

# Semantische Analyse von ALS- und MLS-Daten in urbanen Gebieten mittels Conditional Random Fields

MORITZ STEINSIEK<sup>1</sup>, PRZEMYSŁAW POLEWSKI<sup>1</sup>, WEI YAO<sup>1</sup> & PETER KRZYSZEK<sup>1</sup>

*Zusammenfassung: In dieser Arbeit wird ein Verfahren zur semantischen Analyse von ALS- und MLS-Daten in urbanen Gebieten entwickelt, das auf einem Conditional Random Field (CRF) zur kontextbasierten Klassifikation von 3D-Punktwolken aufbaut. Das CRF nutzt einen Random Forest (RF) Klassifikator für die Knotenpotentiale und eine Variante des kontrast-sensitiven Potts-Modells für die Kantenpotentiale. Die Grundlage der Klassifikation bilden verschiedene geometrische und radiometrische Merkmale, die mittels zweier lokaler Punktnachbarschaften abgeleitet werden. Die Genauigkeit des Verfahrens wird anhand von jeweils einem ALS- und einem MLS-Benchmark-Datensatz mit Referenzdaten überprüft. Zusätzlich erfolgt ein Vergleich der Ergebnisse der kontextbasierten Klassifikation mit einer individuellen Klassifikation mittels RF. Bei Anwendung dieses Verfahrens können Gesamtgenauigkeiten von 80,5% für den ALS-Datensatz und 97,7% für den MLS-Datensatz erzielt werden. Von der Nutzung kontextbasierter Klassifikationsverfahren profitieren im Vergleich zu individuellen Verfahren vor allem unterrepräsentierte Objektklassen deutlich, also solche Klassen mit einer geringen Anzahl von Punkten in den Daten.*

## 1 Einleitung

Die automatische Interpretation von 3D-Punktwolken ist ein hochaktuelles Thema und gewinnt im Bereich der Fernerkundung, Photogrammetrie, Computer Vision und der Robotik zunehmend an Bedeutung. Gründe dafür sind unter anderem die in den letzten Jahren vollzogene Entwicklung von mobilen Laserscanning Systemen (MLS), die Weiterentwicklung des luftgestützten Laserscannings (ALS) und die steigende Verbreitung von *Unmanned Aerial Systems* zur Datenakquisition.

MLS werden im Allgemeinen für kleinräumige Gebiete, beispielsweise einzelne Straßenzüge, eingesetzt und liefern genaue und dichte 3D-Daten der Objektoberflächen. Die Punktdichte ist allerdings von der Entfernung des Scanners zu den einzelnen Objekten abhängig, was zu einer inhomogenen 3D-Punktwolke führt. Das hauptsächliche Einsatzgebiet für ALS ist im Gegensatz dazu die Aufnahme von großflächigen Umgebungen, wie zum Beispiel ganze Teile von Stadt- oder Waldgebieten mit weitaus geringeren aber meist homogenen Punktdichten.

Besonders im urbanen Raum finden sich viele Anwendungsbereiche, deren Grundlage 3D-Punktwolken darstellen. Darunter fallen etwa die Generierung von 3D-Stadtmodellen zur Planung von Infrastruktur und der Stadtentwicklung oder die Nutzung zur Fahrzeugnavigation. Ein grundlegender Schritt für die meisten der genannten Anwendungen ist die Durchführung einer Klassifikation der 3D-Punktwolke. Hierbei wird jedem einzelnen 3D-Punkt eine semantische Objektklasse (z.B. Auto, Gebäude, Vegetation oder Straße) zugeordnet.

---

<sup>1</sup> Hochschule München, Fakultät für Geoinformation, Karlstr. 6, D-80333 München,  
E-Mail: moritz.steinsiek@gmx.de, [polewski, yao, krzyszek]@hm.edu

Häufig konzentrieren sich die Untersuchungen auf einzelne Objektklassen wie Gebäude (ROTTENSTEINER et al. 2014), Vegetation (WURM et al. 2014) oder Straßenlaternen (YONGTAO et al. 2014). NIEMEYER et al. (2014) und WEINMANN et al. (2015) nutzen *Conditional Random Fields* (CRF) zur punktweisen Klassifikation von mehreren Objekten in urbanen ALS-beziehungswise MLS-Daten. Bei der Anwendung von CRF zeigt sich eine deutliche Verbesserung der Klassifikationsergebnisse im Vergleich zu herkömmlichen Verfahren, wie beispielsweise *Random Forest* (RF) oder *Support Vector Machines* (SVM). Durch das komplexere Modell der CRF erhöht sich jedoch der Berechnungsaufwand. YANG & FÖRSTNER (2011) und WEGNER et al. (2011) hingegen nutzen zur Klassifikation urbaner Gebiete CRF auf Basis von Rasterdaten statt 3D-Punktwolken.

Die häufigsten Schwierigkeiten bei der Klassifikation von Punktwolken ergeben sich vor allem aufgrund der großen Menge an Daten, Variationen in der Punktdichte, unvollständigen Strukturen durch Verdeckungen oder wegen der Unterschiede zwischen Objekten innerhalb derselben Klasse. Auf Grund von sehr komplexen Strukturen und vielen unterschiedlichen benachbarten Objektklassen, sind 3D-Szenen in urbanen Gebieten besonders herausfordernd. Das Ziel dieser Arbeit besteht darin, ein Verfahren zur semantischen Analyse von 3D-Punktwolken in urbanen Gebieten zu entwickeln, das sowohl für ALS- als auch MLS-Daten anwendbar ist. Die Analyse umfasst dabei alle nötigen Schritte, um jedem 3D-Punkt eine Objektklasse zuzuweisen: (i) die Auswahl einer passenden lokalen Punktnachbarschaft für die (ii) Extraktion von markanten Merkmalen und (iii) die Durchführung der Klassifikation mit einem geeigneten Verfahren.

## 2 Methodik

Das Ziel der Klassifikation von Punktwolken besteht darin, jedem 3D-Punkt  $P_i$  ein Objektklassenlabel zuzuweisen. In vielen Ansätzen werden individuelle Klassifikationsverfahren wie SVM (CORTESA & VAPNIK 1995) oder RF (BREIMAN 2001) eingesetzt. Bei diesen Verfahren wird jeder 3D-Punkt unabhängig von seiner Nachbarschaft klassifiziert. In der letzten Zeit wurden jedoch vor allem kontextbasierte Klassifikationsverfahren vorgestellt, bei denen auch die Objektklassen der Punkte in der Nachbarschaft in den Zuweisungsprozess der Labels miteinbezogen werden. Dies hat zur Folge, dass alle Punkte gleichzeitig klassifiziert werden müssen. In der vorliegenden Arbeit werden eine kontextberücksichtigende Klassifikation basierend auf CRF und eine individuelle Klassifikation mittels RF eingesetzt (Abb. 1).

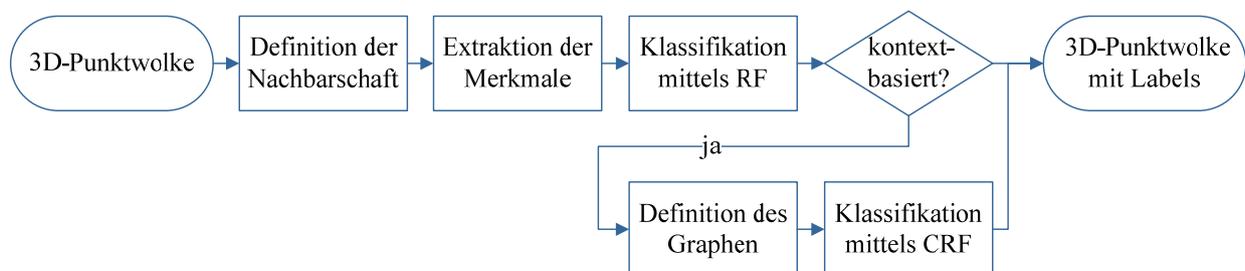


Abb. 1: Ablaufdiagramm mit den einzelnen Schritten zur Klassifikation der 3D-Punktwolke

## 2.1 Definition der lokalen Punktnachbarschaft

Ausgehend von einer Punktwolke mit  $N$  Punkten  $P_i = (X_i, Y_i, Z_i)^T \in \mathbb{R}^3$  mit  $i \in \{1, \dots, N\}$ , besteht der erste Schritt darin, eine geeignete lokale Punktnachbarschaft festzulegen. Die Definition der Nachbarschaft dient zur Beschreibung der lokalen 3D-Struktur um einen gegebenen Punkt  $P_i$  und bildet damit die Grundlage für die Extraktion von geometrischen Merkmalen. In dieser Arbeit werden zwei Nachbarschaftsdefinitionen betrachtet: Sphärische Nachbarschaften  $N_s$  mit fixiertem Radius  $r_s$  und eine optimale sphärische Nachbarschaft  $N_{opt}$  mit variierendem Radius, basierend auf den  $k \in \mathbb{N}$  nächsten Nachbarn und der Eigenentropie (WEINMANN et al. 2014).

Der Wert  $k$  für einen Punkt  $P_i$  hängt unter anderem von der lokalen 3D-Struktur und der lokalen Punktdichte um  $P_i$  ab. Infolgedessen basieren einige vorgeschlagene Verfahren zur automatischen Bestimmung von individuellen Werten für  $k$  auf den lokalen geometrischen Eigenschaften der Punktwolke. Zur Ableitung dieser Eigenschaften wird die 3D-Kovarianzmatrix  $C \in \mathbb{R}^{3 \times 3}$  für einen gegebenen Punkt und seine  $k$  nächsten Nachbarn berechnet. Da es sich bei  $C$  um eine symmetrische, positiv-semidefinite Matrix handelt, existieren ihre drei Eigenwerte. Die Eigenwerte sind positiv und korrespondieren zu einem orthogonalen System von Eigenvektoren. Die drei Eigenwerte  $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ , mit  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$  repräsentieren die Ausdehnung eines dreidimensionalen Kovarianzellipsoids entlang seiner Hauptachsen und eignen sich damit zur Beschreibung der lokalen 3D-Struktur. Die *Shannon Entropie* der normalisierten Eigenwerte  $e_1, e_2, e_3$  dient als Grundlage für die Berechnung der Eigenentropie  $E_{\lambda,i}$ .

$$E_{\lambda,i} = -e_1 \ln(e_{1,i}) - e_2 \ln(e_{2,i}) - e_3 \ln(e_{3,i}) \quad (1)$$

Die Eigenentropie ist ein Maß für die Ordnung beziehungsweise die Unordnung der 3D-Punkte innerhalb der lokalen Nachbarschaft. Die Idee dieses Ansatzes besteht darin, den optimalen Wert von  $k$  zu finden, der die Unordnung der 3D-Punkte innerhalb der Nachbarschaft und damit die *Shannon Entropie* minimiert. Dafür wird der Wert  $k$  für jeden Punkt in einem festgelegten Intervall von  $k_{min}$  bis  $k_{max}$  mit einer Schrittweite von  $\Delta k$  variiert. Der Wert  $k$ , der zur minimalen *Shannon Entropie* korrespondiert, wird als optimaler Wert angenommen.

## 2.2 Extraktion von Merkmalen

Nach der erfolgten Definition der lokalen Punktnachbarschaft für jeden Punkt  $P_i$ , können anschließend Merkmale abgeleitet werden, welche die lokale 3D-Struktur beschreiben. In dieser Arbeit werden die Merkmale adaptiert, die in WEINMANN et al. (2015) vorgeschlagen wurden. Dabei handelt es sich um eine Kombination von 2D- und 3D-Merkmalen, mit der bereits gute Ergebnisse bei der Klassifikation von 3D-Punktwolken erzielt werden konnten.

Die 3D-Merkmale werden für jeden Punkt und seine durch die Nachbarschaft festgelegten Nachbarn abgeleitet. Basierend auf den drei Eigenwerten der 3D-Kovarianzmatrix, werden insgesamt acht geometrische Merkmale bestimmt:

$$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_3} \quad L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_3} \quad (2)$$

$$\begin{aligned}
S_\lambda &= \frac{\lambda_3}{\lambda_{13}} & E_\lambda &= \sqrt[3]{\lambda_1 \lambda_2 \lambda_3} \\
A_\lambda &= \frac{\lambda_1 - \lambda_3}{\lambda_1} & E_\lambda &= - \sum_{i=1}^3 \lambda_i \ln(\lambda_i) \\
\Sigma_\lambda &= \lambda_1 + \lambda_2 + \lambda_3 & C_\lambda &= \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}
\end{aligned}$$

Die Linearität  $L_\lambda$ , Planarität  $P_\lambda$  und die Sphärizität  $S_\lambda$  liefern Informationen darüber, ob es sich um eine lineare 1D-Struktur, eine planare 2D-Struktur oder eine volumetrische 3D-Struktur handelt. Die weiteren eigenwertbasierten Merkmale sind die Omnivarianz  $O_\lambda$ , Anisotropie  $A_\lambda$ , Eigenentropie  $E_\lambda$ , die Summe der Eigenwerte  $\Sigma_\lambda$  und die Änderung der Krümmung  $C_\lambda$ . Als weiteres Merkmal ist die Vertikalität  $V = 1 - |z_{\vec{n}}|$  zu nennen, die sich mit Hilfe der dritten Komponente  $z_{\vec{n}}$  des Normalenvektors  $\vec{n}$  berechnen lässt. Zusätzlich werden der Radius  $r_{knn}$  der Kugel, die durch die  $k$  nächsten Nachbarn gebildet wird, und die resultierende lokale Punktdichte  $D_{3D}$  bestimmt.

$$D_{3D} = \frac{k + 1}{3/4 \pi r_{knn}^3} \quad (3)$$

Weitere markante Merkmale können mit Hilfe der Höhe abgeleitet werden. Eine wichtige Eigenschaft stellt die normalisierte Höhe  $H_{norm}$  dar, welche der über einem digitalen Geländemodell entspricht. Für die ALS-Daten wird die normalisierte Höhe des Punktes  $P_i$  als Merkmal verwendet, da die Höhe in größeren Gebieten stärker variiert. Im Falle von MLS-Daten mit kleinräumiger Ausdehnung wird dagegen die absolute Höhe  $H_{abs}$  verwendet. Weiter dienen die Standardabweichung der Höhe  $\sigma_H$  und die maximale Höhendifferenz  $\Delta H$  innerhalb der lokalen Nachbarschaft als Merkmale. Beim ALS-Datensatz wird darüber hinaus für jeden Punkt der Intensitätswert  $I_i$  genutzt.

Zur Ableitung der 2D-Merkmale werden zunächst alle 3D-Punkte auf die XY-Ebene projiziert. Beruhend auf dieser 2D-Darstellung der Daten, wird der Radius  $r_{knn,2D}$  der kreisförmigen Nachbarschaft, die sich aus  $P_i$  und seinen  $k$  Nachbarn ergibt, bestimmt. Mit  $r_{knn,2D}$  kann anschließend die lokale Punktdichte  $D_{2D}$  berechnet werden.

$$D_{2D} = \frac{k + 1}{\pi r_{knn,2D}^2} \quad (4)$$

Überdies wird die 2D-Kovarianzmatrix  $C_{2D}$  berechnet und daraus die beiden Eigenwerte  $\lambda_{1,2D}$  und  $\lambda_{2,2D}$  abgeleitet. Auf Basis der beiden Eigenwerte werden die Summe  $\Sigma_{\lambda,2D}$  und das Verhältnis  $R_{\lambda,2D} = \lambda_{2,2D} / \lambda_{1,2D}$  gebildet und als weitere Merkmale genutzt.

Weiterhin werden 2D-Merkmale aus einem Akkumulationsbild extrahiert, indem die Projektionsebene in diskrete, quadratische Bins mit einer Seitenlänge von  $s$  unterteilt wird. Für jeden Punkt werden aus dem ihm zugewiesenen Bin folgende Merkmale abgeleitet: Die Anzahl der Punkte  $N_{Acc}$  in dem Bin, die maximale Höhendifferenz  $\Delta H_{Acc}$  und die Standardabweichung der Höhe  $\sigma_{H,Acc}$ .

## 2.3 Conditional Random Field

Die CRF gehören zu den probabilistischen graphischen Modellen und erlauben die Modellierung von Abhängigkeiten zwischen benachbarten Punkten. Die CRF sind diskriminative Modelle und modellieren die bedingte Klassenwahrscheinlichkeit  $P(Y|x)$  direkt (KUMAR & HEBERT 2006):

$$P(Y|x) = \frac{1}{Z(x)} \exp \left( \prod_{i \in n} \phi(x, y_i) \prod_{i \in n} \prod_{j \in N_i} \psi(x, y_i, y_j) \right) \quad (5)$$

$Z(x)$  entspricht einer Normalisierungskonstanten und wird auch als Partitionsfunktion bezeichnet. Da diese nicht von den Klassenlabels abhängt, kann sie in Bezug auf die Klassifikation vernachlässigt werden. Die Gleichung besteht somit aus zwei Funktionen: Dem Assoziationspotential  $\phi(x, y_i)$  und dem Interaktionspotential  $\psi(x, y_i, y_j)$ . Das CRF wird durch einen ungerichteten Graphen  $G = (n, e)$  mit einer Menge von Knoten  $n$  und einer Menge von Kanten  $e$  repräsentiert. Die Knoten  $n_i \in n$  entsprechen den 3D-Punkten der Punktwolke; die Kanten  $e_{ij} \in e$  verbinden zwei benachbarte Knoten  $n_i$  und  $n_j$  und modellieren deren Abhängigkeiten. Innerhalb einer 3D-Punktwolke ist die Definition des Graphen in Bezug auf die lokale Nachbarschaft komplexer als für Bilder. Die 3D-Punkte innerhalb einer Punktwolke haben eine unregelmäßige Anordnung im dreidimensionalen Raum. In Bildern können hier beispielsweise die vier Nachbarn anhand des Bildrasters ausgewählt werden, was bei Punktwolken nicht möglich ist. Häufig basiert die Definition der lokalen Punktnachbarschaft auf den  $k$  nächsten Nachbarn. Dementsprechend erfolgt die Definition der Kanten mittels den in Abschnitt 2.1 beschriebenen Nachbarschaften, die bereits zur Extraktion der Merkmale genutzt werden. Angesichts der hohen Anzahl an Punkten in den Punktwolken, wird die Größe der lokalen Nachbarschaft innerhalb des Graphen aus Gründen der Rechenkapazitäten durch einen Schwellwert  $k_{thr}$  eingeschränkt. Der Vektor  $Y$  enthält die Klassenlabels  $y_i$  für jeden 3D-Punkt.  $x$  entspricht den Zufallsvariablen der Merkmale. Das Ziel der Klassifikation besteht darin, die optimale Konfiguration  $Y$  zu finden, für die  $P(Y|x)$  maximal wird.

### 2.3.1 Definition der Potentiale

Das Assoziationspotential  $\phi(x, y_i)$  verbindet die Daten  $x$  und die entsprechenden Klassenlabels  $y_i$ . Für dieses Potential kann das Ergebnis jedes beliebigen diskriminativen Klassifikators, beispielsweise RF, verwendet werden. Der Klassifikator muss in der Lage sein, eine Verteilung über die möglichen Werte der Klassenlabels zu liefern. Diese Möglichkeit ergibt sich daraus, dass  $\phi(x, y_i)$  für einen Knoten  $n_i$  von den gesamten Daten  $x$  abhängen kann (KUMAR & HEBERT 2006). Infolgedessen wird für jeden Knoten ein Merkmalsvektor  $f_i(x)$  auf Basis der in Abschnitt 2.2 beschriebenen Merkmale erstellt. Die einzelnen Komponenten des Merkmalsvektors sind folglich Funktionen der lokalen Nachbarschaft des Knotens. Das Assoziationspotential kann als die a posteriori Wahrscheinlichkeit eines lokalen diskriminativen Klassifikators beruhend auf  $f_i(x)$  definiert werden (KUMAR & HEBERT 2006):  $\phi(x, y_i) = \log(p(y_i|f_i(x)))$ .

Für das Assoziationspotential  $\phi(x, y_i)$  wird ein RF verwendet. Bei RF wird ein Ensemble von  $N_T$  unabhängigen Entscheidungsbäumen erstellt, die, basierend auf dem *Bagging* Prinzip, mit unterschiedlichen Trainingsmengen trainiert werden. Die anschließende Klassifikation wird mittels einer Mehrheitsentscheidung getroffen, bei der jeder Baum für eine Klasse  $y^l$  stimmt. RF

zählt zu den nicht-probabilistischen Klassifikationsverfahren, die nicht direkt auf Wahrscheinlichkeiten beruhen. Mittels der Anzahl  $N_l$  der Bäume, die für eine Klasse  $y^l$  gestimmt haben, kann allerdings im Anschluss die a posteriori Wahrscheinlichkeit abgeleitet werden:

$$p(y_i = y^l | f_i(x)) = \frac{N_l}{N_T} \quad (5)$$

Das Interaktionspotential  $\psi(x, y_i, y_j)$  beschreibt den lokalen Kontext und verbindet die Klassenlabels  $(y_i, y_j)$  zweier benachbarter Knoten mit den Daten  $x$ . Für das Interaktionspotential kann ebenfalls das Ergebnis jedes beliebigen diskriminativen Klassifikators verwendet werden. In dieser Arbeit wird für dieses Potential eine Variante des häufig verwendeten kontrast-sensitiven Potts-Modells (BOYKOV & JOLLY 2001) angewendet (WEINMANN et al. 2015):

$$\log(\psi(x, y_i, y_j)) = \delta_{y_i y_j} \cdot w_1 \cdot \frac{N_a}{N_{k_i}} \left[ w_2 + (1 - w_2) \cdot \exp\left(-\frac{d_{ij}(x)^2}{2\sigma^2}\right) \right] \quad (6)$$

$\delta_{y_i y_j}$  entspricht dem Kronecker-Delta, welches den Wert eins, falls  $y_i = y_j$  und ansonsten null annimmt.  $d_{ij}(x)$  ist die Euklidische Distanz zwischen den beiden Merkmalsvektoren  $f_i(x)$  des Knoten  $i$  und  $f_j(x)$  des Knoten  $j$ , die durch eine gemeinsame Kante  $e_{ij}$  verbunden sind. Der Wert  $N_a$  ist die durchschnittliche Anzahl der Kanten, mit denen ein Knoten im Graphen verbunden ist und  $N_{k_i}$  die Anzahl der Knoten in der lokalen Nachbarschaft des Knoten  $i$ . Der Parameter  $\sigma$  ist die durchschnittliche Distanz zwischen den Merkmalsvektoren benachbarter Punkte. Bei den beiden Parametern  $w_1$  und  $w_2$  handelt es sich um Gewichtsparameter. Das erste Gewicht bestimmt, welchen Einfluss das Interaktionspotential auf die Klassifikation hat und kann beliebige nichtnegative, reelle Werte annehmen. Das zweite Gewicht  $w_2 \in [0; 1]$  beeinflusst den Grad der Glättung abhängig von den Daten  $x$ .

### 2.3.2 Training und Inferenz

Zunächst müssen die entsprechenden Parameter und Gewichte der unterschiedlichen Klassifikatoren in einem Trainingsprozess bestimmt werden. Dazu werden Trainingsmengen beziehungsweise Referenzdaten der 3D-Punkte mit den entsprechenden Objektklassenlabels benötigt. Für den verwendeten RF müssen im Trainingsschritt die  $N_T$  Entscheidungsbäume angelehrt werden. Dazu wird eine Trainingsmenge  $T$  mit  $N_{tr}$  zufällig ausgewählten 3D-Punkten pro Klasse erzeugt. Durch den Ausgleich der Klassen wird vermieden, dass Entscheidungsbäume entstehen, die mit nur sehr wenigen oder keinen Beispielen für eine unterrepräsentierte Klasse trainiert werden.

Für die Bestimmung der beiden Gewichte  $w_1$  und  $w_2$  kann beispielsweise eine Methode wie die Kreuzvalidierung verwendet werden. Die beiden Parameter werden in dieser Arbeit mittels einer Gittersuche unter Verwendung des gesamten Trainingsdatensatzes ausgewählt. Bei der Klassifikation mittels CRF wird die optimale Belegung der Labels  $Y$  bestimmt, für welche  $P(Y|x)$  maximal wird. Für Graphen mit Zyklen ist die exakte Inferenz nicht in praktischer Zeit lösbar, weshalb im Allgemeinen nur approximative Lösungen verwendet werden. In dieser Arbeit wird die *Quadratic Pseudo-Boolean Optimization* (BOROS & HAMMER 2002) genutzt.

### 3 Experimente und Ergebnisse

#### 3.1 Datensätze

Die Genauigkeit der Methode wird anhand von jeweils einem ALS- und einem MLS-Benchmark-Datensatz mit Referenzdaten überprüft.

Der ALS-Datensatz (NIEMEYER et al. 2014) wird im Rahmen des *3D Semantic Labeling Contest* der ISPRS zur Verfügung gestellt.<sup>2</sup> Dieser wurde über dem Stadtgebiet von Vaihingen in Deutschland aufgenommen. Die Aufnahme erfolgte im August 2008 mit einem *Leica ALS50* System in einer gemittelten Flughöhe von 500 m über dem Boden und einem Sichtfeld von 45°. Der durchschnittliche Überlappungsbereich der Spuren beträgt 30%, die Punktdichte liegt bei ungefähr 8 Punkte/m<sup>2</sup>. Zusätzlich wurden Intensitäten und Echoinformationen aufgezeichnet. Der Datensatz wurde in einen Trainingsdatensatz mit ungefähr 750k Punkten und einen davon unabhängigen Testdatensatz mit ungefähr 412k Punkten aufgeteilt. Der Trainingsdatensatz ist vollständig mit den entsprechenden Referenzlabels versehen. Das Ziel des *3D Semantic Labeling Contest* ist es, jedem 3D-Punkt im Testdatensatz eine von neun Klassen zuzuordnen: *Hochspannungsleitung* (0,07%), *niedrige Vegetation* (23,9%), *undurchlässige Oberfläche* (25,7%), *Auto* (0,61%), *Zaun/Hecke* (1,60%), *Dach* (20,2%), *Fassade* (3,61%), *Strauch* (6,31%) und *Baum* (17,9%).

Bei dem MLS-Datensatz handelt es sich um das *Oakland 3D Point Cloud Dataset* (MUNOZ et al. 2009) mit den fünf Objektklassen *Hochspannungsleitung* (0,3%), *Stange/Mast* (0,6%), *Fassade* (8,4%), *Boden* (70,5%) und *Vegetation* (20,2%).<sup>3</sup> Der Datensatz wurde in der Umgebung des Campus der *Carnegie Mellon University* (CMU) in Oakland, USA aufgenommen. Die Daten wurden mit Hilfe des *Mobile Mapping Systems Navlab11* der CMU aufgenommen. Nach der Aufnahme erfolgte eine Zuordnung der entsprechenden Referenzlabels jedes 3D-Punktes in einem semi-automatischen Prozess. Die Daten liegen aufgeteilt in einem Trainingsdatensatz mit ungefähr 37k Punkten, einem Validationsdatensatz mit 91k Punkten und mehreren Testdatensätzen mit insgesamt ungefähr 1,3M Punkten, vor.

#### 3.2 Experimente

Die Experimente werden mit den beiden in Abschnitt 2.1 beschriebenen lokalen Nachbarschaften durchgeführt. Für die optimale Nachbarschaft  $N_{opt}$  wird das Intervall für  $k$  durch  $k_{min} = 10$  und  $k_{max} = 100$  mit einer Schrittweite von  $\Delta k = 1$  definiert. Für die sphärische Nachbarschaft  $N_s$  werden für den ALS-Datensatz die Radien  $r_s = (0,5m; 1,0m; 2,0m)$  und für den MLS-Datensatz  $r_s = (0,4m; 0,6m; 0,8m)$  untersucht.

Die Extraktion der Merkmale (vgl. Abschnitt 2.2) erfolgt basierend auf den  $k$  nächsten Nachbarn, die sich aus der entsprechenden Nachbarschaftsdefinition ergeben. Die Seitenlängen  $s$  der quadratischen Bins für die Merkmale  $N_{Acc}$ ,  $\Delta H_{Acc}$  und  $\sigma_{H,Acc}$  wurde empirisch bestimmt und im Falle der ALS-Daten auf  $s = 1,25m$  und für die MLS-Daten auf  $s = 0,25m$  gesetzt.

---

<sup>2</sup>Der Datensatz kann unter <http://www2.isprs.org/commissions/comm3/wg4/data-request-form2.html> angefordert werden (Letzter Aufruf: 06.01.2017).

<sup>3</sup>Der Datensatz ist frei verfügbar unter [http://www.cs.cmu.edu/~vmr/datasets/oakland\\_3d/cvpr09/doc/](http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/) (Letzter Aufruf: 06.01.2017)

Für alle Experimente wird für den RF eine gleichbleibende Trainingsmenge  $T$  mit jeweils  $N_{tr} = 1000$  zufällig ausgewählten Beispielen pro Klasse erzeugt. Bei einer Erhöhung des Wertes, konnten keine signifikant besseren Klassifikationsergebnisse erzielt werden. Daraus ergibt sich für die ALS-Daten mit neun Klassen eine Trainingsmenge  $T$  mit  $|T| = 9000$  und für die MLS-Daten mit fünf Klassen eine Trainingsmenge mit  $|T| = 5000$ . Die Anzahl der Entscheidungsbäume wurde empirisch bestimmt und auf  $N_T = 100$  festgelegt.

Die Bestimmung der beiden Gewichtsparameter  $w_1$  und  $w_2$  des Potts-Modells erfolgt mittels einer Gittersuche auf dem gesamten Trainingsdatensatz. Hierzu werden die beiden Parameter jeweils in einem Intervall variiert und das CRF mit allen sich daraus ergebenden Kombinationen trainiert. Anhand der Trainingsgenauigkeiten wird daraufhin die beste Kombination ausgewählt. Dabei ergibt sich für den ALS-Datensatz  $w_1 = 0,6$  und  $w_2 = 0$  und für den MLS-Datensatz  $w_1 = 5$  und  $w_2 = 0,5$ . Der Schwellwert für die maximale Nachbarschaftsgröße innerhalb des Graphen wird auf  $k_{thr} = 25$  gesetzt.

### 3.3 Ergebnisse und Diskussion

Die Tab. 1 zeigt jeweils ein Ergebnis der Klassifikation mit RF und CRF für den ALS-Datensatz mit der Gesamtgenauigkeit ( $O_A$ ), Cohens Kappa Koeffizient  $\kappa$  und dem  $F1$ -Score für jede Klasse. Der  $F1$ -Score ist definiert als der harmonische Mittelwert aus Korrektheit und Vollständigkeit. Für beide Klassifikatoren wurde die Nachbarschaft  $N_{opt}$  genutzt.

Tab. 1: Gesamtgenauigkeit, Cohens-Kappa und F1-Score der einzelnen Klassen für den ALS-Datensatz.

	$O_A$	$\kappa$	Hoch.	n. Veg.	Un. Ober.	Auto	Zaun	Dach	Fass.	Strauch	Baum
RF	71,0	0,71	14,3	65,8	86,1	24,9	19,8	84,8	43,9	40,8	69,5
CRF	80,5	0,75	69,8	73,8	91,5	58,2	29,9	91,6	54,7	47,8	80,2

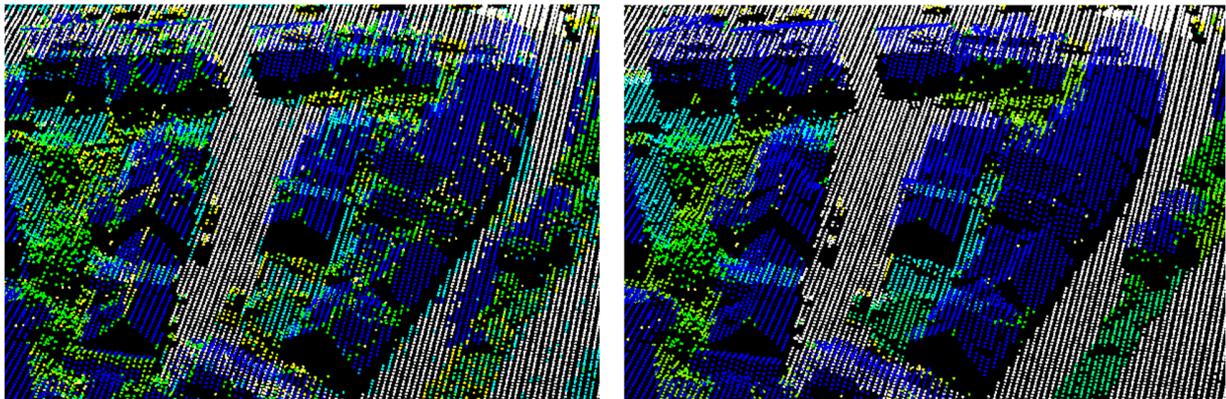


Abb. 2: Vergleich der Ergebnisse von RF (links) und CRF (rechts) anhand eines Ausschnitts der ALS-Daten.

Der Unterschied der beiden Klassifikationsverfahren in der Gesamtgenauigkeit beträgt 9 Prozentpunkte. Die Differenzen im  $F1$ -Score der einzelnen Klassen liegen jedoch zwischen 5 und 55 Prozentpunkten. Besonders die unterrepräsentierten Klassen *Hochspannungsleitung* und *Auto*, die lediglich einen Anteil von 0,07% und 0,6% bezogen auf die Gesamtpunktzahl aufweisen, zeigen Zunahmen von 55 beziehungsweise 33 Prozentpunkten im  $F1$ -Score. Bei beiden Verfahren ist auffällig, dass vor allem die Klassen *niedrige Vegetation*, *Strauch* und

*Zaun/Hecke* schlecht voneinander getrennt werden können. In Abb. 2 werden die beiden Klassifikationsergebnisse gegenüber gestellt. Dabei ist besonders bei der Klasse *Dach* (blau) der Einfluss der kontextbasierten Klassifikation sichtbar.

In Tab. 2 sind die Ergebnisse der Klassifikation mit RF und CRF für den MLS-Datensatz dargestellt. Die besten Resultate konnten mit einem RF mit der Nachbarschaft  $N_{opt}$  und mit einem CRF mit einer sphärischen Nachbarschaft mit fixiertem Radius  $r_s = 0,8\text{m}$  erzielt werden.

Tab. 2: Gesamtgenauigkeit, Cohens-Kappa und F1-Score der einzelnen Klassen für den MLS-Datensatz.

	$O_A$	$\kappa$	Hoch.	Stange	Fass.	Boden	Veg.
<b>RF</b>	92,0	0,83	12,1	32,6	72,9	98,8	87,9
<b>CRF</b>	97,7	0,95	33,0	70,0	94,3	99,5	96,1

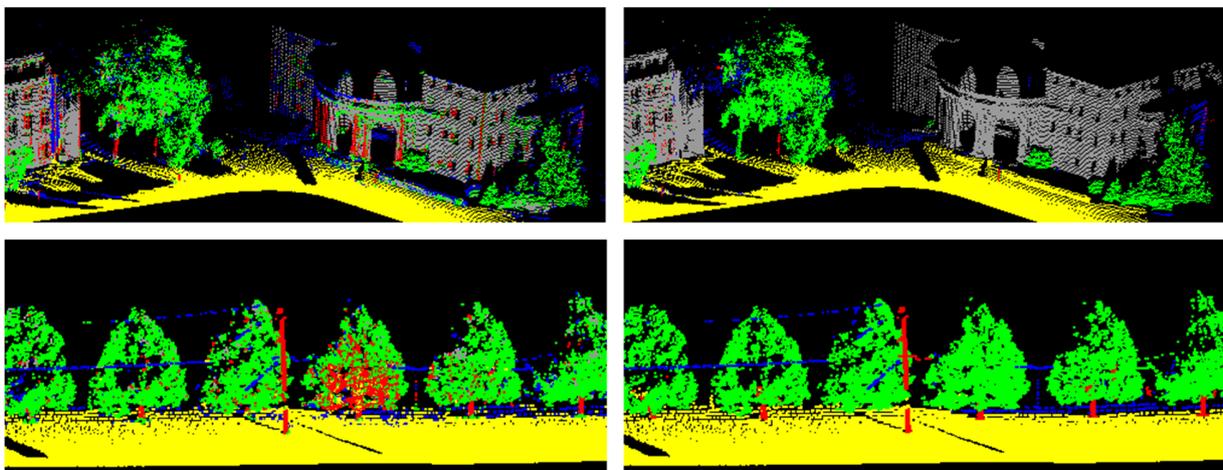


Abb. 3: Vergleich der Ergebnisse von RF (links) und CRF (rechts) anhand eines Ausschnitts der MLS-Daten.

Die hohen Gesamtgenauigkeiten von 92% und 97,7% werden durch die sehr gut klassifizierte Objektklasse *Boden* stark beeinflusst, die 70,5% der Punkte im Testdatensatz ausmacht. Bei beiden Verfahren werden die Klassen *Hochspannungsleitung* und *Stange/Mast* mit nur einer geringen Anzahl an Punkten weniger gut erkannt. Im Vergleich mit RF, können mit Hilfe des CRF vor allem für solche Klassen höhere Werte für den *F1-Score* erreicht werden. Insgesamt ist das CRF-Klassifikationsergebnis bei den MLS-Daten, wie bei den ALS-Daten, glatter als bei RF. Ein Vergleich anhand zweier Ausschnitte der klassifizierten Testdaten ist in Abb. 3 dargestellt. Besonders sichtbar ist dieser Glättungseffekt bei den Klassen *Fassade* (grau) und *Vegetation* (grün).

Zum Vergleich des vorgestellten Verfahrens eignen sich für die ALS-Daten die auf der Seite des *ISPRS Semantic Labeling Contest*<sup>4</sup> eingereichten Ergebnisse. RAMIYA et al. (2016) verwenden für die Klassifikation des Datensatzes eine 3D-Segmentierung und extrahieren aus den Segmenten spektrale und geometrische Merkmale. Hierfür wurde neben der 3D-Punktwolke auch ein *True Orthophoto* verwendet. Ein weiteres Ergebnis wurde von CVIRN et al. (2016)

<sup>4</sup> Die eingereichten Ergebnisse des ISPRS Semantic Labeling Contest sind unter <http://www2.isprs.org/vaihingen-3d-semantic-labeling.html> abrufbar.

eingereicht. Zur Klassifikation wird *Multiclass Machine Learning* mit einer *one-vs-one* Strategie, verbunden mit einem genetischen Algorithmus zur Merkmalsauswahl, angewendet. Als Merkmale werden radiometrische, strukturelle und geometrische Attribute genutzt. Das Verfahren von RAMIYA et al. (2016) resultiert in eine Gesamtgenauigkeit von 76,2%. Allerdings wurde die Objektklasse *Fassade* nicht klassifiziert, was einen negativen Einfluss auf das Ergebnis hat. Mit dem in dieser Arbeit verwendeten Verfahren kann, im Vergleich zu RAMIYA et al. (2016), bei den einzelnen Objektklassen ein etwas besserer *F1-Score* erzielt werden. Die Unterschiede liegen zwischen 0,3 (*Auto*) und 15,4 Prozentpunkten (*Hochspannungsleitung*). CVIRN et al. (2016) erzielten eine Gesamtgenauigkeit von 80,8% für den Datensatz. Auch bei diesem Ansatz werden die unterrepräsentierten Klassen *Hochspannungsleitung*, *Auto* und *Zaun/Hecke*, gemessen am *F1-Score*, schlechter detektiert. Die Klassen *niedrige Vegetation* und *Dach* werden besser erkannt.

Für den MLS-Datensatz werden im Rahmen eines Vergleichs die Ergebnisse von MUNOZ et al. (2009) herangezogen. Die kontextbasierte Klassifikation der Punktwolke geschieht mittels *Functional Max-Margin Markov Networks*. Der Vergleich anhand des *F1-Scores* zeigt lediglich größere Unterschiede bei den Klassen *Hochspannungsleitung* und *Stange/Mast*. Während mit dem in dieser Arbeit verwendeten CRF die Klasse *Stange/Mast* einen um 30,6 Prozentpunkte höheren *F1-Score* aufweist, ist dieser für die Klasse *Hochspannungsleitung* um 31,3 Prozentpunkte geringer.

## 4 Fazit und Ausblick

Das in dieser Arbeit verwendete Verfahren liefert sowohl für den ALS-, als auch für den MLS-Datensatz gute Ergebnisse. Die Anpassungen bezüglich der zwei unterschiedlichen Aufnahmemethoden beschränken sich lediglich auf die Optimierung der beiden Gewichtsparameter. Durch das Einbeziehen von Kontextinformationen kann das Klassifikationsergebnis hinsichtlich der individuellen Klassifikation in allen Belangen verbessert werden. Besonders unterrepräsentierte Klassen mit wenigen Punkten im Datensatz profitieren von der kontextbasierten Klassifikation.

Eine weiterführende Entwicklung des vorgestellten Verfahrens könnte bei der Verbesserung der Klassifikationsgenauigkeit ansetzen, indem weitere Merkmale und Nachbarschaften auf ihre Eignung untersucht würden. Zusätzlich könnte die Übertragbarkeit des trainierten Klassifikators auf unterschiedliche Gebiete und Sensoren überprüft werden. Abschließend könnten die Ergebnisse genutzt werden, um einzelne Objekte wie zum Beispiel Bäume zu extrahieren.

## 5 Literaturverzeichnis

- BOROS, E. & HAMMER, P. L., 2002: Pseudo-boolean optimization. *Discrete Applied Mathematics* **123**(1-3), 155-225.
- BOYKOV, Y. Y. & JOLLY, M.-P., 2001: Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *Proceedings of the Eighth IEEE International Conference on Computer Vision* **1**, 105-112.
- BREIMAN, L., 2001: Random Forests. *Machine Learning* **45**(1), 5-32.

- CORTESA, C. & VAPNIK, V., 1995: Support-Vector Networks. *Machine Learning* **20**(3), 273-297.
- CVIRN, R.; HORVAT, D. & MONGUS, D., 2016. [http://ftp.ipi.uni-hannover.de/ISPRS\\_WGIII\\_website/ISPRSIII\\_4\\_Test\\_results/papers/UM\\_DescriptionOfAlgorithms.pdf](http://ftp.ipi.uni-hannover.de/ISPRS_WGIII_website/ISPRSIII_4_Test_results/papers/UM_DescriptionOfAlgorithms.pdf), letzter Zugriff 20.01.2017.
- KUMAR, S. & HEBERT, M., 2006: Discriminative Random Fields. *International Journal of Computer Vision* **68**(2), 179-201.
- MUNOZ, D.; BAGNELL, J. A.; VANDAPEL, N. & HEBERT, M., 2009: Contextual Classification with Functional Max-Margin Markov Networks. *IEEE Conference on Computer Vision and Pattern Recognition 2009*, 975-982.
- NIEMEYER, J.; ROTTENSTEINER, F. & SOERGEL, U., 2014: Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* **87**, 152-165.
- RAMIYA, A. M.; NIDAMANURI, R. R. & KRISHNAN, R., 2016: 3D semantic labelling of colored LiDAR point cloud using supervoxels based segmentation approach/color based region growing segmentation approach. [http://ftp.ipi.uni-hannover.de/ISPRS\\_WGIII\\_website/ISPRSIII\\_4\\_Test\\_results/papers/IIST\\_3D\\_ext.pdf](http://ftp.ipi.uni-hannover.de/ISPRS_WGIII_website/ISPRSIII_4_Test_results/papers/IIST_3D_ext.pdf), letzter Zugriff 20.01.2017.
- ROTTENSTEINER, F.; SOHN, G.; JUNG, J.; GERKE, M.; BAILLARD, C.; BENITEZ, S. & BREITKOPF, U., 2014: Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* **93**, 256-271.
- WEGNER, J. D.; HÄNSCH, R.; THIELE, A. & SOERGEL, U., 2011: Building Detection From One Orthophoto and High-Resolution InSAR Data Using Conditional Random Fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **4**(1), 83-91.
- WEINMANN, M.; JUTZI, B. & MALLET, C., 2014: Semantic 3D scene interpretation: a framework combining optimal neighborhood size selection with relevant features. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **II-3**, 181-188.
- WEINMANN, M.; SCHMIDT, A. & MALLET, C., 2015: Contextual classification of point cloud data by exploiting individual 3D neighborhoods. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **II-3/W4**, 271-278.
- WURM, K. M.; KRETZSCHMAR, H.; KÜMMERLE, R.; STACHNISS, C. & BURGARD, W., 2014: Identifying vegetation from laser data in structured outdoor environments. *Robotics and Autonomous Systems* **62**(5), 675-684.
- YANG, M. Y. & FÖRSTNER, W., 2011: A hierarchical conditional random field model for labeling and classifying images of man-made scenes. *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 169-203.
- YONGTAO, Y.; JONATHAN, L.; HAIYAN, G. & CHENG, W., 2015: Automated Extraction of Urban Road Facilities Using Mobile Laser Scanning Data. *IEEE Transactions on intelligent transportation systems* **16**(4), 2167-2181.