., 2015) or risk assessment tasks, e.g. input maps for hydrological models in flood management (Vetter et al., 2011).

[1] A taxonomic overview of active remote sensing technologies, organized along the categories of setup, measurement, illumination, modulation, detection, field-of-view, and range, is found in Jutzi (2015).

2.1.1 System Characteristics

In traditional aircraft-mounted scanning applications, there are differences among LiDAR platforms in scanning patterns, such as conic, oscillating or line-wise scanning patterns, all of which depend on the mirror geometry and frequency, the laser pulse rate and the speed of movement of the sensor. The sweeping pattern and sampling rate are therefore crucial components in the representation of the target in the data. There are also differences in the laser wavelength(s) used in different LiDAR platforms. As the reflective properties of target surfaces differ depending on the laser wavelength, different applications use one or several lasers of different wavelength (Morsy et al., 2017). Conventional pulsed LiDAR systems – either discrete return or waveform recording systems – differ in the recording and digitalization of the reflected pulse (cf. Section 2.1.5). Single photon counting systems (Mandelburger and Lehner, 2019; Hong et al., 2024) are currently being developed, which perform a simultaneous measurement of many partial beams, dramatically increasing return density and spatial resolution. Different systems may also have different emitted pulse lengths, pulse energies and beam divergences (which then, in combination with the flying height[2], lead to a difference in so-called footprint size, meaning the size of the area illuminated in the target plane). Some systems even offer different pulse-repetition rates, which lead to different values for the emitted energy per pulse. Traditional aircraft-mounted systems offer point densities of up to 20 pulses per m^2. Recently, with advanced sensor technology, so-called very high density ALS (ALS-HD) data has been recorded from low-flying platforms such as drones or helicopters, with return densities ranging from 500 up to 10000 returns per m^2 (Kellner et al., 2019; Hyyppä et al., 2022).

Those system-specific properties should be kept in mind when working with the produced data, as all of these properties influence the target's representation in the data as well as the applicability of further methods.

2.1.2 Geo-localization of ALS Returns

In ALS and TLS applications, the sensor is usually equipped with a position and orientation system (inertial measuring unit (IMU) and geolocation system (global navigation satellite system (GNSS))), which, given a prior calibration of the setup, allows a coordinate transformation of the registered returns into a georeferenced point cloud with elements $\vec{X}_G$:

$$\vec{X}_G = \vec{X}_0 + \mathbf{R}_{\text{yaw, pitch, roll}} \cdot \vec{P}_G + \mathbf{R}_{\text{yaw, pitch, roll}} \cdot \mathbf{R}_{\Delta\omega,\Delta\phi,\Delta\kappa} \cdot \mathbf{R}_{\alpha,\beta} \cdot \begin{bmatrix} 0 \\ 0 \\ -\rho \end{bmatrix}, \tag{2.1}$$

where $\vec{X}_G$ is a returns coordinates in the ground coordinate system, $\vec{X}_0$ is the vector between the ground coordinate system and the IMU coordinate system, $\vec{P}_G$ is the offset between the LiDAR and IMU coordinate systems and ρ is the range measured by the LiDAR. $\mathbf{R}_{\text{yaw, pitch, roll}}$ is a rotation matrix describing the rotational angles of the IMU relative to the ground coordinate system, $\mathbf{R}_{\Delta\omega,\Delta\phi,\Delta\kappa}$ the boresight rotational angles between the IMU and LiDAR coordinate systems (determined during calibration) and $\mathbf{R}_{\alpha,\beta}$ the rotation matrix describing the mirror scan angles of the laser scanner (Habib et al., 2010). All of these variables, both those measured during operation, as well as those specified during calibration, come with errors, which, by error propagation, influence the positioning accuracy of the measured returns in dependence on acquisition parameters such as sys-

[2] Height is typically referred to as the vertical distance between the point of observation and the Earth's surface, whereas altitude refers to the vertical distance between the point of observation and the mean sea level.

tem parameters, flying height and scan angle (May and Toth, 2007). There are methods to reduce the bias introduced by errors in the calibration parameters through post-calibration methods applied after data acquisition. Those so-called strip adjustment methods use surface elements featured in areas of strip overlay to identify discrepancies in the repeated acquisition of the same target, and then estimate (and correct) the system bias by that (Habib et al., 2010). Further random errors have been simulated to be in an order of magnitude of $0.1 - 0.4\,\mathrm{m}$ in the horizontal and $0.1 - 0.2\,\mathrm{m}$ in the vertical direction, given a flying height of $1.2 - 2.0\,\mathrm{km}$ and a scanning angle of 20° (Habib et al., 2009).

2.1.3 Range Normalization

When aiming to interpret the backscattered intensity in a quantitative way, intensity normalization of the received signal is important, but is faced with practical difficulties.

The received power P_{rec} depends on the sensor-target-range ρ, in a way that can be described by a modified radar equation:

$$P_{\mathrm{rec}} = P_{\mathrm{trans}} \cdot \eta_{\mathrm{sys}}\eta_{\mathrm{atm}} \cdot \frac{D_{\mathrm{r}}^2 r}{\beta^2 \mathit{\Omega}} \cdot \frac{A_{\mathrm{target}}}{\rho^4}, \tag{2.2}$$

where P_{trans} is the transmitted power, D_{r} the receiver aperture size, r the target reflectivity, β the beam divergence, $\mathit{\Omega}$ the bidirectional scattering properties and A_{target} the illuminated area of the target (Wagner et al., 2006). η_{sys} and η_{atm} can be included to denote system and atmospheric attenuation effects (Höfle and Pfeifer, 2007). D_{r}, β, and usually η_{sys} and η_{atm}, can be assumed to be constant during one acquisition. P_{trans}, due to the sensor's electronics, may be subject to some random, target-independent variability (Gatziolis, 2011). The largest influence, apart from the target's properties, is considered to be the sensor-target-range ρ.

Equation 2.2 is sometimes re-written to combine all target parameters, including the illuminated area, in a so-called backscatter cross-section σ (Wagner et al., 2006):

$$P_{\mathrm{rec}} = P_{\mathrm{trans}} \cdot \eta_{\mathrm{sys}}\eta_{\mathrm{atm}} \cdot \frac{D_{\mathrm{r}}^2}{4\pi\rho^4\beta^2} \cdot \frac{4\pi A_{\mathrm{target}} r}{\mathit{\Omega}} \tag{2.3}$$

$$= P_{\mathrm{trans}} \cdot \eta_{\mathrm{sys}}\eta_{\mathrm{atm}} \cdot \frac{D_{\mathrm{r}}^2}{4\pi\rho^4\beta^2} \cdot \sigma. \tag{2.4}$$

This way, it is clear that the backscatter cross-section

$$\sigma = \frac{4\pi}{\mathit{\Omega}} \cdot r \cdot A_{\mathrm{target}} \tag{2.5}$$

combines different backscattering characteristics of the target, namely its illuminated size A_{target}, reflectivity r and the directionality of scattering $\mathit{\Omega}$. As for planar target surfaces of Lambertian reflectance, $\mathit{\Omega}$ can be approximated by the cosine of the angle of incidence α. It has been suggested to use surfaces of in-the-lab measured reflectivity and bidirectional scattering properties for calibration in the field (Kaasalainen et al., 2007). In the case of less well-defined targets, e.g. for targets that overlap or do not fill all of the footprint, the interpretation of the recorded power is difficult due to the mixture of unknown factors.

The footprint area of the beam however also depends on the sensor-target-range. It is approximately

$$A_{\text{laser}} = \pi \cdot \left(\rho \cdot \tan\left(\frac{\beta}{2}\right)\right)^2 , \tag{2.6}$$

and for small angle approximation ($\tan\theta \simeq \theta$)

$$A_{\text{laser}} = \frac{\pi\beta^2}{4} \cdot \rho^2 . \tag{2.7}$$

This means, that the recorded power's range dependence (cf. Equation 2.2) varies, depending on the targets illuminated surface area A_{target}: the recorded power is proportional to ρ^{-2} for homogeneous targets that cover all of the footprint (perpendicular to the beam incidence, $A_{\text{target}} = A_{\text{laser}} \propto \rho^2$),

$$P_{\text{rec, planar target}} \propto \frac{P_{\text{trans}}}{\rho^2} , \tag{2.8}$$

proportional to ρ^{-3} for linear targets ($A_{\text{target}} \propto \rho$) like cables

$$P_{\text{rec, linear target}} \propto \frac{P_{\text{trans}}}{\rho^3} , \tag{2.9}$$

or proportional to ρ^{-4} for point-like targets ($A_{\text{target}} \not\propto \rho$)

$$P_{\text{rec, point target}} \propto \frac{P_{\text{trans}}}{\rho^4} . \tag{2.10}$$

In cases of range variation within one data set, in particular for sloping terrain or different flying heights, but also for slant ranging and overlapping strips, it is of practical interest to normalize the recorded intensities[3] range dependence. This is usually written out as

$$I_{\text{range norm.}} = \left(\frac{\rho}{\rho_{\text{ref}}}\right)^a \cdot I_{\text{raw}} , \tag{2.11}$$

where ρ_{ref} is either a manually defined reference range or a reference range set by a certain moment, usually the mean, of all range recordings. a is the range normalization exponent. As shown in Equations 2.8 to 2.10, a also depends on the target geometry; it ranges in theory from 2 for planar targets, over 3 for linear targets and 4 for point-like targets.

2.1.4 Receiver Effects

However, apart from the constant factor η_{sys}, the effects of the receiver and readout circuits have to be considered. Those vary for different manufacturers and sensors. While most receivers aim to produce a signal output directly proportional to the incoming optical power (linear dependence), technical limitations may result in non-linearities, especially at the high or low end. Those could lead to apparent deviations from $a = 2$ even for planar surface targets, depending on the sensor's emitted power and pulse repetition settings, as well as the flying height. Other (especially full waveform (FW)) systems use an (optional) automatic/active gain control (AGC) circuit to enhance the recording of low-reflectance targets. This changes the recorded intensity values, so that AGC effects, quantified by an AGC value for each recording, have to be corrected for when aiming for a quantitative

[3] The difference between laser power P and intensity I is that the laser power refers to the total optical energy over time [$^{\text{J}}/_{\text{s}}$], while an optical intensity is the power within a certain transverse area within the non-uniform beam profile [$^{\text{J}}/_{\text{sm}^2}$]. The non-uniform irradiance profile of laser beams, however, is typically not considered in the mathematical representation.

analysis (Vain et al., 2010; Korpela et al., 2010a). Correction models typically follow a shape similar to

$$I_{\text{range \& gain norm.}} = I_{\text{raw}} \cdot \left(\frac{\rho}{\rho_{\text{ref}}}\right)^{a} + b \cdot I_{\text{raw}} \cdot (c - \text{AGC}) \ , \tag{2.12}$$

where the values of b and c are determined empirically if AGC is employed (Korpela et al., 2010a).

A number of studies on flat targets could achieve accurate radiometric calibration assuming the theoretical value of $a = 2$ (Ahokas et al., 2006; Kaasalainen et al., 2009). Korpela (2008) used naturally occurring homogeneous targets such as gravel road, asphalt surfaces, a football field and a barley field as calibration surfaces to determine the correction model parameters, and found values of $a = 2.4$ for an ALS50 and $a = 2.5$ for an ALTM3100 sensor. Reitberger et al. (2009) suggested using overlapping LiDAR acquisitions of different flying height to estimate a (in a sensor without AGC). Their results yielded values slightly smaller than two (1.9 and 1.7 for two different data sets acquired by Riegl LMS-Q560 sensor), which they suggested could be explained by non-linearities in the sensor. Korpela et al. (2010a) studied intensity variation among the reflections attributed to crowns of different tree species (and thereby structural target properties) and found that the optimum a values minimizing the variation were higher for conifers than hardwood species, and lower for single returns compared to first or other returns. Gatziolis (2011) aimed to determine optimal a values by minimalizing the difference among normalized intensity in pairs of data, where the same target area was sampled by different flight strips. In grass fields, their computed optimum a value was very close to 2, while in complex vegetation, the optimum a values were 2.04 for single returns, 2.34 for first-of-two returns and much higher (> 2.6) for the first-of-three-or-four returns. Those findings show well how complex target structures (and thus the probability for later returns) influence the accuracy of range normalization and intensity values.

Range-dependent intensity normalization is therefore a difficult task due to different contributing factors. It is ambiguous if the target geometry is unclear or if the target only covers an unknown fraction of the footprint, but it may also be mixed with non-linearity effects of the sensor.

2.1.5 Waveform Analysis

Roughly speaking, for the case of well-defined target surfaces large enough to fill the LiDAR footprint, the normalized signal's characteristics correspond to the sensor-target-range ρ (elapsed time-of-flight), the reflective properties of the target surface (recorded amplitude $P_{\text{rec, max}}$), and the elevation variation along the direction of laser incidence (recorded pulse width or shape) (Jutzi and Gross, 2009b).

The temporal shape of the received signal is typically referred to as a (recorded) waveform. In simple cases like a perpendicular target surface filling the whole laser footprint, the recorded waveform is simply an attenuated replica of the emitted waveform (Wagner et al., 2006), delayed by a factor of $^{2\rho}/_{v_g}$, where v_g corresponds to the group velocity of the travelling laser pulse [4]:

[4] Due to the Heisenberg uncertainty principle, the spectral bandwidth of a laser pulse has to increase to allow for a short pulse in the time domain. Therefore, the different wavelengths of the laser pulse travel at different phase velocities in a dispersive medium. The group velocity is the speed by which the envelope of a wave, consisting of a superposition of different frequencies, propagates through space.

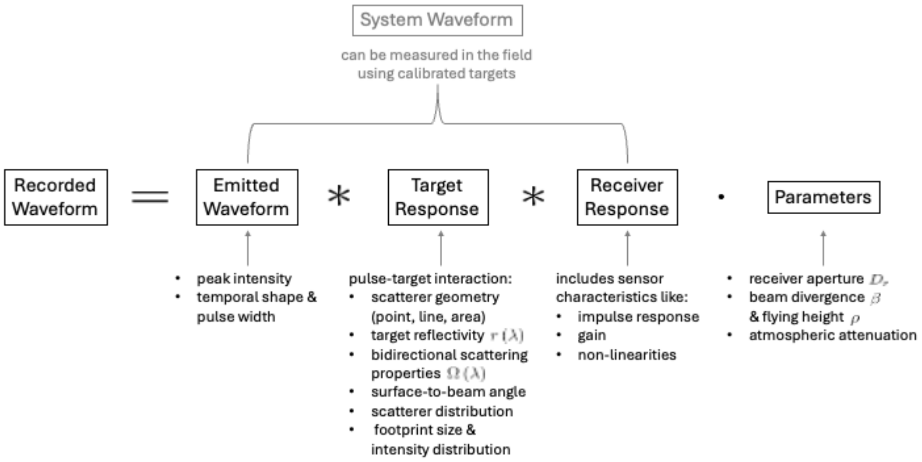

System Waveform
can be measured in the field using calibrated targets
Recorded Waveform
=
Emitted Waveform
*
Target Response
*
Receiver Response
·
Parameters
• peak intensity
• temporal shape & pulse width
pulse-target interaction:
• scatterer geometry (point, line, area)
• target reflectivity $r(\lambda)$
• bidirectional scattering properties $\Omega(\lambda)$
• surface-to-beam angle
• scatterer distribution
• footprint size & intensity distribution
includes sensor characteristics like:
• impulse response
• gain
• non-linearities
• receiver aperture D_r
• beam divergence β & flying height ρ
• atmospheric attenuation

Fig. 2.1: Schematic representation of influences on the recorded waveform. $*$ denotes the convolution operator, and $\cdot$ the multiplication operator. Physically, the emitted waveform is first convoluted with the target response function and then with the receiver response function, but since the convolution operation is commutative, it is more practical to combine emitted waveform and receiver response into one measurable unit, called the system waveform (SWF). The list of influences is non-exhaustive. Graphical illustration in analogy to Hovi (2015).

$$P_{\text{rec}}\left(t\right) = P_{\text{trans}}\left(t - \frac{2\rho}{v_g}\right) \cdot \frac{\eta_{\text{sys}}\eta_{\text{atm}}D_{\text{r}}^2}{4\pi\rho^4\beta^2} \cdot \sigma \ , \tag{2.13}$$

analogously to Equation 2.4.

Temporal distortion of the waveform occurs whenever there is an along-beam variation of the target, which involves cases with tilted surfaces or non-nadir incidence angles, as well as cases with distributed scatterers such as vegetation or sudden edges within the footprint. A general model is to characterize the sensor by a so-called system waveform (SWF), which is a convolution of the emitted waveform and the receiver response function, as shown in Figure 2.1. This is then, in a slightly simplified mathematical representation, convoluted with a differential target backscatter cross-section $\sigma_i\left(\rho\right)$ per range interval $d\rho$ to describe the target-pulse interaction (Wagner et al., 2006):

$$P_{\text{rec, i}}\left(t\right) = \frac{\eta_{\text{sys}}\eta_{\text{atm}}D_{\text{r}}^2}{4\pi\beta^2} \cdot \int_{\rho_i - \Delta\rho}^{\rho_i + \Delta\rho} \frac{1}{\rho^4} \cdot P_{\text{trans}}\left(t - \frac{2\rho}{v_g}\right) \cdot \sigma_i\left(\rho\right) d\rho \tag{2.14}$$

When the spatial extent $2\Delta\rho$ of the distributed cluster of scattering elements is very small compared to the sensor-target-range ρ, that is when $\Delta\rho << \rho$, the dependency on $^1/_{\rho^4}$ can be approximated as constant during the integration. This is practically always the case for manned aircraft ALS, as the flying height is much larger than both the target size and the resolvable distance between distinct target surfaces (Wagner et al., 2006). Equation 2.14 can therefore be approximated by

$$P_{\text{rec, i}}\left(t\right) = \frac{\eta_{\text{sys}}\eta_{\text{atm}}D_{\text{r}}^2}{4\pi\rho^4\beta^2} \cdot P_{\text{trans}}\left(t\right) * \sigma_i'\left(t\right) \ , \tag{2.15}$$

where $*$ is the convolution operator and $\sigma_i'\left(t\right) = \sigma_i\left(\rho\right)$ is the apparent cross-section. Care has to be taken here because the early scatterers along the beam direction may shade potential scatterers later in the signal, so that the second and later pulses along one ray are only generated by those fractions of the target that have not been shaded by previous scatterers. Hence the apparent cross-section is not necessarily a representation of target matter along the beam direction.

Between distinct, non-overlapping target surfaces (acting like Dirac delta functions as target response functions), the range resolution depends on the SWF. Since the refractive index n of air is usually close to 1, the group velocity of the laser pulse is close to the speed of light ($c \approx 3 \cdot 10^8 \mathrm{m/s}$). Due to the fact that the signal has to travel from sensor to target and back, a time-resolution Δt corresponds to a distance $\Delta\rho$ along the ray of approximately

$$\Delta\rho = \frac{c \cdot \Delta t}{2n} \ , \tag{2.16}$$

e.g. 1 ns in the signal corresponds roughly to 15 cm in distance. The minimum distance, in which two SWFs may be resolved, depends on the SWF's pulse width, the rise time and the receiver time resolution. At best, peaks with a minimum distance of half the waveform pulse width may be resolved, but in other cases (e.g. for closely spaced targets with different reflected intensity) the required distance may be larger. Typical SWF pulse widths range from 1 ns for high-resolution ALS systems to 10 ns systems.

Typical LiDAR sensors used in ALS come with either discrete return (DR) or FW recording hardware and data flow. DR sensors can typically record only a few returns per emitted pulse (usually about four to five), sometimes with vertical spacings of several meters due to sensor recovery downtime. However, they usually provide an extremely high horizontal pulse density, which makes them ideal for two-dimensional mapping tasks where vertical profiling is less significant. FW systems, on the other hand, aim to record the full reflection profile with sampling rates of about 1 ns, which has been thought to give a more accurate representation of multi-layered targets such as vegetation. However, this data type requires a more complex, typically custom-made data handling, interpretation and analysis than DR data, which can be delivered in a point cloud format (Ussyshkin and Theriault, 2011).

Concerning the interpretation of the recorded waveform, the main approaches are either to fit a superposition of Gaussian distributions to the shape of the waveform or to split the waveform into return sequences and then to describe the shape of these return sequences by a number of geometric shape descriptors. In the Gaussian decomposition approach, the assumption is that the SWF is usually well described by a Gaussian distribution, and that the apparent cross-section σ_i can be represented by a series of Gaussian functions (Wagner et al., 2006; Li, 2008). As the convolution of one Gaussian with mean μ_1 and variance τ_1 with another Gaussian of mean μ_2 and variance τ_2 results in another Gaussian with mean $\mu_1 + \mu_2$ and variance $\tau_1 + \tau_2$, a fit of Gaussian components to the received waveform and knowledge of the SWF allows one to deduce the parameters of a target response function that is represented by the sum of distinct scatterers, each scatterer influencing the amplitude and temporal stretching of their signal component. This way, a waveform analysis may result in a larger number of detected returns than a comparable DR result. If, however, the apparent backscatter cross section (and hence the target response function) is more than a sum of a finite number of individual scatterers, this approach may reach its limits. This may especially be the case for distributed targets such as vegetation. It has then proven practical to divide the waveform into noise-exceeding amplitude sequence (NEAS), each treated as one return. Those could then be described by a series of geometric attributes, such as peak amplitude A (and the corresponding location), energy E (integral over the waveform shape), full-width-at-half-maximum $FWHM$, length L and 50% energy quantile $EQ50$ (Hovi et al., 2016). Another concept to overcome the weaknesses of Gaussian decomposition in vegetation analysis has been to use skew normal distributions for waveform decomposition in order to allow for the modelling of skewed echoes (Bruggisser et al., 2017). As the same attributes used for the characterization of waveform sequences can be used to describe the characteristics of DR sequences, these representations allow for some comparability among the recoding types. In the remainder of this work, the term return may therefore either refer to a DR return, a decomposition component, or a NEAS, depending on the context of the data used.

Based on individual returns from clear targets, Jutzi and Gross (2009b) developed a method to estimate surface angles within the data and correct the measured intensity of the individual returns by this knowledge. The results showed a much more homogeneous representation of intensity values based on the reflectivity of the target materials in scenes with planar surfaces, such as roof and ground. With complex-structured targets such as vegetation, the scan angle has been found to induce errors in the prediction of vegetation-specific attributes due to the different sampling direction of the vegetation structure (Ahokas et al., 2005; Morsdorf et al., 2008; Liu et al., 2018; Dayal et al., 2022), especially for large scan angles above $15 - 20^{\circ}$.

The overlap-problem, meaning that target components which are shaded by earlier target components along one ray can not be recorded, has given rise to the practical distinction between different return types for individual returns. Returns may either be characterized as *'only returns'* (whenever there is only one return from one pulse, we can assume that the reflective target covers all of the footprint), *'first of many returns'* (in first returns, the reflected power is not compromised by shading, but the fact that there are further returns recorded from the same pulse indicates that the first target component covers only a fraction of unknown proportion of the footprint) or *'other returns'* (in later returns from one pulse, both the fraction of the footprint and shading effects influence the waveform in unresolvable ways) (Holmgren and Persson, 2004; Ørka et al., 2009; Hovi et al., 2016).

2.1.6 Summary of Data Characteristics

In summary, data collected by ALS has unique strengths in terms of absolute height information. However, system- and acquisition-specific parameters such as sampling pattern and pulse density, flying height and footprint size, pulse length, scanning angle, sensor downtime and data acquisition mode, together with atmospheric conditions, all influence the target's representation in the data, and should therefore be considered when inferring conclusions from such data. Importantly, it is not possible to perform an accurate normalization of the intensity's rage dependence for targets of unknown geometrical structure. Furthermore, occlusion effects prohibit an unambiguous representation of target mass distribution and pose additional non-resolvable ambiguities in intensity normalization, when neither the fraction of the footprint covered nor the reflectivity of the target material are known.

However, when aiming for a quantitative signal analysis, the following calibration and normalization techniques should be applied as well as possible: intensity-range-normalization, AGC correction, atmospheric attenuation, and, depending on the processing possibilities, pulse energy or SWF calibration (especially in systems with varying pulse-repetition frequencies) as well as incidence angle corrections (if the scene geometry is suitable). Strip-adjustment strategies furthermore reduce geometric inaccuracies.

2.2 Supervised Classification and Semantic Labelling

As for the strategies used to interpret scenes sampled by ALS, the concepts of supervised classification and semantic labelling are essential. Both are specific tasks within the field of machine learning. A supervised classification task is defined as the assignment of instances like data points or objects to pre-defined categories (classes)[5]. Semantic labelling, more specifically, refers to the assignment of basic data elements such as image pixels to an object class by using classification algorithms and basic per-element features. The semantic labels assigned to the basic data elements therefore form a 'map' which can later aid more complex scene understanding by modelling relationships between the data elements and classes.

In any supervised classification scenario, the data collected has to fulfil certain requirements for successful training of a model as well as for reliable, independent evaluation. Scientific inferences can only be as sound as both the quality and quantity of the data allow. The data used in the classification therefore has to be annotated by manual labelling to provide the so-called ground truth. Care has to be taken, whether this data forms a representative set, or whether it is biased by representing only a certain sub-set. Furthermore, the data has to be divided into independent training and testing data fractions, so that models, developed on the training data, may be tested on independent testing data. Depending both on the complexity of the applied classification model and the complexity of the given class definition, the amount of training data required to train a reliable model varies.

2.3 Traditional Machine Learning Methods for Classification

Supervised classification approaches follow different strategies, such as generative or discriminative approaches. In the following, we will briefly describe the very most popular algorithms, that are either readily implemented in common software or can be found in publicly available code sources.

2.3.1 Generative Classifiers

Generative models describe patterns or distributions among their training data in a way which would allow creating new data of similar characteristics. Among them are probabilistic models, which aim to describe the distribution of the training data via joint probability density functions in a d-dimensional feature space (d being the number of features).

Both linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are examples of probabilistic classifiers. They aim to model the underlying probability distributions of the training data and to infer class-wise labelling probabilities for every novel feature vector. [In LDA], a multivariate Gaussian distribution is fitted to the [] training data, i.e. the parameters of a Gaussian distribution are estimated for each class by parameter fitting. Thereby, the same covariance matrix is assumed for each class and only the means may vary. [In QDA,] not only the means but also the covariance matrices may vary for different classes.

[5] Whereas unsupervised classification is more closely related to clustering. The aim in clustering is to group a set of unlabelled data into clusters based on similarities or patterns in the data, and hence to discover the inherent structure of the data without prior knowledge of the classes. Clusters may not require a semantic meaning, whereas unsupervised classification aims to find interpretable categories.

2.3.2 Discriminative Classifiers

Discriminative models do not assume a joint probability distribution but instead use conditional probabilities. They model the decision boundaries between classes rather than the underlying data distribution, making them less constrained by assumptions about the data's representation or distribution in feature space. This makes them especially well-suited for tasks with a complex distribution of classes or a high number of features.

k-nearest neighbour (k-NN) classification is a well-known example of instance-based classifiers. Given a certain distance metric, a class label is assigned to a testing instance by a majority vote among the k most similar training instances in the feature space. In general, higher values of k reduce the noise on the classification result, whereas too large values of k tend to make the class distinctions less clear. While the simplest and most popular version, known as nearest neighbour classification, sets $k = 1$, there are also many adaptions and specialised versions among this type. Typically, such classifiers depend heavily on the choice of distance metric and are sensitive to noisy or irrelevant features.

Support vector machines (SVMs) follow a max-margin learning approach. Initially designed as binary classifiers, they aim to separate two classes by a linear decision boundary so as to maximize the margin between them. Using the so-called Kernel Trick, they can map data that is not linearly separable into higher-dimensional feature spaces where linear separation becomes possible. Multinomial classification may be achieved by either One-against-One or One-against-All adaptions, both of which combine the results of multiple binary SVMs.

The random forest (RF) classifier is particularly successful by applying the principle of ensemble learning. It strategically combine[s] a set of weak [decision tree] learners to form a single strong learner. [... T]he combination [ ] is realized in a rather intuitive way via bagging (Breiman, 1996), which focuses on training a weak learner of the same type for different subsets of the training data which are randomly drawn with replacement. Accordingly, the weak learners are all randomly different from each other and, hence, taking the majority vote across the hypotheses of all weak learners results in a generalized and robust hypothesis of a single strong learner (Breiman, 2001). It incorporates dual randomness in both the training samples and the selection of features used in constructing each decision tree, which makes it deal well with a large number of features and renders it particularly robust against overfitting.

Classic deep learning, such as multi-layer perceptrons, is also part of discriminative learning. However, because modern deep learning does not follow the traditional feature-based framework, we will describe this in more detail in Section 2.4).

2.3.3 Structured Prediction

Structured prediction refers to a subset of supervised machine learning techniques that aim to predict structured outputs rather than single, independent prediction values. This means that context – such as similarities among neighbouring points in spatial data – is integrated into the classification procedure. By incorporating contextual information, structured prediction methods typically produce less noisy results compared to non-contextual supervised classification.

A common way to model relationships among data elements in structured prediction is through Bayesian networks or random fields. Conditional random fields (CRFs) (Lafferty et al., 2001; Niemeyer et al., 2011a; Weinmann et al., 2015), for example, start with an initial labelling hypothesis, often generated by a SVM or RF classifier in the form of class probabilities. They then infer an association potential from it and propose an improved labelling by enforcing spatial regularity among neighbouring points based on the inferred association potential. However, contextual information may also be considered in other classification approaches, such as deep learning, where it may be embedded through the use of convolutional or recurrent networks.

2.4 Deep Learning

With the availability of very large, annotated data sets and the rapid evolution of powerful GPUs, deep learning has become a powerful and prominent subfield of machine learning. Deep neural networks with many layers are constructed and trained to model complex patterns. Training on massive data sets enables non-manual 'features' or data patterns to be learned automatically from the input data. Lately, different frameworks have become more attainable for non-expert and individual researchers with limited computational resources.

During the active research period of this thesis, deep learning research focused mainly on segmentation and classification tasks in 2D image data. Recently however, deep learning methods have more frequently been adapted to 3D data (Ioannidou et al., 2017), and adaptions for end-to-end application on point cloud data, like PointNet, PointNet++ and other extensions have become available (Qi et al., 2017b,a; Wang et al., 2018; Jiang et al., 2018; Winiwarter et al., 2019; Mao et al., 2022).

However, deep learning is fundamentally different from the traditional feature-based classification procedure. This section therefore aims to give a very rough overview of the working and training principles of neural networks (Nielsen, 2015), in order to be able to draw a comparison among the results of our work by traditional supervised classification methods and recent deep learning enabled advances in the field.

2.4.1 Neural Networks

The basic building block of computational neural networks is a so-called artificial neuron or perceptron (McCulloch and Pitts, 1943), a mathematical model designed to produce a binary decision output from input data x_i, based on a set of weights w_i and a bias b.

$$\text{output} = \begin{cases} 0 & \text{if} \quad \Sigma_j x_j w_j \geq b \\ 1 & \text{if} \quad \Sigma_j x_j w_j < b \end{cases} \tag{2.17}$$

A neuron may be visualized as shown in Figure 2.2.

Neural networks are then built as an architecture of several layers of neurons, where the neurons of one layer take the output of the previous layer neurons as an input. Such networks may be designed as various network architectures and even include loops. A very basic example of network architecture is shown in Figure 2.3.

Depending on the respective weights and biases, such networks can make highly complex decisions. As long as the decisions of each perceptron remain binary however, the change of one value can cause sudden and unforeseen changes in the output. To enable networks to learn by optimising their weights and biases based on training examples, an activation function has to be introduced. The output of

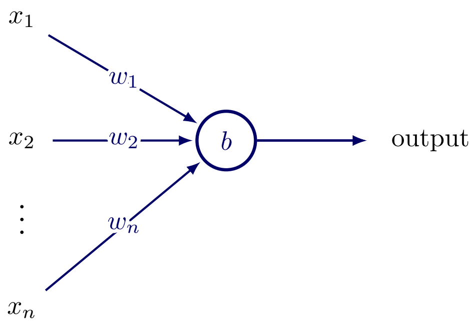


Fig. 2.2: Visualization of the mathematical concept of an artificial neuron.

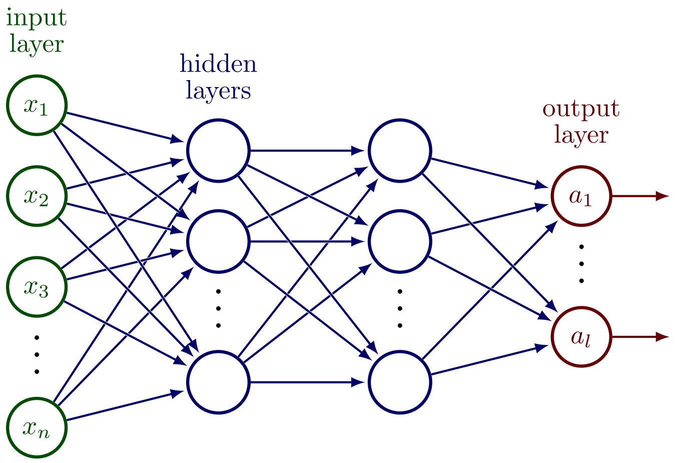


Fig. 2.3: A multi-layer perceptron, serving as an example of a relatively simple neuronal network architecture.

a non-binary neuron is therefore no longer binary, but the original step function is replaced by a smoothed version. A typical choice for example is the sigmoid neuron, which follows a sigmoid activation function:

$$\text{output} = \frac{1}{1 + \exp\left(-\Sigma_j x_j w_j - b\right)} \tag{2.18}$$

Therefore, the neuron's output is identical to the binary case in the far ranges of the input value, but in the transition area, an incremental change in the weight or bias produces an incremental change in the output. This then enables a tuning of the weight and bias parameters towards global optimization.

The training of neural networks does not necessarily involve features calculated manually from the data. A neural network rather forms a complex, yet direct connection between the input and output layer. The input layer may either contain raw data, or, especially in cases where application specific knowledge can be utilized, consist of hand-crafted features. Additionally, intermediate layers or the output of a neural network may also be used as features for subsequent classification tasks, which is commonly referred to as transfer learning.

2.4.2 Learning the Weights and Biases

During the training phase, the weights and biases of a chosen network architecture can be learned by minimising a cost function such as the quadratic cost function, also referred to as the mean squared error,

$$C\left(w,b\right)=\frac{1}{2n}\sum_{x}\left\|a\left(x\right)-y\right\|^{2} \tag{2.19}$$

or the cross-entropy cost function

$$C\left(w,b\right)=-\frac{1}{n}\sum_{x}\left[y\ln\left(a\left(x\right)\right)+\left(1-y\right)\ln\left(1-a\left(x\right)\right)\right]. \tag{2.20}$$

A cost function quantifies the difference between the training output $a\left(x\right)$ based on the input values x and the true output y defined in the training data. It is, by definition, non-negative, and close to 0 when $a\left(x\right)$ is close to y for the given values of w and b. The minimum search can therefore be performed as a gradient descent search, where a value of η, called the learning rate, has to be set to specify the step width of the decent. In order to speed up the process, the gradient search is also not exhaustively performed on all training samples at once, but stochastically. The weight and bias values are updated incrementally by performing the gradient search on randomly sampled subsets of the training data, called batches. During each epoch, which describes a full cycle through the training data by an iteration of batches, incremental changes to the weight and bias values of the network are calculated by the means of backpropagation (Rumelhart et al., 1986). When training a neural network, the number of epochs and their batch size of training examples have to be specified. Those values, as well as the learning rate η, are referred to as hyperparameters, and have to be optimised in order to get good training results.

2.4.3 Towards Deep Neural Networks

Over time, network architecture has become more specialised and increasing computational power has enabled deeper (many-layered) networks. At the same time, it remained crucial to work against overfitting, since larger models offer many parameters. Several regularization techniques can be employed during training to favour more generalized models. Popular options include adding regularization terms to the cost function (Hanson and Pratt, 1988), early training stop by monitoring of the validation error (Morgan and Bourlard, 1990), dropout strategies for introducing more randomness during training (Srivastava et al., 2014) and batch normalization (Ioffe and Szegedy, 2015).

Convolutional neural networks (CNNs) (LeCun et al., 1998; Fukushima, 1980) marked an important development. They use alternating blocks of convolutional and pooling layers. A convolutional layer applies filters over local receptive fields, meaning that only a small, locally connected region of the input is linked to a neuron in the following layer. This enables spatial feature extraction. Successive convolutional and pooling layers form a hierarchy, which allows the network to learn features at increasingly contextual scales, such as simple edges and textures at lower levels and more complex patters or object categories at higher layers. The weights of a filter stay identical regardless of where the filter is applied within one convolutional layer. This weight-sharing mechanism reduces the number of free parameters compared to fully connected layers and contributes to the network's tolerance to small translations of features or objects. Large translations or rotations may however not be recognized by this architecture without additional mechanisms. Pooling layers further reduce the dimensionality by condensing the output of a region of neurons into one output, such as the maximum value or the average value, for example. While this reduction in spacial resolution improves computational efficiency, some spatial information may be lost in the process. Rectified linear unit (ReLU) activation functions improve the efficient propagation of the gradient and therefore enhance training efficiency. CNNs have proven highly effective in image analysis, particularly when trained on large amounts of data with significant computational resources. As the lower levels of a CNN learn universal 'features' applicable to a general task (such as image analysis), transfer learning allows the use of pre-trained nets and re-train only a couple of layers in the end to adapt them to a different specific task, which will then require less data and computational power than training the original net. Moreover, pre-trained CNNs can also be used to extract the output of a late layer and use this as features for traditional machine learning.

However, many-layered deep CNNs often face the vanishing gradient problem, meaning that the partial derivative for optimizing the loss function over many layers gets so small, that early layers can not be updated by training. Skip connections address this issue by allowing the output of one layer to be directly passed to non-adjacent layers. This bypass creates alternate paths for gradient flow, mitigating the vanishing gradient problem, and enabling effective training of deep networks. Visually, their effect can be viewed as smoothing the funnel-shaped gradient landscape (Li et al., 2018). Short skip connections, as implemented in ResNet (He et al., 2016) for example, work by adding up the gradients and learning a residual function rather than a direct mapping. This enables the training of hundreds or thousands of layers, leading to improved accuracy in image recognition. In encoder-decoder architectures (also called auto-encoders), long skip connections concatenate the output of encoder layers with decoder layers of matching dimension. This makes small-scale features with local context accessible at a global scale, and thus helps reduce artefacts at object boundaries in tasks like semantic segmentation. In natural language processing, skip connections are used in transformer architectures, which facilitate modelling of long-range dependencies by maintaining connections between distant positions in a sequence.

2.4.4 3D Deep Learning

The challenge in deep learning on 3D data is the unstructured nature of point cloud data, which is typical for LiDAR or radar scanning, as opposed to the regular grid structure of 2D images. Starting in 2015, attempts have been made to apply deep learning to 3D data by voxel regularization of the data (Maturana and Scherer, 2015), or by multi-view projections onto 2D images followed by 2D semantic segmentation via a CNN and backprojection of the results onto the 3D point cloud (Su et al., 2015). Both types of adaptions faced some drawbacks, like quantization artefacts and an increase in data volume due to sparse voxels, or occlusions and difficulties with complex scenes in the multi-view approach. Both these approaches are therefore limited due to computational complexity and a possible loss in structural resolution. A more detailed review of published approaches in these directions may be found in the work of Xie et al. (2020).

In search of directly applying deep learning on 3D point cloud data, Qi et al. (2017b) proposed a pioneering network architecture named PointNet, which directly accepts point clouds as input and provides either a class label for the complete point cloud or class labels on a per-point basis as output. The key features of this architecture are that it guarantees permutation invariance among the points, as well as invariance to global rotation or translation. PointNet does not include a convolution operator. PointNet++ (Qi et al., 2017a) then added a hierarchical framework architecture, enabling it to learn local features on increasing contextual scales. Moreover, set learning layers were included to combine features from multiple scales, thus making PointNet++ robust with regard to varying point density in the data.

Numerous adaptions have been made since to overcome the limitations of PointNet/PointNet++, such as, e.g. the lack of local features based on the distribution of neighbouring points (pointwise pyramid pooling (Ye et al., 2018), annular convolution (Komarichev et al., 2019), PointSIFT (Jiang et al., 2018) or adaptive feature adjustment (Zhao et al., 2019)) or to provide an alternative way of handling unstructured point cloud data in deep learning. A detailed overview of strategies may again be found in the works of Xie et al. (2020) or Guo et al. (2021).

2.4.5 Characteristics of Deep Learning

Summarizing the advances sketched above, it can be seen how the use of recent neural networks differs from the traditional approach of using hand-crafted, knowledge-based features combined with classifiers that model an underlying distribution (cf. Section 2.3.1), search for similarity among the features or apply discriminative learning (cf. Section 2.3.2). In deep learning, instead, the 'art' is about designing the architecture of a network in a meaningful way. Once a network is set up, the solution is left to an automated optimization process. Neural networks may be set up to generate features and solve specific classification tasks, but the successive steps of traditional machine learning, such as segmentation, feature development and classification, are not necessarily separated in deep learning. This makes it more difficult to interpret the learning process and also makes it more difficult to assess confidence-levels for the produced output. Also, the automated learning process may require huge amounts of training data to optimize a model with many parameters, rendering it inapplicable to some fields, where these large amounts of data cannot be provided. The challenge of possible overfitting is therefore universal to deep learning, so care has to be taken to include strategies that favour more generalized models. Additional workarounds include using pre-trained nets as well as regularization strategies and data augmentation (like rotation, mirroring or coordinate jittering) to synthetically extend the diversity of the training data. Wherever these limitations can be circumvented or accepted, deep learning typically delivers highly accurate results, proves efficient at handling large data sets and skips the difficult task of hand-crafted, application-specific feature design and selection.

2.5 Evaluation

Typically, classification results are evaluated by slightly different yet overlapping evaluation metrics. Depending on the field of research, different names may refer to the same metric (e.g. precision – user's accuracy – correctness), while other metrics (e.g. F_1-score and quality) describe a similar concept but are calculated differently. As a general reference for evaluation of our experiments, we give an overview of the metrics employed throughout this thesis.

2.5.1 Class-Wise Evaluation Metrics

In machine learning, the measures of precision, recall and F_1-score are widely established. They are defined, in a binary case, as

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2.21}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2.22}$$

where classification results are pooled as true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN).

For a confusion matrix $\mathbf{C} = [c_{ij}]$, where $\sum_i c_{ij}$ is the actual number of entities in class j and $\sum_j c_{ij}$ is the number of entities predicted as belonging to class i, the calculation is as follows:

$$n = \sum_{ij} c_{ij}, \tag{2.23}$$

$$n_a = \sum_i c_{ii}, \tag{2.24}$$

$$n_\epsilon = \frac{1}{n} \cdot \sum_i c_{ij} \cdot \sum_j c_{ij}, \tag{2.25}$$

$$\text{precision}_i = \frac{c_{ii}}{\sum_i c_{ij}}, \tag{2.26}$$

$$\text{recall}_i = \frac{c_{ii}}{\sum_j c_{ij}}. \tag{2.27}$$

In the context of remote sensing, precision is sometimes also referred to as user's accuracy and recall as producer's accuracy. Precision describes what proportion of positive identifications was correct, while recall describes what proportion of actual positives was identified correctly, which is also known as the true positive rate.

The F_1-score is the harmonic mean of precision and recall, and thus a combined measure for the predictive performance in one class:

$$F_{1,i} = \frac{2 \cdot \text{precision}_i \cdot \text{recall}_i}{\text{precision}_i + \text{recall}_i} , \tag{2.28}$$

or, in the binary case,

$$F_1 = \frac{2\,\text{TP}}{2\,\text{TP} + \text{FN} + \text{FP}} . \tag{2.29}$$

Furthermore, the evaluation metrics of completeness, correctness and quality have been described by Rutzinger et al. (2009). Completeness is the same values otherwise described as recall, while correctness is the same as precision. Quality is a combined measure calculated as:

$$\text{quality}_i = \frac{1}{{}^1\!/\text{comp.}_i + {}^1\!/\text{corr.}_i - 1} , \tag{2.30}$$

or equivalently, in the binary case, as

$$\text{quality} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} . \tag{2.31}$$

Quality is hence conceptually similar, yet not identical, to the F_1-score.

2.5.2 Overall Evaluation Metrics

Among metrics describing the performance of a classification result over all classes, an intuitive way is to calculate the mean class recall (MCR), which describes how well, on average, instances have been found belonging to the correct class. This does not account for how reliable instances have been found.

$$\text{MCR } = \frac{1}{C} \cdot \sum_{i=1}^{C} \text{recall}_{\,i}. \tag{2.32}$$

Analogously, the mean class precision (MCP) is a measure of how accurate the class assignment is on average.

$$\text{MCP } = \frac{1}{C} \cdot \sum_{i=1}^{C} \text{precision}_{\,i}. \tag{2.33}$$

Furthermore, overall accuracy (OA) can be calculated as the number of correctly labelled instances n_a (as in Equation 2.24) divided by the total number of instances.

$$\text{OA} = \frac{n_a}{n} \ . \tag{2.34}$$

In the binary case, this is

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \ . \tag{2.35}$$

The Cohen's κ coefficient (κ) however, is a more balanced measure of overall class separation, which excludes the rate of correct labelling by chance. It is defined as

$$\kappa = \frac{n_a - n_\epsilon}{n - n_\epsilon}, \tag{2.36}$$

where n is the number of instances, n_a the number of instances labelled correctly (as in Equation 2.24) and n_ϵ the number of instances labelled correctly by chance (as in Equation 2.25).
In the binary case, this is equivalent to

$$\kappa = \frac{2\,(\text{TP} \cdot \text{TN} - \text{FN} \cdot \text{FP})}{(\text{TP} + \text{FP}) \cdot (\text{FP} + \text{TN}) + (\text{TP} + \text{FN}) \cdot (\text{FN} + \text{TN})} \ . \tag{2.37}$$

Note, that while most measures range from 1 to 0, κ ranges from 1 (perfect agreement) over 0 (no agreement beyond chance) to -1 (agreement less than chance).

Chapter 3
Point-Wise Semantic Labelling

3.1 Introduction

Automated scene interpretation is a topic of broad scientific and applicational interest. It describes the process by which computer systems are trained to interpret and analyse scenes, based on various types of input data. Those include images, image sequences, or depth sensor data, which can be recorded by various techniques such as light detection and ranging (LiDAR) systems, stereo cameras, or RGB-D cameras using infrared illumination with structured light, stereo cameras, or time-of-flight technology. Similarly, airborne LiDAR is being used for efficient large-scale land cover analysis, with applications ranging from urban planning, infrastructure and risk management over topographic mapping and land cover classification for flood management, agricultural or forestry planning, environmental monitoring of carbon stocks, coastal processes or glacial retreat up to cultural heritage mapping and archaeology, where structures hidden beneath vegetation or soil can be detected from the airborne LiDAR perspective.

3.1.1 Goals

Most scene interpretation applications rely on semantic labelling. Often, this is implemented as an initial step, where point-wise semantic labelling of the recorded point clouds is achieved via point cloud classification. Semantic labels are assigned to each point of the point cloud, based on a set of per-point features calculated for every instance and a set of training instances for every class (Chehata et al., 2009; Shapovalov et al., 2010; Mallet et al., 2011; Niemeyer et al., 2014; Hackel et al., 2016; Weinmann, 2016; Grilli et al., 2017).

3.1.2 Challenges

To foster scientific exchange on automated scene interpretation in airborne laser scanning (ALS) data and to compare different approaches in the field, a number of benchmark data sets have been released. Those data sets typically provide 3D coordinates of ALS returns as well as a semantic label reference. In urban scenes, the classes typically comprise roofs, façades, trees or high vegetation, low vegetation, sealed surfaces and cars, and sometimes more differentiated classes such as shrubs, fences, and power lines.

Correctly classifying such data is especially challenging, as the structural resolution in ALS data is limited both in footprint diameter and sampling rate compared to point clouds from terrestrial laser scanning (TLS). Moreover, the class definition may not always be clear from a non-contextual, geometric point of view. For example, shrubs, fences or hedges can look geometrically similar to each other, and the semantic difference may not be apparent from the geometric representation. Other classes, such as the common class of roofs, may combine instances of very different geometric appearance, such as pitched roofs and terraced roofs.

3.1.3 Objectives

Our objective is to contribute to the scientific knowledge gain in the field of point-wise semantic labelling of ALS point clouds, based on geometric properties, as stated in research question (RQ)1 of this thesis:

RQ1: How can geometric properties be used for point-wise semantic labelling within ALS point clouds? Can we design a novel geometric feature type, which enables advances compared to existing approaches?

In this chapter, we explicitly aim not to rely on prior knowledge of the class properties. Therefore we follow a generalized framework suited to the task, which breaks down the process into three successive steps:

- neighbourhood definition,
- geometric feature extraction, and
- classification.

Related literature is analysed for each of these steps in Section 3.2. In analysing related literature, we noticed that most existing geometric features focus on describing locally homogeneous neighbourhoods. Those features, however, are highly sensitive to the extent of the neighbourhood considered. Since a complex definition of classes may contain descriptive structures on scales other than those optimal for these geometric features, we propose a novel robust feature type drawn by sampling geometric measures from the returns within the considered neighbourhood. We aim to explore if this feature type is more suited to characterise complex class structures that extend beyond homogeneous local neighbourhoods.

In the methodology section (Section 3.3), we hence describe our implementations of each of the steps listed above in detail, while the materials section (Section 3.4) summarizes the characteristics of the data sets which form the basis of our experiments.

We performed a number of different experiments, using both well established geometric features types, as well as our novel sampled feature type, to address RQ2 of this thesis:

RQ2: What is the influence of different neighbourhood types and scales on the descriptiveness of geometric features with respect to different classes?

Those experiments use some of the material (Section 3.4) and methods (Section 3.3) each.

The results (Section 3.5) are therefore structured along different aspects of our RQs. Section 3.5.1 explores the performance of our novel sampled feature type in comparison to existing geometric features and analyses respective strengths or weaknesses. Section 3.5.2 combines both novel and established features and evaluates, if there are special requirements on the type of classifier used with those features. Section 3.5.3 then combines and compares different feature types, neighbourhood types and scales. This section is subdivided into two parts with different neighbourhood and feature combinations. A detailed analysis explores the benefit on the overall result of combining our novel feature type with well established features, the influence of neighbourhood type and scale, as well as multi-scale, multi-neighbourhood-type combinations. Moreover our method's performance is evaluated on different data sets. In a following step, the differences in performance are linked to the differences among the data. For some results, there is also a direct comparison to other literature published on the same data.

For enhanced clarity, every experimental subsection of the results starts out with a block overview of the neighbourhoods, features, and classifiers used in this experiment. They all start with a first explanation of the experiment, followed by separate paragraphs describing the results and the discussion of these individual results.

The major discussion of all experimental results follows in Section 3.6, along with a comparison to results achieved with deep learning after the publication of our work. A final conclusion is drawn in Section 3.7.

3.2 Related Work

The following section comprises a summary of related work on point-wise semantic labelling of ALS point clouds. Some approaches originate from point-wise semantic labelling of TLS or mobile laser scanning (MLS) point clouds too. In general, those approaches are interchangeable, but due to the different viewing geometries and spatial resolution of the data types, the results may vary or require additional adaptions. In TLS or MLS data, for example, the spatial resolution and sampling density change noticeably for objects at different distances from the sensor. In ALS data, parts of the data are sampled from a nadir perspective, while other parts are sampled from off-nadir scan angles, ranging up to ∼40°, which largely affects the target representation (Liu et al., 2018). Depending on the scanning pattern, scanning angle, and flight path overlap, this results in areas of different target representation throughout one data set. The different viewing geometries lead to different typical occlusion cases. Vertical structures, such as façades, for example, are much better represented in TLS/MLS data or off-nadir viewing directions in ALS data due to the sidewards viewing geometry. Those differences matter when transferring point-wise semantic labelling approaches from one type of data to another.

In accordance with the three steps of the general framework of this task, we structure this summary along the topics of neighbourhood definition (Section 3.2.1), geometric feature extraction (Section 3.2.2) and classification (Section 3.2.3).

3.2.1 Neighbourhood Definition

For point-wise semantic labelling based on geometric features, each return[1] has to be characterized by features describing the spatial distribution of returns within a certain neighbourhood around this return. Both size and shape of this neighbourhood are critical parameters and can be defined in different ways.

[1] As explained in Section 2.1.5, there are different ways of processing ALS data into a point cloud representation. The term return may therefore either refer to a single return in discrete return (DR) data or, in the case of waveform-recording sensors, to a decomposition component of the signal (using either Gaussian or non-symmetric basis functions), or to a noise-exceeding amplitude sequence (NEAS) of the signal, depending on the context of the data used.

3.2.1.1 Neighbourhood Types

Many investigations focus on the representation of local point cloud characteristics at a single scale. For such a single-scale representation, a cylindrical neighbourhood $\mathcal{N}_c$ (Filin and Pfeifer, 2005) or a spherical neighbourhood $\mathcal{N}_s$ (Lee and Schenk, 2002; Linsen and Prautzsch, 2001) is commonly used. Thereby, the scale parameter to describe such a neighbourhood is represented by [] a radius (Filin and Pfeifer, 2005; Lee and Schenk, 2002)[. Sometimes neighbourhoods $\mathcal{N}_k$ are also defined by] the number k of nearest neighbours (Linsen and Prautzsch, 2001)[, resulting in a spherical neighbourhood with a radius dependent on the local point density]. The value of the scale parameter [(radius or k)] is typically selected heuristically based on knowledge about the scene and data.

3.2.1.2 Multiple Scales

In contrast to a representation of local point cloud characteristics at a single scale, a multi-scale representation allows a description of geometric properties at different scales and thereby implicitly accounts for the way in which these properties change across scales. To describe local point cloud characteristics at multiple scales, Niemeyer et al. (2014) and Schmidt et al. (2014) used a collection of cylindrical neighbourhoods with infinite extent in the vertical direction and radii of 1 m, 2 m, 3 m and 5 m, respectively. []

In contrast to these neighbourhood types, it has also been proposed to use a multi-scale voxel representation (Hackel et al., 2016) or even different entities in the form of voxels, blocks, and pillars (Hu et al., 2013), in the form of points, planar segments and mean shift segments (Xu et al., 2014), or in the form of spatial bins, planar segments and local neighbourhoods (Gevaert et al., 2016). [] Yang et al. (2017) considered local point cloud characteristics on the basis of points, segments, and objects as well as local context for analysing point clouds.

3.2.1.3 Optimized Scale

To automatically select a suitable [neighbourhood scale parameter] value in a data-driven approach, [typically in MLS point cloud data,] it has for instance been proposed to select the optimal scale parameter for each individual point via dimensionality-based scale selection (Demantké et al., 2011), where a highly dominant behaviour of one of the dimensionality features (i.e. linearity, planarity, and sphericity) is favoured. A similar approach has been presented with eigenentropy-based scale selection (Weinmann et al., 2015), where the minimal disorder of 3D points is favoured.

3.2.2 Geometric Feature Types

Based on a given neighbourhood definition, each point of the point cloud should be characterized by features describing the geometric properties of the returns within its surrounding neighbourhood. Therefore the point cloud is being filtered for all returns that fall within the local neighbourhood of one seed return. This subset of returns is then used to calculate geometrics features, such as distributions or moments, which characterise this neighbourhood. The following types of geometric features can be found in related literature.

3.2.2.1 Parametric Features

Vosselman et al. (2004) use an approach of fitting geometric primitives, such as planes, spheres, or cylinders to the local point cloud data. The estimated parameters are then used as features.

3.2.2.2 Metrical Features

Many established features describe the local point cloud geometry by evaluating one single geometric property within a local neighbourhood. We summarize such features by the term of metrical features. Often, these features can be understood intuitively (West et al., 2004; Jutzi and Gross, 2009a; Mallet et al., 2011; Weinmann et al., 2015; Guo et al., 2015). A typical group of metrical features are the so-called covariance features, which are based on the eigenvalues of the covariance matrix. Those eigenvalues represent the variability of the given distribution in an orthogonal basis and allow the calculation of features such as linearity (is one eigenvalue way bigger than the other two?) or planarity (are two eigenvalues way bigger than the other(s)?) for example. Other geometric 3D properties that yield metrical features are local point density, verticality, standard deviation of height values or maximum height difference. An exhaustive overview of metrical features and their calculation is given in Section 3.3.2 of the methodology description.

3.2.2.3 Sampled Features

Before our publication (Blomley et al., 2014), there was to the best of our knowledge no published literature on the use of sampled features for ALS point cloud semantic labelling. However, both parametric and metrical features are most discriminative when the distribution of returns within the local neighbourhood they describe follows some degree of local homogeneity, meaning that all of the returns come from one planar surface, one linear structure, or even volumetric scattering. In ALS however, the resolution is usually limited and the number of returns per area may not be sufficient to resolve class-specific geometric primitives.

> Thus probabilistic distributions of geometric properties [] may hold more information than locally calculated parameters. Reaching beyond locally homogeneous neighbourhoods, histogram distributions have already been successfully used in computer vision sciences (Tombari et al., 2010). In image-based keypoint description, the scale-invariant feature transform (SIFT) algorithm is a prominent example [of] robustness and effectiveness achieved by a set of local histograms (Lowe, 2004). In 3D point clouds, existing histogrammetric approaches are limited to surface keypoint description, as they rely on surface normal vectors (Tombari et al., 2013; Rusu et al., 2009).
>
> [Within our research, we therefore aimed] to introduce novel reliable geometric features for volumetric point cloud classification, which perform well at multiple scales. We therefore adopt a proposal from object recognition, using histograms of randomly sampled geometric measures, called shape distributions (Osada et al., 2002), as features within [] local neighbourhood[s].

3.2.2.4 Topographical Features

Normalized height above ground is an important feature for each individual return (Chehata et al., 2009; Mallet et al., 2011). This feature, however, does not require the consideration of the immediate local neighbourhood in the way that the metrical or sampled features types do, but rather relies on a prior step of estimating the scene topography. To do so, a surface created from local minima of the ALS data has to be filtered in order to remove non-ground objects, such as vegetation or buildings, from the scene. This step may therefore include other neighbourhood estimations like rasterization and smoothing operations, as well as ridge detection, region growing, or morphological operations.

Typical methods of estimating a digital terrain model (DTM) include triangulated irregular networks, implemented for example in the LAStools package (Isenburg, 2015), weighted linear least squares (Kraus and Pfeifer, 1998), multi-scale curvature calculation (Evans and Hudak, 2007) or progressive morphological filters (Zhang et al., 2003). Those generally work well on open, vegetated scenes (Silva et al., 2018). If the scene contains artificial objects such as buildings, it can be challenging to extract a suitable DTM by rule-based methods. Therefore, object-based filtering of ground points may be applied (Song and Jung, 2023), which already requires scene understanding. To extract the normalized height above ground as a feature for initial scene understanding though, a suitable compromise has to be sought depending on the level of detail needed and the complexity of the considered scene.

3.2.3 Classification

[With regard to classification], the straightforward solution consists in selecting a standard approach for supervised classification, e.g. a support vector machine (SVM) classifier (Mallet et al., 2011; Lodha et al., 2006), a random forest (RF) classifier (Chehata et al., 2009; Guo et al., 2011; Steinsiek et al., 2017), an AdaBoost(-like) classifier (Lodha et al., 2007; Guo et al., 2015) or a Bayesian discriminant analysis classifier (Khoshelham and Oude Elberink, 2012). However, as these classifiers treat each point of the point cloud individually, they do not take into account a spatial regularity of the derived labelling, i.e. a visualization of the classified point cloud might reveal a 'noisy' behaviour.

To enforce spatial regularity, local context information can be taken into account. This means that, instead of treating each point individually by considering only its corresponding feature vector, the feature vectors and labels of neighbouring points are taken into account as well. In many cases, such a contextual classification involves a statistical model of context. where particular attention has been paid to the use of a conditional random field (CRF) (Niemeyer et al., 2014; Schmidt et al., 2014; Steinsiek et al., 2017; Landrieu et al., 2017a).

In the scope of our work, we focus on standard approaches for supervised classification, as respective classifiers are [] available in numerous software tools and rather easy-to-use by non-expert users.

3.3 Methodology

This section describes the methods which were implemented for each step of our point-wise semantic labelling framework. There are three types of neighbourhoods (Section 3.3.1), four types of geometric features (covariance features (Section 3.3.2.1), other 3D geometric properties (Section 3.3.2.2), shape distribution features (Section 3.3.3) and normalized height (Section 3.3.4)). Feature values are normalized in order to ensure improve classification performance, model stability and ensure a balanced feature contribution (Section 3.3.5) before classification. Our experiments use the following classifiers: nearest neighbour (NN) classification (Section 3.3.6.1), linear discriminant analysis (LDA) classification (Section

3.3.6.2), quadratic discriminant analysis (QDA) classification (Section 3.3.6.3), SVM (Section 3.3.6.4), and a RF classifier (Section 3.3.6.5).

3.3.1 Neighbourhoods for Feature Extraction

We argue that cylindrical and spherical neighbourhoods have the benefit that they rely only on one scale parameter independent of the local point distribution, but we also advocate that in this case of neighbourhoods with fixed scale parameters, multiple sizes [] should be considered. In addition to the cylindrical neighbourhoods proposed by Niemeyer et al. (2014) and Schmidt et al. (2014), we hence also use a collection of spherical neighbourhoods as proposed by Brodu and Lague (2012) in the scope of an investigation focusing on terrestrial laser scanning data. As we focus on ALS data with a significantly lower point density, we do not consider neighbourhoods with radii in the centimetre scale. Instead, we select the same radii as used by Niemeyer et al. (2014) and Schmidt et al. (2014) for cylindrical neighbourhoods. Consequently, we consider a collection of spherical neighbourhoods with radii of 1 m, 2 m, 3 m and 5 m, respectively.

In addition to [fixed scale] neighbourhoods, [we] also use a spherical neighbourhood of locally adaptive size for each [return]. Thereby, the local adaptation is achieved via eigenentropy-based scale selection (Weinmann et al., 2015), where the optimal scale parameter is directly related to the minimal disorder of 3D points within a local neighbourhood.

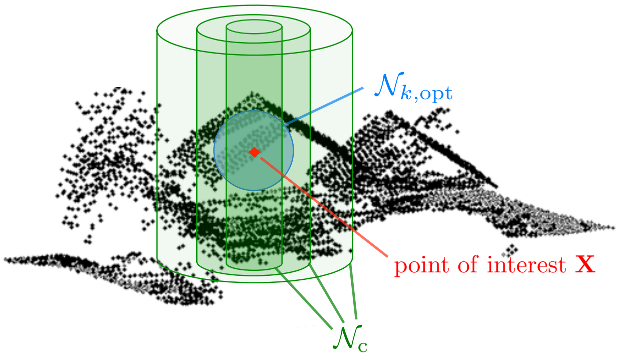


Fig. 3.1: [] Neighbourhood definitions used in this work as the basis for extracting features for a considered 3D point [(return)] X [(red)]: cylindrical neighbourhoods $\mathcal{N}_c$ [with different radii and infinite extent in the vertical direction (green)] and a spherical neighbourhood $\mathcal{N}_{k,opt}$ formed by an optimal number k_{opt} of nearest neighbours [(blue). For the sake of clarity, other spherical neighbourhoods are left out in this visualization, even though spherical neighbourhoods $\mathcal{N}_s$ of different fixed radii are sometimes also used.]

The different sets of neighbourhoods considered either individually or as a multi-scale multi-neighbourhood combination throughout this work are therefore as follows:

- cylindrical neighbourhoods $\mathcal{N}_c$ with fixed horizontal radii of 1 m, 2 m, 3 m and 5 m and an infinite extent in the vertical direction,

- spherical neighbourhoods $\mathcal{N}_s$ with fixed radii of 1 m, 2 m, 3 m and 5 m,

- and one spherical neighbourhood type $\mathcal{N}_{k,\mathrm{opt}}$ formed by an optimal number k_{opt} of nearest neighbours. This last neighbourhood type has a varying size for each return in the point cloud, depending both on the local return density and distribution around this point of interest.

- 2D rasters of coarse grid spacing (20 m) and fine grid spacing (0.5 m) are used as neighbourhoods for the normalized height feature.

3.3.2 Established Features for Comparison

In the scope of this work, we did not use any parametric features as we did not imply knowledge about the underlying geometric primitives. Instead, we focused on a more generic description of the geometric distribution of returns within a given neighbourhood. The features most common in related literature were:

- covariance features, belonging to the group of metrical features and
- other metrical geometric 3D properties, as suggested by Weinmann et al. (2015).

In order to evaluate the benefit of our novel feature type, those feature types were implemented for comparison, as explained throughout this section.

3.3.2.1 Covariance Features

Most present approaches using 3D geometric features employ features derived from the local covariance matrix representing second-order invariant moments within the [local distribution of returns]. The covariance matrix is calculated from N observations $A_{x,y,z}$ as follows:

$$[c]_{ij} = \frac{\sum_{l=1}^{N} (A_i - \overline{A_i}) \cdot (A_j - \overline{A_j})}{N} , \tag{3.1}$$

where $i, j \in [x, y, z]$ and $\overline{A_i}$ holds the mean of all observations in the respective dimension. Subsequent principal component analysis is used to determine linearly uncorrelated second-order moments in an orthogonal eigenvector space. The [] eigenvalues $\lambda_{1,2,3}$ [corresponding to the eigenvectors $\vec{e}_{1,2,3}$] then hold a great potential to calculate local features including dimensionality (linearity, planarity, and sphericity) and other measures such as omnivariance, anisotropy and eigenentropy. The eigenvalues, sorted as $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ and the measures listed in Equation 3.2, will be referred to as covariance features:

$$\begin{array}{rl}
\text{linearity} & L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1}, \\
\text{planarity} & P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1}, \\
\text{sphericity} & S_\lambda = \frac{\lambda_3}{\lambda_1}, \\
\text{omnivariance} & O_\lambda = \sqrt[3]{\lambda_1 \lambda_2 \lambda_3}, \\
\text{anisotropy} & A_\lambda = \frac{\lambda_1 - \lambda_3}{\lambda_1}, \\
\text{eigenentropy} & E_\lambda = -\sum_{i=1}^{3} \lambda_i \ln(\lambda_i), \\
\text{sum of } \lambda\text{s} & \Sigma_\lambda = \lambda_1 + \lambda_2 + \lambda_3, \\
\text{change of curvature} & C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}.
\end{array} \tag{3.2}$$

Yet it is especially important for these features to be derived from a suitably chosen neighbourhood size. It is the nature of second-order moments that the distance of one element from the mean contributes quadratically (cf. Equation 3.1) and therefore elements in the vicinity are far less important than those further away. Since the principal component analysis is an orthogonal and thereby unitary transformation, the resulting eigenvalues are sensitive to the original scaling. Demantké et al. (2011) and Gressin et al. (2012) show evidence that a suitable spherical neighbourhood size can be found by minimization of the Shannon entropy, based on dimensionality-features. Yet it remains to be shown if this optimum neighbourhood size for covariance features corresponds to the characteristic scale of structure[s characteristic of the given classes]. To advance further research in the field, other geometrical features more suited to multiple scales are indispensable.

Dittrich et al. (2017) also studied the effect of discretization and noise in point cloud data on the covariance features in well-defined primitive shape cases (such as a line, end-of-line, plane, half-plane, etc.). Some covariance features, especially eigenentropy and the sum of eigenvalues, were found to be subject to significant relative errors when applied to discretized data. Moreover, when the individual point measurements were subject to variance (such as noise resulting from the measurement device, surface properties or scanning geometry), variance propagation especially affects the features of linearity, planarity and sphericity in a strong way.

3.3.2.2 Other Geometric 3D Properties

The geometric 3D properties proposed by (Weinmann et al., 2015) are derived from the spatial arrangement of points within the considered cylindrical or spherical neighbourhood. The respective features are represented by the

$$\begin{aligned} \text{local point density} \quad & D, \\ \text{verticality} \quad & V = \vec{e}_{3z}, \\ \text{maximum height difference} \quad & \mathit{\Delta} H, \\ \text{standard deviation of height values} \quad & \sigma_H \end{aligned} \tag{3.3}$$

[of] those points within the local neighbourhood.

For the spherical neighbourhood [whose scale parameter has been] determined via eigenentropy-based scale selection [$N_{k,\text{opt}}$], the radius [R] of the local neighbourhood is considered as an additional feature.

Similarly to the covariance features described above, these features are sensitive to outliers.

3.3.3 Shape Distributions

This section describes how we adapted shape distributions (Osada et al., 2002) as features for semantic ALS point cloud labelling. They constitute a sampled feature type descriptive of complex local geometries or repeating patterns.

The characteristic scale of complex and partially random structures may not always be identical to the optimum neighbourhood size of covariance features. Yet to reveal such patterns, a statistical distribution of randomly sampled values may be more suitable than single values such as covariance [or other geometric 3D property] measures. [] The key idea [by (Osada et al., 2002)] is to use random sampling of simple geometric measures to obtain a signature of the neighbourhood around each point [] as a histogrammetric shape distribution. [As in said] reference, we investigate the following geometric measures[, which are visualized in Figure 3.2]:

- D1: distance between any random point and the centroid of all considered points,

- D2: distance between two random points,
- D3: square root of the area of a triangle between any three random points,
- D4: cubic root of the volume of a tetrahedron between any four random points,
- A3: angle between any three random points.

[] The resulting histogram therefore represents the probability distribution of the taken geometric measures within the [considered neighbourhood] and should reveal repeating structures by a more frequent occurrence of some values. By this approach, feature extraction is reduced to a simple random sampling procedure. Such features are fast to calculate, mirror- and rotation-invariant, and robust regarding outliers, noise and varying point density due to application-specific scanning or flight patterns.

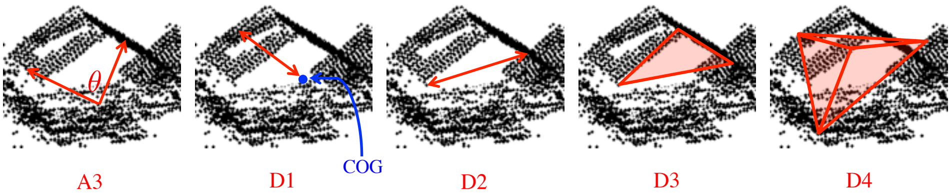


Fig. 3.2: Visualization of the geometric measures taken for the sampling of shape distributions.

As for the number of random samples taken within each neighbourhood, a compromise has to be sought between computational effort and descriptiveness, taking into account the typical data type and application case considered. We chose to limit ourselves to 255 pulls of geometric measures within each neighbourhood. In the Vaihingen data set (Cramer, 2010), where we first tested the shape distributions, 50 % of all points had 255 neighbours in cylindrical neighbourhoods of ~2.5 m radius, suggesting that this number of pulls would allow for a representative sample of geometric measures within a neighbourhood. For larger neighbourhoods, 255 samples might represent a random subset, but we consider this to reduce the danger of over-fitting when reference areas are small with respect to the considered neighbourhood scale.

Dissenting from the original shape distribution proposal, we use an adaptive histogram binning approach to achieve maximum variance of significant observations from the gross of the total data set. Above all, this step ensures a scale-independent performance. For this purpose, a simple histogram equalization procedure, known from image processing applications (Gonzales and Wood, 2002), is adapted. For all measured values at a linear binning scope m_k, with $k = 0, ..., L-1$ and L the number of bins, a transformation function $T(m_k)$ to a non-linear binning scope is found in such a way that a histogram of any large number of random samples is equally distributed. The transformation function is defined as

$$T(m_k) = \sum_{j=0}^{k} p_m(m_j)\,, \tag{3.4}$$

where p_m is defined as the probability of occurrence of a value within m_k from a large number of samples n:

$$p_m(m_k) = \frac{n_k}{n}, \tag{3.5}$$

where n_k is the number of occurrences within m_k.

[However,] the number of bins per shape histogram has to be specified. A large number of bins will allow for sophisticated neighbourhood descriptions, but [given the limited point density of ALS data] the signature may then not be descriptive. Therefore we decided to use 10 bins per shape histogram in [our initial] proposal. We did not find it necessary to further optimize this later. To determine the histogram binning thresholds, 500 random samples were drawn from the training data for each neighbourhood type. No significant change of the adapted binning thresholds was observed after this point.

3.3.4 Normalized Height

Normalized height is conceptually simple, and the difficulty in implementation varies depending on the types of objects in the scene and the accuracy required. It has shown to be a significant factor in distinguishing between similar structures at different height above ground, such as impervious surface and building in urban scenes (Gerke, 2015). This section describes our simple but useful implementation.

The normalized height feature is derived from an approximation of the scene topography and estimated from the point cloud itself, as shown in Figure 3.3. First, absolute height minima are determined on a large grid with a sampling distance of 20 m. Afterwards, a linear interpolation is performed among those coarsely gridded minimum values and evaluated on a fine grid of 0.5 m sampling distance. Finally, a normalized height value is assigned to each 3D point by calculating the difference of the point's height value and the topographic height value of the corresponding grid cell.

3.3.5 Feature Normalization

It is obvious that – by definition – the considered features address different quantities and may therefore be associated with different units as well as a different range of values. This, in turn, might have a negative impact on the classification results as the distribution of single classes in the feature space might be suboptimal. Accordingly, it is desirable to introduce a normalization which allows the transfer of the given feature vectors to a new feature space where each feature contributes approximately the same, independent of its unit and its range of values. For this purpose, we conduct a normalization of all features.

For the covariance features, the geometric 3D properties and the normalized height, we use a linear mapping to the interval [0, 1]. To reduce the effect of outliers, the range of the data is determined by the 1st-percentile and the 99th-percentile of the training data. Only if the absolute minimum is zero, the lower range value is set to zero too. For shape distributions, normalization is achieved by dividing each histogram count by the total number of pulls from the local neighbourhood.

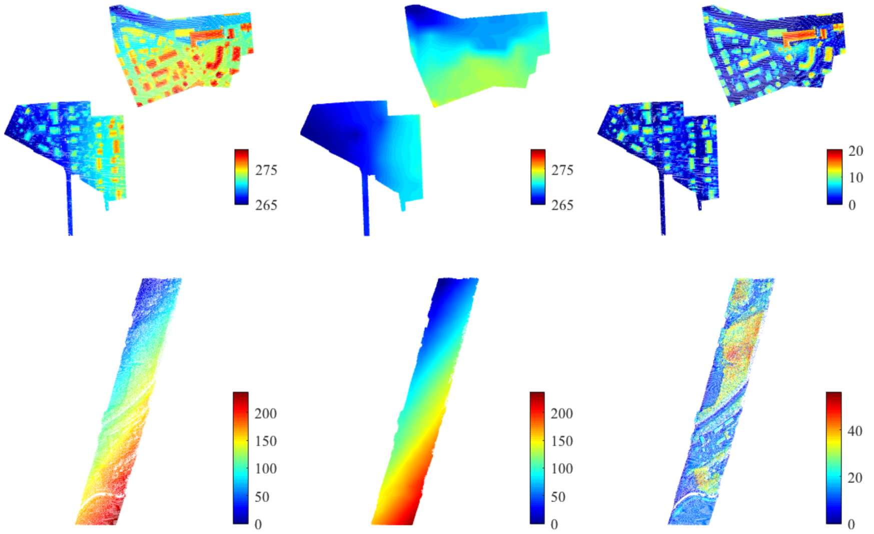


Fig. 3.3: Effects of the scene topography[, colour values indicate the unit of meters in the legend bar]. The point clouds' height minima on a 0.5 m grid are shown on the **left**, the approximation of the scene topography is plotted in the **middle**, and the normalized minima are shown on the **right**. The **top** row depicts [area 1 & 3] of the **Vaihingen** data set, while the **bottom** row shows the test area of the **GML** data set A.

3.3.6 Classification

As explained in Section 2.2, classification approaches in machine learning may be divided into generative or discriminative approaches. Although discriminative approaches (which focus on the best separation of classes) enable separating more closely intertwined classes than generative approaches (which focus on a modelling of underlying probability density functions), it is of interest for feature development to see if the devised features enable class separation on a less sophisticated level. The individual classifiers used in the scope of this work are described in the following subsections.

3.3.6.1 Nearest Neighbour Classification

This simple NN classification approach (a version of k-nearest neighbour (k-NN) with $k = 1$) compares the feature vectors of the test set to those of the training set. Each feature vector in the test set is directly compared to the feature vectors in the training set, and the class label of the most similar training example is assigned. As similarity metric, we use the Euclidean distance. To reduce the detrimental effect of unbalanced training examples per class, we reduced the training set to a certain number of examples per class when using this classifier and duplicated training examples for classes with less than this number of training instances.

3.3.6.2 Linear Discriminant Analysis

LDA assumes that the probability density functions for all classes follow multivariate Gaussian distributions, and that the class covariances are identical, meaning that the variance within a class is the same across different classes. This leads to a linear decision boundary between classes. LDA is sensitive to outliers, and the predictive power decreases if some of the features are correlated. The size of the smallest class must be larger than the number of features. In our experiments, we thus used the same methods as for the nearest neighbour classification to produce a balanced training set.

3.3.6.3 Quadratic Discriminant Analysis

QDA uses the same approach as LDA in assuming a multivariate Gaussian distribution for each class. However, QDA allows for different class covariances. Thus it leaves many more free parameters for the classification problem to be fitted, making this model more flexible, while at the same time rendering it more prone to overfitting compared to LDA. QDA thus requires more training data to produce reliable results compared to LDA. We used this model with the same balanced training set as the nearest neighbour classification and LDA.

3.3.6.4 Support Vector Machines

Support vector machines (SVMs) belong to the group of discriminative learning approaches. When testing the shape distributions in 2014, SVMs were a state-of-the-art classifier with readily available implementations. [We used a] SVM classifier provided by the LIBSVM package (Chang and Lin, 2011). [This] classifier uses a radial basis function kernel and depends on two parameters, namely γ, representing the width of the Gaussian kernel function and C, a soft margin parameter allowing for some mis-classifications. A grid search for optimal values of γ and C is completed [for every classification task] by evaluation of the cross-validation accuracy on a threefold partition of the training data.

3.3.6.5 Random Forest Classifier

Random forests (RFs) are an ensemble learning method used for classification and regression, and belong to the group of discriminative learning approaches. They use the principles of bootstrapping and bagging, meaning they draw multiple bootstrapped samples (random sampling of data subsets drawn with replacement from the training data) to produce a weak decision tree classifier from each random subset, and later aggregate the predictions of all weak classifiers by a majority vote to determine a final strong prediction. Thus, the technique has some free parameters, such as the number of bootstrapped decision trees N_T and the maximum depth of each decision tree d_T, as well as splitting criteria like the minimum number of samples required for a split $n_{\min}$, or the number of active variables used for the test in each tree node n_a. Random forest decision models are particularly robust, since they do not require assumptions about the distribution of the data and prevent overfitting by training on different random subsamples.

3.4 Material

With the Methods described above, several experiments were conducted on two different ALS benchmark data sets, each comprising urban landscapes.

3.4.1 ISPRS Vaihingen Data Set

The Vaihingen data set, provided by the International Society for Photogrammetry and Remote Sensing (ISPRS) (Cramer, 2010) is an airborne laser scanning data set acquired in August 2008 over Vaihingen, a small village in Germany[2]. It was recorded with a Leica ALS50 system at a mean flying height of 500 m with a 45^o viewing angle, 30 % overlap of flight strips and a median point density of 6.7 pts./m^2, where regions without strip overlay have a mean point density of 4 pts./m^2.

The data set is split into three areas displaying different characteristics. Area 1 contains historic buildings with complex roof shapes and some trees. Area 2 contains a few high-rise residential buildings surrounded by trees as well as a patch of small detached houses, while Area 3 is a residential area with small detached houses only.

The Vaihingen data set has been presented in the scope of the ISPRS Test Project on Urban Classification and 3D Building Reconstruction (Rottensteiner et al., 2012), and it meanwhile serves as benchmark data set for the ISPRS benchmarks on 2D and 3D semantic labelling. More details about this data set are provided on the ISPRS webpages [3]. The data set is now fully available for download, whereas at the time of this research, reference data for Area 3 was withheld, so that participants in the benchmark had to submit their results for external evaluation.

In the scope of the ISPRS benchmark on 3D semantic labelling, nine semantic classes have been defined for the Vaihingen data set, and these classes are given by *Powerline*, *Low Vegetation*, *Impervious Surfaces*, *Car*, *Fence / Hedge*, *Roof*, *Façade*, *Shrub* and *Tree*. The point-wise reference labels have been determined based on (Niemeyer et al., 2014). The Vaihingen data set is split into a training set and a test set (see Table 3.1). The training set is visualized in Figure 3.4 and contains the spatial XYZ-coordinates, reflectance information, the number of returns and the reference labels. For the test set, only the spatial XYZ-coordinates, reflectance information and the number of returns are provided.

class	training set		test set	
	num.	perc.	num.	perc.
Powerline	546	0.07	600	0.15
Low Vegetation	180 850	24.0	98 690	24.0
Impervious Surface	193 723	25.7	101 986	24.8
Car	4 614	0.6	3 708	0.9
Fence / Hedge	12 070	1.6	7 422	1.8
Roof	152 045	20.2	109 048	26.5
Façade	27 250	3.6	11 224	2.7
Shrub	47 605	6.3	24 818	6.0
Tree	135 173	17.9	54 226	13.2
Σ	753 876		411 722	

Table 3.1: Number of 3D points per class in the **Vaihingen** data set. At the time of publication, the reference labels were only available for the training set, but not for the test set. Evaluation was therefore conducted externally. The full information about the test set has only become available later.

[2] `https://www2.isprs.org/media/komfssn5/complexscenes_revision_v4.pdf` (Accessed in May 2024)

[3] `https://www.isprs.org/education/benchmarks/UrbanSemLab/default.aspx` (Accessed in May 2024)

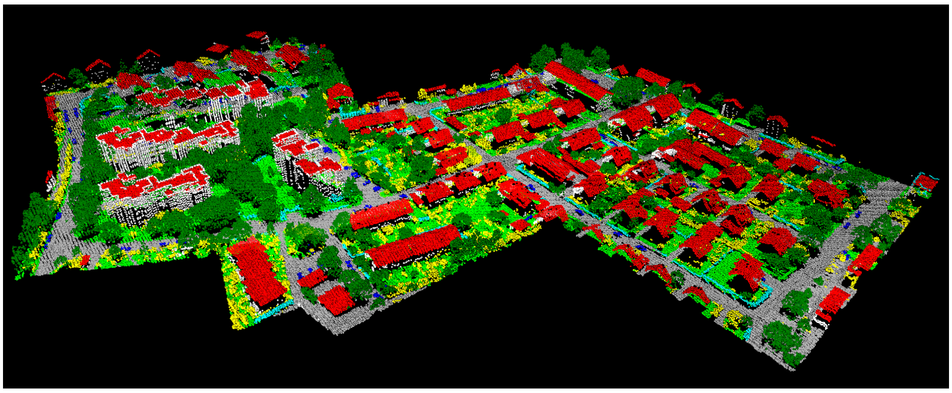


Fig. 3.4: Reference point cloud of the **Vaihingen** data set Area 2, coloured with respect to nine semantic classes (*Roof*: red; *Façade*: white; *Impervious Surfaces*: grey; *Car*: blue; *Tree*: dark green; *Low Vegetation*: bright green; *Shrub*: yellow; *Fence / Hedge*: cyan; *Powerline*: black)

3.4.2 GML Data Set A

The GML data set A (Shapovalov et al., 2010) is provided by the Graphics and Media Lab (GML), Moscow State University, and used to be publicly available [4]. This data set has been acquired with an ALTM 2050 system (Optech Inc.) and contains about [two million] labelled 3D points, whereby the reference labelling has been performed with respect to five semantic classes represented by *Ground*, *Building*, *Car*, *Tree*, and *Low Vegetation*. For this data set, a split into a training scene and a test scene is provided as indicated in [Table 3.2. The reference data is visualized in Figure 3.5. Note, that the distribution of classes (and hence the colour scheme) is different to that of the Vaihingen data set.]

class	training set		test set	
	num.	perc.	num.	perc.
Ground	557 142	51.8	439 989	43.9
Building	98 244	9.1	19 592	2.0
Car	1 833	0.2	3 235	0.3
Tree	381 677	35.5	531 852	53.0
Low Vegetation	35 093	3.2	7 758	0.8
Σ	1 074 569		1 002 668	

Table 3.2: Number of 3D points per class in the **GML** data set A.

The GML data set A used to be complemented by a second set, the GML data set B, which is generally similar to data set A, slightly bigger, but less detailed. The class *Car*, for example, does not occur in data set B. We considered the GML data set B together with data set A in our respective publication (Blomley et al., 2016b). However, the results on the GML data set B did not yield additional information compared to the results on data set A. For a more streamlined presentation, the additional results on data set B are therefore not included here.

[4] `https://graphics.cs.msu.ru/en/science/research/3dpoint/classification`, (Accessed in April 2017)

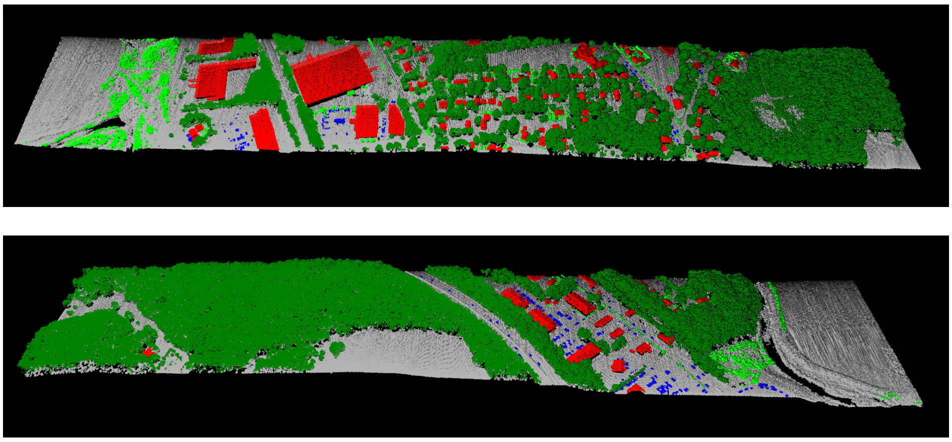


Fig. 3.5: Reference point clouds of the **GML** A data set A ((**top**: training area; **bottom**: testing area), coloured with respect to five semantic classes (*Building*: red; *Ground*: grey; *Car*: blue; *Tree*: dark green; *Low Vegetation*: bright green)

3.5 Results

In line with the RQs of this chapter, different experiments have been carried out following different objectives. Those objectives are reflected in the three following subsections. The first subsection (Section 3.5.1) is intended in an investigative style and explores how different feature types interact with different classes on different scales. The following subsection (Section 3.5.2) deals with the influence of different classifier types. After those experiments yielded promising results, the third subsection (Section 3.5.3) finally combines different complementary feature types extracted on different neighbourhood scales in order to perform a joint classification on two benchmark semantic labelling challenges. Results in this section are evaluated either numerically, according to the evaluation metrics given in Section 2.5 or visually, based on manual inspection of point clouds colour-coded according to the class labels assigned.

3.5.1 Comparison of Shape Distribution and Covariance Features in Binary Classification of Four Classes on Varying Scales

Neighbourhoods

separate $\mathcal{N}_c$ at radii of $2^{n/2}$ m with $n \in \mathbb{N}$ and $-4 \leq n \leq 11$

Features

either
- covariance features

or
- shape distributions

Classification

- balanced training $^{1000\,\text{pts}}/_{\text{class}}$
- binary one-against-all SVM
- four selected classes

The aim in this section was to investigate, how different feature types, namely shape distributions and covariance features, are suited to describe class-specific structures or patterns. As those descriptive structures or patterns may occur on different class-specific scales, each feature type is extracted separately on cylindrical neighbourhoods of different scales. The experiments in this section are therefore structured along an investigative line of thought, rather than aiming at producing coherent point cloud semantic labelling results of good quality.

For the sake of clarity, we focus on the four main classes of *Roof*, *Tree*, *Vegetated Ground*, and *Sealed Ground*. The tests are being performed on the Vaihingen data set (Section 3.4.1). Due to the plausible class-dependence of optimal neighbourhood scales, each feature type is tested individually in binary one-against-all classifications for each class using cylindrical neighbourhoods $\mathcal{N}_c$ of varying radius. The neighbourhood radii are chosen as $2^{n/2}$ m with $n \in \mathbb{N}$ and $-4 \leq n \leq 11$. Thereby the radius ranges between 0.25 m (within the lateral placement accuracy of the laser scanner) and 45 m (above most object sizes), which should cover all possibly resolved structural scales. The results of the binary one-against-all classifications are presented in Section 3.5.1.1. The results achieved for covariance features are further supported by an analysis of class-wise mean Shannon entropy. In Section 3.5.1.2, a filter-based feature relevance assessment is performed to compare the different types of shape distributions as well as the covariance features. To strengthen a qualitative and visual understanding of the shape distributions' capabilities, the results of the class-wise binary classifications at their respective best neighbourhood scales are combined in Section 3.5.1.3.

3.5.1.1 Class-Wise Analysis of Neighbourhood Scale Impact

To investigate whether some classes are particularly well described by features from certain scales, separate one-against-all distinctions are better suited than one multi-class classification. The identification of each separate class is performed using a one-against-all binary SVM classifier [as described in Section 3.3.6.4]. To ensure a smooth classification procedure, all feature data [except for 1 % of outliers] are scaled to a range between zero and one [as described in Section 3.3.5.] The grid search and subsequent training of a classifier with the best respective parameters is performed on a subset containing 1 000 data points of each class to avoid a bias by unbalanced reference data distribution. Afterwards, the performance of any selected classifier is tested on all labelled training data ($4.1 \cdot 10^5$ points). Each classification result is then evaluated according to completeness / recall, correctness / precision, and quality as described in Section 2.5. Results are plotted in Figures 3.6 and 3.7.

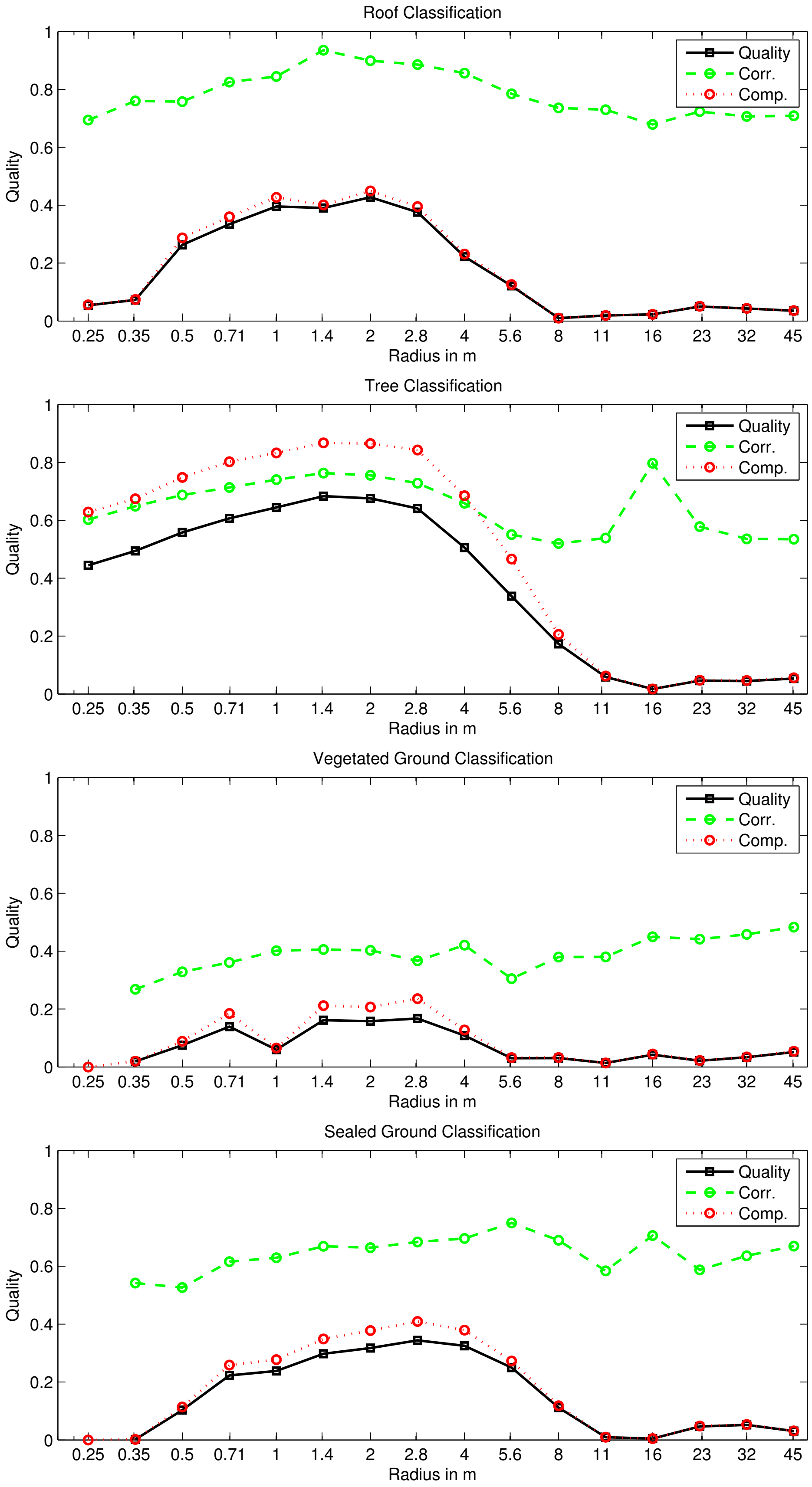
Results

[For the shape distributions,] the resulting graphs [shown in Figure 3.6] are smooth and generally display an even peak-like distribution[. This indicates,] that shape distribution features are a suitable choice to evaluate geometrical properties of point clouds over a wide spatial neighbourhood scale. No prominent peaks occur to suggest a strong pattern or scale preference for individual classes in this data. Optimal results are achieved at the following neighbourhood radii:

2.0 m	for *Roof*
1.4 m	for *Tree*
2.8 m	for *Vegetated Ground*
2.8 m	for *Sealed Ground*

Interestingly all those maxima fall within a similar range. This generally descriptive size for shape distribution features is significantly higher than the [typical] neighbourhood size used for covariance features in ALS. The values of 1.0 m used by Jutzi and Gross (2009a) and 0.75 m used by Niemeyer et al. (2012) agree well with [our findings for covariance features, as shown in Figure 3.7].

Using exclusively covariance features, it is not possible to conduct a cohesive analysis spanning the same scale range as shown for [the shape distributions]. For small neighbourhoods, the covariance features are not separable by the SVM classifier. As, at these radii, more than 25 % of all points have four or less neighbours, this is not surprising, since with less than four elements no three invariant moments may be calculated. This finding of generally better classification performance at


Roof Classification

Tree Classification

Vegetated Ground Classification

Sealed Ground Classification

Fig. 3.6: Evaluation of [SVM] classification results exclusively employing shape distributions for different neighbourhood sizes. [Quality measures are calculated according to Section 2.5 and plotted against the radius of a cylindrical neighbourhood.]

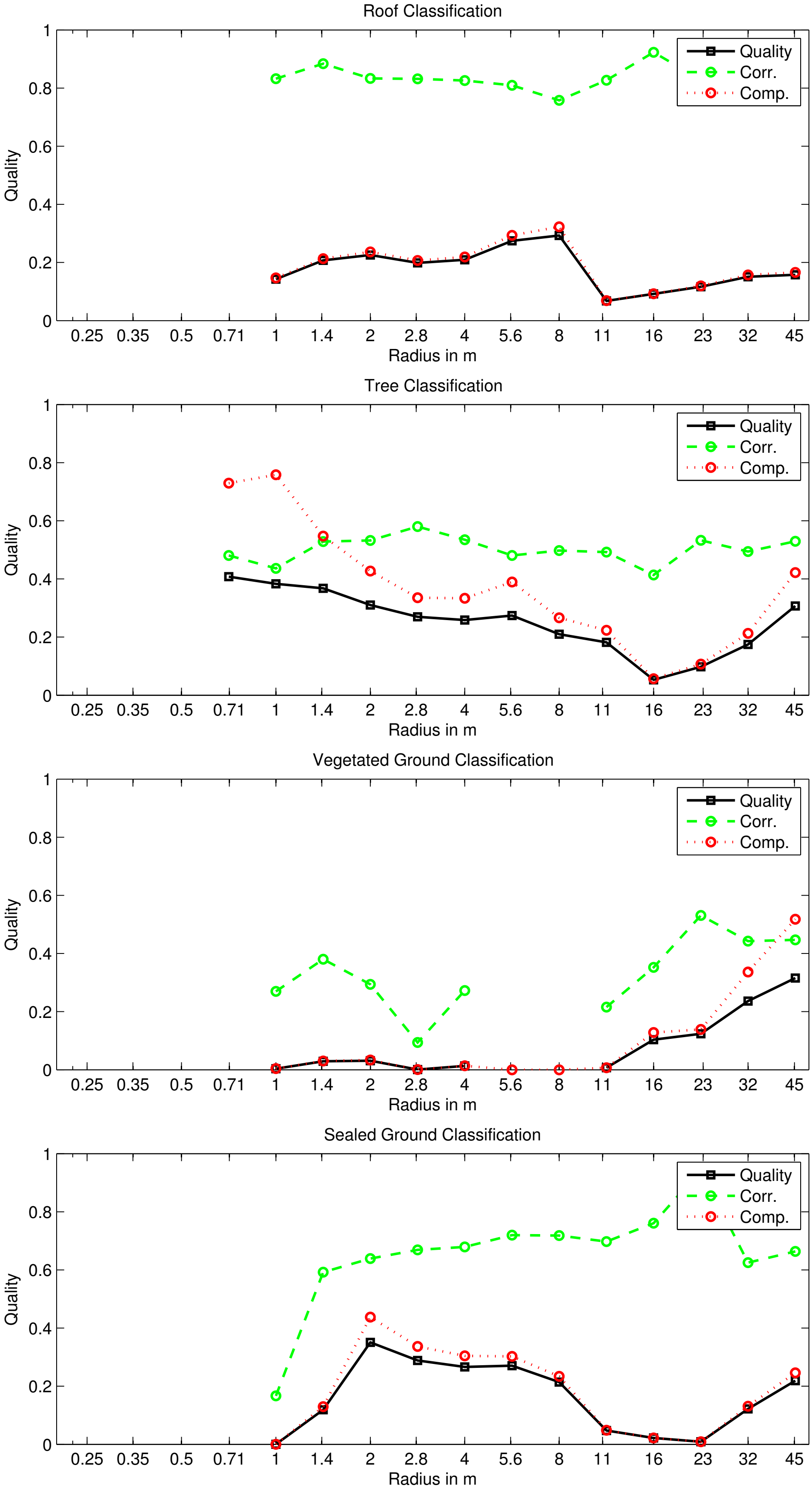


Fig. 3.7: Evaluation of [SVM] classification results exclusively employing covariance features for different neighbourhood sizes. [Quality measures are calculated according to Section 2.5 and plotted against the radius of a cylindrical neighbourhood].

higher numbers of neighbouring elements agrees well with findings published [by] Niemeyer et al. (2011b), where covariance features are used among others.

A very intriguing observation is to be seen at large neighbourhood sizes, where quality increases dramatically. Yet this increase cannot be said to result from a generally better classification performance. Both C and γ parameters of the SVM are very high for these results, indicating an over-fitting (Kotsiantis, 2007). This can be explained by a high overlap between the neighbourhood of points within the same reference area, as the reference areas are much smaller than the neighbourhood for feature calculation. As explained in Section 3.3.2.1, covariance features are highly influenced by elements far away from the mean of all observations. For a homogeneous point distribution in the area, the extra number of points in a bigger circle increases more than quadratically by the increase in radius, and the distance of all those extra points in turn contributes quadratically to the covariance matrix. Therefore the neighbourhood of relatively close points has a [considerable] overlap and the resulting features are nearly identical, causing the observed overfitting. Further studies, taking only a reduced random subset per neighbourhood for covariance feature calculation, did not display an increase in completeness and quality for high neighbourhood radii but were otherwise identical. Therefore increased classification quality for radii above 16 m is regarded as erroneous.

Generally the classification results based on covariance features are less smooth than those determined from shape distributions. *Roof*[s] could perform best at 8 m, but not as good as the [respective] shape distributions result. *Sealed Ground* shows a spike at 2 m, reaching a similar result to the shape distributions. *Tree*s are best detected at lower radii, but not as well as when using shape distributions, and *Vegetated Ground* is virtually undetectable.

[To further examine what the ideal neighbourhood size would be for covariance features, we followed] an argument stated [by] Demantké et al. (2011) [on dimensionality-based scale selection. According to this,] the optimum local neighbourhood size [for covariance features] may be found by minimizing the absolute value of the Shannon entropy []:

$$E_{Shannon} = -L_\lambda \cdot \ln(L_\lambda) - P_\lambda \cdot \ln(P_\lambda) - S_\lambda \cdot \ln(S_\lambda) \tag{3.6}$$

For radii below 16 m the class-wise mean Shannon entropy [of the given data set], ignoring ill-defined values due to $\lambda_i = 0$, is shown in Figure [3.8]. For trees, there is no minimum to be found, whereas for [*Roof* and *Sealed Ground*, which are distinguished in their planarity,] a slight minimum occurs at a neighbourhood radius of 0.5 m. Due to the limited point density of the data [the mean entropy decreases further for smaller neighbourhoods, as there are only few returns within those neighbourhoods (so there is little entropy). But precisely due to this lack of returns] no feature separation [could] be achieved on this scale [in the class-wise binary classification experiments reported in Figure 3.7].

Discussion

Given these experimental findings, we deduce that covariance features are not ideally suited to describe the examined classes of *Roof*, *Tree*, *Vegetated Ground*, and *Sealed Ground* in ALS data of the given point density. Both the neighbourhood scale studies shown in Figure 3.7 and the minimization of the class-wise mean Shannon entropy shown in Figure 3.8 indicate, that the best neighbourhood size for these features would be relatively small, possibly around or below 0.5 m. In neighbourhoods that small however, only relatively few neighbouring returns ($4\,{}^{\text{pts}}/_{\text{m}^2} \cdot \pi \cdot (0.5\,\text{m})^2 \approx 3\,\text{pts}$) are captured to describe the geometry around one given return. Shape distributions on the other hand seem to provide a similar if not superior characterization of geometric structures over a wider range of big-

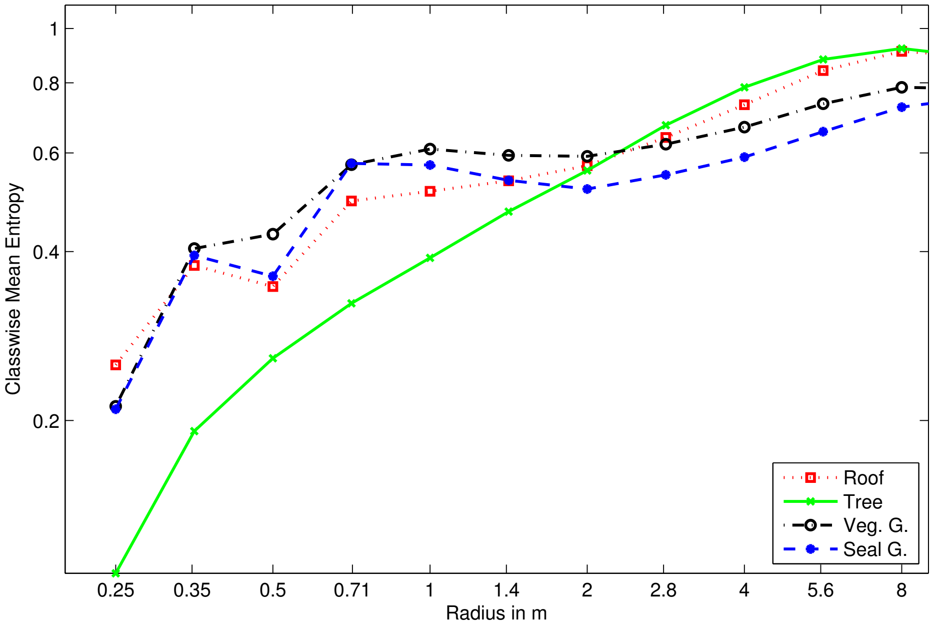


Fig. 3.8: Class-wise mean Shannon entropy calculated according to Equation [3.6] for different neighbourhood sizes. [This indicates, that the optimum scale of covariance features is typically small, around 0.50 m.]

ger neighbourhood size scales, as shown in Figure 3.6, peaking at similar sizes around 2.0 − 2.8 m, where the neighbourhoods' geometry is described by more $(4\,\mathrm{pts}/\mathrm{m}^2 \cdot \pi \cdot (2.8\,\mathrm{m})^2 \approx 99\,\mathrm{pts})$ returns. The strength of shape distributions compared to covariance features therefore is closely tied to the point density of the given data, as shape distributions are rather designed to describe more varied geometrical structures (which occur in larger neighbourhood sizes) rather than locally homogeneous distributions (which occur in smaller neighbourhood sizes), for which covariance features are well suited.

This finding is also linked to an effect described later by Dittrich et al. (2017). They compared the values of covariance features for simulated, discretized data with their theoretical value in certain geometric distributions, such as planes, edges, corners, etc.. As mentioned in Section 3.3.2.1, they found that some covariance features are subject to significant relative errors (compared to their theoretical expectation value) as soon as they are calculated from discretized data points. Also, possible noise in the data, which could stem from measurement inaccuracies or surface properties, affected some covariance features in a strong way.

3.5.1.2 Comparison of Feature Relevance

To [] evaluate [the different] features' performance independent of the used classification scheme, a filter-based feature relevance assessment is performed. The procedure follows (Weinmann et al., 2013). Seven filter-based feature relevance measures are evaluated, each resulting in a relevance rating for all elements of the feature vector. In this case, 61 feature vector elements have to be compared (5 shape distribution types with 10 feature values each and 11 covariance features). The applied score functions evaluate the relation between the values of a feature vector element for all observations and the respective class labels. Tested measures are c_χ from a χ^2 independence test, the Fisher score c_{Fisher} describing the ratio of interclass and intraclass variance, the Gini Index c_{Gini} as a statistical dispersion measure, the Information Gain measure c_{IG} revealing the dependence in terms of mutual information, the Pearson correlation coefficient c_{Pearson} derived from the degree of correlation between a feature and the class labels, the ReliefF measure c_{ReliefF} revealing the contribution of a certain feature to the separability of different classes, and the c_t measure derived from a t-test for checking how effective a feature is for separating different classes.

Results

Since all relevance measures follow different metrics, the value for relative importance was deduced from the ranking order among all feature vector elements. Afterwards, the mean of all importance values was taken for every feature vector element, resulting in a mean importance. A value of one would be achieved if a feature vector element was rated the most important feature by all relevance measures, and zero if it was always rated least important. The mean of all mean importance values belonging to one feature type group is plotted in Figure 3.9. To avoid any bias by unbalanced reference data, a subset containing 1 000 points per class is investigated.

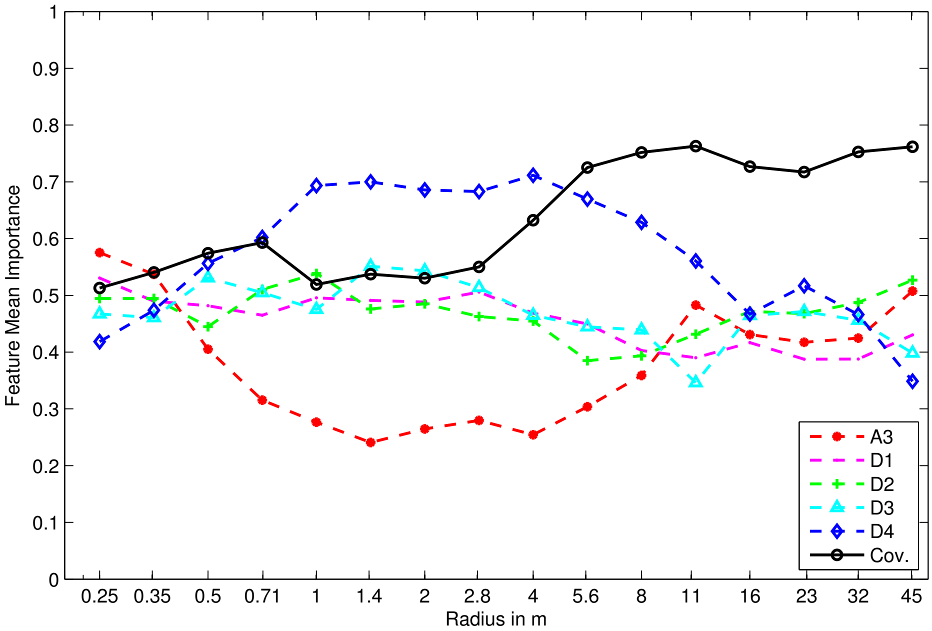


Fig. 3.9: Mean rank of mean feature relevance per feature group for different neighbourhood sizes. A3, D1, D2, D3 and D4 are the shape distribution[s], and Cov. the covariance features [].

Discussion

Comparing the five different shape distribution types, all printed as slashed lines, it is clearly seen that the angle between any three random points A3 is only weakly descriptive at those neighbourhood radii that showed the best classification performance in Section 3.6. The volume between any four random points D4 is of great importance here. Obviously the different classes in this test could be best separated by distinctive probability distributions of random volumetric measures. However, at very small scales, angular and lower dimensional measures like D1, D2 and D3 are of more importance.

As for covariance features, a different behaviour can be observed. Below ~3 m the importance is roughly the same as for the D1, D2 and D3 shape distributions. The slight peak in importance at 0.71 m corresponds well to the optimum neighbourhood size derived from entropy measures (cf. Figure 3.8). For higher radii a steep increase followed by high constant importance is measured. This corresponds directly to the scale at which the performance of shape distributions decreases in the SVM classifications (cf. Figure 3.6).

3.5.1.3 Combined Results of Class-Wise Binary Classifications

As the approach with varying scales implies that distinctive neighbourhood sizes may be class-dependent, four separate classification results from different best neighbourhood sizes ([Section 3.5.1.1]) have to be considered. Combining these separate results necessitates the choice between complete labelling and higher label accuracy. Since the chosen subset of classes may not be complete, we choose only to regard elements with a [SVM-based] label probability higher than 50 % as labelled. Therefore some elements may not belong to any class. [Moreover], some points may be found belonging to two or more classes. In this case, the label probability is weighted by the quality of the respective binary classifier before choosing the maximum [score].

Results

[This combination procedure is followed] for both shape distribution and covariance feature [results]. Rejection rate (percentage of unidentified elements) and overall accuracy of both results each combined from four binary classifications are shown in Table 3.3.

For a more extensive analysis of the different classes' performance, the resulting confusion matrices as well as completeness, correctness, and quality are shown in Table 3.4 and 3.5.

	rejection rate	OA
shape distributions	38.4 %	75.6 %
covariance features	56.6 %	63.1 %

Table 3.3: Rejection rate and overall accuracy (OA) for [combined binary] classifications based on shape distributions and covariance features.

[In addition to the quantitative evaluation given by the metrics above,] Figures 3.10 and 3.11 depict shape distribution classification results as coloured point clouds for qualitative analysis in particular areas. Both the gable roof and tree visible in Figure 3.10 are generally well classified. Only minor errors occur at the ridge of the roof, where some points are mistaken for vegetated ground, and at the rim of the roof, where some points are misclassified as tree.

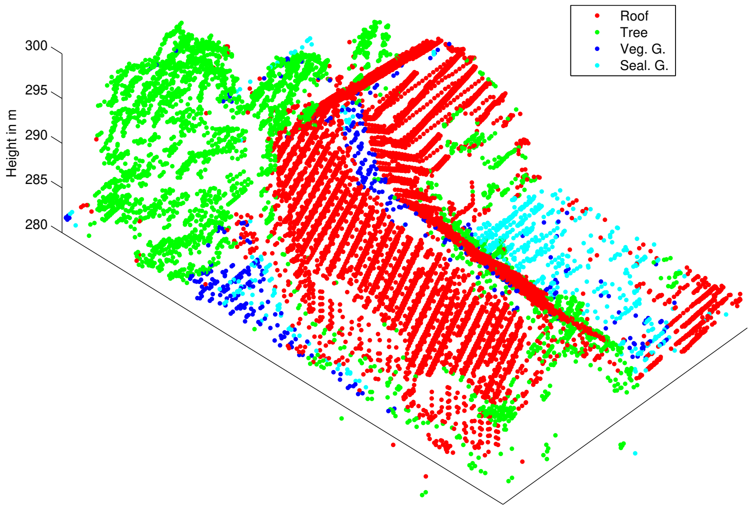


Fig. 3.10: Gable roof and adjacent trees, generally well labelled [by the combined binary classification results of shape distributions on four class-specific neighbourhood sizes.]

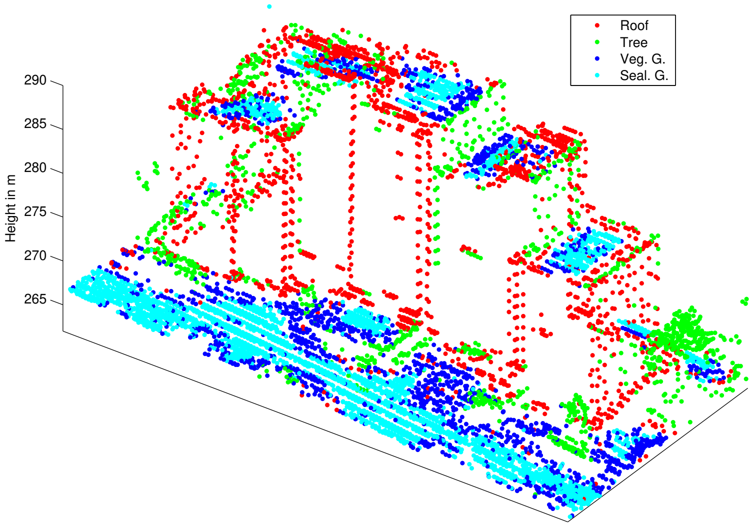


Fig. 3.11: High-rise buildings[, labelled by the combined binary classification results of shape distributions or four class-specific neighbourhood sizes.] Without the use of further features, flat roofs are mistaken as ground [due to the rotation-invariance of shape distributions.]

known \ pred.	*Roof*	*Tree*	*Veg. G.*	*Seal. G.*	Comp.
Roof	77 530	10 094	4 036	6 749	78.8 %
Tree	1 052	61 474	80	47	98.1 %
Veg. G.	4 983	4 271	8 576	11 493	29.2 %
Seal. G.	6 994	3 771	8 634	45 032	69.9 %
Corr.	85.6 %	77.2 %	40.2 %	71.1 %	
Quality	69.6 %	76.1 %	20.4 %	54.4 %	

Table 3.4: Confusion matrix of combined classification results using shape distributions [on four scales optimal for one class each,] ignoring all cases in which an element could not be detected in any of the classes.

known \ pred.	*Roof*	*Tree*	*Veg. G.*	*Seal. G.*	Comp.
Roof	31 544	37 225	0	10 993	39.5 %
Tree	1 040	46 941	0	78	97.7 %
Veg. G.	2 867	7 073	0	13 101	0 %
Seal. G.	4 699	10 306	0	41 953	73.7 %
Corr.	78.6 %	46.2 %	-	63.4 %	
Quality	35.7 %	45.7 %	-	51.7 %	

Table 3.5: Confusion matrix of combined classification results using [covariance features on four scales optimal for one class each,] ignoring all cases in which an element could not be detected in any of the classes.

Discussion

[With regard to rejection rate and overall accuracy (OA),] the shape distributions [] obviously outperform the covariance features, since the rejection rate is significantly lower whilst the overall accuracy is increased. Not only can more elements be identified, but also more of these elements are identified correctly.

[With regard to completeness, correctness and quality, the] most significant increase is observed for the detection of [] *Roofs*, where quality almost doubles [for shape distributions compared to covariance features], mainly due to an increase in completeness. The significant quality increase for *Trees* is mainly due to increased correctness, whereas *Sealed Ground* is detected with comparable quality. *Vegetated Ground* lacks a comparison, as it cannot be detected at all by covariance features.

[The qualitative analysis according to Figures 3.10 and 3.11] is in good agreement with the findings of Figure 3.6, indicating that trees are generally covered by great completeness, but correctness is lacking. The high-rise buildings depicted in Figure 3.11 show a misclassification of flat roofs as sealed and vegetated ground. This is not surprising, as shape distributions do not incorporate knowledge about a predominant direction. Therefore, a flat roof and its edge to the ground have the very same characteristics as flat ground and the adjoining edge of a house. Except for the existing confusion between *Sealed* and *Vegetated Ground*, those examples explain all main off-diagonal contributions to the confusion matrix.

In conclusion, the findings presented above show, that shape distributions have a huge potential as features for ALS point cloud classification. The most frequent misclassification cases pointed out above should be overcome when combining shape distributions with other geometric features including verticality information, which is contained in the geometric 3D properties described in Section 3.3.2.2, or normalized height above ground, as described in Section 3.3.4.

3.5.2 Analysis of Class Separability by Simple Classifiers

Neighbourhoods

- $\mathcal{N}_c$ (1 m, 2 m, 3 m & 5 m)
- $\mathcal{N}_{k,\mathrm{opt}}$

Features

- covariance features
- geom. 3D properties
- shape distributions
- absolute height

Classification

- balanced training $^{10\,000\,\mathrm{pts}}/_{\mathrm{class}}$
- multinomial classifications using NN, LDA, QDA & RF

Since we focus on feature development, it is interesting to see if those features already enable a class separation on a simple level, such as by instance-based or probabilistic classifiers, or if the boundaries between classes are so intertwined that they must rely on more complex algorithms so as to be separated. Therefore we perform a test comparing the results of a NN classifier as an example of instance learning, LDA and QDA as examples of probabilistic learning and a RF classifier as an example of ensemble learning.

For this case, we use cylindrical neighbourhoods $\mathcal{N}_c$ of 1 m, 2 m, 3 m and 5 m radius respectively, and one optimized spherical neighbourhood $\mathcal{N}_{k,\mathrm{opt}}$. For each of these neighbourhoods we calculate covariance features, geometric 3D properties and shape distributions. This amounts to a total number of 316 feature values per ALS return.

For the training phase, we take into account that an unbalanced distribution of training examples per class might have a detrimental effect on the training process (Chen et al., 2004; Criminisi and Shotton, 2013). Accordingly, we introduce a class re-balancing by randomly sampling the same number of training examples per class to obtain a reduced training set. For our experiments, a reduced training set comprising 10 000 training examples per class has proven to yield results of reasonable quality. Note that this results in a duplication of training examples for those classes represented by less than 10 000 training examples. For the RF classifier, several parameters had to be determined via a heuristic grid search. We use $N_T = 2\,000$, $n_{\min} = 1$, $n_a = 3$. The NN classifier as well as LDA and QDA do not require manually set parameters.

Results

The classification performance for all four classifiers is evaluated according to the class-wise metrics of precision and recall (cf. Table 3.6) and the F_1-score (cf. Table 3.7), as well as the overall evaluation metrics of OA, Cohen's κ coefficient (κ), mean class recall (MCR), and mean class precision (MCP) (cf. Table 3.8). The classification results obtained with all four considered classifiers are also visualized in Figure 3.12.

[The four different classifiers] clearly reveal a different behaviour. [] The classification metrics of [OA, κ, MCR and MCP] indicate that the LDA classifier achieves the best performance [(OA = 50.2%, κ = 38.3%, MCR = 49.1%, and MCP = 39.7%)] for our application.

The class-wise evaluation indicates that some classes (like *Impervious Surface*, *Roof* and *Tree*) are better recognized across all classifiers than others (like *Powerline* or *Fence / Hedge*). It is interesting that *Car*s in particular are better recognized by the LDA classifier than by other classifiers.

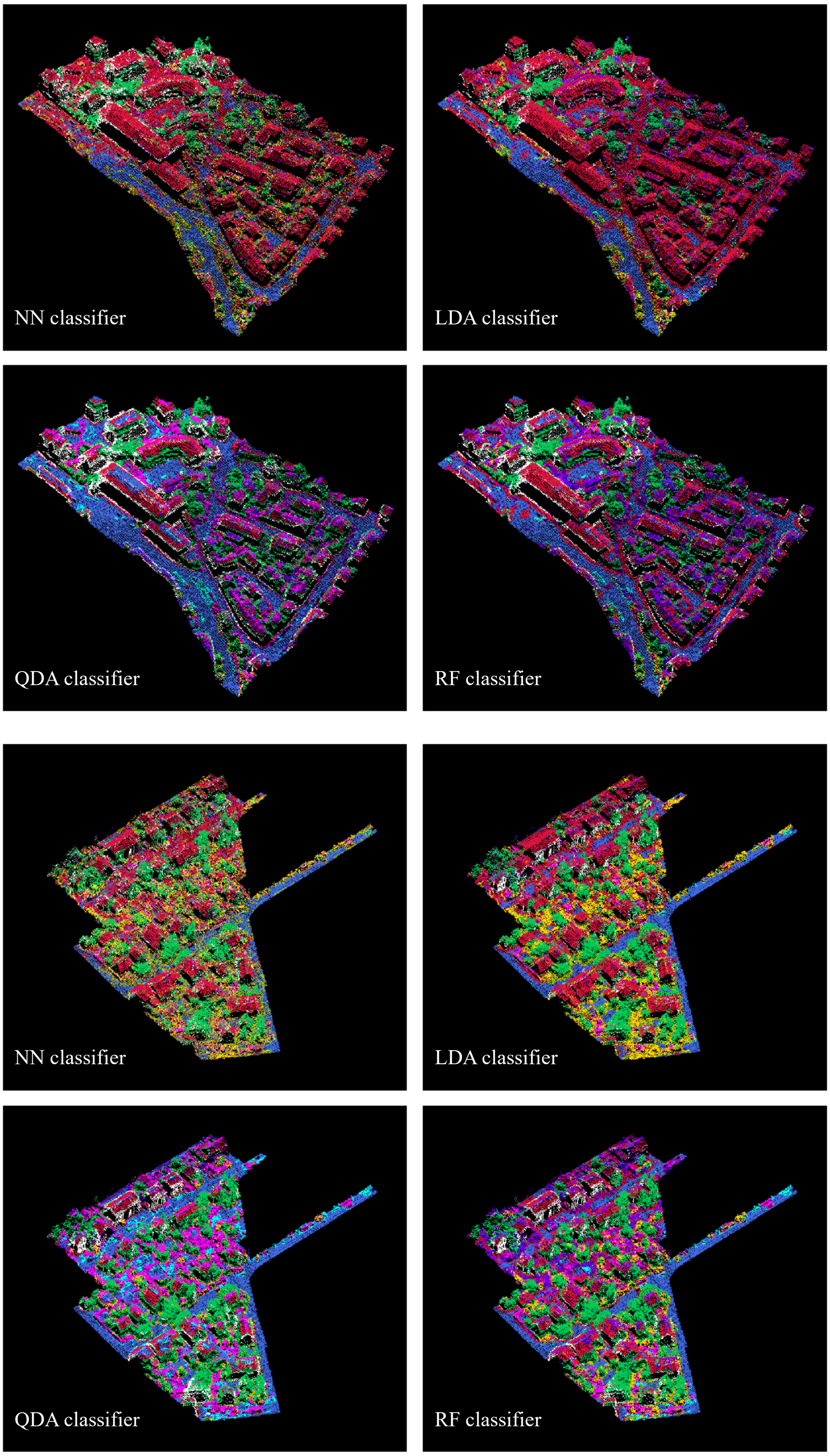


Fig. 3.12: Visualization of classification results using a NN classifier, a LDA classifier, a QDA classifier and a RF classifier respectively on Area 1 (**top**) and Area 3 (**bottom**) of the **Vaihingen** data set. For clearer discriminability among the classes, a non-natural colour encoding is chosen (*Powerline*: violet; *Low Vegetation*: yellowish green; *Impervious Surfaces*: royal blue; *Car*: cyan; *Fence / Hedge*: pink; *Roof*: crimson; *Façade*: white; *Shrub*: gold; *Tree*: emerald green)

class	NN		LDA		QDA		RF	
	prec.	rec.	prec.	rec.	prec.	rec.	prec.	rec.
Powerline	7.6	50.3	3.0	89.3	1.5	40.3	0.9	74.3
Low Vegetation	40.7	16.8	53.3	12.4	44.4	2.3	50.6	4.5
Impervious Surface	74.3	40.2	84.98	47.6	65.5	69.4	78.2	54.3
Car	12.7	11.4	31.4	28.9	6.1	35.0	13.1	22.1
Fence / Hedge	8.3	18.4	13.2	20.4	5.4	45.5	7.8	21.8
Roof	47.9	69.9	48.6	80.7	59.9	35.8	47.9	56.1
Façade	17.5	34.9	36.8	51.3	16.1	50.9	19.3	50.5
Shrub	23.6	31.1	28.3	38.4	20.9	10.2	33.6	21.6
Tree	49.0	70.1	57.5	72.8	37.3	58.4	44.4	66.6

Table 3.6: Class-wise precision and recall values (in %) for classifications by the NN, LDA, QDA classifiers, and a RF classifier on the **Vaihingen** data. Results published in a similar representation in (Blomley et al., 2016a).

class	NN	LDA	QDA	RF
Powerline	13.2	5.9	2.8	1.7
Low Vegetation	23.8	20.1	4.4	8.2
Impervious Surface	52.2	61.1	67.4	64.1
Car	12.0	30.1	10.4	16.5
Fence / Hedge	11.4	16.0	9.6	11.4
Roof	56.8	60.7	44.8	51.7
Façade	23.3	42.8	24.4	28.0
Shrub	26.9	32.6	13.7	26.3
Tree	57.7	64.2	45.5	53.3

Table 3.7: Class-wise F_1-scores (in %) for classifications by the NN, LDA, QDA classifiers, and by a RF classifier on the **Vaihingen** data. Results published in a similar representation in (Blomley et al., 2016a).

metric	NN	LDA	QDA	RF
OA	45.1	50.2	38.1	41.5
κ	32.1	38.3	27.7	30.3
MCR	38.1	49.1	38.7	41.3
MCP	31.3	39.7	28.6	32.9

Table 3.8: Overall classification results (in %) by the NN, LDA, QDA classifiers, and by a RF classifier on the **Vaihingen** data set, evaluated according to the metrics of OA, κ, MCR, and MCP. Results published in a similar representation in (Blomley et al., 2016a).

Discussion

Our results with relatively low numbers for different evaluation metrics indicate that the Vaihingen data set with a labelling with respect to nine semantic classes represents a rather challenging data set when focusing on a 3D semantic labelling. To a certain degree, this might be due to the fact that some classes might not be representatively covered in the training data – e.g. the class *Powerline* with only 546 given training examples and the class *Car* with 4 614 given training examples (cf. Table 3.1 in Section 3.4.1) – which, in turn, yields poor classification results for these classes.

Furthermore, the derived results indicate that [considering only] geometric features might not be sufficient for obtaining adequate classification results for all considered classes, since some of these classes might have a quite similar geometric behaviour, e.g. the classes *Low Vegetation*, *Fence / Hedge*, and *Shrub*. This indeed becomes visible in Figure [3.12] where particularly misclassifications among these three classes may be observed for different classifiers. Yet, also the extracted geometric features may not be optimal as some of the neighbourhoods used as the basis for feature extraction are relatively large, e.g. the cylindrical neighbourhoods with radii of 3 m and 5 m which have also been used by Niemeyer et al. (2014) and Schmidt et al. (2014). This, in turn, results in misclassifications [particularly at] those locations where the cylindrical neighbourhood includes 3D points associated to the classes *Roof* and *Impervious Surfaces*.

A closer look [at] the classification results provided in Figure [3.12] [] reveals seam effects where borders between roofs and façades or between façades and ground are largely categorized into the class *Façade*, particularly for the QDA classifier and the RF classifier. Furthermore, the QDA classifier provides the best recognition of *Impervious Surfaces*, while the classification results are rather poor for the classes *Low Vegetation*, *Fence / Hedge* and *Shrub*. In contrast, the LDA classifier provides a good recognition for the classes *Roof* and *Tree*, while problems in the separation between *Impervious Surfaces* and *Roof* become visible. [Those could be related to the 'error' (realized later, cf. Section 3.5.3) of using absolute height as a feature, and not having a normalized height feature instead.]

Overall, LDA performed well in comparison to both QDA and the RF classifier. This is unexpected, because our total number of features is relatively high compared to other approaches, which typically do not use shape distributions (which contribute 50 values per neighbourhood, whereas covariance features and 3D geometric properties only constitute 13 or 14 values per neighbourhood), and because we use a range of different neighbourhoods which multiplies the number of feature values for each return. High numbers of feature values are typically associated with a series of disadvantages:

- The 'curse of dimensionality': in a high-dimensional feature space, the sampling density is sparse, which reduces the statistical significance of patterns and typically leads to a loss of predictive power.
- An increased risk of overfitting: redundant or irrelevant features typically cause the model to memorize the training data, which results in poor generalization to unseen data.
- Sensitivity to irrelevant features: irrelevant or redundant features can increase the complexity of the decision boundary, leading to reduced generalization.
- Dimensionality mismatch: if the number of features exceeds the number of samples, models may fail to converge at all.

Generative classifiers (like k-NN, LDA, and QDA) are generally more prone to those difficulties, while discriminative classifiers such as SVMs or RFs are more robust with regard to high feature dimensionality. The unexpected relative strength of the LDA classifier in this experiment indicates that the relatively high number of features and their possible redundancy due to repeated characteristics across different neighbourhood scales do not seem to influence the classification badly.

Concerning the class-wise evaluation in Tables 3.6 and 3.7, we see that the classes *Impervious Surface*, *Roof* and *Tree* – which are geometrically characteristic within an urban environment – were detected with an acceptable accuracy. Classes of rather similar geometric appearance, such as *Low Vegetation*, *Fence / Hedge*, and *Shrub* are not appropriately assigned in the derived classification results. This could probably be improved by including further information, such as reflected intensity, return number, or number of returns as additional features.

3.5.3 Comparison and Combination of Different Feature Types, Neighbourhood Types, and Scales

After working solely on the Vaihingen data set up to this point, we study our method's performance on the GML data set A in Section 3.5.3.1. Here, we also focus on evaluating the benefits of combining features from different neighbourhood types, and scales, as well as the combination benefit of complementing feature types.

After doing so however, we realized, that the absolute height value we had been using as a feature (in legacy of adapting the geometric 3D properties from Weinmann et al. (2015), who had been working on TLS data) introduces errors in variegated terrain and causes confusion among geometrically similar classes at different normalized heights in the scene, such as flat ground and flat roofs, if they are at similar absolute height by the effect of topography. We therefore replaced absolute height by normalized height in Sections 3.5.3.2 and 3.5.3.3. Also, since the results of Section 3.5.2 did not indicate problems connected to an overall high number of feature values per entity, and since the multi-scale, multi-feature-type studies in Section 3.5.3.1 showed improvements for every added group, we extended our multi-scale, multi-feature-type approach to multiple-neighbourhood-types in Section 3.5.3.2 and 3.5.3.3.

With these changes in place, we thoroughly compared our method's performance on both data sets in parallel in Section 3.5.3.2. For a more compact overview, the reader may jump directly to this Section.

In the end, we noticed that we could achieve even better results when training our classifier with more training entities at the cost of higher duplication rates for small classes. These results are shown in Section 3.5.3.3, where we also draw a comparison to comparable work in the field.

3.5.3.1 Tests on the GML Data Set using Multi-Scale Cylindrical Neighbourhoods and an Optimized Spherical Neighbourhood

Neighbourhoods

- $\mathcal{N}_c$ (1 m, 2 m, 3 m & 5 m)
- $\mathcal{N}_{k,\text{opt}}$

Features

- covariance features
- geom. 3D properties
- shape distributions
- absolute height

Classification

- balanced training $1\,000\ ^{\text{pts}}/_{\text{class}}$
- multinomial RF

In this section, we aim for a comprehensive analysis of the gains in classification performance either by combining complementary feature types or by combining different neighbourhood types and scales. As feature types, we use a group of metrical features (namely covariance features, 3D geometric properties and the absolute height of the measured return) and shape distributions as a sampled feature type. As neighbourhoods we used cylindrical neighbourhoods $\mathcal{N}_\text{c}$ with different radii of 1 m, 2 m, 3 m and 5 m respectively, and a locally optimized spherical neighbourhood $\mathcal{N}_{k,\text{opt}}$. Therefore a total of 21 combinations are tested and analysed. We used the GML data set A described in Section 3.4.2 as an application scenario. Using the RF implementation available with Liaw and Wiener (2002), we trained RF classifiers for each combination using a balanced subset of 1 000 training entities per class, chosen randomly from the training scene. The number of trees N_T for the RF was determined via a standard grid search testing 50, 100, 200 or 500 trees, where classifications using 200 trees yielded the best results.

Results

First, we focus on a classification based on [the] distinct feature [types for all individual neighbourhoods and two combinations ($\mathcal{N}_{\mathrm{c,all}}$ and $\mathcal{N}_{\mathrm{all}}$)] and, subsequently, we consider them in combination for the classification task. In order to compare the classification results obtained with the different approaches on point-level, we consider a variety of measures for evaluation on the respective test data: (*i*) Cohen's κ coefficient (κ) [and] (*ii*) overall accuracy (OA) [cf. Table 3.9], (*iii*) mean class recall (MCR) and (*iv*) mean class precision (MCP) [cf. Table 3.10]. Furthermore, we involve different measures for class-wise evaluation: (*i*) recall, (*ii*) precision and (*iii*) F_1-score [cf. Table 3.11]. Figure 3.13 shows a visual representation of the classification result with both features types and all neighbourhoods $\mathcal{N}_{\mathrm{all}}$ combined.

General observations of the presented results reveal that the combination of [metrical features] and [sampled features] produces improved classification results compared to both separate groups. Furthermore, it may be observed that features extracted from multi-scale neighbourhoods of the same type tend to lead to improved classification results. The combination of features derived from multi-scale, multi-type neighbourhoods does, in general, even lead to further improved classification results compared to features derived from multi-scale neighbourhoods of the same type.

A more detailed view on the derived results reveals that, when considering the evaluation among the single scales, there is no [universal] best neighbourhood scale among the cylindrical neighbourhoods $\mathcal{N}_{\mathrm{c,1m}}$, $\mathcal{N}_{\mathrm{c,2m}}$, $\mathcal{N}_{\mathrm{c,3m}}$ and $\mathcal{N}_{\mathrm{c,5m}}$. For [the metrical features], $\mathcal{N}_{\mathrm{c,5m}}$ [] perform[s] well in class separability [and in overall accuracy (Table 3.9)] . For shape distributions, $\mathcal{N}_{\mathrm{c,3m}}$ [] show[s] the best results. The class-wise classification results reveal that different classes favour a different neighbourhood size [(Table 3.11). Especially small structures such as the class *Car* favour smaller cylindrical neighbourhoods among the individual neighbourhood results.]

When considering features extracted from multiple scales, we [] observe that the features extracted from multiple cylindrical neighbourhoods $\mathcal{N}_{\mathrm{c,all}}$ are usually similar to or slightly improved over the best classification result from the individual neighbourhoods (Table [3.9]). However, since there is a large variation among which neighbourhood size performs best [], it seems worthwhile to [include] all scales.

When considering multi-scale, multi-type neighbourhoods, we may state that – even though the spherical neighbourhood selected via eigenentropy-based scale selection does not always perform very well on its own – there is usually a notable performance increase for the multi-type combination $\mathcal{N}_{\mathrm{all}}$ over [the combination of cylindrical neighbourhoods only $\mathcal{N}_{\mathrm{c,all}}$].

When inspecting the visual representation of the classification result in Figure 3.13 the most notable observation is a frequent misclassification between *Ground* and *Building*. Both classes are similar in geometry, but usually distinct in height above ground. In this experiment, there was however no approximation of scene topography leading to estimated values of height above ground, but only the absolute height values, which are no indicator of height above ground in this particular scene due to the landscape's topography. Hence, in the following experiments, we introduced the normalized height feature.

$\mathcal{N}$	metrical features		sampled features		combination	
	κ	OA	κ	OA	κ	OA
$\mathcal{N}_{\mathrm{c,1m}}$	25.2	50.2	34.1	56.6	35.7	57.9
$\mathcal{N}_{\mathrm{c,2m}}$	30.6	55.5	42.4	64.1	44.5	65.7
$\mathcal{N}_{\mathrm{c,3m}}$	32.2	57.7	44.8	66.1	48.6	69.1
$\mathcal{N}_{\mathrm{c,5m}}$	41.6	65.1	42.1	63.9	50.4	70.8
$\mathcal{N}_{k,\mathrm{opt}}$	23.3	41.6	32.5	51.8	28.3	47.1
$\mathcal{N}_{\mathrm{c,all}}$	35.5	60.4	48.4	68.9	49.6	69.9
$\mathcal{N}_{\mathrm{all}}$	57.3	74.3	53.9	72.2	61.2	76.8

Table 3.9: κ and OA (in %) for different neighbourhood definitions and different feature sets on the **GML** data set A. Results published in (Blomley et al., 2016b).

$\mathcal{N}$	metrical features		sampled features		combination	
	MCP	MCR	MCP	MCR	MCP	MCR
$\mathcal{N}_{\mathrm{c,1m}}$	30.9	45.9	36.1	53.0	35.6	52.7
$\mathcal{N}_{\mathrm{c,2m}}$	32.3	49.5	38.2	58.8	37.9	58.0
$\mathcal{N}_{\mathrm{c,3m}}$	32.5	49.3	38.7	59.5	38.9	60.3
$\mathcal{N}_{\mathrm{c,5m}}$	34.3	46.5	38.4	56.6	39.3	58.2
$\mathcal{N}_{k,\mathrm{opt}}$	35.8	36.4	37.5	47.4	37.1	39.2
$\mathcal{N}_{\mathrm{c,all}}$	33.9	52.0	39.8	64.8	39.3	65.0
$\mathcal{N}_{\mathrm{all}}$	41.6	63.5	41.4	68.0	43.6	70.2

Table 3.10: MCP and MCR (in %) for different neighbourhood definitions and different feature sets on the **GML** data set A. Results published in (Blomley et al., 2016b).

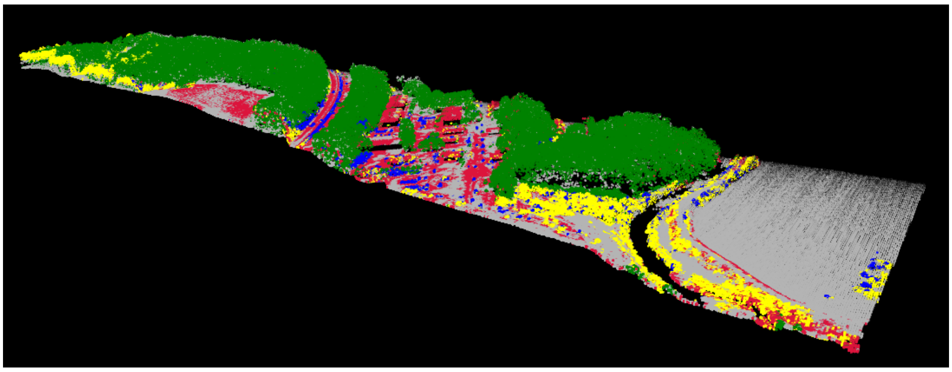


Fig. 3.13: Visualization of classification results for the **GML** data set A using $\mathcal{N}_c$ and $\mathcal{N}_{k,\mathrm{opt}}$ with covariance features, geometric 3D properties, shape distributions and absolute height. The colour encoding refers to the classes *Ground* (grey), *Building* (red), *Car* (blue), *Tree* (dark green), and *Low Vegetation* (bright green).

class	$\mathcal{N}$	metrical features			sampled features			combination		
		prec.	rec.	F_1	prec.	rec.	F_1	prec.	rec.	F_1
Ground	$\mathcal{N}_{\text{c,1m}}$	62.5	50.8	56.1	87.5	34.0	49.0	82.8	43.2	56.7
	$\mathcal{N}_{\text{c,2m}}$	64.0	51.3	57.0	92.4	40.3	56.1	88.9	46.1	60.8
	$\mathcal{N}_{\text{c,3m}}$	62.8	55.4	58.9	93.5	42.5	58.4	89.9	49.6	64.0
	$\mathcal{N}_{\text{c,5m}}$	70.7	57.1	63.2	94.3	37.5	53.7	91.1	48.9	63.7
	$\mathcal{N}_{k,\text{opt}}$	91.5	43.1	58.6	89.1	40.1	55.3	93.3	44.4	60.1
	$\mathcal{N}_{\text{c,all}}$	63.6	57.1	60.2	94.0	46.4	62.1	82.4	55.6	66.4
	$\mathcal{N}_{\text{all}}$	85.8	64.6	73.7	95.1	52.4	67.5	93.9	62.9	75.3
Building	$\mathcal{N}_{\text{c,1m}}$	6.8	50.8	12.0	5.3	43.2	9.5	6.8	50.6	12.0
	$\mathcal{N}_{\text{c,2m}}$	6.5	50.8	11.5	6.3	42.5	10.9	7.4	48.6	12.9
	$\mathcal{N}_{\text{c,3m}}$	8.0	53.2	13.9	6.7	38.3	11.4	9.0	48.1	15.1
	$\mathcal{N}_{\text{c,5m}}$	10.2	56.4	17.2	7.5	39.3	12.6	10.8	53.8	18.0
	$\mathcal{N}_{k,\text{opt}}$	7.1	62.9	12.8	5.7	48.9	10.3	7.0	60.1	12.5
	$\mathcal{N}_{\text{c,all}}$	9.7	60.5	16.7	8.1	45.7	13.8	10.0	53.8	16.8
	$\mathcal{N}_{\text{all}}$	11.6	63.8	19.7	9.6	53.6	16.2	11.6	64.8	19.7
Car	$\mathcal{N}_{\text{c,1m}}$	2.8	60.0	5.4	3.1	62.0	5.9	3.4	69.0	6.5
	$\mathcal{N}_{\text{c,2m}}$	4.7	56.5	8.6	3.9	66.8	7.3	5.2	68.2	9.6
	$\mathcal{N}_{\text{c,3m}}$	4.7	49.9	8.6	3.3	63.3	6.3	5.5	62.5	10.0
	$\mathcal{N}_{\text{c,5m}}$	2.3	18.9	4.2	2.4	52.5	4.5	4.1	38.6	7.4
	$\mathcal{N}_{k,\text{opt}}$	0.6	33.4	1.2	1.2	45.8	2.3	0.8	31.6	1.5
	$\mathcal{N}_{\text{c,all}}$	6.9	37.6	11.7	6.6	73.5	12.1	11.2	71.6	19.3
	$\mathcal{N}_{\text{all}}$	11.5	45.2	18.3	7.8	74.3	14.0	12.8	71.0	21.7
Tree	$\mathcal{N}_{\text{c,1m}}$	81.1	50.1	62.0	81.3	75.9	78.5	82.8	70.6	76.2
	$\mathcal{N}_{\text{c,2m}}$	83.6	59.5	69.5	82.3	84.6	83.4	83.6	82.8	83.2
	$\mathcal{N}_{\text{c,3m}}$	83.7	60.2	70.1	82.4	86.8	84.5	83.8	86.3	85.0
	$\mathcal{N}_{\text{c,5m}}$	84.1	72.9	78.1	82.1	86.6	84.3	83.1	89.8	86.3
	$\mathcal{N}_{k,\text{opt}}$	79.7	40.1	53.4	89.3	61.8	73.1	84.1	49.5	62.3
	$\mathcal{N}_{\text{c,all}}$	84.9	63.6	72.7	82.9	88.4	85.5	85.8	82.5	84.1
	$\mathcal{N}_{\text{all}}$	93.2	83.2	87.9	87.4	89.3	88.3	92.8	88.9	90.8
Low Vegetation	$\mathcal{N}_{\text{c,1m}}$	1.4	17.8	2.5	3.5	49.7	6.5	2.2	30.2	4.1
	$\mathcal{N}_{\text{c,2m}}$	2.9	29.6	5.3	6.2	59.7	11.2	4.5	44.1	8.1
	$\mathcal{N}_{\text{c,3m}}$	3.3	28.0	5.9	7.6	66.6	13.6	6.1	55.1	11.0
	$\mathcal{N}_{\text{c,5m}}$	4.1	27.3	7.1	5.7	66.9	10.4	7.3	59.9	13.0
	$\mathcal{N}_{k,\text{opt}}$	0.1	2.6	0.2	2.2	40.6	4.2	0.5	10.4	0.9
	$\mathcal{N}_{\text{c,all}}$	4.6	41.2	8.3	7.6	70.3	13.7	7.1	61.7	12.7
	$\mathcal{N}_{\text{all}}$	6.2	60.7	11.2	7.3	70.5	13.1	6.9	63.1	12.5

Table 3.11: Class-wise precision, recall and F_1 (in %) for different neighbourhood definitions and different feature sets on the **GML** data set A. Results published in similar presentation in (Blomley et al., 2016b).

Discussion

We realized later, that the majority of misclassifications in the results presented comes from confusion among the geometrically similar classes of (flat) ground and (flat) roofs, as they occur to a large proportion of the training data in this set, as can be seen to the left in the top image in Figure 3.5. These misclassifications come from the use of absolute height as a feature value. This feature came along when we adapted the geometric 3D properties from Weinmann et al. (2015), who had been working with TLS data. In variegated terrain, which is more frequent in larger-scale ALS data, a normalized height value would be preferable to indicate the height above ground at a certain point of the topography.

However, the conclusions drawn with the given results are still relevant. The results accomplished [] are [generally] comparable to th[e] results of existing research. The most important comparison is to [the work of] Shapovalov et al. (2010), where the same [] data set[has] been used as well as a combination of metrical features and distribution features. While our methodology focuses on [a] characterisation of 3D points via feature extraction from local neighbourhoods of different scale and type, the methodology presented [by] Shapovalov et al. (2010) focuses on the use of non-Associative Markov Networks [which perform] a contextual classification.

Shapovalov et al. (2010) use 7 metrical features which describe the local geometry of the point cloud in a similar way to the covariance features used in our approach, as they are described in the work of Munoz et al. (2008). Furthermore, they use 27 sampled feature values deduced from spin images (Johnson and Hebert, 1999), 27 sampled feature values deduced from angular spin images (Endres et al., 2009), and 7 sampled feature values corresponding to a height histogram of the local neighbourhood. While this collection of feature values is different from the ones implemented in our approach, it is still comparable as they contain both metrical and sampled features. However it is important to note, that those features are not computed per point for local neighbourhoods, but for segments produced in the first step of their contextual classification procedure.

As our approach only performs point-wise individual classification, we expect that [the] results of the contextual classification may [not] be matched. Comparing values of recall and precision [shown in Table 3.12], we find that *Ground* performs similar[ly] ([] our inferior recall values are compensated for by a higher precision []), *Buildings* perform slightly better in recall, but worse in precision [], *Car* [] is detected with much higher recall, but lower precision, *Tree* is generally comparable [] and *Low Vegetation* again shows higher recall, but lower precision values. Overall, the comparison of the combined F_1-score turns out as expected: *Ground* and *Building* suffer from the confusion due to the lack of normalized height information, while the classification results for the remaining classes are comparable and show a slight advantage of the contextual classification.

Furthermore, it is interesting to compare the results from the individual $\mathcal{N}_{\mathrm{c}}$ neighbourhood scales to the findings for shape distributions in Section 3.5.1.1, conducted on the Vaihingen data set. There, the class-specific studies of classification performance across different cylinder radii indicated that radii of 1-2 m are suitable for *Building* and *Tree*, while slightly larger radii of about 3 m are more suited for *Ground* and *Low Vegetation*. [The] F_1-scores for the single-scale neighbourhoods $\mathcal{N}_{\mathrm{c,1m}}$, $\mathcal{N}_{\mathrm{c,2m}}$, $\mathcal{N}_{\mathrm{c,3m}}$ and $\mathcal{N}_{\mathrm{c,5m}}$ [in this experiment] show[] best results for *Ground* and *Low Vegetation* at radii of 3m[, which is in accordance to the findings on the Vaihingen data set,] but best results for *Building* and *Tree* at radii of 3 to 5m [], which [is larger than those values performing well on the Vaihingen data set. This could be due to the fact that the landscape in the GML data set A is more open than the dense urban scene of the Vaihingen data set.]

> A comparison of the classification results derived for cylindrical single-scale neighbourhoods and multi-scale neighbourhoods of the same (cylindrical) type reveals that the behaviour of the local 3D structure across different scales provides information which is relevant for the classification task. This becomes visible in improved classification results for multi-scale neighbourhoods of the same type. Furthermore, we [see] that the different neighbourhood types capture complementary information about the local 3D structure. This clearly becomes visible in the improved classification results obtained for multi-scale, multi-type neighbourhoods in comparison to multi-scale neighbourhoods of the same type. Despite the weak performance of the $\mathcal{N}_{k,\text{opt}}$ neighbourhood (which has originally been developed for MLS data) on its own, its combination with the features from cylindrical neighbourhoods provides a significant improvement over the combined result of all cylindrical neighbourhoods $\mathcal{N}_{\text{c,all}}$.

class	this method $\mathcal{N}_{\text{all}}$			Shapovalov 2010		
	prec.	rec.	F_1	prec.	rec.	F_1
Ground	93.9	62.9	74.6	89.8	96.2	92.7
Building	11.6	64.8	19.7	86.8	58.5	69.9
Car	12.8	71.0	21.7	37.0	16.1	22.4
Tree	92.8	88.9	90.8	92.3	99.7	95.9
Low Vegetation	6.9	63.1	12.5	71.6	8.9	15.8
OA			76.8			N.A.
$\bar{F}_1$			43.9			59.3

Table 3.12: Comparison of class-wise precision, recall, and F_1 values (in %), alongside with OA and $\bar{F}_1$ (in %) of our results on the **GML** data set A in Section 3.5.3.1 to those achieved by Shapovalov et al. (2010).

3.5.3.2 Tests on the GML A and Vaihingen Data Sets Using Multi-Scale Cylindrical & Spherical as well as Optimized Neighbourhoods and the Normalized Height Feature

Neighbourhoods

- $\mathcal{N}_c$ (1 m, 2 m, 3 m & 5 m)
- $\mathcal{N}_s$ (1 m, 2 m, 3 m & 5 m)
- $\mathcal{N}_{k,\text{opt}}$
- horizontal binning for normalized height

Features

- covariance features
- geom. 3D properties
- shape distributions
- normalized height

Classification

- balanced training 10 000 pts/class
- multinomial RF

The first multi-scale, multi-neighbourhood-type and multi-feature-type results presented in Section 3.5.3.1 were promising in a way that they yielded results comparable to existing research in the field, even though we focused on individual point classification without spatial regularization in the classification process. However, our detailed analysis revealed critical points for further improvement. Therefore, in this section, we aim to improve our method by the following aspects:

- We suspected that the huge influence of topography on the absolute height could explain the frequent misclassifications between *Building* and (sealed) *Ground*, since those man-made surfaces otherwise share a similar geometry. Therefore we implemented a rough estimation of the scene's topography to produce a normalized height above ground feature.

- In our analysis, we found that every additional neighbourhood and feature type further improved the classification result. Hence in this section, we decided to add features from spherical neighbourhoods from different scales ($\mathcal{N}_s$ (1 m, 2 m, 3 m & 5 m)) too.

- As we noticed that similar classes behaved differently across the neighbourhood scales in the GML data set A results in Section 3.5.3.1 compared to earlier results for the Vaihingen data set in Section 3.5.1.1, we decided to evaluate our method on both data sets in this section.

Therefore, in these experiments, we used features from both cylindrical and spherical neighbourhoods ($\mathcal{N}_\mathrm{c}$ and $\mathcal{N}_\mathrm{s}$ with radii of 1 m, 2 m, 3 m and 5 m each) and a spherical neighbourhood of locally optimized scale $\mathcal{N}_{\mathrm{k,opt}}$. Since the k-optimized neighbourhood is of spherical shape too, we include it in $\mathcal{N}_{\mathrm{s,all}}$. For each of these neighbourhoods, we calculated covariance features, geometric 3D properties and shape distributions. Furthermore, we implemented an approximation of the scene's topography to deduce a point-wise normalized height feature as described in Section 3.3.4. This feature does not rely on the chosen neighbourhood around the point of interest. However, the rasterization of the point cloud on a large sampling distance (20 m) for minima calculation and subsequent interpolation on a fine sampling distance (0.5 m) does give some lateral 'neighbourhood effect' for this feature's calculation. We trained multinomial random forest classifiers on a balanced subset containing 10 000 training entities of each class (chosen randomly, resampled if necessary).

Results

The achieved values for the global evaluation metrics represented by OA and $\bar{F}_1$ are provided in Table [3.15] for the Vaihingen data set and the GML data set A. It can be observed that the combination of features extracted from all neighbourhoods yields the best classification results. [This confirms the respective finding in Section 3.5.3.1.] For the combined cylindrical neighbourhoods [($\mathcal{N}_{\mathrm{c,all}}$)], the combined spherical neighbourhoods [($\mathcal{N}_{\mathrm{s,all}}$)], and the combination of all defined neighbourhoods ($\mathcal{N}_{\mathrm{all}}$), the class-wise evaluation metrics of recall, precision and F_1-score are provided in Table [3.13] for the Vaihingen data set and in Table [3.14] for the GML data set A. For the Vaihingen data set, it can be observed that the classes *Impervious Surfaces*, *Roof*, and *Tree* can be well-detected, whereas particularly the classes *Powerline* and *Fence / Hedge* are not appropriately identified. For the GML dataset A, the classes *Ground* and *Tree* can be well-detected, whereas particularly the classes *Car* and *Low Vegetation* are not appropriately identified. The classification results relying on the use of all defined neighbourhoods ($\mathcal{N}_{\mathrm{all}}$) are visualized in Figure [3.14] for the Vaihingen data set and in Figure [3.15] for the GML data set A.

Discussion

A comparison of the derived classification results with the ones presented in Section 3.5.3.1 on the GML data set A reveals a significant gain. We can observe an improvement $\geq 10\%$ in OA and $\bar{F}_1$ as a result from including the normalized height feature and the additional spherical neighbourhoods. Table 3.17 summarizes both the overall measures and the class-wise F_1 results. Both the classes *Roof* and *Ground* experience a significant gain, which is likely attributed to the normalized height feature, while there is additional improvement across all classes, which could be attributed either to the benefits of the normalized height feature or to the broadening of the neighbourhoods considered.

The visualization in Figure 3.15 shows that some slight errors remained in the classification result, such as small patches labelled as *Building* within an other-

	class	rec.	prec.	F_1	OA	$\bar{F}_1$
$\mathcal{N}_{\text{c,all}}$	*Powerline*	68.7	3.7	7.1	62.2	45.2
	Low Vegetation	49.2	62.0	54.9		
	Imp. Surfaces	72.4	83.0	77.4		
	Car	51.2	27.2	35.5		
	Fence / Hedge	23.4	12.7	16.5		
	Roof	66.2	84.5	74.2		
	Façade	49.2	32.0	38.7		
	Shrub	51.7	27.9	36.2		
	Tree	72.0	61.0	66.1		
$\mathcal{N}_{\text{s,all}}$	*Powerline*	88.5	22.5	35.8	67.4	51.9
	Low Vegetation	52.6	68.9	59.7		
	Imp. Surfaces	78.2	84.5	81.2		
	Car	58.0	33.1	42.1		
	Fence / Hedge	23.9	15.2	18.6		
	Roof	73.8	86.4	79.6		
	Façade	59.0	28.6	38.5		
	Shrub	59.3	32.7	42.2		
	Tree	73.1	65.8	69.3		
$\mathcal{N}_{\text{all}}$	*Powerline*	92.0	19.5	32.1	68.1	52.6
	Low Vegetation	50.7	67.5	57.9		
	Imp. Surfaces	77.6	82.7	80.0		
	Car	57.5	35.7	44.1		
	Fence / Hedge	23.0	14.1	17.5		
	Roof	77.8	86.3	81.8		
	Façade	58.6	39.9	47.5		
	Shrub	58.4	32.2	41.5		
	Tree	75.4	66.9	70.9		

Table 3.13: Class-wise recall, precision and F_1-score (in %) as well as OA and $[\bar{F}_1]$ (in %) for the **Vaihingen** data set.

	class	rec.	prec.	F_1	OA	$\bar{F}_1$
$\mathcal{N}_{\text{c,all}}$	*Ground*	84.0	94.8	89.1	87.6	53.3
	Building	58.1	36.7	45.0		
	Car	71.4	12.7	21.5		
	Tree	92.0	98.4	95.1		
	Low Vegetation	73.8	8.9	15.9		
$\mathcal{N}_{\text{s,all}}$	*Ground*	84.0	98.6	90.7	88.3	57.1
	Building	74.9	47.8	58.4		
	Car	82.6	15.8	26.5		
	Tree	92.4	98.6	95.4		
	Low Vegetation	82.1	8.0	14.6		
$\mathcal{N}_{\text{all}}$	*Ground*	86.3	97.5	91.6	90.5	58.5
	Building	73.7	47.2	57.5		
	Car	76.2	17.2	28.1		
	Tree	94.9	98.7	96.8		
	Low Vegetation	76.0	10.8	18.8		

Table 3.14: Class-wise recall, precision and F_1-score (in %) as well as OA and $[\bar{F}_1]$ (in %) for the **GML** data set A.

	Vaihingen data set		GML data set A	
$\mathcal{N}$	OA	$\bar{F}_1$	OA	$\bar{F}_1$
$\mathcal{N}_{\mathrm{c,1m}}$	56.5	40.3	81.0	45.4
$\mathcal{N}_{\mathrm{c,2m}}$	57.9	41.3	82.7	46.8
$\mathcal{N}_{\mathrm{c,3m}}$	54.4	37.3	84.3	47.7
$\mathcal{N}_{\mathrm{c,5m}}$	52.9	34.9	86.6	49.3
$\mathcal{N}_{\mathrm{s,1m}}$	60.4	42.8	78.4	45.2
$\mathcal{N}_{\mathrm{s,2m}}$	62.4	44.5	81.5	48.3
$\mathcal{N}_{\mathrm{s,3m}}$	60.3	42.9	84.4	50.5
$\mathcal{N}_{\mathrm{s,5m}}$	55.9	37.5	86.7	51.7
$\mathcal{N}_{\mathrm{kopt}}$	61.7	43.2	83.0	49.1
$\mathcal{N}_{\mathrm{c,all}}$	62.2	45.2	87.7	53.3
$\mathcal{N}_{\mathrm{s,all}}$	67.4	51.9	88.3	57.1
$\mathcal{N}_{\mathrm{all}}$	68.1	52.6	90.5	58.5

Table 3.15: OA and mean $\bar{F}_1$-score (in %) achieved for different neighbourhood definitions on the **Vaihingen** data set and the **GML** data set A.

wise empty ground surface (which appears to be a field in the front middle) as well as the blue stripe in the middle of what is perceived as a main road to the right of the field, which is labelled as *Car* but appears to be a central reservation (maybe *Low Vegetation* or another kind of fence or barrier) on the road.

A comparison of the now improved results to those of the contextual classification achieved by Shapovalov et al. (2010) is shown in Table 3.16. Despite the fact that no value of OA is available for Shapovalov's results, they seem to be almost on par. This is remarkable, as we would expect the contextual classification to perform better than a point-wise classification.

For the Vaihingen data set, we can compare the results of this section with those from Section 3.5.2. Compared to the RF classifier results presented there, the experiments in this section have added additional neighbourhoods $\mathcal{N}_s$ (1 m, 2 m, 3 m & 5 m) and the normalized height feature. The results improve dramatically by $\geq$ 20% in OA and $\bar{F}_1$ as seen in Table 3.17. In this case it is difficult to reason

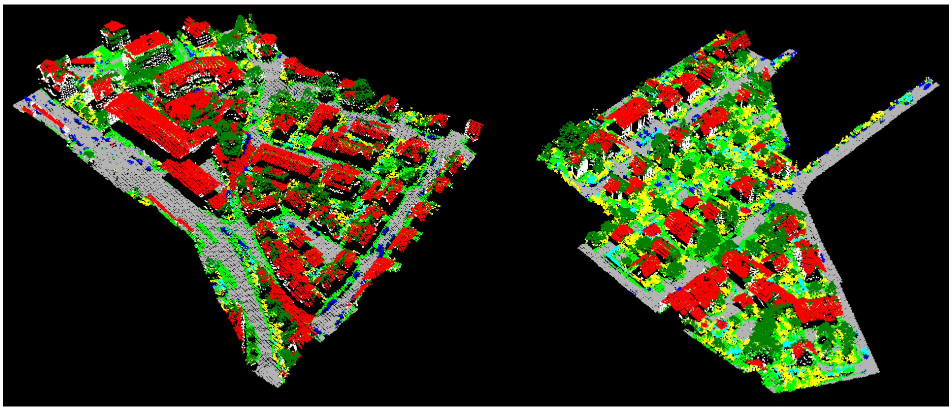


Fig. 3.14: Visualization of classification results for the **Vaihingen** data set Area 1 (**left**) and Area 3 (**right**) using $\mathcal{N}_c$, $\mathcal{N}_s$ and $\mathcal{N}_{k,\mathrm{opt}}$ with covariance features, geometric 3D properties, shape distributions, and normalized height. The colour encoding refers to the classes *Roof* (red), *Façade* (white), *Impervious Surfaces* (grey), *Car* (blue), *Tree* (dark green), *Low Vegetation* (bright green), *Shrub* (yellow), *Fence / Hedge* (cyan), and *Powerline* (black).

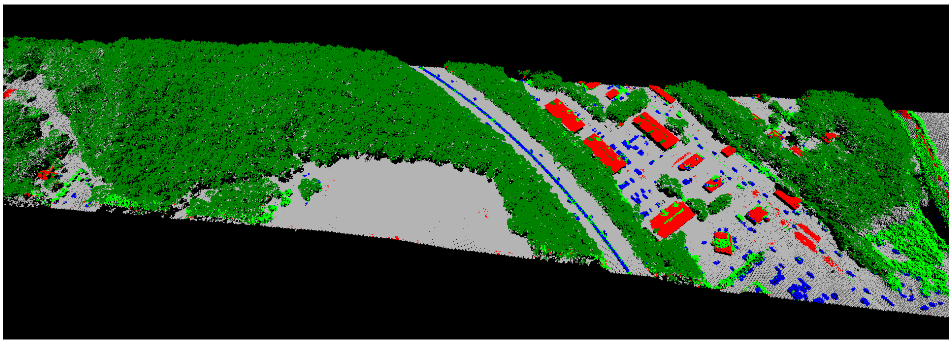


Fig. 3.15: Visualization of classification results for the **GML** A using $\mathcal{N}_c$, $\mathcal{N}_s$, and $\mathcal{N}_{k,\text{opt}}$ with covariance features, geometric 3D properties, shape distributions, and normalized height. The colour encoding refers to the classes *Ground* (grey), *Building* (red), *Car* (blue), *Tree* (dark green), and *Low Vegetation* (bright green).

class	this method $\mathcal{N}_{\text{all}}$			Shapovalov 2010		
	prec.	rec.	F_1	prec.	rec.	F_1
Ground	97.5	86.3	91.6	89.8	96.2	92.7
Building	47.2	73.7	57.5	86.8	58.5	69.9
Car	17.2	76.2	28.1	37.0	16.1	22.4
Tree	98.7	94.9	96.8	92.3	99.7	95.9
Low Vegetation	10.8	76.0	18.8	71.6	8.9	15.8
OA			90.5			N.A.
$\bar{F}_1$			58.5			59.3

Table 3.16: Comparison of class-wise precision, recall, and F_1 values (in %), alongside with OA and $\bar{F}_1$ (in %) of our results on the **GML** data set A in Section 3.5.3.2 to those achieved by Shapovalov et al. (2010).

about which classes profit most of those additional features, since the improvements are rather significant throughout all classes.

[Comparing the results on the two data sets, the] derived classification results reveal that the GML data set A with five semantic classes is not too challenging, as an overall accuracy of about 87-91% can be achieved when using the combined neighbourhoods. This is due to the fact that the dominant classes *Ground* and *Tree* can be accurately classified, whereas the problematic classes *Car* and *Low Vegetation* do not occur that often. In contrast, the Vaihingen data set with nine semantic classes is much more challenging, which can be verified by an overall accuracy of about 62-68%. The reason for the lower numbers is that most of the classes occur rather often, and they are furthermore characterized by a higher geometric similarity. Particularly the classes *Low Vegetation*, *Shrub*, and *Fence / Hedge* exhibit a similar geometric behaviour and misclassifications among these classes therefore occur [rather frequently]. However, this is in accordance with other investigations involving the Vaihingen data set (Blomley et al., 2016a; Steinsiek et al., 2017). Furthermore, the classes *Powerline* and *Car* reveal lower detection rates, but this might partly also be due to the fact that they are probably not covered representatively in the training data, where they are represented by 546 and 4 614 examples, respectively.

Furthermore, Table 3.17 provides a chance to compare the results among the two data sets with respect to similar classes, although care has to be taken due to the different proportions of occurrence in the different data sets. *Cars* are naturally the same class in both data sets, and so are *Trees*. *Cars* occur to roughly the

	GML data set A		Vaihingen data set	
	Section 3.5.3.1	here	Section 3.5.2	here
OA	76.8	90.5	41.5	68.1
$\bar{F}_1$	44.0	58.5	29.0	52.6
F_1 *Car*	21.7	28.1	16.5	44.1
F_1 *Tree*	90.8	96.8	53.3	66.9
F_1 *Low Veg.*	12.5	18.8	8.2	57.9
F_1 *Fence / Hedge*	—	—	11.4	17.5
F_1 *Shrub*	—	—	26.3	41.5
F_1 *Ground*	75.3	91.6	—	—
F_1 *Imperv. Surf*	—	—	64.1	80.0
F_1 *Building*	19.7	57.5	—	—
F_1 *Roof*	—	—	51.7	81.8
F_1 *Façade*	—	—	28.0	47.5
F_1 *Powerline*	—	—	1.7	32.1

Table 3.17: Comparison of the improvements in this section over previous experiments. Class-wise F_1-scores (in %) as well as OA and $\bar{F}_1$ (in %) for the **GML** data set A in comparison to Section 3.5.3.1 (**left**) as well as the **Vaihingen** data set in comparison to the RF results from Section 3.5.2 (**right**).

same proportion in both data sets (0.2 and 0.3% on the GML data set A and 0.6 and 0.9% on the Vaihingen data set), but are better identified on the Vaihingen data set. This might be due to the fact that objects, which could be the source of misclassifications, are better defined as additional classes like *Fence / Hedge* or *Shrub* on the Vaihingen data set. *Trees* occur more frequently on the GML data set A (35.5 and 53.0% vs. 17.9 and 13.2% on the Vaihingen data set), and are better detected on the GML data set A. *Ground* makes out a large proportion of the GML data set A (51.8 and 43.9%), and could be compared to the class of *Impervious Surfaces* (25.7 and 24.8%) on the Vaihingen data set. Both yield comparable results. *Buildings* occur much less frequent on the GML data set A (9.1 and 2.0%) and are subdivided into the classes *Roof* (20.2 and 26.5%) and *Façade* (3.6 and 2.7%) on the Vaihingen data set. *Roof* is better identified and *Façade* slightly worse on the Vaihingen data set than the common group of *Building* on the GML data set A.

3.5.3.3 Tests on the Vaihingen Data Set with All Features and Neighbourhoods and an Increased Number of Training Examples

Neighbourhoods

- $\mathcal{N}_c$ (1 m, 2 m, 3 m & 5 m)
- $\mathcal{N}_s$ (1 m, 2 m, 3 m & 5 m)
- $\mathcal{N}_{k,\text{opt}}$
- horizontal binning for normalized height

Features

- covariance features
- geom. 3D properties
- shape distributions
- normalized height

Classification

- balanced training $100\,000$ pts/class
- multinomial RF

We later noticed (Weinmann et al., 2018), that the choice to use only 10 000 training points per class apparently posed an unnecessary restriction on the Vaihingen data set with the given methodology. When allowing for 100 000 instead of 10 000 training entities per class, we could observe a significant improvement in the classification results. For classes with less than 100 000 training entities, the available training entities were duplicated as in the previous experiments.

Results

Table 3.18 shows the class-wise precision, recall and F_1 improvements. Since Section 3.5.3.2 indicated that misclassifications were more frequent in classes with few elements in the training data, training data incidence was denoted in an additional column. It can be seen, that the four smallest classes of *Powerline*, *Car*, *Fence / Hedge* and *Façade* actually show a decrease in recall due to the increase in training examples used, while precision is improved throughout all classes. The class-wise F_1-score improved in all classes except *Car*.

class	training incidence	rec.	prec.	F_1
Powerline	0.07 %	92.0 / 72.0	19.5 / 64.0	32.1 / 67.8
Low Vegetation	24.0 %	50.7 / 56.8	67.5 / 68.1	57.9 / 61.9
Imp. Surfaces	25.7 %	77.6 / 78.6	82.7 / 83.5	80.0 / 81.0
Car	0.6 %	57.5 / 29.2	35.7 / 67.4	44.1 / 40.7
Fence / Hedge	1.6 %	23.0 / 16.3	14.1 / 22.1	17.5 / 18.8
Roof	20.2 %	77.8 / 83.4	86.3 / 86.6	81.8 / 84.9
Façade	3.6 %	58.6 / 50.5	39.9 / 54.5	47.5 / 52.4
Shrub	6.3 %	58.4 / 61.9	32.2 / 33.2	41.5 / 43.2
Tree	17.9 %	75.4 / 80.3	66.9 / 68.0	70.9 / 73.6

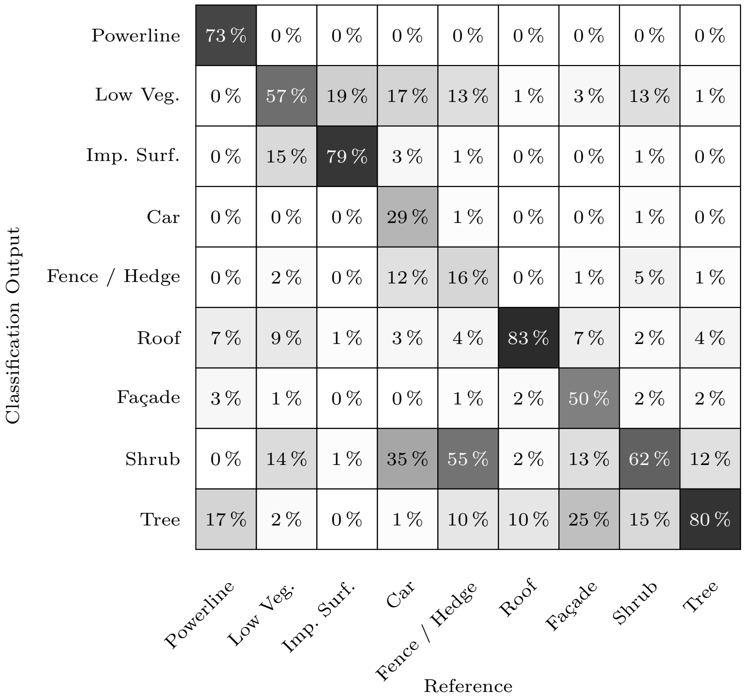
Table 3.18: Class-wise recall, precision and F_1-score (in %), comparing the results of a RF with 10 000 (first value) and a RF with 100 000 (second value) training examples per class on the **Vaihingen** data set. OA increased from 68.1% to 71.5%, while $\bar{F}_1$ improved from 52.6% to 58.3%.

The confusion matrix shown in Figure 3.16 indicates the typical misclassification cases. Most notably, elements of both the classes of *Car* and *Fence / Hedge* are often mislabelled as belonging to the class *Shrub*. While the class *Tree* is generally well detected, elements of many other classes such as *Façade*, *Powerline* or *Shrub* are often misclassified as *Tree* too. Similarly, elements of the classes *Impervious Surface*, *Car*, *Fence / Hedge* or *Shrub* are prone to be misclassified as *Low Vegetation*.

Discussion

As the comparison among the GML A and Vaihingen data sets in Section 3.5.3.2 made clear, the geometric and ontological similarities among classes such as *Fence / Hedge*, *Shrub* and *Low Vegetation* are certainly a contribution for misclassifications. The improvement observed in this section however shows that on the Vaihingen data set, annotated with the nine given classes, our fine-grained features reproduce the high degree of within-class variability well that big classes such as *Roof*, *Tree* and *Impervious Surface* show. The increased number of training examples apparently allows for a better representation of this variability by the classifier, which results in relatively few misclassifications for entities of these classes.

[The] results of [our] point-wise classification [achieved in this section] are comparable to the ones presented in [recent literature at the time of writing by] Steinsiek et al. (2017) for RF-based classification. While our results are [only 0.5% higher in OA, they are 8.3% higher in $\bar{F}_1$]. The latter indicates that our framework allows for a better classification of the different classes, while the approach presented



	Powerline	Low Veg.	Imp. Surf.	Car	Fence / Hedge	Roof	Façade	Shrub	Tree
Powerline	73 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Low Veg.	0 %	57 %	19 %	17 %	13 %	1 %	3 %	13 %	1 %
Imp. Surf.	0 %	15 %	79 %	3 %	1 %	0 %	0 %	1 %	0 %
Car	0 %	0 %	0 %	29 %	1 %	0 %	0 %	1 %	0 %
Fence / Hedge	0 %	2 %	0 %	12 %	16 %	0 %	1 %	5 %	1 %
Roof	7 %	9 %	1 %	3 %	4 %	83 %	7 %	2 %	4 %
Façade	3 %	1 %	0 %	0 %	1 %	2 %	50 %	2 %	2 %
Shrub	0 %	14 %	1 %	35 %	55 %	2 %	13 %	62 %	12 %
Tree	17 %	2 %	0 %	1 %	10 %	10 %	25 %	15 %	80 %

Fig. 3.16: Confusion matrix of the RF classification result using 100 000 training entities per class on the **Vaihingen** data set. Percentage and colour coding are given as percentage of the ground truth reference. The confusion matrix shows, that elements of both the classes of *Car* and *Fence / Hedge* (column 4 and 5) are often mislabelled as belonging to the class *Shrub* (row 8).

by Steinsiek et al. (2017) allows for [a similarly good classification result for a majority of returns. A detailed comparison is shown in Table 3.19.]

Further improvements in the classification results can be achieved by spatial regularization (cf. Section 2.3.3). This has been taken into account by Steinsiek et al. (2017) by employing a CRF, which refines the RF's initial probabilistic labelling output by enforcing spatial regularity among neighbouring data elements via an interaction potential learned from the initial labelling.

Steinsiek et al. (2017) report that an increase in the number of training examples would in their case not enhance the classification results. However, they used a slightly different set of features and neighbourhoods. Apart from the eight covariance features and five other geometric 3D properties that were calculated identically to our approach, and a normalized height feature similar to ours, they used three features produced from a 2D projection of each neighbourhood and three features from a 2D projection into quadratic, spatial bins of 1.25 m width that we did not use. They did not however implement shape distribution features. They considered three spherical neighbourhoods ($\mathcal{N}_s$ with fixed radii of 0.5 m, 1.0 m and 2.0 m) as well as an eigenentropy-based optimized spherical neighbourhood $\mathcal{N}_{k,\text{opt}}$. We used a similar slightly wider range of neighbourhoods (1 m, 2 m, 3 m and 5 m) as well as cylindrical neighbourhoods $\mathcal{N}_c$ they did not use. We can therefore only assume, that the benefit from further training examples that we could observe for our RF results must stem from our additional shape distribution features and/or our additional cylindrical neighbourhoods $\mathcal{N}_c$. The effect of different neighbourhood types however is likely smaller than that of the additional shape distribution features, as could be seen in Table 3.13 in the comparison of different neighbourhood combinations.

Class	this method $\mathcal{N}_{\text{all}}$		Steinsiek 2017		Niemeyer 2016
	$\text{RF}_{10\,000}$	$\text{RF}_{100\,000}$	$\text{RF}_{1\,000}$	CRF	H-CRF
Powerline	32.1	67.8	14.3	69.8	59.6
Low Vegetation	57.9	61.9	65.8	73.8	77.5
Impervious Surfaces	80.0	81.0	86.1	91.5	91.1
Car	44.1	40.7	24.9	58.2	73.1
Fence / Hedge	17.5	18.8	19.8	29.9	34.0
Roof	81.8	84.9	84.8	91.6	94.2
Façade	47.5	52.4	43.9	54.7	56.3
Shrub	41.5	43.2	40.8	47.8	46.6
Tree	70.9	73.6	69.5	80.2	83.1
OA	68.1	71.5	71.0	80.5	81.6
$\bar{F}_1$	52.6	58.3	50.0	66.4	68.4

Table 3.19: Class-wise F_1-scores (in %) as well as OA and $\bar{F}_1$ (in %) for the **Vaihingen** data set, comparing our RF results with 10 000 and 100 000 training examples per class respectively. Furthermore we list the results achieved by Steinsiek et al. (2017) in point-wise classification (RF) or with contextual classification (CRF), as well as the results achieved by Niemeyer et al. (2016) with a hierarchical CRF. Some of these values were presented in a similar comparison in (Weinmann et al., 2018).

Niemeyer et al. (2016) also proposed a hierarchical two-layer CRF. The first layer of the CRF performs point-wise labelling using a 2D-k-connected graph and the following 12 features per node (return): return intensity, echo ratio, a subset of covariance features (linearity, planarity, scatter, anisotropy), verticality, the ratio of the sums of the eigenvalues in a 2D and a 3D neighbourhood, and normalized height. This output is used to produce segments via clustering, which then form the entities for the second layer of the CRF. This second layer uses features computed for each segment, and thus incorporates a larger scale context among the segments. As seen in Table 3.19, this hierarchical CRF setup yields improved results in comparison to the single-layer CRF result presented by Steinsiek et al. (2017).

3.6 Discussion

As explained in the Personal Framing (cf. page 1) at the beginning of this thesis, the work reported in Section 3.5 was conducted before 2018 with traditional, non-contextual classification methods. Due to significant advances in the field of deep learning, especially concerning its' application on unstructured 3D data, and a widened availability of frameworks designed for end users, this section will discuss the results achieved by the traditional methods first (Section 3.6.1), before contrasting it with more recent developments in the field enabled by deep learning methods (Section 3.6.2).

3.6.1 Own Work

In Section 3.5.1, we analysed the weaknesses of traditional covariance features in their application on ALS data (the low ALS point density requires larger neighbourhoods, which are then less likely to be homogeneous and more likely to include elements of other classes too) and instead proposed an implementation of a sampled feature type, shape distributions (similar to the original proposal by Osada et al. (2002), designed for shape matching of 3D polygonal object models). We tested the shape distributions' respective behaviour across different neighbourhood scales for four classes in Section 3.5.1.1 and evaluated the respective feature relevance in Section 3.5.1.2. Finally, we combined those class-wise binary classification results in Section 3.5.1.3 to pin down applicational strengths and weaknesses of this feature type. We found that shape distributions are better suited than traditional covariance features to deal with the relatively sparse sampling density that prevails in many ALS data sets. While the results are very promising, it is still important to combine shape distributions with other feature types, as they are naturally rotation invariant and do not capture directional information such as height, height distribution, or verticality.

In Section 3.5.2 we analysed whether the relatively high number of feature values produced by shape distributions relies on a complex classifier for separability. This is not the case. A simple LDA classifier performed well in our experiments compared to both QDA and a RF classifier. We deduce that our features provide a good discriminability in the given application case, which does not rely on powerful modelling of complex decision boundaries by the classifier.

In Section 3.5.3, we finally put the different feature types, neighbourhood types and neighbourhood scales together to evaluate their joint performance. Section 3.5.3.1 gives a fine-grained comprehensive analysis of the combination-effect for multiple feature types, neighbourhood types, and scales. These findings indicate that multi-feature-type performs better than single-feature-type, multi-scale better than single-scale and multi-neighbourhood-type better than single-neighbourhood-type. Each addition results in a further improvement of the overall result, antagonizing the 'curse of dimensionality'. A comparison with the results from Section 3.5.1.1 also shows, that no clear 'best neighbourhood scale' can be found for a class in general, since there are noticeable differences between different data sets for similarly defined classes. The results in this section also show, how detrimental the use of an ill-defined feature such as the absolute height value can be, which has little real-live correlation to the desired classes compared to a normalized height approximation.

In Section 3.5.3.2, we used this finding and added a normalized height feature instead of the absolute height value. This additional feature seems to be of particular importance, as the results are drastically improved in comparison to Section 3.5.3.1. However, the normalized height estimation requires an additional continuous neighbourhood consideration apart from the per-point neighbourhoods used for all other features.

The approximation method for the scenes' topography used in the results provided in Section 3.5.3.2 is very rough. It relies on the fact that in the scenes considered, no objects exceeded the rough sampling distance of 20 m in size. In further (unpublished) experiments we enhanced this method further by employing edge detection to identify elevated regions. However, this did not provide additional benefits for the given application case here.

Further enhancements of the given methodology can be achieved by employing a regularized, contextual classification scheme like a CRF (Niemeyer et al., 2014; Weinmann et al., 2015; Steinsiek et al., 2017) instead of the RF classifier used throughout this work. A comparison to the results achieved by Steinsiek et al. (2017) and Niemeyer et al. (2016) can be found in Table 3.19 of Section 3.5.3.2. Alternatively, the results of non-contextual individual semantic labelling may be refined by a structured regularization or smoothing framework after the initial labelling (Landrieu et al., 2017b).

3.6.2 Comparison to Deep Learning

As elaborated in Section 2.4, traditional machine learning and modern deep learning approaches are not directly comparable, primarily for three reasons:

- Feature design: Modern deep learning does not rely on manually hand-crafted features designed to be relevant for the given application case, which is the focus of our approach within the traditional machine learning framework. Instead, 'features' or 'patterns in the data' are automatically learned via gradient optimization and backpropagation.

- Integration of steps: In modern deep learning, the traditional steps of feature extraction, relevance assessment, and classification are inseparable and jointly optimized, making it harder to interpret the results with respect to these components.

- Data dependency: Modern deep learning generally requires larger amounts of training data due to reduced reliance on application-specific knowledge. Consequently, there is an increased need to mitigate overfitting and ensure model generalization. This is sometimes achieved by generating synthetic training data from existing training data.

A number of deep learning approaches have been studied on the Vaihingen benchmark data set. Zhihai and Zhishuang (2018) as well as Zhao et al. (2018) used projections onto a horizontal raster to produce artificial-valued images from certain properties of the point cloud, and to produce features by training a convolutional neural network (CNN) on the basis of these images. Those were then used as input for a different classification method. Others, like Winiwarter et al. (2019), Yu et al. (2022) and Nong et al. (2023), employed deep learning architectures like PointNet and PointNet++, which are designed to handle 3D point cloud data directly, while integrating own adaptions to make those methods more fit to the characteristics of LiDAR point cloud data. In the following, we will briefly describe each of these approaches and compare their results in Table 3.20.

Zhihai and Zhishuang (2018) extracted five features for every return as in traditional feature extraction: intensity (1), eigenvalue-based features (2), normal-vector-based feature (1) and height above ground (1), using a $\mathcal{N}_{k,opt}$ neighbourhood. They then produced a 2D image projection from the point cloud, where every pixel was assigned artificial RGB values calculated from a mapping of a combination of the above features assigned to the return closest to the centre of each 2D image pixel. They then used a CNN for training and classification and re-mapped the output onto the point cloud.

Similarly, **Zhao et al. (2018)** used normalized height, intensity and roughness attributes to produce square 2D projection images of varying extent for the surroundings of each return. They then employed a CNN to produce features from each set of images, which were then used in a softmax regression classifier for point-wise classification.

PointNet (Qi et al., 2017b) is a pioneering network architecture, which enables a semi-convolution operation on unstructured point clouds. However, it cannot capture local context at different scales, as it uses a single max-pooling operation to aggregate the whole point set. **PointNet++** (Qi et al., 2017a) is an encoder-decoder extension to PointNet, which during the encoding layers, aggregates features on different neighbourhood scales by repeated subsampling (which produces increasingly sparse point clouds), grouping, and feature extraction. Afterwards, the features of the different scales are backpropagated (decoded) using inverse-distance-weighted interpolation, and subsequently concatenated for point-wise labelling.

Winiwarter et al. (2019) performed point-wise semantic labelling by applying the strategy of mini-batch training to a PointNet++-based network, while the batches were chosen as spatially aggregated subsets (rather than random subsets) so as to preserve neighbourhood information. Furthermore, they also append further sensor information, such as return intensity (or full waveform attributes in another data set) to the feature vector.

Yu et al. (2022) added a double self-attention mechanism to PointNet++, consisting of an efficient channel attention bock in the encoder layers and a context-guided aggregation module in the decoder layers. Besides improvements in the point-wise semantic labelling result compared to their PointNet++ result, this approach is reported to improve point cloud segmentation. Furthermore, they also provide an additional CRF optimization for their result.

Nong et al. (2023) enrich PointNet++'s local neighbourhood information by additional features (describing the relationship among neighbours as well as the centroid point information) throughout the encoder layers, while embedding elevation information (apart from the inverse distance information used in PointNet++) in the upsampling interpolation of the decoder layers. They also introduce a class-balancing to the loss function to deal better with the uneven class distribution in typical ALS semantic labelling challenges.

For all the approaches listed above that have been applied to the Vaihingen benchmark data set, OA, $\bar{F}_1$ and class-wise F_1 are listed in comparison to our results as well as the results of traditional feature-based contextual classification (Steinsiek et al., 2017; Niemeyer et al., 2016) in Table 3.20.

Interestingly, Yu et al. (2022) and Nong et al. (2023) achieved different results when applying the PointNet or PointNet++ networks respectively on the Vaihingen data set. This may be due to the differences in training procedure, as Yu et al. (2022) used training data augmentation on densely sampled training blocks to reduce overfitting of the model. This apparently improved results for PointNet especially, while the differences in the PointNet++ results are less significant. Their own adaptions of PointNet++ perform similarly well compared to each other.

Overall, the results listed in Table 3.20 indicate that the improvements deep learning achieves over the contextual feature-based approach (to which we compared our results in Table 3.19) are not always that significant. PointNet alone does not match the Steinsiek et al. (2017) and Niemeyer et al. (2016) CRF contributions, but is comparable if not inferior to the results of our method. PointNet++, Zhihai and Zhishuang (2018) and Winiwarter et al. (2019) perform similarly well as the Steinsiek et al. (2017) CRF contribution. Contributions by Zhao et al. (2018), Yu et al. (2022) and Nong et al. (2023) do mark an improvement over the feature-based contextual classification approach. The gain however is not as big as that achieved by contextual CNN classification compared to point-wise RF results in the feature-based methods. This indicates the importance of contextual relationships for reliable interpretations of ALS data. Consequentially, the result provided by Yu et al. (2022) for an additional CRF optimization of their deep learning network result is the overall leader in this comparison, while the improvement over their already excellent PointNet++ adaption is only by one percent in OA and none in $\bar{F}_1$. This hints that most of the contextual information is already exploited by the encoder-decoder structured network equipped with double self-attention.

It is noteworthy however, that contextual classification, such as CRFs applied to the output of a probabilistic classifier supplied with traditional hand-crafted features, can produce results which are comparable to those of a number of deep learning applications on the same problem.

Some core concepts that have proven to be significant in ALS data interpretation throughout both manual feature design and deep learning include:

- The importance of distance metric interaction among neighbouring points to describe point cloud structures alongside with a general permutation invariance among the elements of the point cloud. This is the case for covariance and shape distribution features, as long as calculated from a symmetrical neighbourhood, and these are also the core prerequisites considered in the development of the PointNet architecture.
- An aspiration for generalized models, as in bootstrap aggregating classifiers such as the RF classifier, or as in the batch training of neural networks.
- The distinction between transformation invariance of point cloud elements within the horizontal context, but the importance of preserved verticality or normalized height information, as found in Section 3.5.3.2 compared to Section 3.5.3.1, or as considered in the contributions of Zhihai and Zhishuang (2018), Zhao et al. (2018), and Nong et al. (2023).
- Increasing scales for feature extraction (as in our multi-scale feature extraction) or a hierarchical contextual setup (as in all PointNet++-based approaches compared to PointNet).

Compared to the traditional feature-based approach, modern deep learning however has the main advantage that neural networks can adapt extraordinarily well to variable structures, more so than individually targeted, hand-crafted features can do. The 'features' extracted by deep learning on hierarchical contextual levels vary in their degree of abstraction, meaning that lower levels learn basic structures or patterns while higher layers can recognize more complex shapes. The manual feature-based approach on the other hand is confined to ideas and concepts which can be applied to different neighbourhood scales, but can not usually be adapted to complex class definitions and their representation across varying scales in the data to any similar degree.

3.7 Conclusion

In conclusion, we were able to answer the research questions stated for this chapter (cf. Section 3.1.3) in a satisfactory way.

To answer RQ1, we analysed existing approaches in the 'Related Work' section of this chapter (Section 3.2) and proposed a comprehensive framework in the 'Methodology' section (Section 3.3) for point-wise semantic labelling of ALS point clouds. Our framework consists of 10 neighbourhood definitions (cylindrical neighbourhoods on four scales, spherical neighbourhoods on four scales, a spherical neighbourhood on an adaptively chosen scale and a neighbourhood defined by spatial binning and interpolation to approximate the scene's topography), 4 feature types (covariance features, geometric 3D properties, our novel feature type of shape distributions, and normalized height, with $8 + 4(5) + 50 + 1 = 63(64)$ feature values) and subsequent classification and evaluation. Both the different neighbourhood and novel feature type definitions comprise scientific innovations compared to the literature that already existed at the time of writing, and the feature type of shape distributions is a valuable addition in the field, complementing earlier feature types.

RQ2 is about a more detailed analysis of the strengths and weaknesses of the chosen approach and is answered by a number of thorough experiments presented in Section 3.5. In Section 3.5.1 we were able to show considerable strengths of our novel shape distribution feature type compared to covariance features in terms of descriptiveness in four typical major classes in an urban environment as well as

class	feature-based			deep learning		deep learning					deep learning
	point-wise	contextual		via 2D projection		directly on unstructured point cloud					added CRF
	this method, $\mathcal{N}_{\text{all}}$	Steinsiek 2017	Niemeyer 2016	Zhihai 2018	Zhao 2018	PointNet *	PointNet++ *	Winiwarter 2019	Yu 2022	Nong 2023	Yu 2022
Powerline	67.8	69.8	59.6	43.3	62.0	43.6 / 52.6	57.9 / 58.8	70.1	70.6	77.6	71.2
Low Vegetation	61.9	73.8	77.5	81.7	88.8	80.4 / 70.0	79.6 / 81.8	80.5	84.2	82.7	85.1
Impervious Surfaces	81.0	91.5	91.1	91.1	91.2	87.8 / 83.2	90.6 / 91.4	90.2	94.8	91.7	94.9
Car	40.7	58.2	73.1	68.3	66.7	47.3 / 11.2	66.1 / 73.2	45.7	69.7	79.2	72.2
Fence / Hedge	18.8	29.9	34.0	25.9	40.7	21.6 / 7.5	31.5 / 36.3	7.6	38.7	38.9	39.2
Roof	84.9	91.6	94.2	95.7	93.6	81.9 / 74.8	91.6 / 89.1	93.1	93.4	92.2	94.8
Façade	52.4	54.7	56.3	43.3	42.6	28.7 / 7.8	54.3 / 51.1	47.3	53.3	61.3	45.3
Shrub	43.2	47.8	46.6	45.2	55.9	38.5 / 24.6	41.6 / 44.4	34.7	45.6	43.2	46.3
Tree	73.6	80.2	83.1	73.2	82.6	64.5 / 45.4	77.0 / 75.0	74.5	80.6	79.2	81.3
OA	71.5	80.5	81.6	82.2	85.2	75.1 / 65.7	81.2 / 81.0	80.6	84.8	83.5	85.4
$\bar{F}_1$	58.3	66.4	68.4	63.7	69.3	54.9 / 41.9	65.6 / 66.7	60.4	70.1	71.8	70.0

Table 3.20: Class-wise F_1-scores (in %) as well as OA and mean $\bar{F}_1$ (in %) for the **Vaihingen** data set. A comparison of our results, feature-based contextual learning (Steinsiek et al., 2017) and deep learning implementations, either via horizontal 2D projections or directly on the unstructured point cloud by adaptions of PointNet/PointNet++. (* first value according to Yu et al. (2022), second value according to Nong et al. (2023))

in applicability over multiple neighbourhood scales. Meanwhile we could show in Section 3.5.2, that this did not come at the expense of difficult separability due to the higher overall number of features, but that the problem remained solvable by traditional classification methods such as LDA. In Section 3.5.3 we demonstrated the beneficial effect of multi-scale and multi-neighbourhood-type combinations, and showed how a combination of shape distributions with well established features clearly and markedly improved the results achieved. We compare our results with other publications in the field and found comparable if not superior results, except for approaches utilizing contextual classification instead of point-wise classification. Both contextual classification, which implies spatial regularization during the classification process, as well as object-based approaches could further improve the level of detail and quality in scene interpretation. A comparison among Sections 3.5.3.1 and 3.5.3.2 showed the paramount importance of a normalized height feature for individual returns. Last but not least, we analysed the differences among our results on two different data sets, which showed that the definition of classes is generally of importance to the classification performance, while some of the differences between identically defined classes could be related to their different degree of prevalence in the training data sets.

We acknowledge the improved classification potential in applying contextual classification strategies such as CRFs (Steinsiek et al., 2017; Niemeyer et al., 2016), especially in challenging scenes. Due to the time elapsed between the scientific work and publishing, and the writing of this thesis, we are also able to evaluate our work in hindsight by comparing it to the modern advances in the field enabled by deep learning. Concerning the interpretation of unstructured 3D point clouds in the context of semantic labelling in urban scenes, multi-view-projection-based approaches could only produce results comparable or inferior to those of traditional approaches applying structured prediction. Point-based deep learning strategies show a higher potential, but have to be adapted to the given case. Especially adaptions focusing on attention mechanisms or the propagation of neighbourhood relationship throughout the layers of an encoder-decoder architecture have proven effective on the Vaihingen 3D semantic labelling challenge. This highlights the importance of contextual learning that is implemented both in the structured prediction and the best-performing deep learning implementations.

Chapter 4

Object-Wise Classification of Tree Species for Individual Tree Segments

4.1 Introduction

While Chapter 3 focused on discriminating different classes within a scene, this Chapter focuses on analysing vegetation, and forest trees in particular, at a higher level of detail. In Europe, forests have become more heterogeneous over the past decades, partly due to increasing disturbances such as windthrows, bark beetle infestations or wildfires, but partly also because forest planning is increasingly taking an ecological, diversity-oriented perspective into account, which further contributes to structural heterogeneity. This challenges established manual forest inventory practices (Latifi et al., 2015). Gradually, remote sensing technologies are therefore becoming more important to support the assessment of existing resources while keeping the manual effort manageable. Airborne small-footprint pulsed light detection and ranging (LiDAR) is an important measurement tool for forest canopies with good estimation of biomass and height (Vauhkonen et al., 2014a).

4.1.1 Goals

Generally, forest inventories aided by LiDAR data may either use area-based methods or single tree identification (Latifi et al., 2015). In both cases, however, [] tree species identification is important since different species have different allometric dependencies [1] , which influence the accuracy of timber volume estimation (Packalén and Maltamo, 2008). [Moreover, species information as such is also important for forestry decisions such as treatment schedules and growth predictions.] Accurate tree species identification [in LiDAR-based forest inventories] however is very challenging. It is to some degree possible from structural and intensity features calculated from discrete-return LiDAR data (Ørka et al., 2009), but is more accurate in combination with aerial imagery (Holmgren et al., 2008; Ørka et al., 2012), hyperspectral data (Kandare et al., 2017) or with attributes derived from waveform-recording LiDAR data (Yao et al., 2012; Hovi et al., 2016). If LiDAR data alone could provide a reliable species discrimination, acquisition costs would be reduced and economic feasibility stimulated. Prospectively, precision forestry could be stimulated, whereby forestry decisions are made and valuable timber is collected at demand based on precise inventory data from remote sensing surveys.

[1] Allometry is the study of the relative size of parts of organisms. In forestry, it is very important to study allometric relationships in order to estimate tree measurements, such as timber volume, from an easily measured attribute such as diameter at breast height (DBH) or height.

4.1.2 Challenges

For tree-level inventories, the detection of individual trees traditionally relies on a raster-based canopy height model (CHM) interpolated from the airborne laser scanning (ALS) height data (Persson et al., 2002; Popescu et al., 2003), individual stem detection (Reitberger et al., 2007), point-based segmentation techniques (Strîmbu and Strîmbu, 2015) or layer-wise analysis (Ayrey et al., 2016; Hamraz et al., 2017). Since trees with interlacing crown or trees below the dominant canopy can be difficult to detect (Vauhkonen et al., 2012), there have also been approaches of combining the tree-level inventory scheme with area-based approaches to minimize errors in timber volume estimation, caused by undetected trees as well as errors in the stand-specific allometric model predictions (Lindberg et al., 2010). Recently, deep learning methods have also been applied successfully on drone-recorded very high density LiDAR point clouds (Zhang et al., 2023; Xiang et al., 2024), which solve the segmentation task alongside with the automated retrieval of tree parameters and stand structure.

4.1.3 Objective

During the main research period of this thesis (cf. Personal Framing on page 1), tree-level inventories could only be envisaged as a succession of tree detection and segmentation, followed by a subsequent description and classification of each segment. This is why, throughout this Chapter, a prior step of individual tree segmentation is assumed. Some tree and timber attributes are directly linked to the geometrical distribution of the returns among a segment (like height distributions, alpha-shapes, crown characteristics, etc.), while other tree characteristics, especially tree species, are more difficult to infer from the LiDAR point cloud. Up until the recent advances brought by higher resolution LiDAR sensors and deep learning, full-waveform analysis of the returns (Hovi et al., 2016; Bruggisser et al., 2017) has been the best option for this attribution so far.

The idea in this thesis was to combine geometric and waveform information per segment and see if this combination improves the tree species class separability. Hence, we formulated the following research questions (RQs)3-5 for this thesis:

- **RQ3:** Is it possible to design a feature type which can be used to improve tree species classification of individual tree segments by capturing the geometric distribution of waveform properties (generated by below-footprint-scale structures) within tree crowns?

- **RQ4:** Given the baseline accuracies in tree species classification by detailed waveform analysis (Hovi et al., 2016), can the accuracy be improved even further by considering the localization of the waveform attributes within the tree crown? If so, how big is the gain?

- **RQ5:** How are the failure cases distributed among tree sizes? Trends are indicative of practical relevance.

4.2 Related Work

Different approaches have been taken in tree species identification so far, yet the comparison between different studies is intricate. Different feature types may be more or less descriptive, depending on the context such as the local biome (boreal, temperate), the species composition, the number of species and the classification depth (like a limitation on the number of classes to be distinguished, e.g. classifying tree genera rather than species, or classifying deciduous vs. coniferous trees (Heinzel and Koch, 2011)), the season (Kim et al., 2009; Hovi et al., 2016), age distribution, site fertility, sensor characteristics and settings and stand type (homogeneous or mixed (Korpela et al., 2010b)). The final results in classification performance may furthermore depend on the amount of and the variability within the validation data as well as on the validation scheme, and are therefore [quite] case-specific. Furthermore, many algorithms require parameter selections that may have to be optimized. Thus, generic methods are eligible.

In the following, the current state of the art is assessed. Sections [4.2.1, 4.2.2 and 4.2.3] cover the general topics of segmentation, classification and feature design for LiDAR point clouds, while different approaches in tree species classification are summarized according to their scale of operation in Sections [4.2.4, 4.2.5 and 4.2.6]. Finally, multi-scale approaches are reviewed in Section [4.2.7].

4.2.1 Single-Tree Segmentation

Despite the fact, that there are area-based approaches for species-specific timber volume estimation (Räty et al., 2016), we focus on single-tree species classification in order to obtain [a] reliable evaluation []. Thus, crown segmentation is a crucial prerequisite, and failures of the single-tree extraction reduce the quality of the species classification result (Vauhkonen et al., 2014b). Solutions include, but are not limited to, crown-surface-based approaches (Persson et al., 2002), k-means clustering (Morsdorf et al., 2004), algorithms using both the crown surface and volumetric normalized cuts (Reitberger et al., 2009) and graph-based solutions (Strîmbu and Strîmbu, 2015). In (Vauhkonen et al., 2012), different single-tree detection algorithms were compared on different testing sites, showing that the stand density and spatial complexity (clustering or regular patterns) affect the quality of the segmentation more than differences among the algorithms and that most algorithms perform better in the environment they were developed in. This sensitivity of segmentation algorithms to local allometry indeed holds true to the current date, as Cao et al. (2023) show. Also, tree segmentation algorithms are still only accurate for canopy trees, while detecting understory trees has remained a common difficulty.

4.2.2 Tree Species Classification

In single-tree species classification, the different methods of discrimination include both statistical classifiers and [discriminative classifiers]. Since one cannot usually assume linear separability or an easy-to-model distribution of features, nonlinear and non-parametric classification algorithms are especially popular (Vauhkonen et al., 2014b). Highly efficient algorithms such as support vector machines (Schölkopf, 1997) and random forest classifiers (Breiman, 2001) are available in a variety of software packages. With these classifiers, the most important prerequisite for a successful classification is a good feature design that condenses class-specific [characteristics] in the data into discriminant features, so that the different classes are represented differently in the feature space. Lately, deep learning techniques such as [multi-view] convolutional neural networks (CNNs)[(Hamraz et al., 2019; Marinelli et al., 2022), sparse 3D CNNs (Xiang et al., 2024) or PointNet-based approaches (Briechle et al., 2019)] are becoming more and more popular, where the filtering for class-specific patterns in the data is autonomously learned via backpropagation. However, [streamlined application of these techniques remains] intricate.

4.2.3 Feature Design

Generally, [the] design [of hand-crafted features] for point cloud classification is highly application-specific and depends on both the data and object properties [such as point density, geometry or surface structure, occlusions and many more]. However, some general concepts are noteworthy. For objects with densely sampled surfaces, signatures of histograms of orientations (SHOT) (Tombari et al., 2010) and spin images (Johnson and Hebert, 1999) have been used as descriptors of local surface patches in object recognition (Velizhev et al., 2012). Since these descriptors require a surface to define a normal vector, more general descriptors such as 3D shape contexts (Frome et al., 2004) or point feature histograms (Rusu et al., 2009) have been designed. Furthermore, shape distributions (Osada et al., 2002) enable a parametrization of the overall object shape. [] In ALS, the sparsity of the data (usually 5-[50] $^{\text{pts}}/_{\text{m}^2}$), the geometric imprecision and object complexity limit the feature choice. Apart from waveform attributes describing the shape of the recorded signal, geometric features such as point height and [statistical distribution measures] as well as the eigenvalues of the 3D covariance matrix are usually applied (Mallet et al., 2011; Chehata et al., 2009; West et al., 2004). [During the course of our work in point-wise semantic labelling], shape distributions [describing the local neighbourhood of individual returns] have also been shown to be successful in urban scene classification of airborne LiDAR data (Blomley and Weinmann, 2017). Apart from point-wise classification, object-based strategies in remote sensing include shape parametrizations similar to 3D Hough transformations (Vosselman et al., 2004) and model-based segmentation (Polewski et al., 2014).

In tree species classification, [] existing research can be grouped by three scales of interest. The spatial distribution of backscattering within a tree segment has been studied on a large (overall) or medium (within-crown) scale. Small structures below the footprint size contribute to the shape and amplitude of the recorded waveform (small scale) (see Figure 4.1).

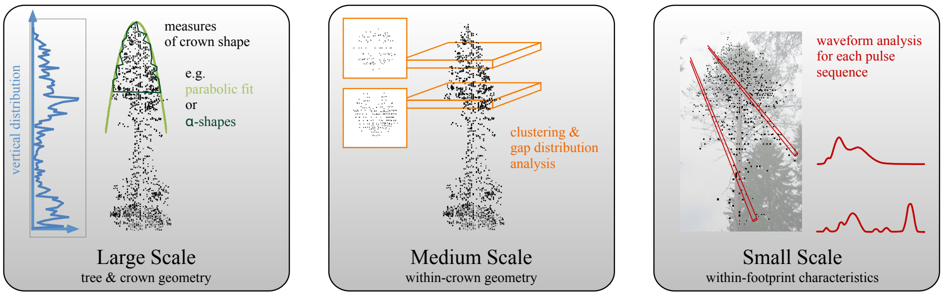


Fig. 4.1: Concepts used in related work on tree species classification.

4.2.4 *Large-Scale Structures*

Both on individual tree and stand level, the vertical distribution of backscattering is a key observation. For individual tree species classification, independence from the naturally varying absolute tree height is often sought by normalization of the vertical profile (Brandtberg, 2007). However, factors such as age and site fertility influence the species-specific rate and cessation of apical growth and thus the crown morphology. A single species (e.g. Scots pine) may change its shape and appearance throughout its lifetime (young pines have tapered crowns, while older pines have a rounded apex) or take entirely different shapes depending on their site's conditions (pines can be small and bushy in peat bogs or mountainous regions, while they can grow tall under favourable conditions). Height normalization can therefore not eliminate the height-related effects entirely. Mostly, statistical metrics such as moments, deciles or ratios above and below thresholds are chosen to describe the height and intensity (peak amplitude) distribution of LiDAR returns (Ørka et al., 2009; Suratno et al., 2009). In (Kim et al., 2011), length-to-width ratios of upper percentiles of the tree segment point cloud are added as features describing the overall shape of the crown. There are, however, more specialized methods to assess the shape and volume of the crown. Holmgren and Persson (2004) used features from a 3D parabolic surface fit together with statistical height and intensity features. [] Barilotti et al. (2009) [also fit] the crown surface of individual trees [] by second-order polynomials. Dong (2009) characterized crown shapes by the D2 shape distribution and found no noticeable difference in the distribution when using all returns in the segments or only those that lay on a (15 cm) surface model. Vauhkonen et al. (2009) successfully used α-shape metrics, a technique carving out empty space from a point cloud by a 'spherical eraser' of radius α, for individual tree species classification. α-shape features performed well compared to statistical height and intensity features as well as to textural features of the crown surface model. However, it must be noted the trees in this study were from sparse stands. α-shape features (Vauhkonen et al., 2010) and implicit surface reconstruction (Kato et al., 2009) were also used to estimate other individual tree parameters such as crown base height (CBH) or crown volume. CBH has also been determined by a voxel-based approach (Popescu and Zhao, 2008) and may, under certain conditions, be informative of the tree species (Holmgren and Persson, 2004).

In 2018, Axelsson et al. (2018) also published a geometric feature definition to describe the large-scale distribution of returns within a tree crown by defining concentric ellipsoidal layers within the tree crown with a fixed layer thickness of 0.5 m. Relative return densities within ellipsoidal layers, the percentage of all returns that fall outside of the outermost ellipsoid, and the ratio of horizontal and vertical ellipsoid radii were then used as feature values. This approach is motivated as identifying the variation between species with large leaves and a dense crown of limited permeability, and other species with thinner foliage constituting a sparser crown.

4.2.5 *Medium-Scale Structures*

Within the tree crown, the spatial distribution of the 3D return coordinates may be analysed in densely sampled LiDAR data. Ko et al. (2012) identified two within-crown feature groups that perform well in tree species classification. One feature group relies on segments generated by a merge-and-split algorithm, which approximately follows linear branching structures, while the other feature group is based on cluster analysis by 3D buffering (Ko et al., 2013). Both these feature groups acquire results similar to those of large-scale approaches such as convex hull, α-shape and vertical distribution features. Li et al. (2013) described four groups of within-crown structural features. First, they used 3D Haralick texture features deduced from the co-occurrence matrix of the number of returns in a sub-meter voxel representation. The remaining three groups analyse clustering and gap distribution within horizontal slices of the tree crown. The relative degree of clustering was quantified by the variance-to-mean ratio of the number of returns in sub-meter quadrants. The relative scale of clustering was quantified by characteristics of the L-function, which maps the deviation of the number of returns within a certain radius from even distribution. Finally the gap distribution was quantified by the variance of the frequency distribution of Delaunay-triangulation edge distances. Feature selection indicated that features from the top layers (roughly within 6 m from the tree top) showed the strongest structural differences between mature and over 15 m height trembling aspen, sugar maple, jack pine and white pine.

Lin and Hyyppä (2016) suggested a crown-internal feature type, that is based on eight vertical projection profiles, aligned with the vertical centre of the tree crown and covering a 45° rotational segment each. Partitioning each projection into 1 m spaced grid, they then calculated 11 measures based on these values for the tree crown.

Using ALS data of higher density, generated by repeated flight over the same area with a typical ALS laser scanner, Harikumar et al. (2017) even managed to model the internal branch structure of conifers and produce features such as average branch slope, branch length, branch compactness, branch width, branch symmetry and branch density. Those features, combined with convex hull features describing the external shape of the crown, enabled a classification among different coniferous species.

4.2.6 Small-Scale Structures

Below the footprint diameter, small structures such as the size, shape, orientation and spatial arrangement of the leaves and needles affect the differential backscatter cross-section. These target properties largely determine the time-dependent signal from the vegetation entering the receiver, while the receiver characteristics further influence the recorded waveform (Korpela et al., 2013) [cf. Section 2.1.5]. Thus, the interpretation of the waveform is highly non-trivial in vegetation. Geometric optical models have been employed to explore relationships between the waveform shape, sensor and acquisition settings and the canopy structure (Ni-Meister et al., 2001; Morsdorf et al., 2009; Disney et al., 2010; Hancock et al., 2012; Hovi and Korpela, 2014).

Many studies use the intensity statistics of the raw recorded data for tree species classification (Ørka et al., 2009; Kim et al., 2009; Yu et al., 2014). In the presence of range variation however, the raw intensities need a range normalization (Korpela, 2008; Suratno et al., 2009; Yao et al., 2012). In linear receivers, the normalization, using the radar equation, can be done for well-defined intersection geometries such as even surfaces and linear or point-like targets (Wagner et al., 2006). In vegetation however, the non-uniform irradiance field of the laser beam falls on tilted surfaces of varying size, reflectance and spatial arrangement. The vegetation structure varies for example with tree species, which results in different normalization coefficients, making range normalization ambiguous (Brandtberg, 2007; Korpela et al., 2010a). Also, LiDAR sensors need a high dynamic range, because the signal is strongly influenced by range variation (Wagner et al., 2006). Therefore, sensor-specific receiver gain and linearity corrections may have to be considered in order to reach accurate radiometry [cf. Section 2.1.3].

When analysing the shape of a recorded waveform, the most physically motivated approach is to use a deconvolution technique using both the transmitted and the received waveform for retrieval of the target's differential backscatter cross-section, sometimes also called surface response (Jutzi and Stilla, 2006). Using this approach, the discrimination of surfaces closely spaced in depth is improved, but still limited by the bandwidth of the receiver unit [cf. Section 2.1.5]. There are also numerous approaches that aim at modelling the received waveform by fitting functions into it (see Mallet and Bretar (2009) and Hancock et al. (2015) for reviews of different methods). This enables deconvolution analytically, but [also has] other benefits. Peaks corresponding to individual reflectors can be detected accurately, and attributes describing [the] target's shape can be calculated from the fitted functions. The method most commonly referred to in literature is to decompose the waveform into a sum of n Gaussian components (Persson et al., 2005; Wagner et al., 2006; Reitberger et al., 2008). Because the assumptions made for decomposition models are not necessarily valid in complex tree canopies or with sensors that have varying shapes of emitted pulse and/or receiver response, others prefer to directly use the raw waveform (Litkey et al., 2007; Yu et al., 2014; Hovi et al., 2016).

Furthermore, when scanning tree canopies, many partially overlapping reflectors occur within one beam path. Disambiguations arise when the illumination incidence, target reflectance and the fraction of the footprint covered are unknown. Upper 'preceding' canopy may attenuate or even shield the laser scanners light before reaching lower layers. To estimate the amount of pulse reflecting vegetation from the waveform, Lindberg et al. (2012) found it was best to assume a constant ratio between reflectance and attenuation and normalize the shape of the waveform by an iterative algorithm based on the Beer-Lambert attenuation law, while Romanczyk (2015) and Richter et al. (2015) later developed both a discrete attenuation correction, assuming individual interactions with clusters of scattering material, and an integral attenuation correction, assuming a continuous non-linear attenuation for the beam's propagation through the canopy. For rays penetrating to the ground, a constant ratio of canopy and ground reflectance has to be assumed (Armston et al., 2013; Richter et al., 2015), which may not always be valid, depending on the ground cover, tree species composition and laser wavelength.

Another solution that has been proven beneficial in tree species classification is a practical distinction between only, first[,] and other (subsequent) reflections along a beam path (Ørka et al., 2009). First and only echoes, as opposed to subsequent echoes, are less influenced by attenuation, and only echoes, which are always strong, as opposed to first echoes, can be assumed to come from dense foliage that covers the whole laser beam diameter. However, observing only one echo can also be due to dark targets further along the beam absorbing the signal without triggering the receiver.

Waveform attributes inferred directly from the signal shape have been compared to Gaussian fitting (Neuenschwander et al., 2009; Lindberg et al., 2012). [It] has been found that descriptive waveform attributes of noise exceeding sequences may significantly contribute to the distinction of individual tree species (Yu et al., 2014; Hovi et al., 2016).

Lately waveform decomposition has also been performed using a skew normal distribution function that allows for higher order statistical moments such as skewness and kurtosis (as compared to Gaussian decomposition, which is restricted to a symmetrical basis function). The benefit observed for waveform attributes calculated directly from the signal shape (which implicitly describe skewness and kurtosis) is retained in this type of waveform decomposition (Bruggisser et al., 2017). This indicates the importance of distributed scattering elements (such as foliage) and multiple scattering along the beam path for tree species discrimination in ALS data.

4.2.7 Multi-Scale Approaches

In applications concerned with natural targets, geometric multi-scale approaches show promising results. Using a terrestrial laser scan, Brodu and Lague (2012) showed that the classification of natural structures benefits from evaluating covariance-based eigenvalue features at multiple scales, arguing that the characteristics of different structures (twigs, ground) are best captured at different scales. [Our earlier work in the scope of this thesis] show[s] that land cover and urban classification tasks also benefit from the combination of features from multiple scales (Niemeyer et al., 2014; Schmidt et al., 2014; Blomley et al., 2016b).

4.3 Methodology

Overall, we follow a standard classification procedure. First, the raw recorded waveforms are processed into individual returns, which are each described by a position and a set of six waveform attributes as described in Section [4.3.1]. Similar to previous studies (Holmgren and Persson, 2004; Ørka et al., 2009; Hovi et al., 2016), we distinguish between *only*, *first-of-many* and *subsequent* returns along the beam path, and group them into three return type groups: *only*, *first-or-only* or *all* returns. *Only* returns provide the clearest data, as they occur when the full laser footprint is covered by pulse-reflecting matter, yet they include little spatial distribution information as they mostly occur at the dense top of the crowns. The groups of *first-or-only* and *all* returns are incrementally more permissive and could possibly lead to a loss of information due to the mixture of different return types, yet they might also lead to an information gain due to the improved spatial coverage of canopy parts, that cannot be measured without preceding occlusions. The three groups of return types used in this study are chosen to reflect this natural trade-off. Secondly, [both established feature types as well as our novel spin image feature type] are calculated separately for all three return type groups per tree segment, as summarized in Sections [4.3.2 and 4.3.3]. Finally, we employ the classifier and evaluation scores described in Sections [4.3.4 and 4.3.5] respectively.

4.3.1 Waveform Processing

Waveform attributes are extracted using the same implementation documented in (Hovi et al., 2016), who implemented a waveform decomposition strategy which achieved a very good species separability for pine, spruce and birch in a boreal forest environment. In short, the recorded waveforms are divided into returns, each of which represents a continuous noise-exceeding amplitude sequence (NEAS). The noise threshold was determined from empty tail-parts of the waveform recordings that fall below ground level. The parameters are the degree of low-pass filtering, the noise threshold and the minimum plausible length, which were adopted from Hovi et al. (2016), except for the length, which was reduced from 5 ns to 1 ns, which slightly increased the number of *first-of-many* and *subsequent* returns. Each return is then spatially assigned to the coordinate of the highest amplitude in the sequence. The coordinates are used in assigning the return to a tree segment as well as in calculating geometric features. Six descriptive attributes are calculated from the shape of the sequence. Those attributes include the amplitude A of the highest peak, the total energy E integrated over the sequence, the length L of the sequence, the full-width half maximum [] $FWHM$ [of the sequence], the length $EQ50$ of the sequence after which 50% of the total sequence's energy has been deposited at the sensor, and the number $\#P$ of peaks in the sequence.

4.3.2 Established Features for Comparison

Pursuing [*RQ3*], [we aimed to design a feature type which would capture the geometric distribution of waveform attributes among a tree crown. For this purpose we chose to adapt the feature type of spin images (Johnson and Hebert, 1999), proposed originally as surface point descriptors on densely meshed object model surfaces. Our implementation of this] feature type can [produce] both [] purely geometric feature[s (as in their original implementation)], when the local return density is evaluated (spin images of the local return density), as well as [] feature[s] describing the geometric distribution of waveform attributes within the crown (spin images of waveform attributes). The spin image features are compared both to geometric α-shape features and to non-spatial waveform attribute features.

4.3.2.1 α-Shape Features

The α-shape features are calculated according to Vauhkonen et al. (2009) as the volumes inside and outside of α-shape components divided by the cubic of the tree's height and as the number of α-shape components divided by the tree's height. We used all relative height sections (except for those falling below 10 % and only above 98 %) and α values of the above-mentioned reference, resulting in 486 feature values per tree segment. Due to the large number of parameter variations, some redundancy is to be expected among the feature values.

4.3.2.2 Non-Spatial Waveform Attribute Features

Non-spatial statistical metrics of waveform attributes are calculated according to Hovi et al. (2016) from those returns that fall within the top 40% of the tree segment. They comprise the first four statistical moments and the deciles. If calculated separately for *only*, *first-or-only* or *all* returns within the tree segment, they amount to 42 feature values per waveform attribute. Some of them may hold redundant information, as the moments and deciles are alternate descriptors of the shape of the distribution.

4.3.3 Spin Image Features

The aim in developing spin image features for tree species classification was to capture the geometric distribution of values within the crown in relatively few features, while maintaining a high descriptiveness. The number of feature values is adaptively chosen according to the degree of variability in a set of representative data.

The procedure of calculating spin image features is adapted from a technique, which was originally proposed as descriptors of local properties for meshed object model surfaces (Johnson and Hebert, 1999). In the following paragraph, we will first discuss model assumptions about the tree crown structure and explain the motivation of adopting spin image features for tree species classification. Afterwards, the three main processing steps of the method, visualized in Figure [4.2], are explained in detail. As a first step, an image plane is spun around a defined axis, collecting the number of returns or the mean value of waveform attributes per pixel as respective values. In the second step, a principal component analysis (PCA) is used to identify the components of highest variability that this representation has among a set of library trees. Those components are given by the eigenvectors $\vec{e_i}$ of the PCA, that may be visualized as eigen-spin images. As a third step, feature values are extracted for every individual tree segment by projecting the individual tree's data onto the most relevant eigen-spin images.

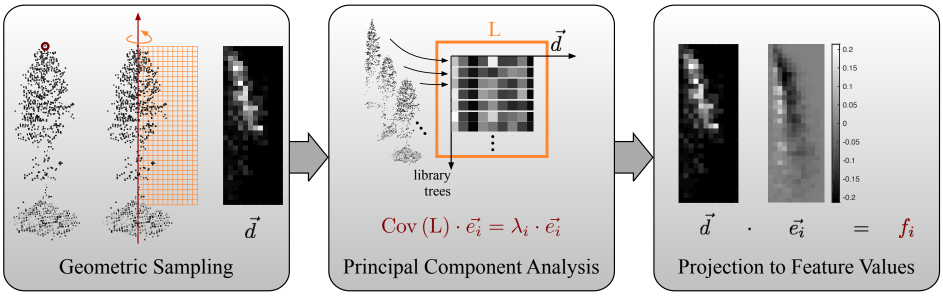

$\vec{d}$
Geometric Sampling
L
$\vec{d}$
library trees
$\mathrm{Cov}\,(\mathrm{L}) \cdot \vec{e_i} = \lambda_i \cdot \vec{e_i}$
Principal Component Analysis
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
$\vec{d} \quad \cdot \quad \vec{e_i} \quad = \quad f_i$
Projection to Feature Values

Fig. 4.2: Overview of the [proposed] spin image method. Returns are first sampled to a rotating image plane for each tree, producing a data vector $\vec{d}$. Sampled data is then collected from sample trees to form a library **L**. By principal component analysis, variable components within the library are identified as eigenvectors $\vec{e_i}$. Finally, the original data $\vec{d}$ is projected into the eigenvector space, producing a single feature value (linear combination) f_i for each eigenvector $\vec{e_i}$.

4.3.3.1 Motivation

A tree's shape and internal structure are the result of both intrinsic growth behaviour and reactions to external conditions and forces, such as local lighting or mechanical stress. While external conditions are highly variable throughout individual cases, feature development is interested in those structural traits which apply universally due to intrinsic growth behaviour and which may hold information relevant for species distinction.

Generally, tree growth is governed by a set of botanical mechanisms. Negative gravitropism (growth against the direction of gravity) and phototropism (growth towards the light) cause branches to grow upwards, but they are counteracted by epinastical suppression of subordinate branches (hormone-induced outward and downward bend of older branches lower down the tree). Therefore, the optimum compromise of branch angles, and thus the internal structure of the tree crown changes with both the vertical distance from the tree apex and the horizontal distance from the stem (Mattheck, 1991, page 10). Structural traits should therefore be on average invariant under a rotation transformation, while the axis of rotational symmetry is vertical due to the effect of negative gravitropism.

It is known, that there are differences among tree species concerning their strategy of biomass allocation, crown structure and leaf morphology (Menalled and Kelty, 2001; Alves and Santos, 2002; Koike et al., 2001). Those differences will affect both the shape of the recorded waveform – and thus the attributes derived (Korpela et al., 2013; Hovi and Korpela, 2014) – as well as the returns' distribution among the crown. It should therefore prove beneficial for species identification to evaluate the returns' waveform attributes with regard to their relative position to the origin of symmetry. This is equivalent to the evaluation on a rotated image plane, where the spinning axis corresponds to the natural axis of symmetry. Note, that every individual tree is not assumed to fulfil the rotational symmetry, as uneven lighting conditions or mechanical stress may cause each individual tree to grow differently. The species-specific structural traits however are on average symmetrical.

By exploiting the rotational symmetry, the spatial complexity is reduced by one dimension and the observations are presented in a dense representation without a general loss of their descriptiveness regarding the tree species. This dense representation is particularly valuable for ALS data, as the tree may be partially occluded from some directions and sparsity is a general constraint, given that lower flying height and more overlap increase acquisition costs.

4.3.3.2 Geometric Sampling

The first step of the spin image method is therefore to use a vertical spinning axis and sample the attributes within each individual tree segment to the pixels of an image plane rotated around the axis (Figure [4.2], left box). This results in a data vector $\vec{d}$ of length $d = \#$pixels for each waveform attribute per tree.

In our implementation, the horizontal position of the vertical spinning axis is determined by the highest return in the segment. Quality measures for the goodness of this positioning are not generally attainable. Holmgren and Persson (2004), in a similarly rotation-symmetric approach, used the highest return as an origin for their parabolic crown surface fit and validated these positions against field-measured stem positions. However, it is to be noted that the position of the symmetry axis does not necessarily coincide with the stem position measured at the ground, since possible curvature of the stem would have to be considered. Therefore we compare the placement of our spinning axes to positions of tree tops measured manually in airborne images during our method's evaluation in Section [4.5.1].

Parameters, which have to be set for the geometric sampling on a rotating image plane, are both the image's horizontal and vertical dimension as well as its pixel size. In the original spin image approach, Johnson and Hebert (1999) evaluated the descriptiveness of spin images of varying pixel size. Their findings suggested that spin images were most descriptive when the pixel size was approximately equal to the surface mesh resolution of the model. Experiments, presented is Section [4.5.2], evaluate if an equivalent relation holds true in the case of ALS data for tree segments. The horizontal image dimension is kept at a fixed value of 5 m, because very few trees are expected to exceed this range. If crowns are generally smaller, the peripheral areas of the spin images will sample very few returns due to the previous tree segmentation. Therefore, this parameter is non-critical. The influence of the image's vertical dimension is evaluated in Section [4.5.2] too.

4.3.3.3 Principal Component Analysis

As a second step (Figure [4.2], middle box), the most variable components of the data on the spinning image planes are identified by PCA. To do so, a library of n representative sample trees is required, whilst the variability among the library data should cover the species-specific differences in a representative manner. Therefore, the library trees are chosen as a balanced sample of species and age classes.

The mean of all library data vectors $\vec{d}_1, ..., \vec{d}_n$ is first subtracted from each data vector $\vec{d}$:

$$\vec{d'} = \vec{d} - \frac{1}{n}\sum_{i=1}^{n}\vec{d}_i. \tag{4.1}$$

All n mean-subtracted data vectors $\vec{d'}$ of length d of the library trees then form a $d \times n$ library data matrix $\mathbf{L}$:

$$\mathbf{L} = \left[\vec{d'_1}, ..., \vec{d'_n}\right]. \tag{4.2}$$

The eigenvalue decomposition of the $d \times d$ covariance matrix $\mathrm{Cov}\,(\mathbf{L}) = \mathbf{L} \cdot \mathbf{L}^T$ now yields a set of linearly uncorrelated eigenvectors $\vec{e}_i$ and eigenvalues λ_i:

$$\mathrm{Cov}\,(\mathbf{L}) \cdot \vec{e}_i = \lambda_i \cdot \vec{e}_i. \tag{4.3}$$

Each eigenvector $\vec{e}_i$ represents one dimension of variability in the library data. The magnitude of the eigenvalues λ_i indicates the proportion of variance in $\mathbf{L}$ that is covered by the corresponding eigenvector $\vec{e}_i$. Since the eigenvectors of the library data matrix are of the same dimension as the spinning images, they may be visualized as an image (as in Figure [4.2], right box) and are hereafter referred to as eigen-spin images.

4.3.3.4 Projection to Feature Values

Finally, feature values are generated by projecting the data $\vec{d}$ of an individual spinning image onto the most relevant eigen-spin images (Figure [4.2], right box). As projections onto those eigen-spin images, which represent only little variability in the library data – indicated by small corresponding λ_is – may be omitted, the total number of feature values is greatly reduced. We use only those eigen-spin images for feature calculation, whose λ_i, starting with the largest λ_i, sum up to 50% of the sum of all λ_is. The projection

$$\vec{d} \cdot \vec{e}_i = \sum_{\mathrm{px}} \left(d_{\mathrm{px}} \cdot e_{i\,\mathrm{px}}\right) = f_i \tag{4.4}$$

is the sum over all image pixels weighted with the corresponding entry of the eigen-spin image. Image areas which generally lie outside of the crowns or return constant values have eigen-spin image entries close to zero and therefore contribute less to the feature's values than areas which experience high variability among the library data.

4.3.4 Classification

Segment-wise classification is performed by a random forest (RF) classifier (Breiman, 2001). [As explained in Section 2.3.2, the] RF is a representative of [] ensemble learning methods (Schindler, 2012), that provides a good trade-off between classification accuracy and computational effort (Weinmann et al., 2015). [... In our experiments,] neither the depth of the individual decision trees nor the total number of decision trees show[ed] a significant influence on the classification performance[. Therefore,] no exhaustive parameter optimization [was] required.

4.3.5 Evaluation

Evaluation is performed with respect to standard evaluation metrics, as given in Section 2.5.

4.4 Material

The proposed method is tested for the classification of individual trees in a boreal forest environment. This section introduces the study site (Section [4.4.1]), field data (Section [4.4.2]) and two ALS datasets (Section [4.4.3]) alongside the radiometric corrections applied to them (Section [4.4.4]). Finally, statistics of the individual tree segments used are given (Section [4.4.5]).

4.4.1 Study Area

The study area is located in Hyytiälä (62° N, 24° E), Finland and represents boreal, mostly even-aged and commercially managed forests. Scots pine (*Pinus sylvestris L.*) and Norway spruce (*Picea abies (L.) Karst.*) are the prevailing species, mixed with a small degree of deciduous trees, most of which are birches (*Betula pendula Roth.* and *Betula pubescens Ehrh.*). The study area extends over 2 km × 6 km and contains 175 permanent forest plots, within which each tree has been mapped and described. Among the plots, there are 35 % pine, 49 % spruce and 13 % birch trees. The plots vary in complexity and represent a variety of forest types: managed and pristine forests, mineral and peat soils as well as pristine and drained mires. More information on the site can be found in related publications (Korpela, 2006; Korpela et al., 2010b).

4.4.2 Field Data

The initial mapping of the trees was done either by real-time kinematic positioning (RTK) [] or by a combination of photogrammetric treetop positioning, followed by field triangulation/trilateration (Korpela et al., 2007). The positioning accuracy of stems is generally better than 25-30 cm (Hovi et al., 2016). The tree data used in this study belongs to 115 plots, for which species, crown status and stem DBH were re-measured in May-July 2013. An estimate of tree age was available from historic records and bore core samples (Korpela et al., 2010b).

4.4.3 ALS Data

For our investigations, two leaf-on early summer waveform-recording LiDAR datasets with only two and a half weeks between the acquisitions are used. Table [4.1] summarizes the LiDAR data characteristics. Both sensors differ in receiver and amplifier technology, which is important as it affects the likelihood to detect *first-of-many* and *subsequent* returns. The different pulse lengths further determine if consecutive targets along the ray will be recorded as one or several returns. Furthermore, the different scanning technology results in a different sampling geometry, which may also affect the within crown distribution of waveform returns. Thus, the sensor setup has a significant effect on both the geometrical distribution of waveform returns and the deduced waveform attributes [as explained in Section 2.1].

The Riegl LMS-Q680i sensor uses a rotating polygonal mirror setup, which results in a line-wise scanning pattern. To record the waveform information, this sensor samples returning pulses at a 1 ns rate. The triggering mechanism of the waveform storage is unknown for this instrument, but up to four 80 ns-long sequences were observed per pulse. The LMS-Q680i has two fixed receiver gain channels, and for the lower range of intensities (below 150 units) the amplitude scale is linear with respect to power entering the receiver. As the data were recorded from a relatively high [flying height] with high pulse repetition frequency, all recorded data were within the lower half of the receiver range. The full width half maximum (FWHM) of the system waveform (SWF) (as defined in (Hovi and Korpela, 2014)) was measured to be 4.5 ns at perpendicular incidence on a flat target.

The Leica ALS60 uses an oscillating mirror setup, which results in a scanning pattern of sinusoidal shape. Here, a constant fraction discriminator is employed to detect the first echo within each pulse. This first echo then triggers a 256 ns-long continuous waveform recording at a 1 ns sampling rate. The ALS60 employs an automatic/active gain control (AGC) circuit that changes the receiver gain on the fly by up to 3 dB. Therefore the resulting amplitude values have to be corrected for the varying gain. For this instrument, the FWHM of the SWF was measured as 7.8 ns, while that of the transmitted pulse is only 4 ns according to the system manufacturer.

	Riegl LMS-Q680i	Leica ALS60
date	May 28th, 2013	June 15th, 2013
time	08-09 GTM	21-00 GTM
flying height	760 m	700 m
divergence ($^1/_{e^2}$)	≤0.5 mrad	0.22 mrad
footprint diameter ($^1/_{e^2}$)	40 cm	15 cm
scan zenith angle	30°	15°
strip overlap	75 %	55 %
pulse density	20 $^1/_{m^2}$	10 $^1/_{m^2}$
laser wavelength	1550 nm	1064 nm
$\mathrm{FWHM}_{\mathrm{SWF}}$	4.5 ns	7.8 ns
waveform sampling rate	1 ns	1 ns
# of samples per sequence	multiples of 80	single 256
# of receivers, type	2, low & high gain	1, AGC controlled
scanning pattern	lines	sinusoidal

Table 4.1: Main characteristics of the two datasets.

4.4.4 Radiometric Correction

Both a range-dependent physical [correction] and [a] sensor-specific correction were employed to achieve accurate radiometry.

Physical Range-Dependent Correction

Since the scanning laser beam is divergent and the receiver aperture is of constant size, the incident power is dependent on the range distance ([ρ]) between laser scanner and target. We therefore normalize the recorded signal S_{inc} across the dataset to obtain range-normalized data S_{corr}:

$$S_{\text{corr}} = (S_{\text{inc}} - c) \cdot \left(\frac{\rho}{\rho_{\text{ref}}}\right)^{a}. \tag{4.5}$$

c is a constant, set to the average noise level in the waveform recording data, and [ρ_{ref}] is the average scanning range. The exact relation depends on the target geometry [cf. Section 2.1.3]. According to theory, the exponent a can take values between 2 and 4, where 2 is valid for flat surfaces, 3 for linear structures and 4 for point targets (Wagner et al., 2006). Studies by Gatziolis (2009) and Korpela et al. (2010a) in discrete return intensity data indicated that a value between 2 and 3 provided best fits for different vegetation types. We therefore use $a = 2.5$. The exact optimum for a is in fact species-dependent (Korpela et al., 2010a), but since there is less than 5 % range variation in our data, this effect is small.

Sensor-Specific Amplitude Correction

Furthermore, sensor-specific amplitude corrections are necessary if the amplitude response is non-linear. For the LMS-Q680i, no correction is necessary, as all amplitudes are within the linear part of [the] input range, and therefore the recorded signal S_{rec} is proportional to the incoming signal S_{inc}:

$$S_{\text{inc}} \sim S_{\text{rec}}. \tag{4.6}$$

For the ALS60 however, the impact of the AGC circuit has to be taken into account. We use the model provided by Hovi and Korpela (2014) to correct the recorded signal S_{rec}:

$$S_{\text{inc}} \sim \frac{1}{1 + \text{AGC}_{\text{voltage}} \cdot b} \cdot S_{\text{rec}}, \tag{4.7}$$

where the parameter b has been derived from well-defined homogeneous surfaces of varying reflectance.

4.4.5 Tree Segments

Autonomous individual tree segmentation (Reitberger et al., 2009; Vauhkonen et al., 2012; Strîmbu and Strîmbu, 2015) is not aimed for in the scope of this work. Instead, we use very conservatively chosen segments while also aiming for consistency with a comparable study on the same ALS60 data (Hovi et al., 2016). There, the segments are generated by a watershed algorithm aided by ground truth measurements to optimize the segmentation to produce single tree segments. This process results in a slight selection bias towards larger than average individuals and yields 3630 segments in total. Segments with less than three returns among the group of *only* returns are excluded from classification, since some non-spatial waveform attribute features are ill-defined in these cases. To ensure full comparability to the above-mentioned reference, identical segment boundaries are used for both datasets in this study.

Table [4.2] gives an overview of the number of segments available per species and age class. Furthermore, Table [4.3] gives an overview of the mean number of returns per segment and return type in both datasets. For comparative classification, a fixed set of training segments is used. The same total number of segments is chosen at random from each of the three tree species, corresponding to 75% of all trees in the smallest class (cf. Table [4.2]). In total, 849 tree segments are chosen for training, while the evaluation of the classification results is performed on all 2353 remaining tree segments.

	premature	mature	old	total
Pine	826	157	196	1179
Spruce	740	477	443	1660
Birch	284	45	34	363

Table 4.2: Distribution of the tree segments according to species and three different age classes, representing different periods of tree life (premature: 30-60 years; mature: 60-100 years; old: >100 years).

return type	Riegl LMS-Q680i	Leica ALS60
only	135	104
first-of-many	213	78
subsequent	346	101
only	135	104
first-or-only	348	182
all	694	283

Table 4.3: Mean number of returns per tree segment for different return types (top) and the defined return type groups (bottom).

4.5 Results

To evaluate and optimize the feature design presented in Section [4.3.3], we undertook a series of different experiments. As these experiments are concerned with the way that geometric distribution is captured, they are performed using spin images of local return density and not one particular waveform attribute. First, we challenge the validity of the positioning of the spinning axis (Section [4.5.1]), secondly we evaluate the effect of free parameters (Section [4.5.2]), and finally we evaluate the performance of the spin images of local return density in comparison with α-shape features (Section [4.5.3]).

After this thorough testing and adjustment of the spin image method, we further tested whether the geometric distribution of waveform attributes, captured by spin images of waveform attributes, could indeed provide improvements in tree species classification (Section 4.5.4) and analysed the failure cases which remain in the classification (Section 4.5.5) to see whether there is a certain trend in this method to miss particular types of trees.

4.5.1 Choice of Symmetry Axis Compared to Manual Tree Top Measurement

The placement of the spinning axis is of crucial importance to the spin image method. However, as the tree stems are not necessarily straight and the stem position is not necessarily identical to the centre of the crown, the tree positions at ground are no reliable validation data.

In the Hyytiälä study area, the tree top position of most of the trees, which are used in this study, had been measured semi-manually by photogrammetric monoplotting for a different study in 2011 (Korpela et al., 2011). We assume now that any changes between the acquisition of the aerial image and LiDAR campaigns in 2011 and the LiDAR campaigns in 2013 are mainly related to vertical tree growth, by which the horizontal position is largely unaffected. Figure [4.3] shows species-wise cumulative histograms of the horizontal distance between the 2011 monoplotting positions and the positions of the highest LiDAR returns within a tree segment, which we use to define the rotational axis.

It can be seen, that the median displacements for spruce and pine are 27 cm and 32 cm respectively, while the distributions are almost identical[] in the two datasets. For birch however, the median displacement is 42 cm in the ALS60 data and 58 cm in the LMS-Q680i data. Possible causes, why the results among the two datasets differ only for birches could either be the different prevailing wind conditions on the two acquisition days, as birches sway more in the wind than pine or spruce (Korpela, 2004), the circadian movement of birch branches and foliage (Puttonen et al., 2016), or planimetric offsets in LiDAR data in those parts of the study area, where the birch-rich plots are. The average (per-minute) wind speed was less than $1\,{}^{m}/_{s}$ for the ALS60 and about $4\,{}^{m}/_{s}$ for the LMS-Q680i acquisition (Junninen et al., 2009), which supports the theory of wind sway.

Both the manual tree top measurement and the positioning by the highest LiDAR return are subject to random and systematic errors. The standard deviation of the planimetric monoplotting positions is 10-30 cm, while strip and campaign level offsets in LiDAR have been in the order of 10-20 cm in Hyytiälä. The horizontal tree top displacement in Figure [4.3] is moderately higher than the combination of those two uncertainties. Furthermore, the displacement is mostly below the chosen pixel sizes of the spin images. Therefore, we conclude that the placement of the spinning axis at the highest return of the segment is uncritical and, despite its simplicity, a valid choice.

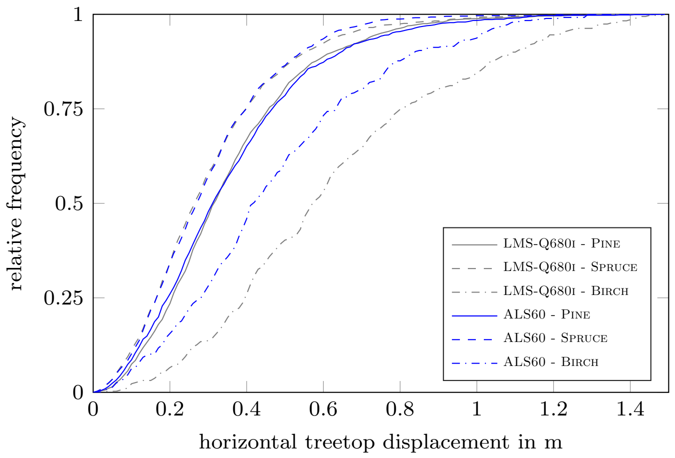


Fig. 4.3: Cumulative histogram of horizontal displacement between tree top position measured manually by photogrammetric monoplotting (2011) and the position of the highest LiDAR return (2013).

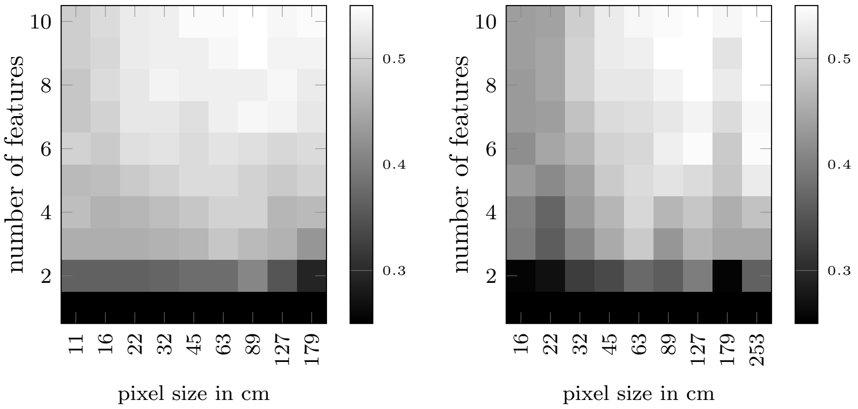


Fig. 4.4: Classification results for spin images of local return density. Results are plotted for varying pixel size and constant vertical reach of 14 m. The diagrams show the κ classification results as grey values for different pixel size and a varying number of features. The **left** diagram shows results based on **LMS-Q680i** data, while the **right** shows those based on **ALS60** data.

4.5.2 Effect of Parameter Choices

As explained in Section 4.3.3.2, the three free parameters of the spin image method are the rotating image's pixel size, horizontal and vertical reach, whereas the horizontal reach is uncritical and can be kept as a fixed value of 5 m. The other two parameters' influence on the classification performance however are evaluated in this section.

The Rotating Image's Pixel Size

The question arose during feature development (cf. Section [4.3.3]), if there is an optimum pixel size which depends on the mean sampling distance (cf. (Johnson and Hebert, 1999)) of the input data. Thus, we define the mean sampling distance $\overline{d_s}$ for ALS data from the 2D pulse density ρ and evaluate the influence of pixel size for a range of $k = 2^{i/2}$ with $i = -2, ..., 6$.

$$\overline{d_s} = {}^1/\sqrt{\rho} \tag{4.8}$$

$$\text{pixel size} = k \cdot \overline{d_s} \tag{4.9}$$

Smaller pixel sizes maintain a higher spatial resolution on the rotating image plane and produce more eigen-spin images, therefore leading to a larger number of spin image features per tree, while larger pixel sizes introduce a higher degree of spatial averaging and condense the distribution information to fewer spin image features per tree. A compromise has to be sought, where the distribution information is captured well without introducing redundancy among the features [or loss of information].

Therefore, the number of features used in classification has to be taken into account in the optimization process. If this was disregarded, an optimization of the total classification accuracy is likely to introduce a bias towards small pixel sizes and a large number of features, while the quality of the single features may actually be reduced. In Figure [4.4], κ classification results are therefore plotted for one to ten features (added in descending λ_i's order) for all pixel sizes in the given range.

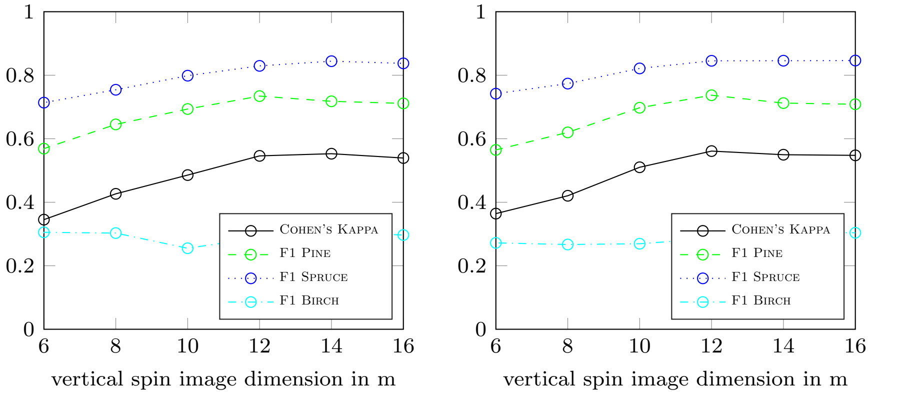


Fig. 4.5: Classification results for spin images based on the number of returns. Results are plotted for a varying vertical dimension of the spin images. The **left** diagram uses data from the **LMS-Q680i** data with a pixel size of ∼89 cm, while the **right** diagram shows **ALS60** data with a pixel size of ∼127 cm.

In Figure [4.4], it can be seen that very small pixel sizes are detrimental, while only a slight decrease is observed for the larger pixel sizes in the given range. These findings are in good agreement with the study of pixel size over mesh resolution by Johnson and Hebert (1999) in Figure 6 ibidem. For the LMS-Q680i (Figure [4.4], left) an optimum in the κ for classifications with ten features is found at a pixel size of 89 cm, while this pixel size also performs well at lower numbers of features, indicating that the individual features are descriptive. For the ALS60 (Figure [4.4], right), an optimum pixel size is found at 127 cm, which again performs well for lower numbers of features. Therefore, pixel sizes of 89 cm and 127 cm are chosen throughout our other experiments.

The Rotating Image's Vertical Dimension

As the rotating image's vertical dimension is of lesser influence to the number of features available (the dependence is only linear, as compared to the quadratic dependence in pixel size), the effect of this parameter is presented in Figure [4.5] by classification results for ten features per classification only. κ increases notably for an increase in the vertical dimension from 6 m to 12 m, while a slight decline is observed again for 16 m. The optimum, which is used in further experiments, is 14 m for the LMS-Q680i and 12 m for the ALS60.

4.5.3 Geometric Descriptiveness of Spin Images Compared to α-Shape Features

To finally evaluate if the spin image method proposed is a good choice to capture geometric distributions within the tree crown, we compare them to α-shape features, which are commonly viewed as potent descriptors of crown geometry (Ko et al., 2012; Vauhkonen et al., 2010). Table [4.4] compiles the respective classification results. It can be seen that the classification results using all available features (all 70/40 spin image features or all 486 α-shape features) are on par. While the spin image features perform slightly better than the α-shape features on the LMS-Q680i data, the situation is the opposite on the ALS60 data. This is probably due to the fact that the LMS-Q680i data, compared to the ALS60 data, has more *subsequent* returns that fall within the tree crown and spin image

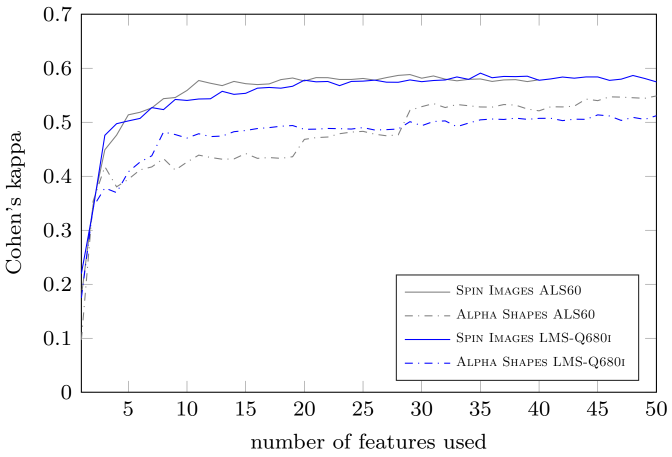


Fig. 4.6: κ classification results, when spin image features (solid lines) of local return density or alpha shape features (dashed lines) are added to the classification one by one. The order is given by an importance measure estimated by the RF classifier.

	Riegl LMS-Q680i		Leica ALS60	
features	#	κ in %	#	κ in %
spin image	best 3	47.6	best 3	44.9
	best 10	54.0	best 10	55.9
	all 70	59.0	all 40	57.6
α-shape	best 3	37.8	best 3	41.8
	best 10	47.1	best 10	42.6
	all 486	56.8	all 486	61.7

Table 4.4: Comparison of κ classification results using either the three best, ten best or all spin image and α-shape features.

features evaluate the return density throughout the crown, while α-shape features are more descriptive of the hull of the point cloud.

However, the aim in using spin image features for tree species classification is not only to build a good descriptor of tree crown geometry, but to utilize the geometric distribution of waveform attributes. As there are always several waveform attributes to be considered, this multiplies the total number of features, while very large numbers of features are known to be potentially detrimental to the classification performance (Hughes, 1968). Therefore, it is important that the geometric distribution is characterized well by a small number of features. Figure [4.6] shows, how the classification result develops, when features are added one by one in the order of importance estimated by the RF classifier. It can be seen both from the results in Table [4.4] and Figure [4.6], that the spin image features excel in comparison to the α-shape features by a higher descriptiveness when only a few features are allowed in the classification.

4.5.4 Species Classification Improvement

To evaluate if the tree species classification accuracy can be improved by considering the spatial distribution of waveform attributes within the crown as captured by spin image features, a detailed comparison is conducted. Classifications are performed using spin image features, non-spatial statistical features and the combination of both feature groups. Each of these are tested for one waveform attribute at a time, as well as for the combination of all waveform attributes. Cohen's κ coefficient (κ) classification results for *only*, *first-or-only* and *all* returns as well as for the combination of features from all three groups are reported in Table [4.6]. Meanwhile, Table [4.5] gives the overall accuracy (OA) for classification results with all waveform attributes.

These results show that the classification with all 70 (ALS60) / 40 (LMS-Q680i) spin images of waveform attributes alone does not perform as well as the classification with all 42 non-spatial waveform attribute features, but that the combination of both feature types yields an overall improvement. The spin image features therefore hold some information complementary to the non-spatial [per segment metrics].

Among the results of individual waveform attributes, the variability among the spin image results is lower than that among the non-spatial results. The gain however, from the combination result of all waveform attribute features, over the mean of the individual results, is always higher for the non-spatial features. This suggests, that the spin image features of the different waveform attributes contain some common or correlated information (e.g. the geometric distribution).

To evaluate the performance of different species in the classification results, Table [4.7] gives an overview of precision, recall and F_1-score. Those results indicate that for pine there is no improvement in the LMS-Q680i data and a small improvement in the ALS60 data. For spruce, some improvement is seen for both datasets. For birch, which is generally much more difficult to classify than the other species, a substantial improvement is observed when spin image features are added. Among both datasets, this improvement is mainly due to an increased precision, but recall values are also slightly improved.

	Riegl LMS-Q680i				Leica ALS60			
	Spin Images	Non-spatial	Combination	Improvement	Spin Images	Non-spatial	Combination	Improvement
only, *first-or-only* and *all*	82.6	94.5	95.1	**0.6**	82.9	94.8	95.9	**1.1**
only	79.6	92.9	92.9	**0.0**	79.7	93.9	95.0	**1.1**
first-or-only	78.0	92.1	93.5	**1.4**	81.1	94.1	95.4	**1.3**
all	80.2	91.0	93.6	**2.6**	81.0	93.6	94.7	**1.1**

Table 4.5: OA in % for classification with waveform attributes for different groups of returns. Improvements are marked in bold.

	Riegl LMS-Q680i				Leica ALS60			
	Spin Images	Non-spatial	Combination	Improvement	Spin Images	Non-spatial	Combination	Improvement
	only, *first-or-only* and *all*							
All	69.6	90.1	91.2	**1.1**	70.0	90.5	92.4	**2.0**
A	63.1	67.5	77.4	9.9	65.0	78.6	84.3	5.7
E	63.9	64.1	78.9	14.8	63.3	87.0	89.1	2.1
L	66.4	68.6	76.7	8.1	51.3	57.1	67.0	9.9
$EQ50$	58.2	70.5	77.1	6.5	50.4	43.2	63.7	20.5
$FWHM$	64.9	71.0	77.6	6.6	54.6	69.8	74.6	4.8
$\#P$	53.1	35.2	60.8	25.7	51.6	23.5	52.5	29.0
	only							
All	65.2	87.1	87.1	**0.0**	64.7	88.9	91.6	**2.8**
A	57.9	51.9	70.6	18.7	56.8	72.7	80.4	7.8
E	58.5	57.6	73.8	16.2	55.5	84.1	87.7	3.6
L	58.9	56.6	68.8	12.2	48.2	49.7	62.8	13.1
$EQ50$	52.5	63.7	71.3	7.6	43.6	36.5	59.6	23.1
$FWHM$	60.1	62.6	71.8	9.2	46.9	59.9	68.2	8.3
$\#P$	52.8	21.2	54.6	33.4	43.6	23.4	48.4	25.1
	first-or-only							
All	61.7	85.8	88.2	**2.4**	67.0	89.3	91.6	**2.3**
A	55.2	62.4	73.5	11.1	59.3	76.5	82.1	5.7
E	52.2	56.1	70.5	14.4	60.2	85.2	87.5	2.3
L	53.5	63.5	72.0	8.5	45.2	49.5	61.1	11.6
$EQ50$	45.3	56.3	68.3	11.9	39.6	35.4	56.1	20.8
$FWHM$	53.2	57.7	71.1	13.3	42.4	56.6	68.4	11.7
$\#P$	44.8	38.2	51.0	12.8	37.9	14.0	43.6	29.6
	all							
All	65.2	83.8	88.3	**4.6**	66.9	88.4	90.3	**1.9**
A	57.0	54.3	73.0	18.7	51.6	75.7	81.9	6.3
E	58.1	50.3	71.6	21.3	58.1	84.1	87.6	3.5
L	58.0	60.8	73.5	12.7	45.3	45.5	58.3	12.8
$EQ50$	33.4	53.2	66.1	12.9	27.6	31.2	49.2	18.0
$FWHM$	42.3	57.0	68.5	11.5	39.3	54.7	66.0	11.3
$\#P$	37.3	37.0	52.7	15.7	32.7	14.0	39.9	25.9

Table 4.6: κ in % for classification with waveform attributes for different groups of returns. The improvements using all attribute types are marked in bold.

	Riegl LMS-Q680i				Leica ALS60			
	Spin Images	Non-spatial	Combination	Improvement	Spin Images	Non-spatial	Combination	Improvement
Pine								
F_1	79.1	94.4	94.4	**0.0**	84.3	95.6	96.3	**0.7**
precision	83.5	95.0	95.3	0.3	84.1	95.5	96.0	0.5
recall	75.1	93.9	93.6	-0.3	84.6	95.7	96.7	1.0
Spruce								
F_1	91.8	97.1	97.8	**0.7**	87.4	95.8	96.8	**1.0**
precision	94.9	98.2	98.6	0.4	92.0	96.5	97.4	0.9
recall	88.8	96.0	97.1	1.1	83.3	95.1	96.3	1.2
Birch								
F_1	47.3	76.8	79.1	**2.3**	51.5	82.5	85.6	**3.1**
precision	35.0	69.1	71.7	2.6	40.5	78.3	83.1	4.8
recall	73.0	86.5	88.2	1.7	70.8	87.1	88.2	1.1

Table 4.7: Precision, recall and F_1-score in % for all three species in classifications with all waveform attributes of *only*, *first-or-only* and *all* returns. Improvements in the F_1-score are marked in bold.

4.5.5 Failure Cases

The given approach is chosen to be sensitive towards the overall crown shape and size, which of course also varies within a species. Indicators linked to this variability may be age as well as environmental factors like stand density or climate. To check if the overall tree size has an effect on the misclassification likelihood, species-wise histograms of correctly and incorrectly classified trees are plotted over the tree's DBH in Figures [4.7] and [4.8]. Furthermore, the percentages of these trees classified as either of the three species are denoted by red lines (spin image features), green lines (non-spatial features) and blue lines (combination of spin image and non-spatial features).

For spruce, the classification performs generally very well across all DBH sizes in both data sets, while occasionally small spruces between 11 and 14 cm are mistaken for pine. Pines are generally classified correctly at those DBH sizes that occur most frequently in the given data (approximately 11-35 cm). Small pines are, in this case, often misclassified as spruce, which is hardly significant though given the very small number of individuals (only 8 pines in total [have a DBH] smaller than 11 cm). Larger pines, however, tend to be mistaken for birches, an effect that grows significantly larger for pines above 29 cm DBH in both data sets, while being slightly more severe in the LMS-Q680i data compared to the ALS60 data.

Birches are generally classified correctly, while there is a stronger confusion with pine among very small trees under 14 cm DBH. Large birches, even though not so frequent in the given data, are largely classified correctly. However, all of the misclassification cases are largely improved in the classification scenario with both spin image and non-spatial waveform attribute features combined. Compared to

	Riegl LMS-Q680i			Leica ALS60		
	as Pine	as Spruce	as Birch	as Pine	as Spruce	as Birch
Spin Images:						
Pine	75.10	5.75	19.15	84.58	8.37	7.06
Spruce	7.64	88.81	3.55	8.87	83.29	7.84
Birch	19.66	7.30	73.03	16.29	12.92	70.79
Non-spatial:						
Pine	93.85	2.32	3.83	95.67	3.53	0.81
Spruce	1.91	95.98	2.11	2.52	95.09	2.39
Birch	11.8	1.69	86.52	4.49	8.43	87.08
Combination:						
Pine	93.55	1.81	4.64	96.67	2.42	0.91
Spruce	1.84	97.07	1.09	2.18	96.25	1.57
Birch	10.67	1.12	88.20	4.49	7.30	88.20

Table 4.8: Confusion matrices normalized to ground truth in % in classifications with all waveform attributes of *only*, *first-or-only* and *all* returns.

the non-spatial waveform attribute features alone, the combination with spin image features especially improves the misclassification cases of medium to large pines in both data sets.

4.6 Discussion

During the first part of our Discussion in Section 4.6.1, we discuss the work presented so far. We start out with the thorough testing of our method and its applicational performance, and continue to compare our results to other publications that were released shortly after (Bruggisser et al., 2017; Shi et al., 2018). These partly covered similar ground in terms of (waveform) feature analysis and relevance assessment. We were able to find congruent results and could relate differences to differences in the respective study sites and sensor technology. We could also integrate our work with other studies in the field on the basis of a comprehensive review paper published in 2021 (Michałowska and Rapiński, 2021). In Section 4.6.2 we further researched the field from a current perspective, trying to evaluate the possible benefit that could be achieved by deep learning technology. Other than in our chapter on point-wise semantic labelling (cf. Section 3.6.2), we could only find two publications that used deep learning strategies on the problem of tree species classification in data of similar characteristics as ours. Neither of them could show a clear advantage compared to studies using advanced application-specific manual features with an appropriate discriminative classifier. However, the years that passed since the initial publication of our work did also bring about new advances in sensor technology, which in turn facilitate the use of more detailed strategies of data analysis. Those are discussed later during Section 4.6.2.

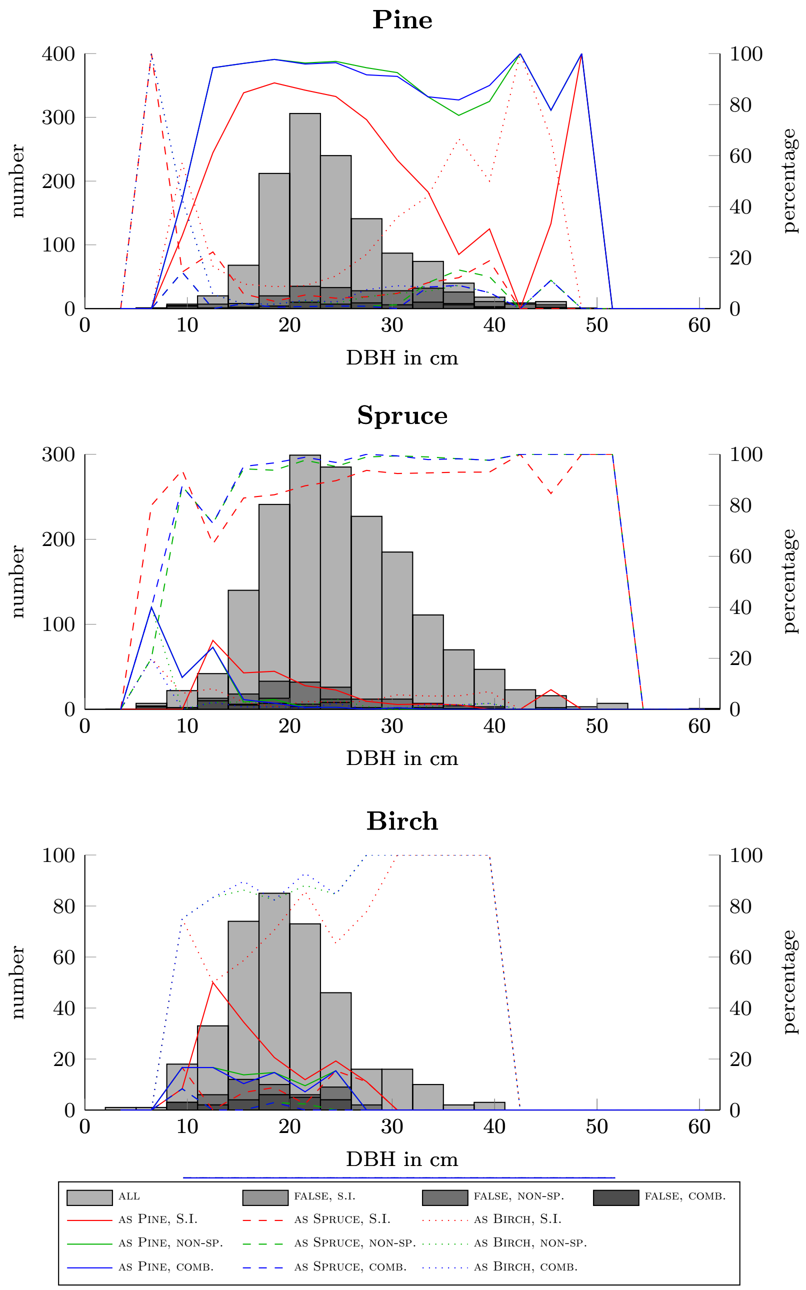


Fig. 4.7: DBH histogram of trees classified according to *only*, *first-or-only* and *all* returns of the **LMS-Q680i** data. Light grey bars denote the total number of trees, medium grey those classified incorrectly by spin image features alone, dark grey those classified incorrectly by non-spatial features alone and black those classified incorrectly by the combination of both feature types.

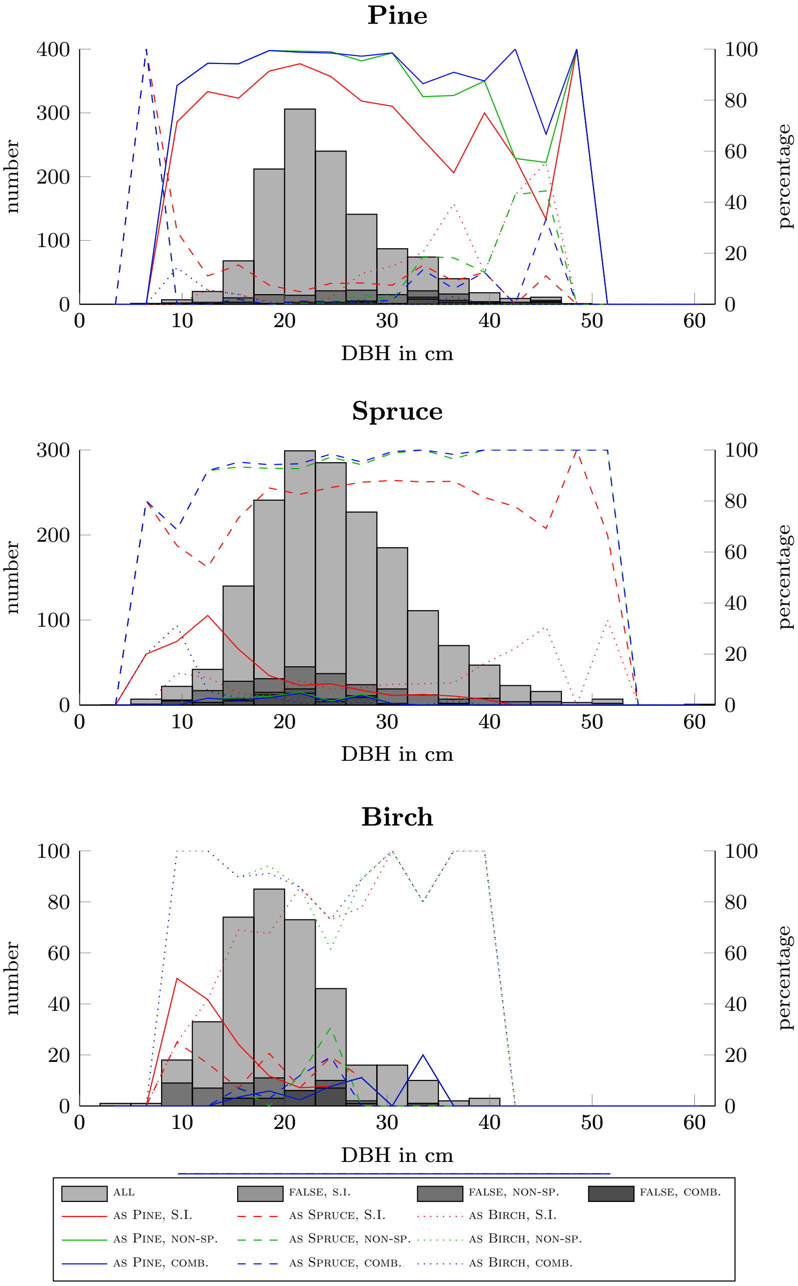


Fig. 4.8: DBH histogram of trees classified according to *only*, *first-or-only* and *all* returns of the **ALS60** data. Light grey bars denote the total number of trees, medium grey those classified incorrectly by spin image features alone, dark grey those classified incorrectly by non-spatial features alone and black those classified incorrectly by the combination of both feature types.

4.6.1 Own Work

In the light of our results presented in [Section 4.5.3], we can say that we succeeded [in developing] a feature type, which is efficient at condensing the geometrical distribution information within tree crowns [into] very few feature[values]. It may be used with both discrete data (return positions) as well as semi-continuous data ([such as] waveform attribute[s]) and performs well compared to α-shape features, which are generally regarded as very potent geometrical features to describe the distribution of discrete data.

[As shown in Section 4.5.2], our feature type is robust with regard to the choice of its two parameters. The parameter values for the best performing pixel sizes correspond to the same value of $k = 4$ in [Equation 4.9], indicating that the difference in pixel size between the two data sets is likely attributed to the different pulse densities (Table [4.1]). For pixel sizes below 45 cm, a decline in classification accuracy is seen in both data sets. This decline is more significant for the less dense ALS60 data, in which the longer SWF also results in a lower number of returns compared to the LMS-Q680i (Table [4.3]). The values of the vertical dimension parameter are relatively large compared to the findings reported by (Li et al., 2013), who state that the structural features from medium-scales were most descriptive within the top 6 m of the tree crown. However, these features were generated from horizontal crown slices, while our features capture the vertical and radial distribution.

Furthermore, Section 4.5.1 shows that the assumptions made when positioning the axis of rotational averaging at the highest return from the tree segment appears to be a valid choice by comparison to manually selected tree top positions in photogrammetric monoplotting. While we found that Holmgren and Persson (2004) and Lin and Hyyppä (2016) used similar assumptions about the general rotation symmetry, our critical evaluation of this assumption as well as that of the grid size and dimension parameters has remained unparalleled.

Last but not least, our tests confirmed that [our feature type] yields similar results on two datasets of different [FWHM_{SWF}], footprint diameter, and pulse density, ergo is robust with regard to those sensor characteristics. This also is not common in the field due to limited availability of such data.

[With regards to tree species classification (cf. Section 4.5.4),] the results for the [non-spatial waveform attribute features of the] ALS60 data are in good agreement with results for the same tree segments and LiDAR data in (Hovi et al., 2016), where a κ of 91 % was reported. The difference is mainly due to differences in the training and validation procedure (leave-one-plot-out), but there were also differences in the classifier (quadratic discriminant analysis [(QDA)]) and the features' choice (restricted exhaustive feature search) and the parameters of the waveform processing. In our experiments, a κ of 95.8 % is achieved for ALS60 data with the non-spatial metrics using all waveform attributes and return type groups by a RF classifier, when we apply the same leave-one-plot-out scenario. The relative performance of the waveform attributes also matches the findings in (Hovi et al., 2016), where E is the most important and A the second most important attribute for the ALS60 data. The waveform attribute calculation and the general training and validation procedure therefore compare well to existing research. However, one has to bear in mind that the performance of the waveform attributes may vary, e.g. with changes in the tree phenology throughout the seasons (Hovi et al., 2016).

The baseline given by the non-spatial metrics is already very high (90.1 % for the LMS-Q680 and 90.5 % for the ALS60) and since κ naturally saturates at 100 %, an overall improvement in the combined case of 1.1 % for the LMS-Q680i data and of 2.0 % for the ALS60 data is significant. This shows that the localization information contained in the spin image features is of additional value to the classification, [but] it cannot replace the non-spatial metrics, which give a detailed statistical frequency analysis of waveform attributes within the top 40 % of the tree crown. The combination of spin images and non-spatial metrics provides an improved classification performance in spruce and birch classification on the LMS-Q680i data and in all three species on the ALS60 data.

By using data acquired by two different sensors, we tried to infer how different sensor characteristics (cf. Table [4.1]) impact our method. Due to the different strip overlay, with which the two campaigns had been flown, the pulse density is about twice as high in the LMS-Q680i data than in the ALS60 data. However, due to a larger footprint diameter and shorter $\mathrm{FWHM}_{\mathrm{SWF}}$, which both increase the likelihood for *subsequent* returns, the mean total number of returns recorded per tree segment (cf. Table [4.3]) is about 2.5 times as high (694 vs. 283) for the LMS-Q680i data to the ALS60 data. The likelihood [of receiving] an *only* return from the LMS-Q680i is thus reduced compared to the ALS60, as *only* returns occur when the footprint diameter is densely filled. Since these dense foliage patches generally occur at the top of trees only, their location information is less valuable, explaining why there is no improved classification accuracy for *only* returns in the LMS-Q680i data in Tables [4.6 and 4.5]. The likelihood for *first-of-many* and *subsequent* returns, however, is higher in the LMS-Q680i data compared to the ALS60 data, making their location information more valuable, which is why *all* returns show a higher improvement in this data. Note however, that the combined classification results using the return type group of *all* returns is lower in the LMS-Q680i data compared to the ALS60 data. This could either be due to the more diverse conditions within the footprint under which returns are recorded, or due to the different reflectivity of the species' material at the respective scanner wavelengths. In summary, the trends of classification improvements achieved among different return type groups follow the respective strengths of the two sensors: the strength of the LMS-Q680i instrument with the wider footprint and narrower $\mathrm{FWHM}_{\mathrm{SWF}}$ is to record more returns, in particular more *first-of-many* and *subsequent* returns, leading to higher classification improvements by considering spin image features in the *first-or-only* and *all* return type groups. The strength of the ALS60 instrument with the narrower footprint is to record more *only* returns (which stem from more standardized conditions within the footprint compared to *first-of-many* or *subsequent* returns), leading to a larger classification improvement in the *only* return type group compared to *first-or-only* or *all.* When combining the different return type groups, these effects seem to even out, while the total classification improvement of considering spin image features is higher for the ALS60 data set.

We expect that the magnitude of the gain is dependent on a number of additional factors, such as the forest complexity. Species and age composition do have an effect as shown in Section [4.5.5], and the density of the stand is important for both the shape and structure of the tree crowns as well as for the segmentation quality. Since the segmentation quality influences the detection rate, it is of crucial applicational importance, but beyond the scope of this study to explore.

The misclassification of large pines [found in Section 4.5.5] is a severe issue from the applicational point of view, since those large trees provide valuable timber. This shortcoming in the performance of the spin images of waveform attributes may be due to the different shape that old pines have compared to younger ones, as they have shorter and more round[ed], dense and asymmetric crowns compared to less than 80-100-year-old pines that [are] still grow[ing] in height. However, as seen in Table [4.8] and in Figures [4.7 and 4.8] (blue lines), the misclassification of (large) pines is significantly reduced in both data sets when the combination of spin image and non-spatial waveform attribute features is used in the classification.

Around the time of publication of our results Bruggisser et al. (2017) found, that in their data (recorded with a LMS-Q680i scanner) the mean energy of the first returns, the mean amplitude of the first returns, and the mean skewness (in a skew normal distribution waveform decomposition) of all returns from one tree crown contribute most significantly to the tree species classification result. In our experience, recorded in Table 4.6, this depends on the sensor used: while for the ALS60 the total energy E of the waveform sequence produced best results in *only*, *first-or-only* and *all* returns per segment, this was less clear for the LMS-Q680i data. There, $EQ50$, $FWHM$, and a combination of spin images and non-spatial features of E showed the best results for *only* returns, A and L showed the best results for *first-or-only* returns, and L and E, as well as a combination of spin images and non-spatial features of A, performed particularly well when considering *all* echo classes. The difference between our findings and those reported by Bruggisser et al. (2017) might stem from a different species composition in the data set. Their study area contained a majority of deciduous trees (90%), while the Hyytiälä study area only contained very few deciduous trees (11%). In their classification (based on A alone, the combination of A, E, and $FWHM$, or the full set) between deciduous and coniferous trees, Bruggisser et al. (2017) reported much better precision and recall for the deciduous class than for the coniferous class.

Similarly, Shi et al. (2018) analysed the importance and correlation among different geometric and radiometric metrics from airborne LiDAR under leaf-on and leaf-off conditions for individual tree species classification. They, too, found that radiometric features contributed a higher accuracy compared to geometric features, and that the combination of complementary features and data, such as data from leaf-on and leaf-off conditions, improved results. They specifically mentioned the intensity of *first-or-only* returns as well as echo width (using a LMS-Q680i laser scanner) as robust features for tree species classification. In our LMS-Q680i data, A was particularly successful in *first-or-only* returns too, and $FWHM$ produced good results throughout all echo class combinations too. These findings are therefore in agreement. Their site contained 69% deciduous trees.

In 2021, Michałowska and Rapiński (2021) published a review of 44 tree species classification studies on ALS data (including our work) in order to identify the most efficient group of LiDAR derived features. Our results scored high throughout their comparisons. They found, that features extracted from full-waveform data yielded the highest overall accuracies, and that both RF and support vector machine (SVM) classifiers produced good results. Concerning geometric features for tree species classification, their review confirmed former analyses (Suratno et al., 2009; Li et al., 2013; Yu et al., 2014) according to which the effectiveness of geometric features depends largely on the point density, and that species classification by geometric features alone is rarely very successful. In general, Michałowska and Rapiński (2021) state that there is no specific group of features, that, when used with a suitable classifier, guarantees high overall accuracy. Instead, it is always required to combine multiple features for good classification results. This agrees with our findings too.

4.6.2 Comparison to Deep Learning and High-Resolution Data

In using data of similar characteristics to ours, we only know of two publications that study the effect of deep learning. As forests are very variable depending on their location and species composition, a quantitative comparison of our method to their results is not necessarily possible due to the differences in study site and data characteristics. Therefore, relevant approaches have to be discussed in higher detail to attempt a qualitative comparison. In the following, we will list and describe the relevant publications that apply deep learning to individual tree species classification in ALS data.

Hamraz et al. (2019) produced multi-view projections for pre-segmented tree crowns to train CNNs on, in order to classify tree crowns as either coniferous or deciduous. Their projection approach actually shares some qualities of our spin image approach. After segmenting individual tree segments and co-registering them with field data, they used one leaf-on and one leaf-off data set of the same region to produce the following projections and supporting features for each segment:

- Horizontal projections onto a 16 m × 16 m square with 12.5 cm pixel size, recording the height above ground of the highest return and the normalized return intensity of the highest return (both for leaf-on and leaf-off conditions). The horizontal projections were complemented by the 2D crown area inferred from the segmentation result.
- Two side profile projections of a 'slice' of 75 cm thickness through the crown apex (positioned in the top middle pixel), measuring 16 m × 16 m with a pixel size of 25 cm, recording the mean intensity in the profile pixel (both for leaf-on and leaf-off conditions). These projections were complemented by the features of tree height and crown width.

In producing these projections, Hamraz et al. (2019) share assumptions we made for our spin images about the theoretically rotation-invariant nature of the tree crown and the accuracy of apex location. To increase the training data size for deep learning however, they did not perform rotational averaging as we did, but sampled over 180 rotational variations of their projections. This might be a crucial point, as we at some point tried to apply a CNN to the spin image representation of our tree segment data too, but without success.

After applying five-to-six-layered CNNs to their different projections, which produce 16 output units, they combined those with the supplementary features for the respective projection, and passed this combination through two more dense layers and a final softmax layer. After excluding cases of mismatch between the LiDAR segments and the field data, they compared different combinations of their approach to traditional feature-based classifiers such as LDA, QDA, SVMs or RFs. The hand-crafted features chosen for this comparison were tree height, crown width, mean intensity for both leaf-on and leaf-off conditions, and the proportion of leaf-on returns to leaf-off returns.

In this comparison the deep learning methods showed only slightly better accuracies for conifers, and statistically more significant improvements for deciduous trees. This could be an effect of conifers being highly underrepresented in the data set, as the authors suggest. It is noteworthy however, that no geometric features for the tree crowns were being used in the comparison to hand-crafted features and traditional classifiers.

Briechle et al. (2019) adapted the PointNet++ architecture to perform a tree species mapping for spruce (coniferous) and beech (deciduous). The basis of their work is a well-established normalized cut segmentation algorithm (Reitberger et al., 2009). In order to generate sufficient training data, they had to use a RF classifier using traditional features (height dependent features, density dependent features and crown shape features) to label a set of 97 000 tree segments based on a manually labelled reference of 918 trees. The RF classification was evaluated on a test data set of 529 tree segments with manual reference. This yielded values of precision = 93%, recall = 80% for coniferous trees, and precision = 82%, recall = 92% for deciduous trees. An application of this classifier to the large set of 97 000 tree segments (without field-measured reference) was used as a reference basis for the PointNet++ application. After hyperparameter adjustment and batch training of 20 m square blocks within epoch blocks of 60 m edge length, this then yielded results of precision = 90%, recall = 79% for coniferous trees and precision = 81%, recall = 91% for deciduous trees. The authors had expected superior results from the PointNet++ application, and attributed its limited performance to errors in the artificially generated training data and edge effects in batch training. In our opinion this is an impressive and interesting feasibility analysis, yet the use of the RF result as basis for training and evaluation of the PointNet++ application limits the quantitative analysis.

Overall, neither of these studies could prove a clear advantage of deep learning approaches over a technically advanced implementation using different subsidiary hand-crafted features with a discriminant classifier on the type of data we were using. There are other data types however, on which deep learning has proven beneficial to tree species classification.

A large body of research has been dedicated to extracting tree attributes from terrestrial laser scanning (TLS) data. A good visualization of the difference in point density and structural detail among data from TLS and ALS ($\sim$10 $^{\text{pulses}}/_{\text{m}^2}$, as in our ALS60 data) can be found in the work of Lin and Hyyppä (2016), Figure 1 *ibid.* There the difference in level of detail is distinctly represented. Due to this deeper level of detail, TLS data has already provided impressive possibilities outside of deep learning approaches for advanced feature extraction of tree structures, e.g. by quantitative structural models (QSMs) that approximate the branching structure and store geometric and topological properties for individual trees (Åkerblom et al., 2017; Terryn et al., 2020). Features, that can be extracted from this kind of representation include not only tree height and crown volume distributions, but also stem characteristics such as volume, verticality and curvature, as well as branching structure or branch characteristics such as length, radii and verticality measures. Xi et al. (2020) extracted 32 such hand-crafted features from TLS scans of 771 individual trees from different sites, assigned to nine different species. They compared an impressive number of different machine learning and deep learning classifiers on this data set. RF and AdaBoost classifiers utilising the hand-crafted features performed similarly well as the highest-ranking deep learning classifiers on the species classification task. Among them were Inception-ResNet-v2, which is a voxel-based network that did not require information other than the raw point cloud, and PointNet++, which was supplied with point-wise training data for a wood vs. foliage classification alongside with the per-segment tree species classification.

Since the early 2000s, unmanned aerial vehicle laser scanning (UAVLS) has emerged as a new option for LiDAR mapping. Down-sized laser scanners, inertial measuring unit (IMU), and global positioning system (GPS), mounted on a drone or low-flying helicopter, enable dense point cloud recordings from an aerial perspective. UAVLS therefore offers a cost-effective alternative to manned aircraft laser scanning for small-to-medium projects, provides higher point density, and is particularly useful in forestry, disaster management and archaeological applications. In forestry, the enhanced structural resolution enables a wider use of deep learning technology or hand-crafted feature extraction. While UAVLS data does not quite reach the accuracy and point density of TLS data, it is much less labour-intensive to acquire over a larger area and in difficult terrain, and still delivers 5-10 times higher point densities than typical ALS. Therefore, this type of data is more suitable for deep learning analysis. Consistent point or pulse density information however is more difficult to achieve for this type of data recording, as the low flying height, the more sideward facing geometry and the less systematic flying patterns are more difficult to quantify. Also, the resulting point density depends largely on the vegetation structure examined, as dense foliage or low stands produce less returns than more permeable or higher stands. A good example of this is Table 2 in an article by Fan et al. (2023), where the number of points recorded per segment for different species is shown. Apart from offering new options by enabling deep learning, this type of high-resolution data is also valuable for explicit modelling and measurement, such as measurements of stem curvature and timber volume (Hyyppä et al., 2022).

Chen et al. (2021) compared tree species classification by deep learning on both TLS and UAVLS recordings of the same sites. They worked directly on the point clouds of semi-manually segmented individual trees, using T-net elements from PointNet and different methods of down-sampling in their network architecture. They also included the radiometric intensity information of each return as a fourth dimension for each point. They found that for TLS data, the overall classification accuracy improved with a down-sampled point cloud, while for UAVLS data, the highest point density available yielded the best results. This indicates that the lower point density of UAVLS compared to TLS might not be a big obstacle compared to the benefits in terms of data acquisition.

Marinelli et al. (2022) too used data acquired by UAVLS. Unfortunately they do not specify the resulting point density in their paper. They apply a multi-view projection CNN approach, based on the projections described in the work of Lin and Hyyppä (2016) (cf. Section 4.2.5). Using the best performing manual features calculated according to Lin and Hyyppä (2016) as a baseline comparison, they reported incremental improvements using PointNet++, a CNN from scratch, and a pre-trained CNN respectively. Unfortunately they did not specify which classifier they used in the traditional machine learning experiment and did not comment on whether those results could be optimized by using more than the unclear number of best-performing hand-crafted features.

Since 2023, a novel UAVLS benchmark data set for semantic labelling and instance segmentation of individual trees is available (Puliti et al., 2023). It comprises UAVLS data from five locations around the globe, representing various forest types. The data is annotated into individual trees (instances) and different semantic classes (e.g. stem, woody branches, live branches, terrain, low vegetation). This data set is intended to foster research in the field by providing easily accessible data which is otherwise costly to acquire and labour-intensive to annotate, and to enhance scientific comparability. Xiang et al. (2024) for example worked on this data and presented a network setup, that would jointly perform stand segmentation (canopy layers, ground), individual tree segmentation and semantic tree component segmentation (stem, live, and dead branches), which allows the retrieval of both tree and stand-wise inventory parameters. Apart from showing impressive proof-of-concept results, they also studied the effect of point density by artificial downsampling and concluded, that their 3D deep learning method is challenged by point densities below $100\,{}^{\mathrm{pts}}/_{\mathrm{m}^2}$.

As of 2024, an even larger benchmark data set for tree species classification has been released (Puliti et al., 2024), aiming to track progress in deep learning model development and to support convergence towards a best practice for species classification. Both point-wise and projection-based frameworks (including data augmentation) have been compared on the basis of this benchmark. The results indicate a general superiority of the projection-based approaches compared to those working directly on the point cloud. The authors promote further efforts to collect an even more extensive database to cover most European species in order to enable very generalized model training, which could prospectively be applied to unknown stands (not included in the model training). There are of course some limitations to this approach, such as possible differences in the quality of the tree segmentation and sensor platform-specific representations.

Last but not least, multispectral LiDAR is promising for species identification. In mounting three laser scanners of different wavelengths on a helicopter, flown at 80 m above ground level, Hakula et al. (2023) acquired a dense multispectral point cloud for tree species classification. They used a novel, hand-crafted layer-wise segmentation algorithm (similar to layer-stacking by Ayrey et al. (2016)) with good results, and a RF classification based on hand-crafted geometric, single-channel reflectance and multi-channel reflectance features. In doing so they could show a substantial gain in overall tree species classification accuracy by the use of multispectral reflectance features, as well as prove that the detection and segmentation of understory trees is also attainable by refining traditional clustering segmentation. Similarly, Axelsson et al. (2023) also showed the benefits of multispectral ALS, while working from an 800 m flying height perspective with point densities around 50 pts/m². They estimated both species composition and species-specific stem volumes at the level of individual trees. Notably, they also followed the traditional machine learning procedure of successive segmentation, feature extraction, feature selection and classification for this purpose. The use of the green laser channel in addition to the more common near-infrared channel proved especially beneficial for the detection and timber volume estimation of deciduous species.

Other approaches of course also focus on the combination of different data types, for example by combining conventional ALS in terms of LiDAR derived metrics per area or pixel with supplementary data such as multispectral aerial imagery. In this case, deep learning is shown to produce improved results compared to traditional machine learning (Gahrouei et al., 2024).

4.7 Conclusion

Overall, the work we did 8 years ago to answer the RQs posed for this thesis was successful. The results document the development of a novel feature type, as proposed in RQ3 for this thesis, that is designed to capture the spatial distribution of laser scanning returns or their waveform attributes within the tree crown to aid in tree species classification in ALS data. To the best of our knowledge, these features [were among] the first to include both the vertical and lateral distribution relative to a central axis of rotational symmetry, as well as [the only ones to date] to allow an analysis of the geometrical distribution of waveform attributes. [(Some modern approaches include intensity as a fourth chanel in the data (Chen et al., 2021), but we have not encountered another encompassing integration of geometry and waveform analysis.)] In a detailed evaluation, we demonstrated that these features perform well in comparison to renowned descriptors of tree crown geometry [at the time], but can also include continuous data values (e.g. waveform attribute values) that are assigned to the geometrical positions. They are robust with regard to sensor characteristics or parameter choices[.]

Our detailed analysis of classification results in Tables 4.6 and 4.5 serves as an answer to RQ4: despite the excellent performance of the non-spatial signal properties captured by waveform attributes derived by (Hovi et al., 2016), our novel spin image feature type could still provide a significant improvement. The detailed per waveform attribute analysis in Table 4.6 also allowed a comparison to independent (later) research which studied the relative importance of different waveform qualities for species classification.

Concerning RQ5, Section 4.5.5 shows how failure cases are distributed across tree size measured in DBH. In both data sets, the correct classification of large pines is fraught with some difficulty when using spin image features only (red lines in Figures 4.7 and 4.8), but this does not show in the combined results (blue lines).

From an applicational point of view, comprising frameworks from field- and ALS data towards a full-stand inventory end-product are desirable. Modern examples towards this end are the work of Axelsson et al. (2023) or Xiang et al. (2024). They comprise both solutions with multispectral ALS or UAVLS data recordings, and techniques from either deep learning or following the hand-crafted feature development approach. The new methods of data acquisition, such as UAVLS and sensor development, as well as the rise of big data techniques, open huge possibilities for automated assessment of vegetation and forest resources. However, our review of current literature in the field indicated, that there were no major improvements that could have been achieved given the kind of data we were using.

Chapter 5
Summary

The layout of this thesis follows along a set of RQs posed in the introduction (page 7). These build upon each other and aim at enhanced scene and data understanding in ALS point clouds of medium pulse density (roughly $5 - 20\,{}^{\text{pulses}}/_{\text{m}^2}$) by the means of geometric features in feature-based supervised classification.

In RQ1 we focused on designing, implementing and testing a geometric sampling feature type for point-wise semantic labelling (inspired by shape distributions (Osada et al., 2002)) that reaches beyond locally homogenous neighbourhoods. Those are presently favoured by the prevailing feature type of covariance features, but the point density of the given data is often not sufficient for a statistically sound representation of homogeneous neighbourhoods in urban surroundings. This became clear in a class-wise analysis of different neighbourhood sizes (cf. Figure 3.7), an analysis of optimum neighbourhood size by minimization of the Shannon entropy (cf. Figure 3.8), and was later supported by complementing literature, which has analysed the weaknesses of covariance features in sampled data on an analytical level (Dittrich et al., 2017). Our sampled feature type of shape distributions instead provides better results as well as an even distribution when tested for different neighbourhood sizes, peaking at a reasonable neighbourhood radius around 2 m (cf. Figure 3.6). Classification accuracy improved throughout our various tests whenever we added shape distribution features (cf. Table 3.9).

In RQ2, we focused on evaluating the effect of different neighbourhood types and scales with respect to different classes. We provided a detailed and insightful analysis which helped optimize classification results (multi-feature-type is better than single-feature-type, multi-scale and multi-neighbourhood-type are better than single-scale or single-neighbourhood-type, and there is no significant preference for a best scale in a multinomial classification task, cf. Table 3.11) and were able to achieve good overall results (cf. Figure 3.16). However, context in the sense of structured prediction (Niemeyer et al., 2016; Steinsiek et al., 2017) can have a substantial impact on the classification results, especially on challenging data such as the Vaihingen benchmark data (cf. Table 3.19). On the less challenging GML data set A however, our results were roughly comparable to those of non-Associative Markov Networks (Shapovalov et al., 2010) (cf. Table 3.16). This could be either due to the different complexity of the data sets and the number of classes considered, or due to differences among the contextual classifiers.

Furthermore, modern data analysis that evolved from deep learning in the meantime has more powerful means of adaptively describing textures and structures on a deeper level of scale understanding than we could model by hand-crafted features from different neighbourhood types and sizes. Our summary of results for the Vaihingen benchmark (cf. Table 3.20) relates our work to these results, while we draw a more detailed qualitative comparison in Section 3.6.2.

On a next level, we explored the possibilities of geometric features in the context of complex vegetation analysis, using tree species classification as a reliable reference. The absence of clear geometric structures in vegetation and a large degree of stand-specific variability make this a challenging field. Also, small-scale structures, such as the quality and distribution of foliage within the laser footprint and its influence on the shape of the reflected waveform, are important indicators of tree species (Hovi et al., 2016; Bruggisser et al., 2017).

Therefore, RQ3 aimed at capturing the geometric distribution of those radiometric waveform properties in few, descriptive features, which led us to implement a feature type inspired by spin images (Johnson and Hebert, 1999). We thoroughly tested the assumptions therein (cf. Section 4.5.1 and 4.5.2) and measured our success in terms of geometric descriptiveness by comparison to α-shape features (Ko et al., 2012; Vauhkonen et al., 2010) (cf. Figure 4.6). From a current point of view, we found other approaches which used similar assumptions on the symmetry of species-related properties for projection-based feature extraction (Lin and Hyyppä, 2016; Marinelli et al., 2022), but could not find a similarly thorough testing of the assumptions elsewhere.

Answering RQ4, we proved the benefits of this feature type for tree species classification (cf. Table 4.5) and analysed the impact both for different return types, waveform attributes and species (cf. Table 4.6 and 4.7). Our review of many sources after the publication of our work (Lin and Hyyppä, 2016; Bruggisser et al., 2017; Michałowska and Rapiński, 2021) indicates that the combination of different feature groups indeed seems to improve tree species classification beyond the scope of any singular feature group.

An analysis of failure cases, as in RQ5, yielded weaknesses related to the rounded shape of old pines compared to young pines, which were therefore mistaken as birch, but only in the absence of non-spatial statistical incidence metrics of waveform attributes (cf. Figures 4.7 and 4.8).

Following along the traditional maxim of feature development for discriminant classifiers, our feature design aimed for a reduction of dimensionality to achieve a more generalized representation. Modern deep learning methods however prefer more variable training data to learn their features, so instead of condensing the information in specific features as we did, they rather work with many different projections or data augmentation to diversify their training set, before having the network learn the generalization. Still, our thorough review of modern literature found clearly superior solutions to the tree species classification problem only for point cloud data of higher pulse densities or for combinations with other data. A study on point density with a successful deep learning framework (Xiang et al., 2024) further indicates that deep learning reaches its limits at the point densities we had available in our data set. Given the data we used, our work would therefore still be difficult to improve on.

Chapter 6
Conclusions and Outlook

Apart from covering its original RQs (cf. Chapter 1 and 5) and at the same time anchoring them within the technical foundations (cf. Chapter 2), this work contains added value by a thorough review of current literature as well as the detailed evaluation of our work from a current hindsight perspective due to the years between publications and writing up.

The original goals of course were set at the time following the traditional maxim of feature development. We brought forward an argument, that sampled geometric features from larger neighbourhoods were more suitable for practical ALS point densities at the time than geometric features targeting homogeneous structures, and supported this argument by experimental evidence. The new sampled feature type allowed us to produce top-performing results in point-wise semantic labelling of urban benchmark data. From today's point of view, our motivation and analysis was later supported by statistical modelling of covariance feature behaviour (Dittrich et al., 2017).

However, the impact of contextual relationships as opposed to point-wise semantic labelling can have a substantial impact in complicated labelling situations, as we found by a comparison of our results to those employing conditional random fields (CRFs) (Niemeyer et al., 2016; Steinsiek et al., 2017). On less challenging data the difference might be less pronounced.

From our in-depth review of deep learning enabled results on the Vaihingen data set we conclude that the main advantage of deep learning algorithms for point-wise semantic labelling seems to be linked to the level of context that can be included via transformer architectures such as an encoder-decoder layout including attention mechanisms. When trying to improve such results further by a subsequent CRF classification, the effect is small (Yu et al., 2022).

It would still be interesting to see for comparison how the result of our multi-scale multi-neighbourhood(type) results, including shape distribution features as well as the more typical feature groups, would perform if combined with a CRF classification.

On our more application-specific topic of tree species classification in ALS data, the review of current literature concludes that good results are best achieved by a combination of data or feature types, and that the performance of geometric features is clearly linked to the point density of the data. This supports our intuition in the setting of our research aims: we aimed at capturing not only the geometric features of the point cloud, but the geometric distribution of waveform properties. This enabled improved species classification results based on a high baseline of statistical waveform attribute features for the pre-segmented dominant trees used in our study. Thorough parameter testing and failure case analysis supported the strength of our method.

Especially our comparison of tree top positioning to photogrammetric monoplotting is highly relevant to date, as other approaches too have followed in projecting returns onto a plane centred around the top of the tree, while not having the means to test and support the tree top positioning accuracy as we did (Lin and Hyyppä, 2016; Marinelli et al., 2022).

A review of current tree species classification approaches, using either deep learning or other methods for data analysis, did not show fundamental improvements compared to our approach when using ALS data of the type we used. In particular, deep learning for tree species classification typically requires higher return densities in the point cloud data (Xiang et al., 2024). In tree species classification and vegetation analysis beyond the individual stand level, deep learning could also provide significant applicational opportunities when large databases are collected throughout a climatic region (Puliti et al., 2024). Practical improvements in LiDAR-based forest inventories are currently being attained using higher resolution laser scanning data with or without deep learning in data analysis (Hyyppä et al., 2022; Xiang et al., 2024) or using multispectral LiDAR recordings (Axelsson et al., 2023; Hakula et al., 2023).

The challenge of tree segmentation, which is the second most important challenge in the field, was not covered by our work. Substantial improvements of classification accuracy in our test cases for the echo class of *all* returns (as opposed to those of early scatterers only) are promising. But since the tree segments used in our study were largely dominant trees, it would require further testing if the geometry of understory trees would support our approach. Difficulties in segmentation, however, pervade and challenge most approaches in the field. In our view, the possibility to train for both segmentation and classification by a joint deep learning architecture is a promising development.

References

Ahokas, E., S. Kaasalainen, J. Hyyppä, and J. Suomalainen, 2006: Calibration of the Optec ALTM3100 laser scanner intensity data using brightness targets. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Marne-la-Vallee, France, Vol. XXXVI-1, 36(A1).

Ahokas, E., X. Yu, J. Oksanen, J. Hyyppä, H. Kaartinen, and H. Hyyppä, 2005: Optimizatioin of the scanning angle for countrywide laser scanning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Enschede, The Netherlands, Vol. XXXVI-3/W19, 115–119.

Åkerblom, M., P. Raumonen, R. Mäkipää, and M. Kaasalainen, 2017: Automatic tree species recognition with quantitative structure models. *Remote Sensing of Environment*, **191**, 1–12.

Alves, L. F. and F. A. Santos, 2002: Tree allometry and crown shape of four tree species in Atlantic rain forest, south-east Brazil. *Journal of Tropical Ecology*, **18**, 245–260.

Armston, J., M. Disney, P. Lewis, P. Scarth, S. Phinn, R. Lucas, P. Bunting, and N. Goodwin, 2013: Direct retrieval of canopy gap probability using airborne waveform LiDAR. *Remote Sensing of Environment*, **134**, 24–38.

Axelsson, A., E. Lindberg, and H. Olsson, 2018: Exploring multispectral ALS data for tree species classification. *Remote Sensing*, **10 (2)**, 183.

Axelsson, C. R., E. Lindberg, H. J. Persson, and J. Holmgren, 2023: The use of dual-wavelength airborne laser scanning for estimating tree species composition and species-specific stem volumes in a boreal forest. *International Journal of Applied Earth Observation and Geoinformation*, **118**, 103 251.

Ayrey, E., S. Fraver, J. A. Kershaw, L. S. Kenefic, D. Hayes, A. R. Weiskittel, and E. E. Roth, 2016: Layer stacking: a novel algorithm for individual forest tree segmentation from LiDAR point clouds. *Canadian Journal of Remote Sensing*, **43 (1)**, 16–27.

Barilotti, A., F. Crosilla, and F. Sepic, 2009: Curvature analysis of LiDAR data for single tree species classification in alpine latitude forests. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Paris, France, Vol. XXXVIII-3/W8, 129–134.

Blomley, R., B. Jutzi, and M. Weinmann, 2016a: 3D semantic labeling of ALS point clouds by exploiting multi-scale, multi-type neighborhoods for feature extraction. *Proceedings of the International Conference on Geographic Object-Based Image Analysis (GEOBIA)*, Enschede, The Netherlands, 1–8.

———, 2016b: Classification of airborne laser scanning data using geometric multi-scale features and different neighbourhood types. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Prague, Czech Republic, Vol. III-3, 169–176.

Blomley, R. and M. Weinmann, 2017: Using multi-scale features for the 3D semantic labeling of airborne laser scanning data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Wuhan, China, Vol. II-4, 1–8.

Blomley, R., M. Weinmann, J. Leitloff, and B. Jutzi, 2014: Shape distribution features for point cloud analysis – a geometric histogram approach on multiple scales. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Zürich, Switzerland, Vol. II-3, 9–16.

Brandtberg, T., 2007: Classifying individual tree species under leaf-off and leaf-on conditions using airborne LiDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, **61 (5)**, 325–340.

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

———, 2001: Random forests. *Machine Learning*, **45 (1)**, 5–32.

Briechle, S., P. Krzystek, and G. Vosselman, 2019: Semantic labelling of ALS point clouds for tree species mapping using the deep neural network PointNet++. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Enschede, The Netherlands, Vol. XLII-2/W13, 951–955.

Brodu, N. and D. Lague, 2012: 3D terrestrial LiDAR data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS Journal of Photogrammetry and Remote Sensing*, **68**, 121–134.

Bruggisser, M., A. Roncat, M. E. Schaepman, and F. Morsdorf, 2017: Retrieval of higher order statistical moments from full-waveform LiDAR data for tree species classification. *Remote Sensing of Environment*, **196**, 28–41.

Cao, Y., J. G. C. Ball, D. A. Coomes, L. Steinmeier, N. Knapp, P. Wilkes, M. Disney, K. Calders, A. Burt, Y. Lin, and T. D. Jackson, 2023: Benchmarking airborne laser scanning tree segmentation algorithms in broadleaf forests shows high accuracy only for canopy trees. *International Journal of Applied Earth Observation and Geoinformation*, **123**, 103 490.

Chang, C.-C. and C.-J. Lin, 2011: LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, **2 (3)**, 27:1–27:27, software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Chehata, N., L. Guo, and C. Mallet, 2009: Airborne LiDAR feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Paris, France, Vol. XXXVIII-3/W8, 207–212.

Chen, C., A. Liaw, and L. Breiman, 2004: *Using random forest to learn imbalanced data.* Technical Report, University of California, Berkeley, USA.

Chen, J., Y. Chen, and Z. Liu, 2021: Classification of typical tree species in laser point cloud based on deep learning. *Remote Sensing*, **13**, 4750.

Cramer, M., 2010: The DGPF test on digital aerial camera evaluation - overview and test design. *Photogrammetrie - Fernerkundung - Geoinformation*, **2**, 99–115.

Criminisi, A. and J. Shotton, 2013: *Decision forests for computer vision and medical image analysis.* Advances in Computer Vision and Pattern Recognition, Springer, London, UK.

Dayal, K. R., S. Durrieu, K. Lahssini, S. Alleaume, M. Bouvier, J. Monnet, J. Renaudd, and F. Revers, 2022: An investigation into LiDAR scan angle impacts on stand attribute predictions in different forest environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, **193**, 314–338.

Demantké, J., C. Mallet, N. David, and B. Vallet, 2011: Dimensionality based scale selection in 3D LiDAR point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Calgary, Canada, Vol. XXXVIII-5/W12, 97–102.

Disney, M., V. Kalogirou, P. Lewis, A. Prieto-Blanco, S. Hancock, and M. Pfeifer, 2010: Simulating the impact of discrete-return LiDAR systems and survey characteristics over young conifer and broadleaf forests. *Remote Sensing of Environment*, **114**, 1546–1560.

Dittrich, A., M. Weinmann, and S. Hinz, 2017: Analytical and numerical investigations on the accuracy and robustness of geometric features extracted from 3D point cloud data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **126**, 195–208.

Dong, P., 2009: Characterization of individual tree crown using three-dimensional shape signatures derived from LiDAR data. *International Journal of Remote Sensing*, **30 (24)**, 6621–6626.

Endres, F., C. Plagemann, C. Stachniss, and W. Burgard, 2009: Unsupervised discovery of object classes from range data using latent Dirichlet allocation. *Robotics: Science and Systems*, Seattle, WA, USA, Vol. V.

Evans, J. S. and A. T. Hudak, 2007: A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments. *IEEE Transactions on Geoscience and Remote Sensing*, **45 (4)**, 1029–1038.

Fan, Z., J. Wei, R. Zhang, and W. Zhang, 2023: Tree species classification based on PointNet++ and airborne laser survey point cloud data enhancement. *Forests*, **14**, 1246.

Filin, S. and N. Pfeifer, 2005: Neighborhood systems for airborne laser data. *Photogrammetric Engineering & Remote Sensing*, **71 (6)**, 743–755.

Frome, A., D. Huber, R. Kolluri, T. Bülow, and J. Malik, 2004: Recognizing objects in range data using regional point descriptors. *Computer Vision - ECCV 2004*, Springer, Heidelberg, Germany, No. 3023 in Lecture Notes in Computer Science, 224–237.

Fuhr, M., E. Lalechère, J.-M. Monnet, and L. Bergès, 2022: Detecting overmature forests with airborne laser scanning (als). *Remote Sensing in Ecology and Conservation*, **8 (5)**, 731–743.

Fukushima, K., 1980: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, **36 (4)**, 193–202.

Gahrouei, O. R., J.-F. Côte, P. Bournival, P. Giguère, and M. Béland, 2024: Comparison of deep and machine learning approaches for Quebec tree species classification using a combination of multispectral and LiDAR data. *Canadian Journal of Remote Sensing*, **50 (1)**, 2359 433.

Gatziolis, D., 2009: LiDAR intensity normalisation in rugged forested terrain. *Proceedings of the Silvilaser 2009 Conference*, College Station, TX, USA, on CD-ROM.

———, 2011: Dynamic range-based intensity normalization for airborne, discrete return LiDAR data of forest canopies. *Photogrammetric Engineering & Remote Sensing*, **77 (3)**, 251–259.

Gerke, M., 2015: Use of the Stair Vision Library within the ISPRS 2D semantic labeling benchmark (Vaihingen). Technical Report, January 2015, DOI: 10.13140/2.1.5015.9683.

Gevaert, C. M., C. Persello, and G. Vosselman, 2016: Optimizing multiple kernel learning for the classification of UAV data. *Remote Sensing*, **8 (12)**, 1025.

Gonzales, R. C. and R. E. Wood, 2002: Histogram equalization. *Digital Image Processing*, Prentice-Hall, Upper Saddle River, NJ, 2d ed., 91–95.

Gressin, A., C. Mallet, and N. David, 2012: Improving 3D LiDAR point cloud registration using optimal neighborhood knowledge. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. I-3, 111–116.

Grilli, E., F. Menna, and F. Remondino, 2017: A review of point clouds segmentation and classification algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Nafplio, Greece, Vol. XLII-2/W3.

Guo, B., X. Huang, F. Zhang, and G. Sohn, 2015: Classification of airborne laser scanning data using JointBoost. *ISPRS Journal of Photogrammetry and Remote Sensing*, **100**, 71–83.

Guo, L., N. Chehata, C. Mallet, and S. Boukir, 2011: Relevance of airborne LiDAR and multispectral image data for urban scene classification using random forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, **66 (1)**, 56–66.

Guo, Y., H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, 2021: Deep learning for 3D point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43 (12)**, 4338–4364.

Habib, A., K. I. Bang, A. Kersting, and D.-C. Lee, 2009: Error budget of LiDAR systems and quality control of the derived data. *Photogrammetric Engineering & Remote Sensing*, **16 (9)**, 1093–1108.

Habib, A., A. Kersting, and K. I. Bang, 2010: Impact of LiDAR system calibration on the relative and absolute accuracy of the adjusted point cloud. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Castelldefels, Spain, Vol. XXXVIII-I/5.

Hackel, T., J. D. Wegner, and K. Schindler, 2016: Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Prague, Czech Republic, Vol. III-3, 177–184.

Hakula, A., L. Ruoppa, M. Lehtomäki, X. Yu, A. Kukko, H. Kaartinen, J. Taher, L. Matikainen, E. Hyyppä, V. Luoma, M. Holopainen, V. Kankare, and J. Hyypä, 2023: Individual tree segmentation and species classification using high-density close-range multispectral laser scanning data. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, **9**, 100 039.

Hamraz, H., M. A. Contreras, and J. Zhang, 2017: Vertical stratification of forest canopy for segmentation of understory trees within small-footprint airborne LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, **130**, 385–392.

Hamraz, H., N. B. Jacobs, M. A. Contreras, and C. H. Clark, 2019: Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, **158**, 219–230.

Hancock, S., J. Armston, Z. Li, R. Gaulton, P. Lewis, M. Disney, F. M. Danson, A. Strahler, C. Schaaf, K. Anderson, and K. J. Gaston, 2015: Waveform LiDAR over vegetation: An evaluation of inversion methods for estimating return energy. *Remote Sensing of Environment*, **164**, 208–224.

Hancock, S., P. Lewis, M. Foster, M. Disney, and J.-P. Muller, 2012: Measuring forest with dual wavelength LiDAR: A simulation study over topography. *Agricultural and Forest Meteorology*, **161**, 123–133.

Hanson, S. J. and L. Y. Pratt, 1988: Comparing biases for minimal network construction with back-propagation. *Proceedings of the 1st International Conference on Neural Information Processing Systems (NeurIPS)*, 177–185.

Harikumar, A., F. Bovolo, and L. Bruzzone, 2017: An internal crown geometric model for conifer species classification with high-density LiDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, **55 (5)**, 2924–2940.

He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 770–778.

Heinzel, J. and B. Koch, 2011: Exploring full-waveform LiDAR parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, **13**, 152–160.

Höfle, B. and N. Pfeifer, 2007: Correction of laser scanning intensity data: Data and model-driven approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, **62**, 415–433.

Holmgren, J. and Å. Persson, 2004: Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, **90 (4)**, 415–423.

Holmgren, J., Å. Persson, and U. Söderman, 2008: Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing*, **29 (5)**, 1537–1552.

Hong, Y., S. Liu, Z. Li, X. Huang, P. Jiang, Y. Xu, C. Wu, H. Zhou, Y. Zhang, H. Ren, Z. Li, J. Jia, Q. Zhang, C. Li, F. Xu, J. Wang, and J. Pan, 2024: Airborne single-photon LiDAR towards a small-sized and low-power payload. *Optica*, **11 (5)**, 612–618.

Hovi, A., 2015: Towards an enhanced understanding of airborne LiDAR measurements of forest vegetation. *Dissertationes Forestales*, Finnish Society of Forest Science, Vol. 200.

Hovi, A., L. Korhonen, J. Vauhkonen, and I. Korpela, 2016: LiDAR waveform features for tree species classification and their sensitivity to tree- and acquisition related parameters. *Remote Sensing of Environment*, **173**, 224–237.

Hovi, A. and I. Korpela, 2014: Real and simulated waveform-recording LiDAR data in juvenile boreal forest vegetation. *Remote Sensing of Environment*, **140**, 665–678.

Hu, H., D. Munoz, J. A. Bagnell, and M. Hebert, 2013: Efficient 3-D scene analysis from streaming data. *Proceedings of the IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, 2297–2304.

Hughes, G. F., 1968: On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, **14 (1)**, 55–63.

Hyyppä, E., A. Kukko, H. Kaartinen, X. Yu, J. Muhojoki, T. Hakala, and J. Hyyppä, 2022: Direct and automatic measurements of stem curve and volume using a high-resolution airborne laser scanning system. *Science of Remote Sensing*, **5**, 100 050.

Ioannidou, A., E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, 2017: Deep learning advances in computer vision with 3D data: A survey. *ACM Computing Surveys*, **50 (2)**, 20.

Ioffe, S. and C. Szegedy, 2015: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, Lille, France, 448–456.

Isenburg, M., 2015: LAStools - efficient tools for LiDAR processing, version 150304. `http://lastools.org`, accessed: 5.1.2016.

Jiang, M., Y. Wu, and C. Lu, 2018: PointSIFT: A SIFT-like network module for 3D point cloud semantic segmentation. `https://arxiv.org/abs/1807.00652`, accessed: 10.12.2024.

Johnson, A. E. and M. Hebert, 1999: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21 (5)**, 433–449.

Junninen, H., A. Lauri, P. Keronen, P. Aalto, V. Hiltunen, P. Hari, and M. Kulmala, 2009: Smart-SMEAR: on-line data exploration and visualization tool for SMEAR stations. *Boreal Environment Research*, **14**, 447–457.

Jutzi, B., 2015: Methoden zur automatischen Szenencharakterisierung basierend auf aktiven optischen Sensoren für die Photogrammetrie und Fernerkundung, Habilitation, veröffentlicht durch das Karlsruher Institut für Technologie (KIT). `https://publikationen.bibliothek.kit.edu/1000049140`, accessed: [18.01.2025].

Jutzi, B. and H. Gross, 2009a: Nearest neighbor classification on laser point clouds to gain object structures from buildings. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Hannover, Germany, Vol. XXXVIII-1-4-7/W5.

———, 2009b: Normalization of LiDAR intensity data based on range and surface incidence angle. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Paris, France, Vol. XXXVIII-3/W8, 213–218.

Jutzi, B. and U. Stilla, 2006: Range determination with waveform recording laser systems using a Wiener filter. *ISPRS Journal of Photogrammetry and Remote Sensing*, **61 (2)**, 95–107.

Kaasalainen, S., H. Hyyppä, A. Kukko, P. Litkey, E. Ahokas, J. Hyyppä, H. Lehner, A. Jaakkola, J. Suomalainen, A. Akujärvi, M. Kaasalainen, and U. Pyysalo, 2009: Radiometric calibration of LiDAR intensity with commercially available reference targets. *IEEE Transactions on Geoscience and Remote Sensing*, **47 (2)**, 588–598.

Kaasalainen, S., J. Hyyppä, P. Litkey, H. Hyyppä, E. Ahokas, A. Kukko, and H. Kaartinen, 2007: Radiometric calibration of ALS intensity. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Espoo, Finland, Vol. XXXVI-3/W52, 201–205.

Kandare, K., M. Dalponte, H. O. Ørka, L. Frizzera, and E. Næsset, 2017: Prediction of species-specific volume using different inventory approaches by fusing airborne laser scanning and hyperspectral data. *Remote Sensing*, **9 (5)**, 400.

Kato, A., L. M. Moskal, P. Schiess, M. E. Swanson, D. Calhoun, and W. Stuetzle, 2009: Capturing tree crown formation through implicit surface reconstruction using airborne LiDAR data. *Remote Sensing of Environment*, **113 (6)**, 1148–1162.

Kellner, J. R., J. Armston, M. Birrer, K. C. Cushman, L. Duncanson, C. Eck, C. Falleger, B. Imbach, K. Král, M. Krůček, J. Trochta, T. Vrška, and C. Zgraggen, 2019: New opportunities for forest remote sensing through ultra-high-density drone LiDAR. *Surveys in Geophysics*, **40**, 959–977.

Khoshelham, K. and S. J. Oude Elberink, 2012: Role of dimensionality reduction in segment-based classification of damaged building roofs in airborne laser scanning data. *Proceedings of the International Conference on Geographic Object Based Image Analysis*, Rio de Janeiro, Brazil, 372–377.

Kim, S., T. Hinckley, and D. Briggs, 2011: Classifying individual tree genera using stepwise cluster analysis based on height and intensity metrics derived from airborne laser scanner data. *Remote Sensing of Environment*, **115 (12)**, 3329–3342.

Kim, S., R. J. McGaughey, H.-E. Andersen, and G. Schreuder, 2009: Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*, **113 (8)**, 1575–1586.

Ko, C., G. Sohn, and T. K. Remmel, 2012: A comparitive study using geometric and vertical profile features derived from airborne LiDAR for classifying tree genera. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. I-3, 129–134.

———, 2013: Tree genera classification with geometric features form high-density airborne LiDAR. *Canadian Journal of Remote Sensing*, **39 (sup1)**, S73–S85.

Koike, T., M. Kitao, Y. Maruyama, S. Mori, and T. T. Lei, 2001: Leaf morphology and photosynthetic adjustments among deciduous broad-leaved trees within the vertical canopy profile. *Tree Physiology*, **21**, 951–958.

Komarichev, A., Z. Zhong, and J. Hua, 2019: A-CNN: Annularly convolutional neural networks on point clouds. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 7421–7430.

Korpela, I., 2004: Individual tree measurements by means of digital aerial photogrammetry. *Silva Fennica Monographs*, **3**, 1–93.

———, 2006: Geometrically accurate time series of archived aerial images and airborne LiDAR data in a forest environment. *Silva Fennica*, **40**, 109–126.

Korpela, I., V. Heikkinen, E. Honkavaara, F. Rohrbach, and T. Tokola, 2011: Variation and directional anisotropy of reflectance at the crown scale – Implications for tree species classification in digital aerial images. *Remote Sensing of Environment*, **115 (8)**, 2062–2074.

Korpela, I., A. Hovi, and L. Korhonen, 2013: Backscattering of individual LiDAR pulses from forest canopies explained by photogrammetrically derived vegetation structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, **83**, 81–93.

Korpela, I., H. O. Ørka, J. Hyyppä, V. Heikkinen, and T. Tokola, 2010a: Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, **65 (4)**, 369–379.

Korpela, I., H. O. Ørka, M. Maltamo, T. Tokola, and J. Hyyppä, 2010b: Tree species classification using airborne LiDAR – Effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fennica*, **44**, 319–339.

Korpela, I., T. Tuomola, and E. Välimäki, 2007: Mapping forest plots: An efficient method combining photogrammetry and field triangulation. *Silva Fennica*, **41**, 457–469.

Korpela, I. S., 2008: Mapping of understory lichens with airborne discrete-return LiDAR data. *Remote Sensing of Environment*, **112 (10)**, 3891–3897.

Kotsiantis, S. B., 2007: Supervised machine learning: A review of classification techniques. *Informatica*, **31**, 249–268.

Kraus, K. and N. Pfeifer, 1998: Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **53**, 193–203.

Lafferty, J., A. McCallum, and F. Pereira, 2001: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the 18th International Conference on Machine Learning (ICML)*, Williamstown, MA, USA, 282–289.

Landrieu, L., C. Mallet, and M. Weinmann, 2017a: Comparison of belief propagation and graph-cut approaches for contextual classification of 3D LiDAR point cloud data. *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, Fort Worth, TX, USA, 1–4.

Landrieu, L., H. Raguet, B. Vallet, C. Mallet, and M. Weinmann, 2017b: A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, **132**, 102–118.

Latifi, H., F. E. Fassnacht, J. Müller, A. Tharani, S. Dech, and M. Heurich, 2015: Forest inventories by LiDAR data: A comparison of single tree segmentation and metric-based methods for inventories of a heterogeneous temperate forest. *International Journal of Applied Earth Observation and Geoinformation*, **42**, 162–174.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998: Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86 (11)**, 2278–2324.

Lee, I. and T. Schenk, 2002: Perceptual organization of 3D surface points. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Graz, Austria, Vol. XXXIV-3A, 193–198.

Li, H., Z. Xu, G. Taylor, C. Studer, and T. Goldstein, 2018: Visualizing the loss landscape of neural nets. *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 31, 6389–6399.

Li, J., B. Hu, and T. L. Noland, 2013: Classification of tree species based on structural features derived from high density LiDAR data. *Agricultural and Forest Meteorology*, **171-172**, 104–114.

Li, Q., 2008: Decomposition of airborne laser scanning waveform data based on EM algorithm. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XXXVII-B1, 211–217.

Liaw, A. and M. Wiener, 2002: Classification and regression by randomForest. *R News*, **2/3**, 18–22.

Lin, Y. and J. Hyyppä, 2016: A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, **46**, 45–55.

Lindberg, E., J. Holmgren, J. W. K. Olofsson, and H. Olsson, 2010: Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. *International Journal of Remote Sensing*, **31 (5)**, 1175–1192.

Lindberg, E., K. Olofsson, J. Holmgren, and H. Olsson, 2012: Estimation of 3D vegetation structure from waveform and discrete return airborne laser scanning data. *Remote Sensing of Environment*, **118**, 151–161.

Linsen, L. and H. Prautzsch, 2001: Local versus global triangulations. *Proceedings of Eurographics*, Manchester, UK, 257–263.

Litkey, P., P. Rönnholm, J. Lumme, and X. Liang, 2007: Waveform features for tree identification. *International Archives of Photogrammetry and Remote Sensing*, Espoo, Finland, Vol. XXXVI-3/W52, 258–263.

Liu, J., A. Skidmore, S. Jones, T. Wang, M. Heurich, X. Zhu, and Y. Shi, 2018: Large off-nadir scan angle of airborne LiDAR can severely affect the estimates of forest structure metrics. *ISPRS Journal of Photogrammetry and Remote Sensing*, **136**, 13–25.

Lodha, S. K., D. M. Fitzpatrick, and D. P. Helmbold, 2007: Aerial LiDAR data classification using AdaBoost. *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*, Montreal, Canada, 435–442.

Lodha, S. K., E. J. Kreps, D. P. Helmbold, and D. Fitzpatrick, 2006: Aerial LiDAR data classification using support vector machines (SVM). *Proceedings of the International Symposium on 3D Data Processing, Visualization, and Transmission*, Chapel Hill, NC, USA, 567–574.

Lowe, D. G., 2004: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, **60**, 91–110.

Mallet, C. and F. Bretar, 2009: Full-waveform topographic LiDAR: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**, 1–16.

Mallet, C., F. Bretar, M. Roux, U. Soergel, and C. Heipke, 2011: Relevance assessment of full-waveform LiDAR data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **66 (6)**, S71–S84.

Mandelburger, G. and H. Lehner, 2019: Single photon LiDAR - Grundlagen und erste Evaluierungsergebnisse. *Dreiländertagung der DGPF, der OVG und der SGPF*, Wien, Österreich, Vol. 28, 443–457.

Mao, Y., K. Chen, W. Diao, X. Sun, X. Lu, K. Fu, and M. Weinmann, 2022: Beyond single receptive field: A receptive field fusion-and-stratification network for airborne laser scanning point cloud classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **188**, 45–61.

Marinelli, D., C. Paris, and L. Bruzzone, 2022: An approach based on deep learning for tree species classification in LiDAR data acquired in mixed forest. *IEEE Geoscience and Remote Sensing Letters*, **19**, 7004 305.

Mattheck, C., 1991: *Trees: the mechanical design.* Springer, Berlin, Germany.

Maturana, D. and S. Scherer, 2015: VoxNet: A 3D convolutional neural network for real-time object recognition. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, 922–928.

May, N. C. and C. K. Toth, 2007: Point positioning accuracy of airborne LiDAR systems: a rigorous analysis. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Munich, Germany, Vol. XXXVI-3/W49B, 107–111.

McCulloch, W. and W. Pitts, 1943: A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, **5 (4)**, 115–133.

Menalled, F. D. and M. J. Kelty, 2001: Crown structure and biomass allocation strategies of three juvenile tropical tree species. *Plant Ecology*, **152**, 1–11.

Michałowska, M. and J. Rapiński, 2021: A review of tree species classification based on airborne LiDAR data and applied classifiers. *Remote Sensing*, **13 (3)**.

Morgan, N. and H. Bourlard, 1990: Generalization and parameter estimation in feedforward nets: Some experiments. *Neural Networks*, **3 (5)**, 519–526.

Morsdorf, F., O. Frey, E. Meier, K. I. Itten, and B. Allgöwer, 2008: Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*, **29 (5)**, 1387–1406.

Morsdorf, F., E. Meier, B. Kötz, K. I. Itten, M. Dobbertin, and B. Allgöwer, 2004: LiDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, **92 (3)**, 353–362.

Morsdorf, F., C. Nichol, T. Malthus, and I. H. Woodhouse, 2009: Assessing forest structural and physiological information content of multi-spectral LiDAR waveforms by radiative transfer modelling. *Remote Sensing of Environment*, **113**, 2152–2163.

Morsy, S., A. Shaker, and A. El-Rabbany, 2017: Clustering of multispectral airborne laser scanning data using gaussian decomposition. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Wuhan, China, Vol. XLII-2/W7, 269–276.

Munoz, D., N. Vandapel, and M. Hebert, 2008: Directional associative Markov network for 3-D point cloud classification. *Proceedings of the International Sym-*

posium on 3D Data Processing, Visualization and Transmission, Atlanta, GA, USA, 63–70.

Neuenschwander, A. L., L. A. Magruder, and M. Tyler, 2009: Landcover classification of small-footprint full-waveform LiDAR data. *Journal of Applied Remote Sensing*, **3 (1)**, 033 544.

Ni-Meister, W., D. L. B. Jupp, and R. Dubayah, 2001: Modeling LiDAR waveforms in heterogenous and discrete canopies. *IEEE Transactions on Geoscience and Remote Sensing*, **39 (9)**, 1943–1958.

Nielsen, M. A., 2015: *Neural Networks and Deep Learning.* published online, accessed: [30.10.2024].

Niemeyer, J., C. Mallet, F. Rottensteiner, and U. Sörgel, 2011a: Conditional random fields for the classification of LiDAR point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Hannover, Germany, Vol. XXXVIII-4/W19, 209–214.

Niemeyer, J., F. Rottensteiner, and U. Sörgel, 2012: Conditional random fields for LiDAR point cloud classification in complex urban areas. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. I-3, 263–268.

———, 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., F. Rottensteiner, U. Sörgel, and C. Heipke, 2016: Hierarchical higher order CRF for the classification of airborne LiDAR point clouds in urban areas. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Prague, Czech Republic, Vol. XLI-B3, 655–662.

Niemeyer, J., J. Wegner, C. Mallet, F. Rottensteiner, and U. Sörgel, 2011b: Conditional random fields for urban scene classification with full waveform LiDAR data. *ISPRS Conference on Photogrammetric Image Analysis*, Munich, Germany, 233–244.

Nong, X., W. Bai, and G. Liu, 2023: Airborne LiDAR point cloud classification using PointNet++ network with full neighbourhood features. *PLoS ONE*, **18 (2)**, e0280 346.

Ørka, H. O., T. Gobakken, E. Næsset, L. Ene, and V. Lien, 2012: Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. *Canadian Journal of Remote Sensing*, **38 (2)**, 125–138.

Ørka, H. O., E. Næsset, and O. M. Bollandsås, 2009: Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, **113 (6)**, 1163–1174.

Osada, R., T. Funkhouser, B. Chazelle, and D. Dobkin, 2002: Shape distributions. *ACM Transactions on Graphics*, **21 (4)**, 807–832.

Packalén, P. and M. Maltamo, 2008: Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research*, **38 (7)**, 1750–1760.

Persson, Å., J. Holmgren, and U. Söderman, 2002: Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, **68 (9)**, 925–932.

Persson, Å., U. Söderman, J. Töpel, and S. Ahlberg, 2005: Visualization and analysis of full-waveform airborne laser scanner data. *International Archives of Photogrammetry and Remote Sensing*, Enschede, Netherlands, Vol. XXXVI-3/W19, 228–233.

Polewski, P., W. Yao, M. Heurich, P. Krzystek, and U. Stilla, 2014: Detection of fallen trees in ALS point clouds by learning the normalized cut similarity function from simulated samples. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Zurich, Switzerland, Vol. II-3, 111–118.

Popescu, S. C., R. H. Wynne, and R. F. Nelson, 2003: Measuring individual tree crown diameter with LiDAR and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, **29 (5)**, 564–577.

Popescu, S. C. and K. Zhao, 2008: A voxel-based LiDAR method for estimating crown base height for deciduos and pine trees. *Remote Sensing of Environment*, **112 (3)**, 767–781.

Puliti, S., E. R. Lines, J. Müüllerová, J. Frey, Z. Schindler, A. Straker, M. J. Allen, L. Winiwarter, N. Rehush, H. Hristova, B. Murray, K. Calders, L. Terryn, N. Coops, B. Höfle, S. Junttila, M. Krůček, G. Krok, K. Král, S. R. Levick, L. Luck, A. Missarov, M. Mokroš, H. J. F. Owen, K. Stereńczak, T. P. Pitkänen, N. Puletti, N. Saarinen, C. Hopkinson, C. Torresan, E. Tomelleri, H. Weiser, and R. Astrup, 2024: Benchmarking tree species classification from proximally-sensed laser scanning data: introducing the FOR-species20K dataset. `https://arxiv.org/abs/2408.06507`, accessed: [01.12.2024].

Puliti, S., G. Pearse, P. Surový, L. Wallace, M. Hollaus, M. Wielgosz, and R. Astrup, 2023: FOR-instance: a UAV laser scanning benchmark dataset for semantic and instance segmentation of individual trees. `https://arxiv.org/abs/2309.01279`, accessed: [01.12.2024].

Puttonen, E., C. Briese, G. Mandlburger, M. Wieser, M. Pfennigbauer, A. Zlinszky, and N. Pfeifer, 2016: Quantification of overnight movement of birch (*Betula pendula*) branches and foliage with short interval terrestrial laser scanning. *Frontiers in Plant Science*, **7**, 222.

Qi, C. R., L. Li, H. Su, and L. J. Guibas, 2017a: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, 5099–5108.

Qi, C. R., H. Su, K. Mo, and L. J. Guibas, 2017b: PointNet: Deep learning on point sets for 3D classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 652–660.

Räty, J., J. Vauhkonen, M. Maltamo, and T. Tokola, 2016: On the potential to predetermine dominant tree species based on sparse-density airborne laser scanning data for improving subsequent predictions on species-specific timber volumes. *Forest Ecosystems*, **3**, 1.

Reitberger, J., M. Heurich, P. Krzystek, and U. Stilla, 2007: Single tree detection in forest areas with high-density LiDAR data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Munich, Germany, Vol. XXXVI-3/W49B, 139–144.

Reitberger, J., P. Krzystek, and U. Stilla, 2008: Analysis of full waveform LiDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing*, **29 (5)**, 1407–1431.

Reitberger, J., C. Schnörr, P. Krzystek, and U. Stilla, 2009: 3D segmentation of single trees exploiting full waveform LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64 (6)**, 561–574.

Richter, K., R. Blaskow, N. Stelling, and H.-G. Maas, 2015: Reference value provision schemes for attenuation correction of full-waveform airborne laser scanner data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, La Grande Motte, France, Vol. III-2, 65–72.

Romanczyk, P., 2015: Extraction of vegetation biophysical structure from small-footprint full-waveform LiDAR signals. Ph.D. thesis, Rochester Institute of Technology, Rochester, NY, USA.

Rottensteiner, F., G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf, 2012: The ISPRS benchmark on urban object classification and 3D building reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. I-3, 293–298.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning representations by back-propagating errors. *Nature*, **323 (6088)**, 533–536.

Rusu, R. B., N. Blodow, and M. Beetz, 2009: Fast point feature histograms (FPFH) for 3D registration. *Proceedings of the IEEE International Conference on Robotics and Automation*, Kobe, Japan, 3212–3217.

Rutzinger, M., F. Rottensteiner, and N. Pfeifer, 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.

Schmidt, A., J. Niemeyer, F. Rottensteiner, and U. Soergel, 2014: Contextual classification of full waveform LiDAR data in the Wadden Sea. *IEEE Geoscience and Remote Sensing Letters*, **11 (9)**, 1614–1618.

Schölkopf, B., 1997: *Support Vector Learning*. Oldenbourg Verlag, München, Germany.

Shapovalov, R., A. Velizhev, and O. Barinova, 2010: Non-associative Markov networks for 3D point cloud classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Saint-Mandé, France, Vol. XXXVIII-3A, 103–108.

Shi, Y., T. Wang, A. K. Skidmore, and M. Heurich, 2018: Important LiDAR metrics for discriminating forest tree species in Central Europe. *ISPRS Journal of Photogrammetry and Remote Sensing*, **137**, 163–174.

Silva, C. A., C. Klauberg, Â. M. K. Hentz, A. P. Dalla Corte, U. Ribeiro, and V. Liesenberg, 2018: Comparing the performance of ground filtering algorithms for terrain modeling in a forest environment using airborne LiDAR data. *Floresta a Ambiente*, **25 (2)**, e20160 150.

Song, H. and J. Jung, 2023: An object-based ground filtering of airborne LiDAR data for large-area DTM generation. *Remote Sensing*, **15 (16)**, 4105.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 2014: Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, **15**, 1929–1958.

Steinsiek, M., P. Polewski, W. Yao, and P. Krzystek, 2017: Semantische Analyse von ALS- und MLS-Daten in urbanen Gebieten mittels Conditional Random Fields. *Tagungsband der 37. Wissenschaftlich-Technischen Jahrestagung der DGPF*, Würzburg, Germany, Vol. 26, 521–531.

Stephens, P., P. Watt, D. Loubser, A. Haywood, and M. Kimberley, 2007: Estimation of carbon stocks in New Zealand planted forests using airborne scanning LiDAR. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Espoo, Finland, Vol. XXXVI-3/W52, 389–394.

Strîmbu, V. F. and B. M. Strîmbu, 2015: A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **104**, 30–43.

Su, H., S. Maji, E. Kalogerakis, and E. Learned-Miller, 2015: Multi-view convolutional neural networks for 3D shape recognition. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 945–953.

Suratno, A., C. Seielstad, and L. Queen, 2009: Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64 (6)**, 683–693.

Terryn, L., K. Calders, M. Disney, N. Origo, Y. Malhi, G. Newnham, P. Raumonen, M. Åckerblom, and H. Verbeeck, 2020: Tree species classification using structural features derived from terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, **168**, 170–181.

Tombari, F., S. Salti, and L. Di Stefano, 2010: Unique signatures of histograms for local surface description. *Proceedings of the European Conference on Computer Vision (ECCV)*, No. 6313 in Heraklion, Greece, 356–369.

———, 2013: Performance evaluation of 3D keypoint detectors. *International Journal of Computer Vision*, **102**, 198–220.

Ussyshkin, V. and L. Theriault, 2011: Airborne LiDAR: Advances in discrete return technology for 3D vegetation mapping. *Remote Sensing*, **3**, 416–434.

Vain, A., X. Yu, S. Kaasalainen, and J. Hyyppä, 2010: Correcting airborne laser scanning intensity data for automatic gain control effect. *IEEE Geoscience and Remote Sensing Letters*, **7 (3)**, 511–514.

Vauhkonen, J., L. Ene, S. Gupta, J. Heinzel, J. Holmgren, J. Pitkänen, S. Solberg, Y. Wang, H. Weinacker, K. M. Hauglin, V. Lien, P. Packalén, T. Gobakken, B. Koch, E. Næsset, T. Tokola, and M. Maltamo, 2012: Comparative testing of single-tree detection algorithms under different types of forest. *Forestry*, **85 (1)**, 27–40.

Vauhkonen, J., I. Korpela, M. Maltamo, and T. Tokola, 2010: Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, **114 (4)**, 1263–1276.

Vauhkonen, J., E. Næsset, and T. Gobakken, 2014a: Deriving airborne laser scanning based computational canopy volume for forest biomass and allometry studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, **94**, 57–66.

Vauhkonen, J., H. O. Ørka, J. Holmgren, M. Dalponte, J. Heinzel, and B. Koch, 2014b: Tree species recognition based on airborne laser scanning and complementary data sources. *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*, Maltamo, M., E. Næsset, and J. Vauhkonen, Eds., Springer, Dordrecht, Netherlands, No. 27 in Managing Forest Ecosystems, 135–156.

Vauhkonen, J., T. Tokola, P. Packalén, and M. Maltamo, 2009: Identification of Scandinavian commercial species of individual trees from airborne laser scanner data using alpha shape metrics. *Forest Science*, **55 (1)**, 37–47.

Velizhev, A., R. Shapovalov, and K. Schindler, 2012: Implicit shape models for object detection in 3D point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. I-3, 179–184.

Vetter, M., B. Höfle, M. Hollaus, C. Gschöpf, G. Mandlburger, N. Pfeifer, and W. Wagner, 2011: Vertical vegetation structure analysis and hydraulic roughness determination using dense ALS point cloud data - a voxel based approach. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Calgary, Canada, Vol. XXXVIII-5/W12, 265–270.

Vosselman, G., B. G. H. Gorte, G. Sithole, and T. Rabbani, 2004: Recognising structure in laser scanner point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Freiburg, Germany, Vol. XXXVI-8/W2, 33–38.

Wagner, W., A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka, 2006: Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, **60**, 100–112.

Wang, Y., Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, 2018: Dynamic graph CNN for learning on point clouds. `https://arxiv.org/abs/1801.07829`, accessed: [12.12.2024].

Weinmann, M., 2016: *Reconstruction and analysis of 3D scenes – From irregularly distributed 3D points to object classes*. Springer, Cham, Switzerland.

Weinmann, M., R. Blomley, M. Weinmann, and B. Jutzi, 2018: Investigations on the potential of binary and multi-class classification for object extration from airborne laser scanning point clouds. *Tagungsband der 38. Wissenschaftlich-Technischen Jahrestagung der DGPF und PFGK 18 Tagung*, München, Germany, Vol. 27, 1–14.

Weinmann, M., B. Jutzi, and C. Mallet, 2013: Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Antalya, Turkey, Vol. II-5/W2, 313–318.

Weinmann, M., A. Schmidt, C. Mallet, S. Hinz, F. Rottensteiner, and B. Jutzi, 2015: Contextual classification of point cloud data by exploiting individual 3D neighbourhoods. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Munich, Germany, Vol. II-3/W4, 271–278.

West, K. F., J. R. Lersch, S. Pothier, J. M. Triscari, and A. E. Iverson, 2004: Context-driven automated target detection in 3-D data. *Proceedings of SPIE*, **5426**, 133–143.

Winiwarter, L., G. Mandlburger, S. Schmohl, and N. Pfeifer, 2019: Classification of als point clouds using end-to-end deep learning. *PGF Journal of Photogrammetry and Remote Sensing*, **87**, 75–90.

Xi, Z., C. Hopkinson, S. B. Rood, and D. R. Peddle, 2020: See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, **168**, 1–16.

Xiang, B., M. Wielgosz, T. Kontogianni, T. Peters, S. Puliti, R. Astrup, and K. Schindler, 2024: Automated forest inventory: analysis of high-density airborne LiDAR point clouds with 3D deep learning. *Remote Sensing of Environment*, **305**, 114 078.

Xie, Y., J. Tian, and X. Xiang Zhu, 2020: Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geoscience and Remote Sensing Magazine*, **8 (1)**, 38–59.

Xu, S., G. Vosselman, and S. Oude Elberink, 2014: Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, **88**, 1–15.

Yan, W. Y., A. Shaker, and N. El-Ashmawy, 2015: Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment*, **158**, 295–310.

Yang, B., Z. Dong, Y. Liu, F. Liang, and Y. Wang, 2017: Computing multiple aggregation levels and contextual features for road facilities recognition using mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **126**, 180–194.

Yao, W., P. Krzystek, and M. Heurich, 2012: Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sensing of Environment*, **123**, 368–380.

Ye, X., J. Li, H. Huang, L. Du, and X. Zhang, 2018: 3D recurrent neural networks with context fusion for point cloud semantic segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 403–417.

Yu, L., H. Yu, and S. Yang, 2022: A deep neural network using double self-attention mechanism for ALS point cloud segmentation. *IEEE Access*, **10**, 29 878–29 889.

Yu, X., P. Litkey, J. Hyyppä, M. Holopainen, and M. Vastaranta, 2014: Assessment of low density full-waveform airborne laser scanning for individual tree detection and tree species classification. *Forests*, **5**, 1011–1031.

Zachary, J. B. and R. H. Wynne, 2005: Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **59 (6)**, 342–360.

Zhang, K., S. Chen, D. Whitman, M. Shyu, J. Yan, and C. Zhang, 2003: A progressive morphological filter for removing nonground measurements from airborne LiDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, **41 (4)**, 872–882.

Zhang, Y., H. Liu, X. Liu, and H. Yu, 2023: Towards intricate stand structure: a novel individual tree segmentation method for ALS point cloud based on extreme offset deep learning. *Applied Sciences*, **13 (11)**, 6853.

Zhao, H., L. Jiang, C. Fu, and J. Jia, 2019: Pointweb: Enhancing local neighborhood features for point cloud processing. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, California, 5565–5573.

Zhao, R., M. Pang, and J. Wang, 2018: Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network. *International Journal of Geographical Information Science*, **32 (5)**, 960–979.

Zhihai, X. and Y. Zhishuang, 2018: Eigenentropy based convolutional neural network based ALS point cloud classification method. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XLII-3, 2017–2022.

List of Figures

2.1 Schematic representation of influences on the recorded waveform. $*$ denotes the convolution operator, and $\cdot$ the multiplication operator. Physically, the emitted waveform is first convoluted with the target response function and then with the receiver response function, but since the convolution operation is commutative, it is more practical to combine emitted waveform and receiver response into one measurable unit, called the system waveform (SWF). The list of influences is non-exhaustive. Graphical illustration in analogy to Hovi (2015). 14

2.2 Visualization of the mathematical concept of an artificial neuron. . . . 20

2.3 A multi-layer perceptron, serving as an example of a relatively simple neuronal network architecture. 20

3.1 [] Neighbourhood definitions used in this work as the basis for extracting features for a considered 3D point [(return)] X [(red)]: cylindrical neighbourhoods $\mathcal{N}_c$ [with different radii and infinite extent in the vertical direction (green)] and a spherical neighbourhood $\mathcal{N}_{k,opt}$ formed by an optimal number k_{opt} of nearest neighbours [(blue). For the sake of clarity, other spherical neighbourhoods are left out in this visualization, even though spherical neighbourhoods $\mathcal{N}_s$ of different fixed radii are sometimes also used.] 33

3.2 Visualization of the geometric measures taken for the sampling of shape distributions. 36

3.3 Effects of the scene topography[, colour values indicate the unit of meters in the legend bar]. The point clouds' height minima on a 0.5 m grid are shown on the **left**, the approximation of the scene topography is plotted in the **middle**, and the normalized minima are shown on the **right**. The **top** row depicts [area 1 & 3] of the **Vaihingen** data set, while the **bottom** row shows the test area of the **GML** data set A. 38

3.4 Reference point cloud of the **Vaihingen** data set Area 2, coloured with respect to nine semantic classes (*Roof*: red; *Façade*: white; *Impervious Surfaces*: grey; *Car*: blue; *Tree*: dark green; *Low Vegetation*: bright green; *Shrub*: yellow; *Fence / Hedge*: cyan; *Powerline*: black) 41

3.5 Reference point clouds of the **GML** A data set A ((**top**: training area; **bottom**: testing area), coloured with respect to five semantic classes (*Building*: red; *Ground*: grey; *Car*: blue; *Tree*: dark green; *Low Vegetation*: bright green) 42

3.6 Evaluation of [SVM] classification results exclusively employing shape distributions for different neighbourhood sizes. [Quality measures are calculated according to Section 2.5 and plotted against the radius of a cylindrical neighbourhood.] 44

3.7 Evaluation of [SVM] classification results exclusively employing covariance features for different neighbourhood sizes. [Quality measures are calculated according to Section 2.5 and plotted against the radius of a cylindrical neighbourhood]. 45

3.8 Class-wise mean Shannon entropy calculated according to Equation [3.6] for different neighbourhood sizes. [This indicates, that the optimum scale of covariance features is typically small, around 0.50 m.] 47

3.9 Mean rank of mean feature relevance per feature group for different neighbourhood sizes. A3, D1, D2, D3 and D4 are the shape distribution[s], and Cov. the covariance features []. 48

3.10 Gable roof and adjacent trees, generally well labelled [by the combined binary classification results of shape distributions on four class-specific neighbourhood sizes.] 50

3.11 High-rise buildings[, labelled by the combined binary classification results of shape distributions or four class-specific neighbourhood sizes.] Without the use of further features, flat roofs are mistaken as ground [due to the rotation-invariance of shape distributions.] . . . 50

3.12 Visualization of classification results using a nearest neighbour (NN) classifier, a linear discriminant analysis (LDA) classifier, a QDA classifier and a RF classifier respectively on Area 1 (**top**) and Area 3 (**bottom**) of the **Vaihingen** data set. For clearer discriminability among the classes, a non-natural colour encoding is chosen (*Powerline*: violet; *Low Vegetation*: yellowish green; *Impervious Surfaces*: royal blue; *Car*: cyan; *Fence / Hedge*: pink; *Roof*: crimson; *Façade*: white; *Shrub*: gold; *Tree*: emerald green). . . . 53

3.13 Visualization of classification results for the **GML** data set A using $\mathcal{N}_c$ and $\mathcal{N}_{k,\text{opt}}$ with covariance features, geometric 3D properties, shape distributions and absolute height. The colour encoding refers to the classes *Ground* (grey), *Building* (red), *Car* (blue), *Tree* (dark green), and *Low Vegetation* (bright green). 58

3.14 Visualization of classification results for the **Vaihingen** data set Area 1 (**left**) and Area 3 (**right**) using $\mathcal{N}_c$, $\mathcal{N}_s$ and $\mathcal{N}_{k,\text{opt}}$ with covariance features, geometric 3D properties, shape distributions, and normalized height. The colour encoding refers to the classes *Roof* (red), *Façade* (white), *Impervious Surfaces* (grey), *Car* (blue), *Tree* (dark green), *Low Vegetation* (bright green), *Shrub* (yellow), *Fence / Hedge* (cyan), and *Powerline* (black). 64

3.15 Visualization of classification results for the **GML** A using $\mathcal{N}_c$, $\mathcal{N}_s$, and $\mathcal{N}_{k,\text{opt}}$ with covariance features, geometric 3D properties, shape distributions, and normalized height. The colour encoding refers to the classes *Ground* (grey), *Building* (red), *Car* (blue), *Tree* (dark green), and *Low Vegetation* (bright green). 65

3.16 Confusion matrix of the RF classification result using 100 000 training entities per class on the **Vaihingen** data set. Percentage and colour coding are given as percentage of the ground truth reference. The confusion matrix shows, that elements of both the classes of *Car* and *Fence / Hedge* (column 4 and 5) are often mislabelled as belonging to the class *Shrub* (row 8). 68

4.1 Concepts used in related work on tree species classification. 80

4.2 Overview of the [proposed] spin image method. Returns are first sampled to a rotating image plane for each tree, producing a data vector $\vec{d}$. Sampled data is then collected from sample trees to form a library $\mathbf{L}$. By principal component analysis, variable components within the library are identified as eigenvectors $\vec{e}_i$. Finally, the original data $\vec{d}$ is projected into the eigenvector space, producing a single feature value (linear combination) f_i for each eigenvector $\vec{e}_i$. . . 86

4.3 Cumulative histogram of horizontal displacement between tree top position measured manually by photogrammetric monoplotting (2011) and the position of the highest LiDAR return (2013). 93
4.4 Classification results for spin images of local return density. Results are plotted for varying pixel size and constant vertical reach of 14 m. The diagrams show the κ classification results as grey values for different pixel size and a varying number of features. The **left** diagram shows results based on **LMS-Q680i** data, while the **right** shows those based on **ALS60** data. 94
4.5 Classification results for spin images based on the number of returns. Results are plotted for a varying vertical dimension of the spin images. The **left** diagram uses data from the **LMS-Q680i** data with a pixel size of ∼89 cm, while the **right** diagram shows **ALS60** data with a pixel size of ∼127 cm. 95
4.6 κ classification results, when spin image features (solid lines) of local return density or alpha shape features (dashed lines) are added to the classification one by one. The order is given by an importance measure estimated by the RF classifier. 96
4.7 DBH histogram of trees classified according to *only*, *first-or-only* and *all* returns of the **LMS-Q680i** data. Light grey bars denote the total number of trees, medium grey those classified incorrectly by spin image features alone, dark grey those classified incorrectly by non-spatial features alone and black those classified incorrectly by the combination of both feature types. 101
4.8 DBH histogram of trees classified according to *only*, *first-or-only* and *all* returns of the **ALS60** data. Light grey bars denote the total number of trees, medium grey those classified incorrectly by spin image features alone, dark grey those classified incorrectly by non-spatial features alone and black those classified incorrectly by the combination of both feature types. 102

List of Tables

3.1 Number of 3D points per class in the **Vaihingen** data set. At the time of publication, the reference labels were only available for the training set, but not for the test set. Evaluation was therefore conducted externally. The full information about the test set has only become available later. 40
3.2 Number of 3D points per class in the **GML** data set A. 41
3.3 Rejection rate and overall accuracy (OA) for [combined binary] classifications based on shape distributions and covariance features. . 49
3.4 Confusion matrix of combined classification results using shape distributions [on four scales optimal for one class each,] ignoring all cases in which an element could not be detected in any of the classes. 51
3.5 Confusion matrix of combined classification results using [covariance features on four scales optimal for one class each,] ignoring all cases in which an element could not be detected in any of the classes. 51
3.6 Class-wise precision and recall values (in %) for classifications by the NN, LDA, QDA classifiers, and a RF classifier on the **Vaihingen** data. Results published in a similar representation in (Blomley et al., 2016a). 54
3.7 Class-wise F_1-scores (in %) for classifications by the NN, LDA, QDA classifiers, and by a RF classifier on the **Vaihingen** data. Results published in a similar representation in (Blomley et al., 2016a). 54
3.8 Overall classification results (in %) by the NN, LDA, QDA classifiers, and by a RF classifier on the **Vaihingen** data set, evaluated according to the metrics of OA, κ, mean class recall (MCR), and mean class precision (MCP). Results published in a similar representation in (Blomley et al., 2016a). 54
3.9 κ and OA (in %) for different neighbourhood definitions and different feature sets on the **GML** data set A. Results published in (Blomley et al., 2016b). 58
3.10 MCP and MCR (in %) for different neighbourhood definitions and different feature sets on the **GML** data set A. Results published in (Blomley et al., 2016b). 58
3.11 Class-wise precision, recall and F_1 (in %) for different neighbourhood definitions and different feature sets on the **GML** data set A. Results published in similar presentation in (Blomley et al., 2016b). 59
3.12 Comparison of class-wise precision, recall, and F_1 values (in %), alongside with OA and $\bar{F}_1$ (in %) of our results on the **GML** data set A in Section 3.5.3.1 to those achieved by Shapovalov et al. (2010). 61
3.13 Class-wise recall, precision and F_1-score (in %) as well as OA and [$\bar{F}_1$] (in %) for the **Vaihingen** data set. 63

3.14 Class-wise recall, precision and F_1-score (in %) as well as OA and $[\bar{F}_1]$ (in %) for the **GML** data set A. 63
3.15 OA and mean $\bar{F}_1$-score (in %) achieved for different neighbourhood definitions on the **Vaihingen** data set and the **GML** data set A. . . . 64
3.16 Comparison of class-wise precision, recall, and F_1 values (in %), alongside with OA and $\bar{F}_1$ (in %) of our results on the **GML** data set A in Section 3.5.3.2 to those achieved by Shapovalov et al. (2010). 65
3.17 Comparison of the improvements in this section over previous experiments. Class-wise F_1-scores (in %) as well as OA and $\bar{F}_1$ (in %) for the **GML** data set A in comparison to Section 3.5.3.1 (**left**) as well as the **Vaihingen** data set in comparison to the RF results from Section 3.5.2 (**right**). 66
3.18 Class-wise recall, precision and F_1-score (in %), comparing the results of a RF with 10 000 (first value) and a RF with 100 000 (second value) training examples per class on the **Vaihingen** data set. OA increased from 68.1% to 71.5%, while $\bar{F}_1$ improved from 52.6% to 58.3%. 67
3.19 Class-wise F_1-scores (in %) as well as OA and $\bar{F}_1$ (in %) for the **Vaihingen** data set, comparing our RF results with 10 000 and 100 000 training examples per class respectively. Furthermore we list the results achieved by Steinsiek et al. (2017) in point-wise classification (RF) or with contextual classification (CRF), as well as the results achieved by Niemeyer et al. (2016) with a hierarchical CRF. Some of these values were presented in a similar comparison in (Weinmann et al., 2018). 69
3.20 Class-wise F_1-scores (in %) as well as OA and mean $\bar{F}_1$ (in %) for the **Vaihingen** data set. A comparison of our results, feature-based contextual learning (Steinsiek et al., 2017) and deep learning implementations, either via horizontal 2D projections or directly on the unstructured point cloud by adaptions of PointNet/PointNet++. (* first value according to Yu et al. (2022), second value according to Nong et al. (2023)) 74

4.1 Main characteristics of the two datasets. 90
4.2 Distribution of the tree segments according to species and three different age classes, representing different periods of tree life (premature: 30-60 years; mature: 60-100 years; old: >100 years). . . . 92
4.3 Mean number of returns per tree segment for different return types (top) and the defined return type groups (bottom). 92
4.4 Comparison of κ classification results using either the three best, ten best or all spin image and α-shape features. 96
4.5 OA in % for classification with waveform attributes for different groups of returns. Improvements are marked in bold. 97
4.6 κ in % for classification with waveform attributes for different groups of returns. The improvements using all attribute types are marked in bold. 98
4.7 Precision, recall and F_1-score in % for all three species in classifications with all waveform attributes of *only*, *first-or-only* and *all* returns. Improvements in the F_1-score are marked in bold. . . 99
4.8 Confusion matrices normalized to ground truth in % in classifications with all waveform attributes of *only*, *first-or-only* and *all* returns. 100

Acknowledgments

The Vaihingen data set used in Chapter 3 was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (Cramer, 2010).

My research on tree species classification in Chapter 4 would not have been possible without numerous support from researchers at the Department of Forest Sciences, University of Helsinki. I am thankful to the Academy of Finland, Metsähallitus, Metsämiesten säätiö, TEKES, the University of Helsinki, and the strategic funds of the University of Eastern Finland, for sharing valuable data with me and supporting my stay abroad. I thank Harri Hytönen, Timo Ketolainen, and the many forestry students for the field work.

Furthermore, my studies were supported by the Carl-Zeiss-Stiftung (Nachwuchsförderprogramm 2014).

As a member of the GRACE - Graduate School of Climate and Environment at KIT I am also grateful for their educational curriculum of scientific skills and content, as well as financial support during my first unpaid parental leave due to my situation as a scholarship PhD student.

Many thanks I also attribute to the Institute of Photogrammetry and Remote Sensing at KIT throughout the years. I got to know many great colleagues there and enjoyed the mutual scientific exchange despite our often very diverse topics.

Also many thanks to Prof. Markus Gerke for examining this thesis.

I especially thank my supervisors and collaborators: Boris Jutzi and Jens Leitloff for initial supervision. Ilkka Korpela and Aarne Hovi for an exceptional cooperation, scientific support and supervision. Martin Weinmann for continued collaboration, support and supervision over the years. Many thanks to Stefan Hinz, for granting me the possibility to work in the field and at the institute, as well as providing regular supervision and advice. I am very grateful for the support I received from all of you.

Finally, I want to thank my family. Many thanks to my father for proofreading in terms of language. Many thanks to my parents and our family friend Heiko for personal and family support during busy work phases. To my husband. Without your backing and succour, I could not have completed this work. To my children, who were very supportive when I had to work and who ground me in the truly important things of life.