

**Institut für Photogrammetrie und Fernerkundung (IPF)
des Karlsruher Institut für Technologie (KIT)**

**Methoden zur automatischen
Szenencharakterisierung basierend
auf aktiven optischen Sensoren für
die Photogrammetrie und
Fernerkundung**

Habilitationsschrift
genehmigt von der Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften (BGU)

von

Dr.-Ing. Boris Jutzi

Nach § 3 Abs. 1 der Habilitationsordnung der Universität Karlsruhe vom 23.11.1999 anerkannte Habilitationsschrift. Habilitationsschrift, genehmigt von der Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften (BGU) des Karlsruher Institut für Technologie (KIT), 2015.

Datum des Kolloquiums

22.7.2015

Habilitationskommission

Prof. Dr.-Ing. habil. Dr. h.c. Bernhard Heck, Karlsruher Institut für Technologie (KIT)^a

Prof. Dr.-Ing. habil. Stefan Hinz, Karlsruher Institut für Technologie (KIT)^b

Prof. Dr.-Ing. habil. Richard Bamler, Technische Universität München (TUM)^b

Prof. Dr. sc. techn. habil. Hans-Gerd Maas, Technische Universität Dresden^b

Prof. Dr.-Ing. Maria Hennes, Karlsruher Institut für Technologie (KIT)^c

Prof. Dr.-Ing. Sascha Gentes, Karlsruher Institut für Technologie (KIT)^c

Prof. Dr.-Ing. Dr. h.c. mult. Franz Nestmann, Karlsruher Institut für Technologie (KIT)^c

Prof. Dr. rer. nat. habil. Sebastian Schmidlein, Karlsruher Institut für Technologie (KIT)^c

Prof. Dr.-Ing. Peter Vortisch, Karlsruher Institut für Technologie (KIT)^c

doi:10.5445/IR/1000050691

^aVorsitzender

^bReferenden

^cWeitere Kommissionsmitgliederinnen und -mitglieder

Erfahrung ist verstandene Wahrnehmung.
Immanuel Kant

Inhaltsverzeichnis

1	Motivation	1
1.1	Taxonomie der aktiven Sensoren	1
1.2	Aufbau der Habilitationsschrift	7
1.3	Konzeption der Forschungsaktivitäten	10
2	Zusammenfassung der entwickelten Methoden	11
2.1	Modellierung und Analyse von Messsignalen	11
2.1.1	<i>Phase-Unwrapping</i> bei Entfernungsbildern	12
2.1.2	Signaloptimierung bei Full-Waveform-Laserscanner-Daten . . .	18
2.1.3	Radiometrische Kalibrierung bei Laserscanner-Daten	23
2.2	Koregistrierung mit 2D- bzw. 3D-Sensoren	26
2.2.1	3D/2D-Koregistrierung	27
2.2.2	3D-Koregistrierung	32
2.3	Szenenanalyse	37
2.3.1	Bestimmung der optimalen lokalen Nachbarschaften	38
2.3.2	Merkmalsextraktion	40
2.3.3	Auswahl relevanter Merkmale	43
2.3.4	Klassifikation	45
2.4	Objektrekonstruktion	46
2.4.1	Natürliche Objekte	46
2.4.2	Anthropogene Objekte	47
3	Diskussion und Ausblick	49
4	Relevante Publikationen als Teil der Habilitationsschrift	53
A.	Publikationen zu 2.1 Modellierung und Analyse von Messsignalen	55
B.	Publikationen zu 2.2 Koregistrierung mit 2D- bzw. 3D-Sensoren	127
C.	Publikationen zu 2.3 Szenenanalyse	199
D.	Publikationen zu 2.4 Objektrekonstruktion	245
Z.	Nicht referenzierte Publikationen des Verfassers	275
	Literaturverzeichnis	279

1 Motivation

In der Photogrammetrie und Fernerkundung werden unterschiedliche Messsysteme zur Datenerfassung eingesetzt. Die eingesetzten Messsysteme lassen sich in passive und aktive Systeme einteilen. Bei den aktiven Systemen wird elektromagnetische Strahlung verwendet um die Szene zu bestrahlen und damit spezifische Eigenschaften der bestrahlten Oberflächen abzuleiten. Eine Unterscheidung ist bei den aktiven Systemen hinsichtlich der eingesetzten Wellenlänge der Strahlung möglich. *Radio Detection and Ranging* (RADAR) Systeme strahlen mit Wellenlängen im m- bis mm-Bereich, elektro-optische (EO) Systeme [47] hingegen mit Wellenlängen im μ - bis nm-Bereich. Im Folgenden werden diese elektro-optischen Systeme zur Vereinfachung als aktive optische Sensoren oder kurz aktive Sensoren bezeichnet. Solche aktive Sensoren sind beispielsweise Laserscanner [84, 95], strukturierte Lichtprojektion Sensoren oder Entfernungskameras [16].

Die mit aktiven Sensoren (Kapitel 1.1) erfassten 3D-Daten haben spezielle Eigenschaften weswegen für eine automatische Verarbeitung anwendungsspezifische Methoden (Kapitel 2) entwickelt werden müssen um eine Szene zu analysieren. Mittels vier aufeinander aufbauenden Verarbeitungsstufen (Kapitel 1.2) wird diese Zielsetzung angegangen und umgesetzt.

1.1 Taxonomie der aktiven Sensoren

In Abhängigkeit von der Anwendung werden spezielle aktive Sensoren eingesetzt. Die Charakterisierung dieser aktiven Sensoren kann nach unterschiedlichen Kriterien erfolgen [A3]. Die in dieser Habilitationsschrift eingesetzte Taxonomie orientiert sich an den Kategorien: Modulationstechnik, Detektionstechnik, Anordnung, Messtechnik, Beleuchtungsquelle, Sichtfeld und Entfernungsbereich (Abbildung 1.1).



Abbildung 1.1: Taxonomie der aktiven Sensoren nach den Kategorien: Modulationstechnik, Detektionstechnik, Anordnung, Messtechnik, Beleuchtungsquelle, Sichtfeld und Entfernungsbereich.

Modulationstechnik

Die Modulation von elektromagnetischer Strahlung erfolgt bei Aktiven Sensoren durch Dauerstrich- bzw. Pulsmodulation (*Continuous-Wave, CW bzw. Pulse Modulation*). Hierbei ist die eingesetzte Modulationstechnik einer bestimmten Messtechnik zugeordnet, bei Dauerstrichmodulation erfolgt eine Phasendifferenzmessung und bei Pulsmodulation eine Laufzeitmessung.

Detektionstechnik

Die Detektionstechniken sind unterteilbar in Kohärente Detektion (*Coherent Detection*) und Direkte Detektion (*Direct Detection*). Der Unterschied dieser Detektionstechniken besteht darin, dass bei kohärenter Detektion die Kohärenz der elektromagnetischen Strahlung für die Detektion berücksichtigt wird und bei direkter Detektion die Kohärenz unberücksichtigt bleibt. Die kohärenten Detektionsmethoden sind Heterodyne Detektion (*Heterodyne Detection*) und Homodyne Detektion (*Homodyne Detection*). Die direkten Detektionsmethoden sind Mehrfach-Photonen-Detektion (*Multi Photon Detection*) und Einzel-Photon-Detektion (*Single Photon Detection*).

Anordnung

Bei der Anordnung der Sende- und Empfangseinheit erfolgt eine Unterscheidung zwischen monostatischer (*monostatic*) und bistatischer (*bistatic*) Anordnung, wobei der technische Aufwand für ein monostatisches System größer ist, als der technische Aufwand für ein bistatisches System.

Messtechnik

Die bei aktiven Sensoren eingesetzte Messtechnik ermöglicht die Entfernung direkt oder indirekt bestimmt. Bei der Phasendifferenz- und Laufzeitmessung wird direkt aus dem gemessenen Phasendifferenzwert bzw. Laufzeitwert die Entfernung abgeleitet. Bei der Triangulationsmessung wird unter Berücksichtigung der inneren Orientierung und der relativen Orientierung zwischen Sende- und Empfangseinheit indirekt über den gemessenen Disparitätswert die Entfernung bestimmt. Zudem ist der funktionale Zusammenhang zwischen gemessenem Wert und Entfernung linear (direkt) bzw. nichtlinear (indirekt).

Beleuchtungsquelle

Als Beleuchtungsquelle werden bei aktiven optischen Sensoren wegen ihrer hohen Strahlintensität meist *Light Amplification by Stimulated Emission of Radiation* (Laser), Laserdioden oder Leuchtdioden eingesetzt.

Sichtfeld

Das Sichtfeld (*Field of View, FOV*) ist konstruktionsbedingt durch die Anordnung der einzelnen Sensorelemente geprägt, hierbei werden bei Flächensensoren zweidimensionale Abbildungen ermöglicht, hingegen bei Zeilensensoren bzw. Zeilenscanner nur eindimensionale Abbildungen. Zeilensensoren bestehen aus geradlinig angeordneten Sensorelementen die synchron Daten aufnehmen. Bei Zeilenscanner hingegen

wird ein einziges Sensorelement geradeaus bewegt, währenddessen sequentiell durch punktuelle Abtastung gemessen wird.

Entfernungsbereich

Bei aktiven Sensoren ist der operative Entfernungsbereich unterteilbar in einen nahen Entfernungsbereich (*Close Range* oder *Short Range*), mittleren Entfernungsbereich (*Mid Range*) und fernen Entfernungsbereich (*Far Range*). Die Abgrenzung zwischen den Entfernungsbereichen ist unscharf, die Übergänge sind fließend, eine strikte Trennung ist nur schwer möglich. Trotzdem wird im Folgenden eine grobe Einteilung vorgenommen:

Der nahe Entfernungsbereich (Abbildung 1.5, rot) umspannt Entfernungen unterhalb von mehreren Metern bzw. Zehnermetern. Typische Vertreter von aktiven Sensoren, die in diesem Entfernungsbereich eingesetzt werden, sind strukturierte Lichtprojektion Sensoren (z.B. Microsoft Kinect 1.0), *Time-of-Flight* Entfernungskameras (z.B. MESA SR-4000, PMD CamCube 2.0) oder einfache Laserscanner (z.B. Hokuyo UTM-30LX) (Abbildung 1.2).



Abbildung 1.2: Aktive Sensoren die im nahen Entfernungsbereich messen: Strukturierte Lichtprojektion Sensoren (z.B. Microsoft Kinect 1.0 (a)), *Time-of-Flight* Entfernungskameras (z.B. MESA SR-4000 (b)), PMD CamCube 2.0 (c) und einfache Laserscanner (z.B. Hokuyo UTM-30LX (d)).

Der ferne Entfernungsbereich (Abbildung 1.5, blau) umspannt Entfernungen deutlich oberhalb von mehreren hundert Metern bzw. Kilometer. Flugzeuggetragene Laserscanner (*Airborne Laser Scanning, ALS*) und Satellitengetragene Laserscanner (*Satellite Laser Scanning, SLS*) werden diesem Entfernungsbereich zugeordnet. Typische Vertreter von aktiven Sensoren die in diesem Entfernungsbereich operieren sind auf Flugzeugen (z.B. RIEGL LMS-Q780, Toposys Falcon III, Optech Titan, Leica ALS60) (Abbildung 1.3) oder auf Satelliten (z.B. Geoscience Laser Altimeter System, GLAS) montiert.



Abbildung 1.3: Aktive Sensoren die im fernen Entfernungsbereich messen: Flugzeuggetragene Laserscanner (z.B. RIEGL LMS-Q780 (a), Toposys Falcon III (b), Optech Titan (c), Leica ALS60 (d)).

Zwischen nah und fern liegt der mittlere Entfernungsbereich (Abbildung 1.5, grün) bei dem hauptsächlich Mobile Laserscanner (*Mobile Laser Scanning*, MLS) und Terrestrische Laserscanner (*Terrestrial Laser Scanning*, TLS) eingesetzt werden. Typische Vertreter von aktiven Sensoren die in diesem Entfernungsbereich operieren, befinden sich auf mobilen Robotern bzw. Fahrzeugen (z.B. SICK LMS511, Velodyne HDL-64E, Velodyne VLP-16) oder werden für terrestrische Anwendungen (z.B. Leica HDS6000, Z+F Imager 5006h, RIEGL VZ-4000) eingesetzt (Abbildung 1.4).



Abbildung 1.4: Aktive Sensoren die im mittleren Entfernungsbereich messen: Mobile Laserscanner (z.B. SICK LMS511 (a), Velodyne HDL-64E (b), Velodyne VLP-16 (c)) und Terrestrische Laserscanner (z.B. Leica HDS6000 (d), Z+F Imager 5006h (e), RIEGL VZ-4000 (f)).

Die in Kapitel 2 eingesetzten aktiven Sensoren sind alle in Abbildung 1.5 dargestellt und entsprechend der drei Entfernungsbereiche farblich (rot, blau, grün) markiert. Zudem ist für jeden Sensor das vom Hersteller spezifizierte operative Entfernungintervall aufgetragen.

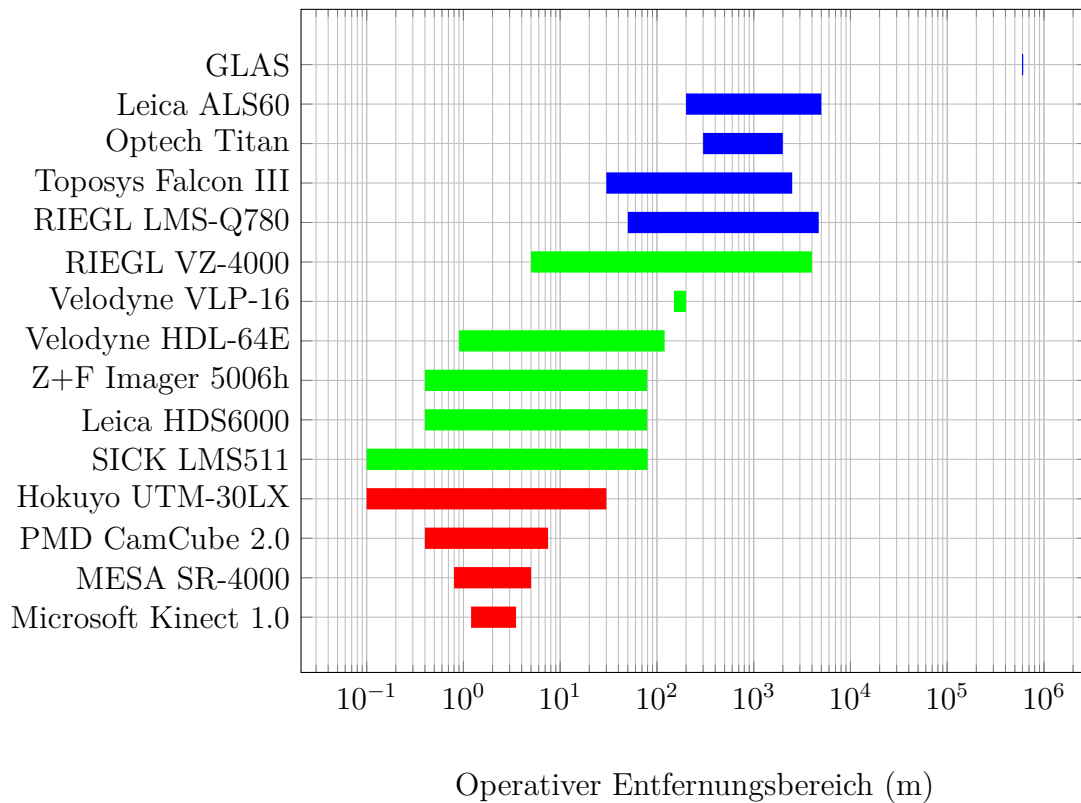


Abbildung 1.5: Vom Hersteller spezifiziertes operatives Entfernungintervall bei verschiedenen aktiven Sensoren. Gruppirt nach dem nahen (rot), mittleren (blau) und fernen (grün) Entfernungsbereich.

1.2 Aufbau der Habilitationsschrift

In diesem Kapitel wird ein Überblick über den Aufbau der Habilitationsschrift gegeben (Abbildung 1.6), der sowohl die wesentliche Ziele als auch die übergeordnete Konzeption darstellt. Dieses Kapitel dient der wissenschaftlichen Einordnung und ist gleichzeitig als nützlicher Hinweis zum Lesen der Habilitationsschrift gedacht.

In **Kapitel 1** wird zum Habilitationsthema hingeführt, indem für die automatischen Szenencharakterisierung in der Photogrammetrie und Fernerkundung mittels neue aktive optische Sensoren motiviert wird. Insbesondere die Charakterisierung von aktiven optischen Sensoren (Kapitel 1.1) hinsichtlich unterschiedlichen Kriterien verdeutlicht deren spezielle Eigenschaften, wodurch die Datenerfassung unterschiedlich beeinflusst wird. Ausgehend von den erfassten Daten werden zielgerichtete Forschungsaktivitäten konzipiert (Kapitel 1.3), die eine automatische Szenencharakterisierung ermöglichen.

Aus dieser Konzeption werden vier aufeinander aufbauenden Verarbeitungsstufen entwickelt. In **Kapitel 2** werden die hierfür relevanten Methoden vorgestellt.

- Beginnend mit der ersten Verarbeitungsstufe **Modellierung und Analyse von Messsignalen** (Kapitel 2.1) sollen die mit aktiven Sensoren erfassten Daten dahingehend aufbereitet werden, dass ein wesentlicher Zugewinn für die nachfolgenden Verarbeitungsstufen erreicht wird. Beispielsweise wurden spezielle Signalverarbeitungsmethoden für Entfernungskameras (Kapitel 2.1.1) und Full-Waveform-Laserscanner (Kapitel 2.1.2) entwickelt bzw. eine radiometrische Kalibrierung (Kapitel 2.1.3) basierend auf verschiedenen Beleuchtungsmodellen konzipiert.
- Nach der Aufbereitung der Daten werden in der zweiten Verarbeitungsstufe durch die **Koregistrierung mit 2D- bzw. 3D-Sensoren** (Kapitel 2.2) die Beobachtungen durch mehrere Sensoren bzw. durch einen bewegten Sensor für eine umfassende Szenenerfassung vereinheitlicht. Hierfür ist beispielsweise die Zusammenführung von geometrischen und radiometrischen Daten bzw. die Überführung von 2.5D-Daten in eine übergeordnete Punktwolke zweckdienlich. Deswegen wurden neue Lösungsansätze zur Bestimmung der relativen Orientierung zwischen aktiven 3D-Sensoren und passiven 2D-Sensoren (Kapitel 2.2.1) sowie bei bewegten aktiven 3D-Sensoren (Kapitel 2.2.2) entwickelt. Im Speziellen wurde hierfür eine Entfernungskamera mit einer thermischen Infrarot-Kamera (Kapitel 2.2.1.1) bzw. ein mobiler Laserscanner mit einer RGB-Kamera (Kapitel 2.2.1.2) kombiniert. Zudem wurden neue Methoden entwickelt, bei denen keine zusätzliche Sensoren zur Bestimmung der Position und der Lage des aktiven 3D-Sensors verwendet werden, sondern ausschließlich die mit scannenden aktiven 3D-Sensoren (Kapitel 2.2.2.1) bzw. bildgebenden aktiven 3D-Sensoren (Kapitel 2.2.2.2) erfassten Daten berücksichtigt werden.
- Nach der Vereinheitlichung der Daten wird in der dritten Verarbeitungsstufe durch **Szenenanalyse** (Kapitel 2.3) der Inhalt einer Szene bezüglich der darin enthaltenen Strukturen effizient analysiert. Da aktive Sensoren primär der geometrischen Erfassung dienen, werden bei der Analyse die geometrische Eigenschaften der Szene in Form von Punktwolken priorisiert. Weil die Charakterisierung von 3D-Punkten sowohl von der nicht bekannten Szenenstruktur als auch von der zu wählenden Nachbarschaft geprägt ist, ist die optimale lokale Nachbarschaft zur Merkmalsberechnung für jeden 3D-Punkten gesondert zu bestimmen (Kapitel 2.3.1). Für jeden 3D-Punkt werden aus der optimalen lokalen Nachbarschaft verschiedene geometrische 3D-Merkmale und 2D-Merkmale extrahiert (Kapitel 2.3.2) und dem 3D-Punkt zugewiesen. Da die Relevanz der geometrischen Merkmale von den Szeneninhalten beeinflusst wird, ist es sinnvoll für eine performante Klassifizierung eine optimale Auswahl von relevanten Merkmalen (Kapitel 2.3.3) zu bestimmen. Abschließend wird basierend auf den relevanten Merkmalen eine Klassifikation (Kapitel 2.3.4) durchgeführt.

- Nach der sehr allgemein gehaltenen Szenenanalyse wird in der vierten Verarbeitungsstufe durch **Objektrekonstruktion** (Kapitel 2.4) eine differenzierte Analyse der Szene hinsichtlich einzelner Objekte vorgenommen. Es werden sowohl natürliche Objekte (Kapitel 2.4.1) als auch anthropogene Objekte (Kapitel 2.4.2) untersucht.

In **Kapitel 3** ist abschließend eine kurze Zusammenfassung und der Ausblick auf zukünftige Forschungsmöglichkeiten gegeben.

Ergänzend sind in **Kapitel 4** die Forschungsaktivitäten in Form von relevanten Publikationen berücksichtigt. Die durchgeführten Forschungsaktivitäten werden zudem hinsichtlich der übergeordneten Konzeption im folgenden Kapitel 1.3 vertieft.

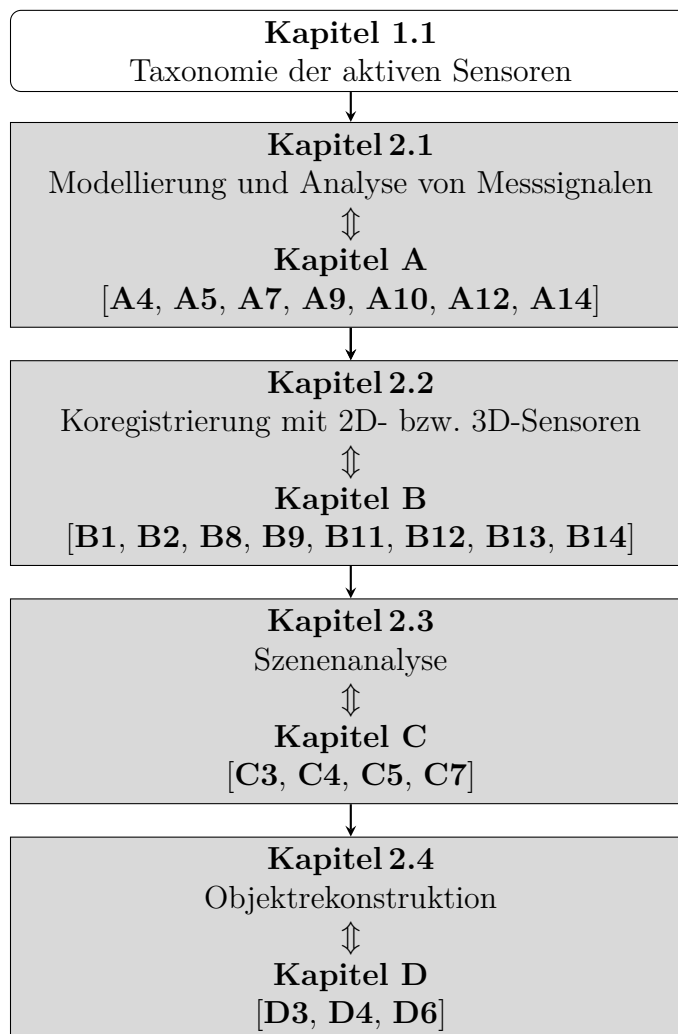


Abbildung 1.6: Aufbau der Habilitationsschrift hinsichtlich aktiver Sensoren, den vier Verarbeitungsstufen und den zugeordneten Schlüsselpublikationen.

1.3 Konzeption der Forschungsaktivitäten

Die vorliegende Habilitationsschrift entstand im Rahmen meiner Forschungsaktivitäten am Institut für Photogrammetrie und Fernerkundung (IPF) des Karlsruher Institut für Technologie (KIT). Die Forschungsaktivitäten waren und sind inhaltlich auf die Thematik **Methoden zur automatischen Szenencharakterisierung basierend auf aktiven optischen Sensoren für die Photogrammetrie und Fernerkundung** ausgerichtet. Für deren Umsetzung sind sowohl mathematisch-analytische, numerische und informationstechnische Grundlagen, die radiometrische und geometrische Modellierung, die Evaluation der automatischen Auswerteverfahren, als auch die eigenständige und innovative Methodenentwicklungen essentiell, um im Bereich der Bild- und Signalverarbeitung innerhalb der Photogrammetrie, Fernerkundung und Geodäsie erfolgreich wissenschaftlich zu wirken.

Forschungsaktivitäten werden durch studentische Arbeiten, Doktorarbeiten und Forschungsk Kooperationen wesentlich unterstützt. Insbesondere die vom Autor betreuten Doktorarbeiten und die Forschungsk Kooperationen ermöglichen hierbei eine weiterführende inhaltliche Vertiefung sowie die zielgerichtete Umsetzung der übergeordneten Konzeption. Die Gesamtkonzeption der inhaltlichen Forschungsrichtungen wurden konsequent durch den Verfasser der Habilitationsschrift geprägt (Kapitel 1.2) und realisiert (Kapitel 2). Die dazu entstandenen 45 relevanten Publikationen sind den vier Verarbeitungsstufen zugeordnet (Kapitel 4). Zudem sind die insgesamt 22 Schlüsselpublikationen¹ (Abbildung 1.6) als Abdruck vollständig in der Habilitationsschrift enthalten.

Vom Verfasser der Habilitationsschrift wurde zudem federführend mit diversen nationalen und internationalen Institutionen gezielt Forschungsk Kooperationen initiiert, um die wesentlichen Forschungsthemen dieser Arbeit zu vertiefen und die angestrebten Ziele zu erreichen, beispielsweise:

- Fachgebiet Photogrammetrie und Fernerkundung (PF) der Technischen Universität München (TUM) zum Thema **Entfernungskamera kombiniert mit thermischer Infrarot-Kamera** (Kapitel 2.2.1.1).
- Institut für Optronik, Systemtechnik und Bildauswertung (IOSB) der Fraunhofer-Gesellschaft (FhG) zum Thema **Mobiler Laserscanner kombiniert mit RGB-Kamera** (Kapitel 2.2.1.2).
- *Laboratoire Méthodes d'Analyses pour le Traitement d'Images et la Stéréorestitution* (MATIS) des *Institut Géographique National France International* (IGN) zum Thema **Auswahl relevanter Merkmale** (Kapitel 2.3.3).
- *Department of Forest Sciences* der *University of Helsinki* (UH) zum Thema **Objektrekonstruktion** (Kapitel 2.4).

¹Weiterführende Information hinsichtlich der Einordnung von Schlüsselpublikationen und relevanten Publikationen sind in Kapitel 4 gegeben.

2 Zusammenfassung der entwickelten Methoden

In diesem Kapitel werden die entwickelten Methoden vorgestellt. Ausgehend von Daten, die mit einem aktiven Sensor erfasst werden, lassen sich folgende vier aufeinander aufbauende Verarbeitungsstufen ausmachen, um die Daten nutzbringend zu analysieren. Beginnend mit der gezielten Aufbereitung der Rohdaten kann bereits durch Modellierung und Analyse von Messsignalen (Kapitel 2.1 mit Kapitel A) ein wesentlicher Zugewinn für die nachfolgenden Verarbeitungsstufen erreicht werden, beispielsweise durch eine verbesserte Erfassung der Szene, die zusätzliche Möglichkeiten für die weiterführende Verarbeitung ermöglicht. In der darauf folgenden Verarbeitungsstufe werden durch Registrierung (Kapitel 2.2 mit Kapitel B) einzelne voneinander unabhängige Beobachtungen in eine übergeordnete vereinheitlichte Repräsentation überführt, beispielsweise durch die Zusammenführung von geometrischen und radiometrischen Daten bzw. die Überführung von 2.5D-Daten in eine Punktwolke. In der nächsten Verarbeitungsstufe wird die Punktwolke hinsichtlich Szeneninhalte analysiert (Kapitel 2.3 mit Kapitel C). Hierbei werden relevante Merkmale ausgewählt und für jeden 3D-Punkt eine optimierte lokale Nachbarschaft zur Berechnung der Merkmale berücksichtigt, um eine erhöhte Gesamtgenauigkeit bei der Klassifikation zu erreichen. Mit der letzten Verarbeitungsstufe wird durch Objektdetektion (Kapitel 2.4 mit Kapitel D) eine differenzierte Analyse der Szene hinsichtlich einzelner Objekte vorgestellt.

2.1 Modellierung und Analyse von Messsignalen

Bei der ersten Verarbeitungsstufe werden die mit aktiven Sensoren erfassten Daten für die nachfolgenden Verarbeitungsstufen aufbereitet. Durch die Modellierung und Analyse von Messsignalen können die Daten dahingehend optimiert werden, dass eine verbesserte Erfassung der Szene ermöglicht wird. Die Verbesserung kann beispielsweise hinsichtlich geometrischer und radiometrischer Aspekte erfolgen. Durch Modellierung des Signals ergeben sich zusätzliche Möglichkeiten bei der Analyse von Messsignalen. Hierfür wurden spezielle Signalverarbeitungsmethoden für Entfernungskameras (Kapitel 2.1.1) und Full-Waveform-Laserscanner (Kapitel 2.1.2) entwickelt bzw. eine radiometrische Kalibrierung (Kapitel 2.1.3) basierend auf verschiedenen Beleuchtungsmodellen konzipiert.

2.1.1 *Phase-Unwrapping* bei Entfernungsbildern

In diesem Kapitel werden zwei Verfahren vorgestellt, die den eingeschränkten Entfernungsmessbereich bei Entfernungskameras (z.B. MESA SR-4000 oder PMD Cam-Cube 2.0; Kapitel 1.1) bild- bzw. signalbasiert über den vom Hersteller spezifizierten Messbereich hinaus zu erweitern (Abbildung 2.1).

Bei der Verwendung von dauerstrichmodulierten (*Continuous-Wave*) Lichtquellen kann durch Messen der Phasendifferenz $\Delta\varphi$ die Entfernung ΔR zwischen Sensor und beleuchteter Objektoberfläche berechnet werden mit

$$\Delta R = \frac{c}{2f_m} \frac{\Delta\varphi}{2\pi}, \quad (2.1)$$

wobei f_m der Modulationsfrequenz und c der Lichtgeschwindigkeit entspricht [16].

Durch die Phasendifferenz $\Delta\varphi$ sind Phasenmehrdeutigkeiten unvermeidbar, welche Mehrdeutigkeiten bei der Entfernungsmessung ΔR verursachen (Abbildung 2.1c und d). In Abhängigkeit von der verwendeten Modulationsfrequenzrate (z.B. 20 MHz) ergibt sich ein eindeutiger Entfernungsmessbereich (z.B. 7,5 m). Absolute Entfernungen R werden fälschlicherweise zu kurz erfasst, wenn die Werte über dem eindeutigen Entfernungsmessbereich R_m mit

$$R_m = \frac{c}{2f_m} \quad (2.2)$$

liegen.

Es werden im Folgenden neue bild- bzw. signalbasierte Methoden für Entfernungskameras vorgestellt mit denen durch Phasenabwicklung (*phase-unwrapping*) der Entfernungsmessbereich um ein Vielfaches erweitert werden kann.

Bildbasierte Erweiterung des Entfernungsmessbereichs

Aus der *Radio Detection and Ranging* (RADAR) Interferometrie [30, 37] sind verschiedene bildbasierte Methoden bekannt, die zum *Phase-Unwrapping* bei Phasenmehrdeutigkeiten verwendet werden können. Um die prinzipielle Funktionsweise zur Phasenrekonstruktion speziell bei Entfernungsbildern, die mit einer Entfernungskamera gemessen wurden, zu untersuchen, wurden sowohl Probabilistische Graphische Modelle [22, 28], als auch der Goldstein-Algorithmus [32, 36] entsprechend für Entfernungsbilder adaptiert und angewandt.

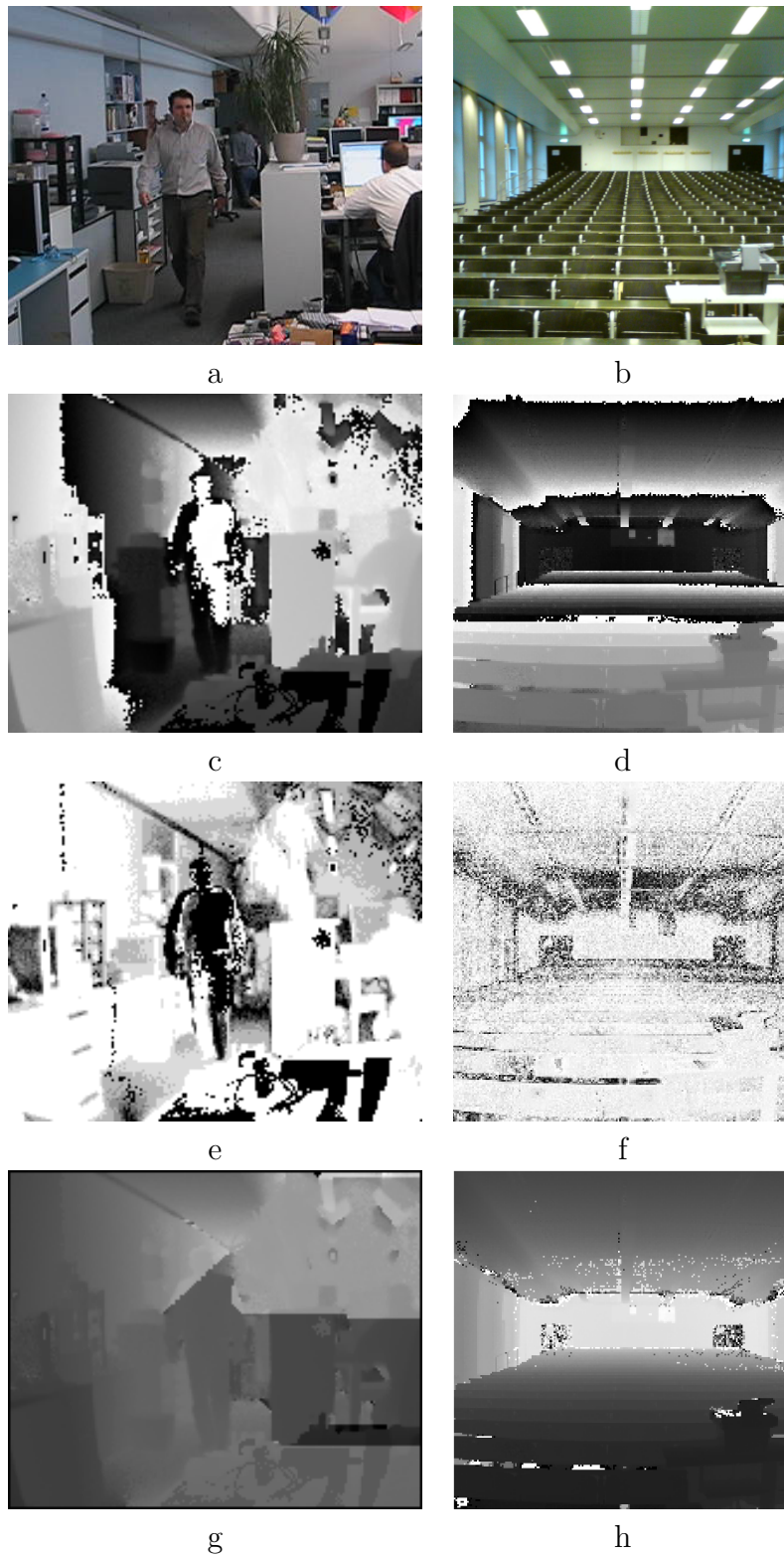


Abbildung 2.1: Zwei Szenen (a,b) erfasst mit verschiedenen Entfernungskameras, Entfernungsbilder gemessen durch MESA SR-4000 (c) und PMD CamCube 2.0 (d), Darstellung von vertrauenswürdigen (helle Grautöne) und nicht vertrauenswürdigen (dunkle Grautöne) Messungen (e,f), Ergebnisse der bildbasierten (g) und signalbasierten (h) Erweiterung des Entfernungsmessbereichs.

Essentiell für die Auflösung von Phasenmehrdeutigkeiten ist die Bestimmung relevanter Inkonsistenzen im Phasenbild. Eine Inkonsistenz, auch Residuum genannt, ist gegeben, falls die Summe der Phasendifferenzen der vier nächstgelegenen und benachbarten Pixel ungleich Null ist. Verursacht werden diese Inkonsistenzen durch (i) szenenbedingte Diskontinuitäten, (ii) den eingeschränkten Entfernungsmessbereich und (iii) Rauschen. Die Residuen sind in Abbildung 2.2a aufgetragen. Zur besseren Unterscheidung wurden die positive Residuen (weiß) und negative Residuen (schwarz) entsprechend eingefärbt.

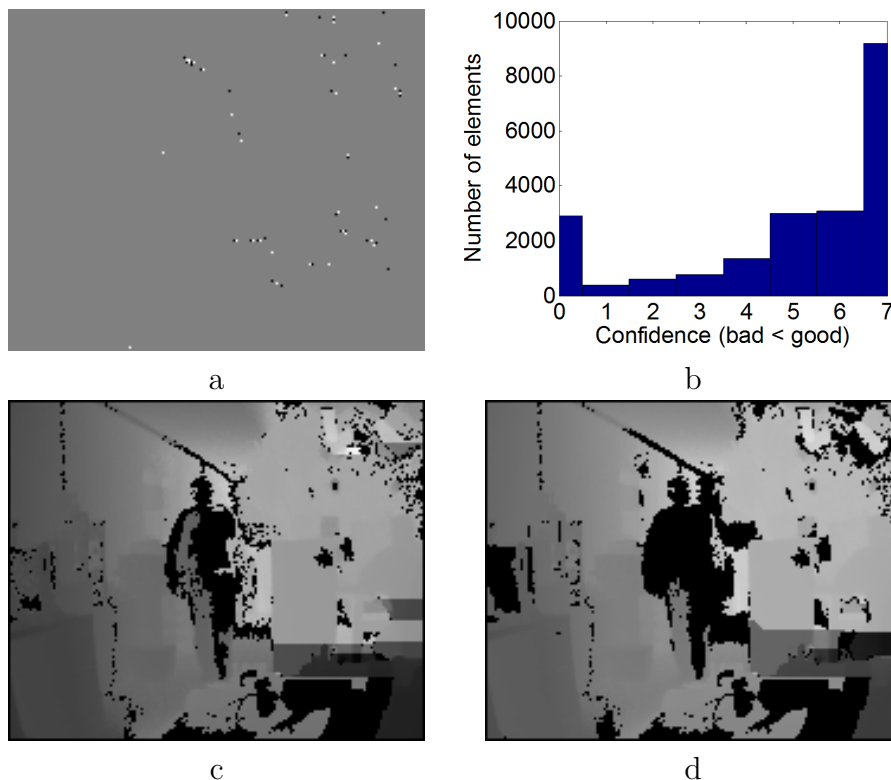


Abbildung 2.2: Berücksichtigung des Vertrauensmaßes bei der Phasenrekonstruktion: a) positive Residuen (weiß) und negative Residuen (schwarz), b) Häufigkeit der Vertrauensmaße der Szene aus Abbildung 2.1a, c) Phasenrekonstruktion (Wert 2 bis 7) und d) Phasenrekonstruktion (Wert 4 bis 7).

Wird ein Vertrauensmaß berücksichtigt bei der Prozessierung, beispielsweise die Signalgüte dargestellt in Abbildung 2.1e, können rauschbedingte und somit nicht vertrauenswürdige Messungen direkt bestimmt werden. Werden alle Messungen bei

der Phasenrekonstruktion berücksichtigt, ergibt sich daraus Abbildung 2.1g. Offensichtlich ist die Phasenrekonstruktion nicht optimal, es sind mehrere Inkonsistenzen gegeben. Im Vergleich dazu kann eine Verbesserung erreicht werden, indem nicht vertrauenswürdige Messungen für die Phasenrekonstruktion unberücksichtigt bleiben und bei der Phasenrekonstruktion ausmaskiert werden [A5]. Werden beispielsweise vertrauenswürdiger Messungen für die Phasenrekonstruktion eingesetzt, in Abbildung 2.2c (Wert 2 bis 7) bzw. in Abbildung 2.2d (Wert 4 bis 7), ergeben sich dadurch weniger Inkonsistenzen, die Rekonstruktion ist durch die Ausmaskierung stellenweise lückenhaft, jedoch qualitativ stimmiger. Abschließend ist zu erwähnen, dass szenenbedingte Diskontinuitäten mit dem vorgestellten Ansatz zur bildbasierten Erweiterung des Entfernungsmessbereichs nicht kompensiert werden können (Abbildung 2.1g).

Signalbasierte Erweiterung des Entfernungsmessbereichs

Unterschiedliche Modulationsfrequenzen werden beispielsweise bei dauerstrichmodulierten TLS Systemen [29, 84] eingesetzt, um mittels niedriger Modulationsfrequenz einen möglichst großen Entfernungsmessbereich abzudecken und mittels hoher Modulationsfrequenz eine möglichst genaue Entfernungsmessung zu ermöglichen. Um schnell, präzise und in großen Entfernungen zu messen, werden die unterschiedlichen Modulationsfrequenzen synchron auf das Signal aufgetragen. Dies ist derzeit bei Entfernungskameras nicht möglich. Moderne Entfernungskameras können mit bis zu vier unterschiedlichen Modulationsfrequenzen im MHz-Bereich messen, jedoch sind die unterschiedlichen Modulationsfrequenzen pro Gerät nicht gleichzeitig einsetzbar. Um eine signalbasierte Phasenrekonstruktion bei dauerstrichmodulierten Entfernungskameras zu ermöglichen, wird ein neues Zeit-Frequenz-Multiplex-Verfahren vorgestellt [A4] mit dem der Entfernungsmessbereich um ein Vielfaches erweitert werden kann.

Die Phase φ ist vollständig beschrieben mit

$$\varphi = 2\pi k + \Delta\varphi, \quad (2.3)$$

wenn die ganzzahlige Periodenanzahl $k = 0, 1, 2, \dots$ und die Phasendifferenz $\Delta\varphi$ bekannt sind.

Das entfernungspezifische Äquivalent hierzu ist die absolute Entfernung R mit

$$R = R_m k + \Delta R, \quad (2.4)$$

bestimmt durch die ganzzahlige Periodenanzahl k , dem eindeutigen Entfernungsmessbereich R_m und der gemessenen Entfernung ΔR .

Werden zwei unterschiedliche Modulationsfrequenzen f_1 und f_2 mit $f_1 < f_2$ bei der Messung verwendet, ergeben sich in Abhängigkeit von der verwendeten Modulationsfrequenz zwei spezifische eindeutige Entfernungsmessbereiche mit R_{m1} und R_{m2} .

Somit ergibt sich Fall A und B für die gemessenen Entfernungen ΔR_1 und ΔR_2 .

Fall A

Sind die gemessenen Entfernungen $\Delta R_1 \leq \Delta R_2$, dann gilt $k_1 = k_2 = k$ und die absolute Entfernung R ist bestimmt durch

$$R = R_{m1}k + \Delta R_1 = R_{m2}k + \Delta R_2, \quad (2.5)$$

mit der ganzzahlige Periodenanzahl $k = (\Delta R_2 - \Delta R_1)/(R_{m1} - R_{m2})$.

Fall B

Sind die gemessenen Entfernungen $\Delta R_1 > \Delta R_2$, dann gilt $k_1 + 1 = k_2 = k$ und die absolute Entfernung R ist bestimmt durch

$$R = R_{m1}(k - 1) + \Delta R_1 = R_{m2}k + \Delta R_2, \quad (2.6)$$

mit der ganzzahlige Periodenanzahl $k = (\Delta R_2 - \Delta R_1 + R_{m1})/(R_{m1} - R_{m2})$.

Exemplarisch sind in Abbildung 2.3 für zwei Modulationsfrequenzen $f_1 = 18$ MHz und $f_2 = 21$ MHz die gemessenen Entfernungen ΔR_1 und ΔR_2 über die absolute Entfernung R aufgetragen. Die berechnete Differenz $\Delta R_2 - \Delta R_1$ zeigen deutlich die vorzeichenbedingte Fallunterscheidung.

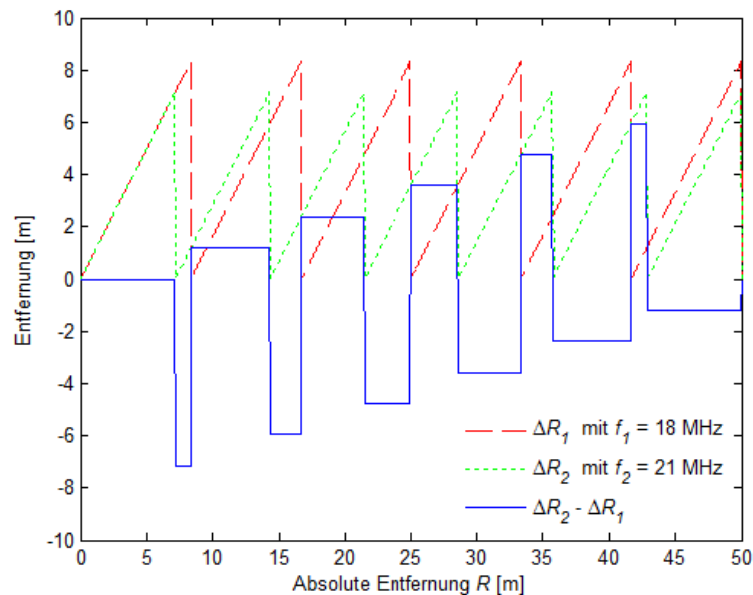


Abbildung 2.3: Gemessenen Entfernungen ΔR_1 (rot gestrichelt) und ΔR_2 (grün gepunktet), die berechnete Differenz $\Delta R_2 - \Delta R_1$ (blaue durchgehend) zur Fallunterscheidung über die absolute Entfernung R aufgetragen.

Durch Messungenauigkeiten bei der Erfassung der gemessenen Entfernungen ΔR_1 und ΔR_2 ergeben sich Abweichungen zum ganzzahligen Erwartungswert der Periodenanzahl k . Entsprechend kann aus der berechneten Abweichung ein Vertrauensmaß q abgeleitet werden, bei dem ganzzahlige Zahlenwerte als vertrauenswürdig und davon abweichende Werte als nicht vertrauenswürdig im Intervall $[0, 1]$ abgebildet werden (Abbildung 2.1f) durch

$$q = 1 - 2 |k - \text{nint}(k)|, \quad (2.7)$$

mit $|\cdot|$ für den Absolutwert und $\text{nint}(\cdot)$ für den nächsten ganzzahligen Wert.

Somit ist es möglich, bei der signalbasierten Erweiterung des Entfernungsmessbereichs jede einzelne Phasenrekonstruktion (Abbildung 2.1h) hinsichtlich der Güte zu quantifizieren. Es werden in Abbildung 2.1f vertrauenswürdige Messungen durch helle Grautöne repräsentiert und nicht vertrauenswürdigen Messungen durch dunkle Grautöne. Deutlich ist zu erkennen, dass bei diesem Beispiel tendenziell viele vertrauenswürdige Phasenrekonstruktionen berechnet werden. Die wenigen Phasenrekonstruktionen die nicht sehr vertrauenswürdig erscheinen sind in großer Entfernung (Hintergrund der Szene), bei spiegelnde Oberflächen (Tischoberfläche im Vordergrund) oder bei stark geneigte Oberflächen (Raumdecke). Eine zusätzliche quantitative Bewertung ist in Abbildung 2.4 dargestellt. Hierfür wurde die Häufigkeit des Vertrauensmaßes von der Szene aus Abbildung 2.1b aufgetragen. Es ist zu erkennen, dass vorwiegend vertrauenswürdige Werte bei der Phasenrekonstruktion bestimmt wurden.

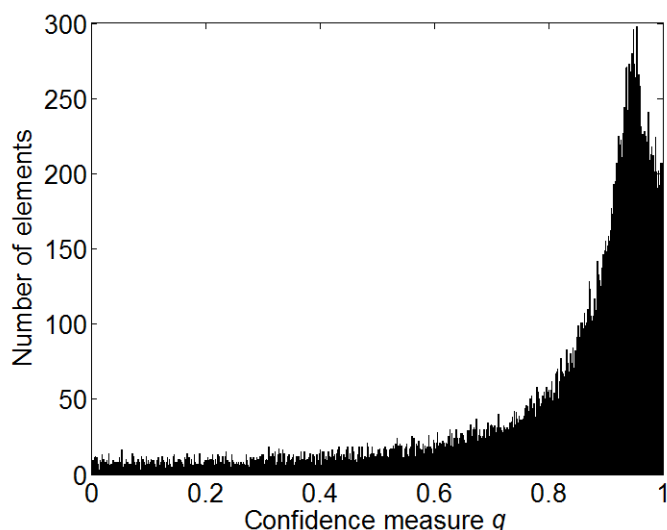


Abbildung 2.4: Häufigkeit der Vertrauensmaße bei der Szene aus Abbildung 2.1b.

2.1.2 Signalsoptimierung bei Full-Waveform-Laserscanner-Daten

In diesem Kapitel wird ein Verfahren vorgestellt, bei dem das empfangene Signal von Full-Waveform-Laserscanner optimiert wird. Hierbei wird durch Inversfilterung das durch die Oberfläche verursachte Signal rekonstruiert, womit beispielsweise eine verbesserte Trennbarkeit bei mehreren in unterschiedlichen Entfernungen beleuchteten Objektoberflächen erreicht werden kann.

Im Gegensatz zur alleinigen Messung charakteristischer Entfernungswerte ergeben sich bei der Aufnahme der kompletten zeitlichen Signalform (*Full-Waveform*) bei Full-Waveform-Laserscanner (z.B. Optech Titan, Leica ALS60 oder RIEGL LMS-Q780; Kapitel 1.1) neue Möglichkeiten bei der Auswertung zur erweiterten Szenenerfassung [A1, A10, 61, 96, A15]. Insbesondere zur Interpretation der empfangenen Signalform des rückgestreuten Laserpulses ist hierbei das Verständnis für den physikalischen Hintergrund von der Laserstrahlungsausbreitung und die Interaktion mit Oberflächenstrukturen relevant [A6, 87]. In Abbildung 2.5 ist schematisch die Laserstrahlungsausbreitung, die Interaktion mit verschiedenen Oberflächenstrukturen (oben) und die daraus resultierende empfangene Signalform des rückgestreuten Laserpulses (unten) skizziert.

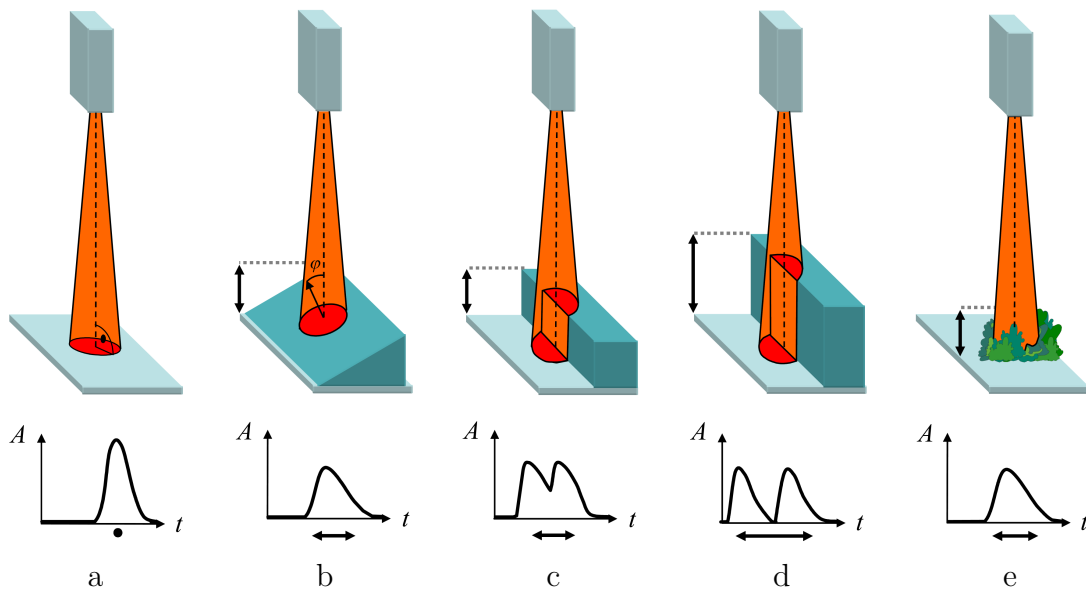


Abbildung 2.5: Laserstrahlungsausbreitung, die Interaktion mit verschiedenen Oberflächenstrukturen (oben) und die daraus resultierende empfangene Signalform des rückgestreuten Laserpulses (unten): a) Ebene Oberfläche ohne Neigung, b) Ebene Oberfläche mit Neigung, c) Zwei Bereiche mit leicht unterschiedlicher Höhe, d) Zwei Bereiche mit deutlich unterschiedlicher Höhe und e) Zufällig angeordnete kleine Objekte (Vegetation).

Um Oberflächenstrukturen unterscheiden zu können, ist eine Einteilung in unterschiedliche Rauigkeitsstufen sinnvoll. Hierbei kann die Größe der angeleuchteten Oberflächengeometrie, bezogen auf den Durchmesser d und die Wellenlänge λ des Laserstrahls, als Unterscheidungsmaß eingesetzt werden [A3], um die unterschiedlichen Oberflächenstrukturen in Makro-, Meso- und Mikro-Strukturen einzuteilen (Abbildung 2.6).

Makro-Strukturen

Mit Makro-Strukturen werden Strukturen bezeichnet, deren Ausmaße die Größe des Strahldurchmessers d (*Footprint*) um ein Vielfaches übersteigen. Bei Lasersystemen sind die Strahldivergenz und die räumliche Abtastung aufeinander abgestimmt. Im Allgemeinen ist hierbei der Strahldurchmesser kleiner oder idealerweise gleich dem Abstand zwischen den Stützstellen. Die Entfernungswerte, die durch die räumliche Abtastung mit einem Lasersystem gemessen werden, ermöglichen die Rekonstruktion von großen Objektstrukturen (beispielsweise Geländetopographie, Dachformen von Gebäuden).

Bei der Analyse von Makro-Strukturen, die bei ausgedehnten Objekten mit einfacher Oberflächengeometrie auftreten, wird vorausgesetzt, dass eine einzelne Reflexion auftritt. Diskontinuitäten innerhalb der beleuchteten Oberflächengeometrie werden vernachlässigt. Makro-Strukturen können durch das Merkmal Entfernung repräsentiert werden.

Meso-Strukturen

Mit Meso-Strukturen werden Strukturen bezeichnet, deren Ausmaße zwischen der Größe des Strahldurchmessers d und der Wellenlänge λ liegen. Unterschiedlich weit entfernte Objektbereiche innerhalb des Strahlengangs bewirken unterschiedliche Entfernungswerte innerhalb einer Messung. Dies kann verursacht werden durch die Beleuchtung von zwei, mehreren oder kontinuierlichen Bereichen mit geringer unterschiedlicher Entfernung (beispielsweise Schornstein, Gaube, Dachschräge, Dachgiebel, Äste, Blätter).

Bei der Analyse von Meso-Strukturen können sowohl deterministische Diskontinuitäten durch anthropogene Objekte als auch zufällige Diskontinuitäten durch natürliche Objekte für die beleuchtete Oberflächengeometrie angesetzt werden. Meso-Strukturen bewirken eine Deformation der gesendeten Signalform des rückgestreuten Laserpulses, wodurch die zeitliche Pulslänge der empfangenen Signalform vergrößert wird. Diese Variation kann als Merkmal Entfernungsvariation bezeichnet werden.

Mikro-Strukturen

Mit Mikro-Strukturen werden Strukturen bezeichnet, deren Ausmaße kleiner beziehungsweise in der Größenordnung der Wellenlänge λ sind. In Abhängigkeit von der Wellenlänge wird das Reflexionsvermögen der Oberfläche durch das Oberflächenmaterial beeinflusst. Das Reflexionsvermögen der Oberfläche kann als Reflexionsstärke durch ein Lasersystem bestimmt werden. Die Reflexionsstärke wird aus der Amplitude oder der Energie der empfangenen Signalform gemessen. Somit kann durch die Amplitude (Energie) auf das Material der Oberfläche geschlossen werden.

Bei der Analyse von Mikro-Strukturen ist das Reflexionsvermögen der beleuchteten Oberflächen relevant, um beispielsweise verschiedene Materialien zu unterscheiden. Mikro-Strukturen verändern die Amplitude der empfangenen Signalform des rückgestreuten Laserpulses, der Verlauf wird hierbei nicht verändert. Die Amplitude kann als Maß für das Merkmal Reflexionsstärke gesehen werden.

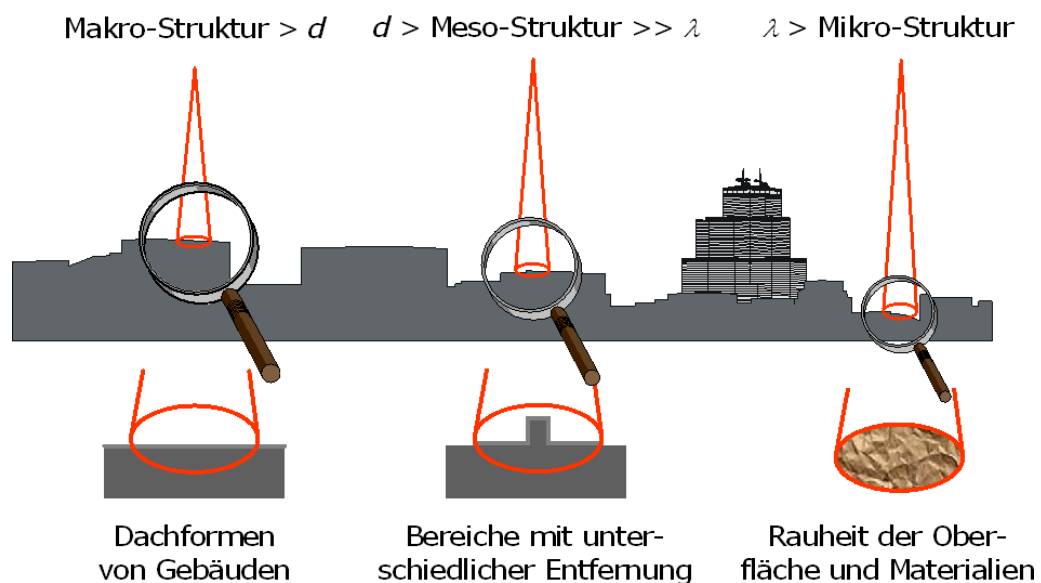


Abbildung 2.6: Unterschiedliche Detaillierungsstufen die mit einem Lasersystem erfasst werden können. In Abhängigkeit von Strahldurchmessers d (*Footprint*) und Wellenlänge λ .

Charakteristische Merkmale zur Beschreibung der Oberfläche sind somit Entfernung, Entfernungsvariation und Reflexionsstärke, die mit dem mittleren Zeitwert, der Pulsdauer und der Amplitude der Signalform (Gleichung 2.8) bei Full-Waveform-Laserscanner gemessen werden können. Die direkte Ableitung dieser Merkmale aus dem Signalverlauf führt jedoch aufgrund der gestörten und diskretisierten Signalform auf nicht repräsentative Parameter, weswegen eine Modellannahme erfolgt. Die

einfache Grundform des Signalverlaufs bei gepulsten Lasersystemen kann durch eine zeitverzögerte Gaußfunktion (Abbildung 2.7) modelliert werden. Hierfür wird zur Vereinfachung aus praktischen Gründen die Dauer des Pulses bei halber Amplitudenhöhe (*Full Width at Half Maximum*, FWHM) für die Gaußfunktion mit w in der folgenden Formel definiert. Für die angepasste Gaußfunktion ergibt sich somit

$$s(t) = \frac{2a}{w} \cdot \sqrt{\frac{\ln(2)}{\pi}} \cdot \exp\left(-4 \cdot \ln(2) \cdot \frac{(t - \tau)^2}{2w^2}\right), \quad (2.8)$$

wobei die zu bestimmenden Parameter der mittlerer Zeitwert τ , die Pulsdauer w und die Amplitude a sind. Die Oberflächeneigenschaften haben direkten Einfluss auf die Pulsdauer und die Amplitude der empfangenen Signalform.

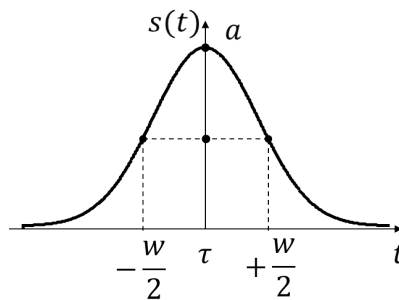


Abbildung 2.7: Signalform $s(t)$ mit mittlerem Zeitwert τ , Pulsdauer w bei *Full Width at Half Maximum* (FWHM) und Amplitude a .

Bedingt durch die Aufzeichnung der empfangenen Signalform eröffnen sich neue Möglichkeiten zur Merkmalsbestimmung (Kapitel 2.3.2). Die Merkmalsbestimmung kann zudem prinzipiell verbessert werden, indem sowohl die gesendete Signalform als auch die empfangene Signalform der Laserpulse für die Analyse berücksichtigt wird. Für die Analyse eignen sich Signalverarbeitungsmethoden [A14], beispielsweise Korrelation [A9, A11, A13] und Inversfilterung [A12, 99]. Es wurde mit der Wiener-Filter-Methode ein neues Verfahren zur besseren Unterscheidbarkeit von mehreren Oberflächen mit geringer unterschiedlicher Entfernung (Abbildung 2.5c) entwickelt [A12].

Im Folgenden wird kompakt die dafür entwickelte Modellierung des Messsignals vorgestellt. Die Oberflächenantwort $h(t)$ repräsentiert hierbei die Oberfläche als 1D-Signal aufgetragen über der Zeit, respektive die Entfernung. Durch Faltung der gesendeten Signalform $s(t)$ mit der Oberflächenantwort $h(t)$ wird die empfangene Signalform $r(t)$ modelliert mit

$$r(t) = s(t) * h(t). \quad (2.9)$$

Das Spektrum der Oberflächenantwort $\underline{H}(f)$ wird bestimmt durch Fouriertransformation \mathcal{F} mit

$$\underline{H}(f) = \frac{\underline{R}(f)}{\underline{S}(f)}. \quad (2.10)$$

Zur Berechnung von $\underline{H}(f)$ müssen die Spektren $\underline{R}(f)$ und $\underline{S}(f)$ bekannt sein, also $r(t)$ und $s(t)$ durch Messung bestimmt werden. Die Messung dieser Signalformen ist immer durch das Rauschen $n(t)$ des Empfängers gestört, welches als additives Rauschen modelliert wird mit

$$s_n(t) = s(t) + n(t) \xrightarrow{\mathcal{F}} \underline{S}_n(f) \quad (2.11)$$

und

$$r_n(t) = r(t) + n(t) \xrightarrow{\mathcal{F}} \underline{R}_n(f). \quad (2.12)$$

Bei starkem Empfängerrauschen kann es unter Berücksichtigung von Gleichung 2.10 zu großen numerischen Fehlern bei der Berechnung von $\underline{H}(f)$ kommen. Dieses Inversproblem kann durch das Wiener Filter [102] gelöst werden, da dieses das Rauschen reduziert, möglichst ohne das Spektrum der Oberflächenantwort nachteilig zu verändern.

Zur Schätzung des Spektrums der Oberflächenantwort $\hat{\underline{H}}(f)$ wird das Wiener Filter $\underline{W}(f)$ berücksichtigt mit

$$\hat{\underline{H}}(f) = \frac{\underline{R}_n(f)}{\underline{S}_n(f)} \cdot \underline{W}(f). \quad (2.13)$$

Eingeführte Randbedingungen [A12] ermöglichen die für die Wiener-Filter-Methode benötigten unbekannt Funktionen durch Messungen direkt zu bestimmen bzw. indirekt abzuschätzen. $\underline{N}(f)$ kann sehr einfach mit dem gemessenen Hintergrundrauschen des Signals geschätzt werden und man erhält das geschätzte Rauschen $\hat{\underline{N}}(f)$. Zudem ist *a priori* das gesendete Signal $s(t)$ nicht bekannt, es kann jedoch das gestörte Signal $s_n(t)$ respektive $\underline{S}_n(f)$ gemessen werden (Gleichung 2.11). Anstelle des im Divisor der Gleichung 2.13 verbleibende Spektrums $\underline{S}(f)$ könnte das durch die Messung gestörte Spektrum $\underline{S}_n(f)$ verwendet werden, was jedoch wegen des Rauschanteils zu Problemen führen kann. Daher wird hierfür das durch Tiefpassfilterung ausreichend geschätzte Spektrum $\hat{\underline{S}}(f)$ eingesetzt und es ergibt sich

$$\hat{\underline{H}}(f) = \underline{R}_n(f) \cdot \frac{\{\hat{\underline{S}}(f)\}^*}{|\hat{\underline{S}}(f)|^2 + |\hat{\underline{N}}(f)|^2} \xrightarrow{\mathcal{F}^{-1}} \hat{h}(t), \quad (2.14)$$

wobei $(\{\})^*$ dem konjugiert komplexen Spektrum entspricht. Zur Berechnung der geschätzten Oberflächenantwort $\hat{h}(t)$ muss abschließend nur noch $\hat{H}(f)$ durch inverse Fouriertransformation \mathcal{F}^{-1} in den Zeitbereich zurück transformiert werden.

Durch die Wiener-Filter-Methode kann eine zeitabhängige Oberflächenrepräsentation bestimmt werden, die eine von der gesendeten Signalform unabhängige Beschreibung der Oberfläche ermöglicht. Es konnte experimentell gezeigt werden, dass für eine typische Pulsdauer w von 5 ns durch das neu entwickelte Verfahren eine Unterscheidbarkeit bei zwei Oberflächen mit einem Entfernungsunterschied von 0,15 m (dies ist äquivalent zu einem Zehntel der Pulsdauer) möglich ist (Abbildung 2.8). Zudem kann mit dieser Methode eine hohe Genauigkeit für die Bestimmung der Entfernung erreicht werden.

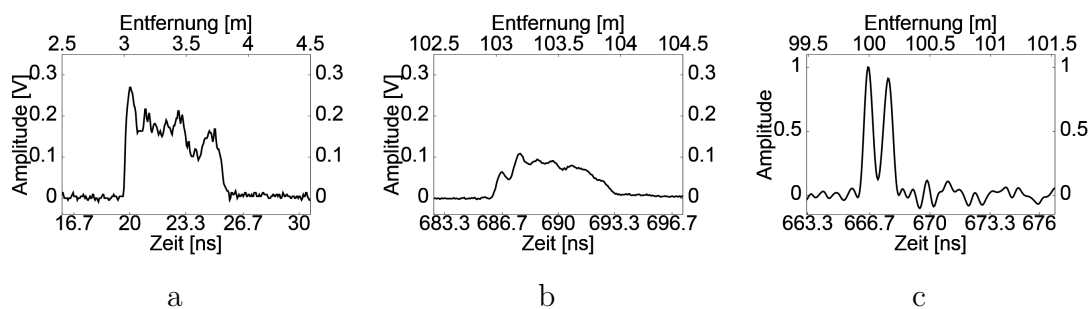


Abbildung 2.8: Gemessene Signalformen mit Rauschen und geschätzte zeitabhängige Oberflächenrepräsentation: a) Gesendete Signalform $s_n(t)$ des emittierten Laserpulses, b) Empfangene Signalform $r_n(t)$ des rückgestreuten Laserpulses, c) Geschätzte Oberflächenantwort $\hat{h}(t)$ bei zwei Oberflächen mit einem Entfernungsunterschied von 0,15 m.

2.1.3 Radiometrische Kalibrierung bei Laserscanner-Daten

Aktive optische Sensoren werden primär zur Erfassung von geometrischen Daten eingesetzt. Darüber hinaus gewinnt zunehmend die Erfassung von radiometrischen Daten an Bedeutung (alle TLS, MLS, ALS; Kapitel 1.1), da diese für verschiedene Anwendungen, wie beispielsweise Registrierung, Segmentierung, Klassifikation und Visualisierung (Kapitel 2.2, Kapitel 2.3 und Kapitel 2.4), zusätzliche Auswertemöglichkeiten eröffnen. Die gemessene Radiometrie unterliegt verschiedenen Einflüssen, beispielsweise Entfernung, Atmosphäre, Inzidenzwinkel und Oberflächeneigenschaften, die bei den zuvor genannten Anwendungen eine direkte Verwendung der gemessenen Daten erschweren bzw. unmöglich machen. Beispiele für die starke Variation der Intensitätswerte sind in Abbildung 2.9 dargestellt. Durch unterschiedliche Beobachtungsrichtungen, die sich bei der Befliegung (Abbildung 2.10) ergeben, beispiels-

weise ist bei Flugpfad 3 und Flugpfad 4 die Beobachtungsrichtung entgegengesetzt, variiert die gemessene Intensität bei den Dachflächen in Abbildung 2.9b und c.

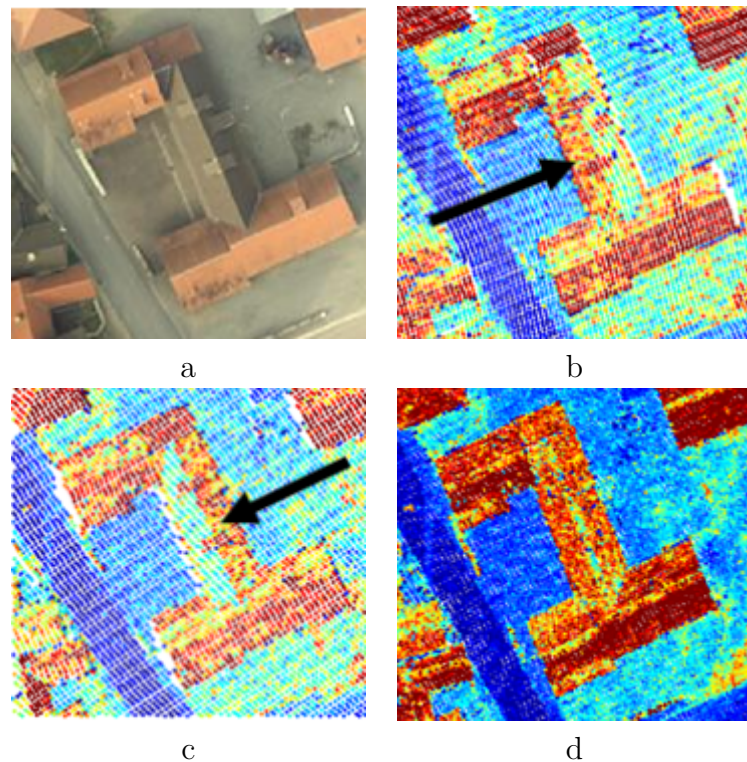


Abbildung 2.9: Radiometrische Kalibrierung bei *Airborne Laser Scanning* (ALS): a) Szene als RGB Bild, b) gemessene Intensität bei Flugpfad 3 für gegebene Beobachtungsrichtung, c) gemessene Intensität bei Flugpfad 4 für entgegengesetzte Beobachtungsrichtung, d) Ergebnis für winkelunabhängige Intensität. Hohe Intensitätswerte sind rot dargestellt.

Um radiometrische Daten sinnvoll nutzen zu können, muss ein eindeutiger Zusammenhang zwischen den beleuchteten Oberflächenstrukturen (Kapitel 2.1.2) und der gemessenen Amplitude (Intensität) des Signals der rückgestreuten Strahlung hergestellt werden. Ist dieser Zusammenhang hergestellt, kann idealerweise mittels einer radiometrischen Kalibrierung das Signal an die Oberflächeneigenschaften angepasst werden, wodurch die Oberflächeneigenschaften über verschiedene Aufnahmepositionen, Beleuchtungswinkel und Beobachtungszeiten hinweg vergleichbar werden. Um den Zusammenhang zu bestimmen, werden meist datengetriebene [42, 45, 46, 52, 74] jedoch auch modellgetriebene [42] Ansätze verfolgt.

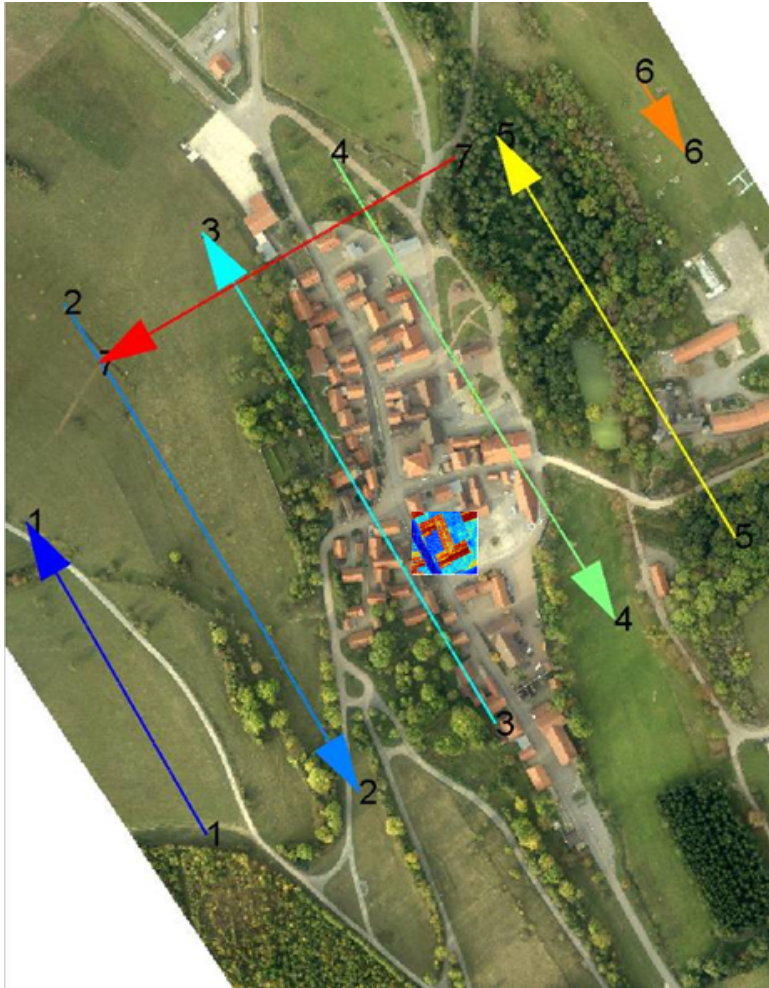


Abbildung 2.10: *Airborne Laser Scanning* (ALS) Befliegung: Flugpfade 1-7 und eingeblendete Laserscanner-Daten aus Abbildung 2.9.

Zur radiometrischen Kalibrierung müssen verschiedene äußere Einflüsse bei der Modellierung berücksichtigt werden. Die empfangene Intensität I_R eines monostatischen Laserscanners kann bei einer gaußförmigen Signalform des rückgestreuten Laserpulses mit der Amplitude a und der Pulsdauer w (Abbildung 2.7) näherungsweise bestimmt werden mit

$$I_R \approx a \cdot w \cdot R^{-2} \cdot e^{-2\alpha R} \cdot \rho_m \cdot \cos(\vartheta), \quad (2.15)$$

unter Berücksichtigung der Entfernung R , der atmosphärischen Transmission $e^{-2\alpha R}$ für Hin- und Rückweg, der Reflektanz des Materials ρ_m , sowie des Inzidenzwinkels ϑ zwischen Strahlausbreitungsrichtung und Oberflächennormale nach dem Lambert'schen Kosinusgesetz. Wird der Inzidenzwinkel ϑ über die lokale Nachbarschaft bestimmt und sind die winkelunabhängigen Komponenten aus der Messung bekannt, dann kann nach Gleichung 2.15 die winkelunabhängige Intensität berechnet werden.

Nach der Signalanpassung ist das Signal idealerweise nur noch von der Oberflächeneigenschaft geprägt. Bei homogenen Oberflächen, beispielsweise Dachflächen, ist nach der Signalanpassung die winkelunabhängige Intensität konstant. Zur Modellierung kann neben dem hier vorgestellten Lambert'schen Beleuchtungsmodell das darauf aufbauende Phong'sche Beleuchtungsmodell [75] herangezogen werden, um nicht nur diffuse, sondern auch spiegelnde Eigenschaften der Oberfläche [A2, A7, A8] zu berücksichtigen.

Bereits durch die Verwendung des Lambert'schen Beleuchtungsmodells kann durch die radiometrische Kalibrierung eine vorteilhafte Signalanpassung erfolgen, wodurch sowohl die Signalvariation bei spezifischen Oberflächen signifikant eingeschränkt werden kann, als auch eine verbesserte Winkelinvarianz bei den Signalen erreicht wird [A7]. Die durch radiometrische Kalibrierung erzielte Verbesserung für eine homogene Oberfläche ist in Abbildung 2.9d qualitativ und in Abbildung 2.11 quantitativ dargestellt.

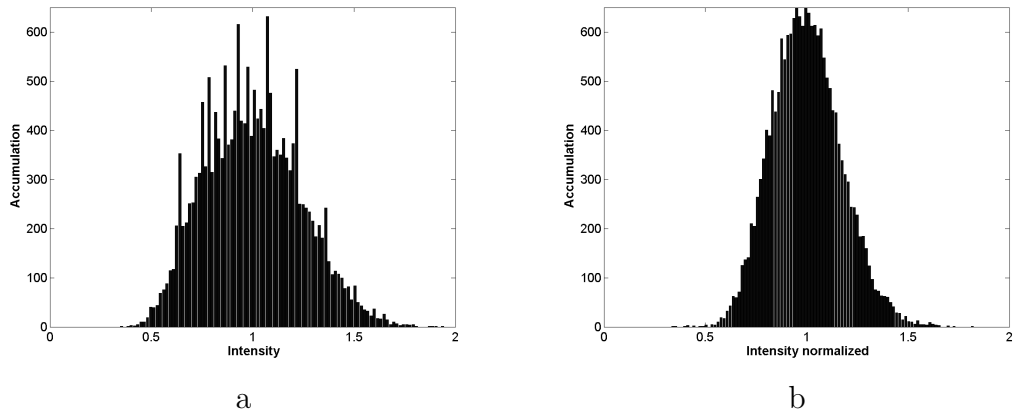


Abbildung 2.11: Homogenen Oberfläche und die dazu korrespondierenden Intensitätswerten: a) starke Variation der gemessenen Intensitätswerten, b) geringe Variation der winkelunabhängigen Intensitätswerten nach radiometrischer Kalibrierung.

2.2 Koregistrierung mit 2D- bzw. 3D-Sensoren

Nach der Aufbereitung der Daten in der vorherigen Verarbeitungsstufe werden mit der zweiten Verarbeitungsstufe unterschiedliche Beobachtungen vereinheitlicht. Bei dem Einsatz von mehreren Sensoren bzw. von einem bewegten Sensor für eine umfassende Szenenerfassung ist, bedingt durch die unterschiedlichen Beobachtungspositionen und -orientierungen, die Bestimmung der relativen Orientierung zwischen

den Beobachtungen essentiell. Ist die relative Orientierung bestimmt, können diese Beobachtungen in ein gemeinsames, übergeordnetes, räumliches, kartesisches Koordinatensystem, auch Weltkoordinatensystem genannt, überführt werden. Die Koregistrierung als Lösungsansatz ermöglicht eine Vereinheitlichung der Beobachtungen und somit wird nachfolgend eine zweckmäßige Analyse der erfassten Daten ermöglicht. Im Folgenden werden neue Lösungsansätze zur Bestimmung der relativen Orientierung zwischen aktiven 3D-Sensoren und passiven 2D-Sensoren (Kapitel 2.2.1) sowie bei bewegten aktiven 3D-Sensoren (Kapitel 2.2.2) vorgestellt.

2.2.1 3D/2D-Koregistrierung

Bei der Kombination von aktiven 3D-Sensoren und passiven 2D-Sensoren ergeben sich unterschiedliche methodische Herangehensweisen für die 3D/2D Koregistrierung. Anhand der gewählten Beispiele, Entfernungskamera kombiniert mit thermischer Infrarot-Kamera (Kapitel 2.2.1.1) sowie mobilen Laserscanner kombiniert mit RGB-Kamera (Kapitel 2.2.1.2), wird die methodische Herangehensweisen in den folgenden Kapitel verdeutlicht.

2.2.1.1 Entfernungskamera kombiniert mit thermischer Infrarot-Kamera

Zur 3D-Szenenerfassung bei dynamischen Szenen ist eine unverzügliche Aufnahme vorteilhaft und wird durch aktive Entfernungskameras ermöglicht. Neben dieser geometrischen Beobachtung kann für bestimmte Anwendungen (z.B. Personendetektion [B3, 64, B13], Gebäudeinspektion [41, 43]) eine zusätzliche radiometrische Beobachtung [55] vorteilhaft sein. Hierfür ist eine Erweiterung der Erfassung mittels Sensoren wie RGB-Kameras oder thermischen Infrarot-Kameras zweckmäßig [2, 13, 93]. Bedingt durch eine separate Sensoranordnung (Abbildung 2.12) ergeben sich für die beiden bildgebenden Sensoren leicht unterschiedliche perspektivische Beobachtungen, die nur dann sinnvoll fusioniert werden können, wenn eine eindeutige Zuordnung zwischen den unterschiedlichen Beobachtungen hergestellt wird. Dafür muss die relative Orientierung der verwendeten Sensoren bestimmt werden.



Abbildung 2.12: Entfernungskamera PMD CamCube 2.0 (links) und thermische Infrarot-Kamera InfraTec VarioCAM hr (rechts).

Um die relative Orientierung einer Entfernungskamera (Kapitel 1.1, z.B. PMD Cam-Cube 2.0) zu einer thermischen Infrarot-Kamera (z.B. InfraTec VarioCAM hr) zu berechnen, müssen zunächst korrespondierende Bildinhalten einander zugeordnet werden. Hierfür werden zuerst die Inneren Orientierungen der Kameras durch eine geometrische Kamerakalibrierung [7, 11, 40] bestimmt. Zur Kalibrierung der Entfernungskamera ist ein planarer Kalibrierkörper mit Schachbrettmuster geeignet. Das Schachbrettmuster ist jedoch im thermischen Infrarot-Bereich kontrastarm und somit für die Kalibrierung der thermischen Infrarot-Kamera unzureichend. Daher wird zur Kalibrierung der thermischen Infrarot-Kamera ein planarer Kalibrierkörper mit regelmäßig verteilt angeordneten strahlenden Glühbirnen (Abbildung 2.13) [54, 60, B5, B6] verwendet, da diese im Intensitätsbild gut zu erkennen sind.

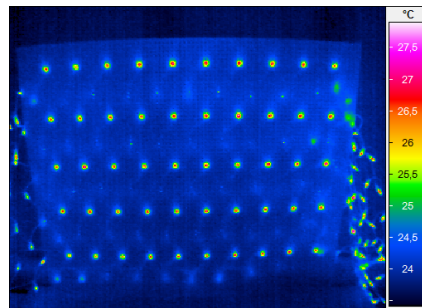


Abbildung 2.13: Planarer Kalibrierkörper mit regelmäßig verteilt angeordneten strahlenden Glühbirnen.

Aufgrund der verschiedenen Spektralbereiche der Kameras und der daraus resultierenden unterschiedlichen Bildeigenschaften (z.B. unscharfe Kanten und unterschiedliche Intensitäten) können zur Ermittlung von korrespondierenden Punkten keine Standardverfahren wie beispielsweise die *Scale Invariant Feature Transform* (SIFT) [59] verwendet werden. Daher wird ein anderer Lösungsweg angegangen, bei dem zuerst mittels Förstner-Operators [27] für das kalibrierte Intensitätsbild der Entfernungskamera als auch für das kalibrierte Intensitätsbild der thermischen Infrarot-Kamera markante Bildbereiche detektiert werden. Basierend auf den markanten Bildbereiche im Intensitätsbild der Entfernungskamera wird unter Berücksichtigung der entsprechenden Entfernungswerte und der gemessenen Basis die zu erwartenden Koordinaten im Intensitätsbild der thermischen Infrarot-Kamera geschätzt, um den Suchraum bei der Korrespondenzsuche einzuschränken. Ausgehend von diesen Koordinaten wird im Intensitätsbild der thermischen Infrarot-Kamera der entsprechende markante Bildbereich mit dem minimalen Abstand ausgewählt und

eine Zuordnung hergestellt [B4]. Eine zusätzliche Verbesserung wird durch die Verwendung von formbasierten Zuordnungsverfahren (*Shape-Based-Matching*) [86, 90] erreicht, da bei diesem Ansatz wellenlängenunabhängige Formeigenschaften für die Zuordnung berücksichtigt werden [B13].

Unter Verwendung der ermittelten korrespondierenden 2D-Punkte und deren 3D-Koordinaten sind verschiedene Lösungsansätze zur Bestimmung der relative Orientierung zwischen Entfernungskamera und thermischen Infrarot-Kamera möglich:

- Unter der Annahme, dass die 2D-Punkte auf einer Ebene (z.B. bei planaren Gebäudefassaden) liegen, kann durch eine Homographie [38] die relative Orientierung geschätzt werden [B10].
- Ist eine gute Anfangsnäherung für die relative Orientierung bekannt, kann über eine Ausgleichung die relative Orientierung durch den räumlichen Rückwärtsschnitt geschätzt werden [B4].
- Unter Berücksichtigung der 3D/2D-Korrespondenzen kann durch die Kombination von *Random-Sample-Consensus* (RANSAC) [25] mit dem nicht-iterative *Efficient Perspective-n-Point* (EPnP) Verfahren [66] die relative Orientierung ohne Anfangsnäherung geschätzt werden [B13].

Die geschätzte relative Orientierung aus Intensitätsbildern einer Entfernungskamera und den Intensitätsbildern einer thermischen Infrarot-Kamera kann beispielsweise zur Erzeugung von infrarot-texturierten Punktwolken (Abbildung 2.14) verwendet werden.

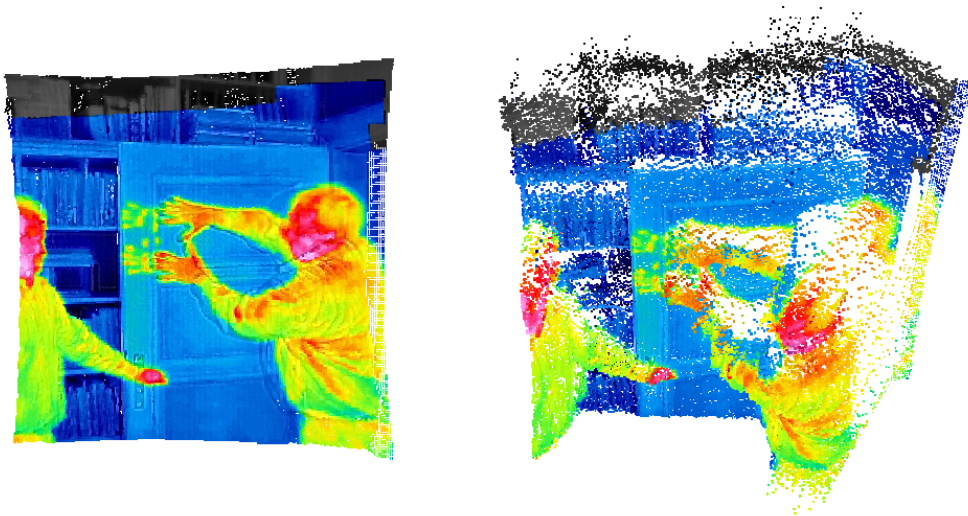


Abbildung 2.14: Zentralperspektivische Ansicht (links) und schräge Ansicht (rechts) der selben infrarot-texturierten Punktwolke zur Verdeutlichung der Qualität der 3D/2D-Korregistrierung.

2.2.1.2 Mobiler Laserscanner kombiniert mit RGB-Kamera

Bei *Unmanned Aerial Systems* (UASs) ist die Verwendung von Kameras zur Szenenerfassung weit verbreitet. Als zusätzliche Beobachtungsmöglichkeit ist der Einsatz eines aktiven Sensors besonders bei der Vegetationsanalyse attraktiv. Die Verwendung eines mobilen Laserscanners (Kapitel 1.1, z.B. Hokuyo UTM-30LX-EW) für *Mobile Laser Scanning* (MLS) kombiniert mit RGB-Kamera (z.B. GoPro 1080p HD HERO2) auf einer gemeinsamen Plattform starr fixiert (Abbildung 2.15) bietet neue Möglichkeiten bei der 3D-Szenenerfassung durch geometrische und radiometrische Datenfusion. Beispielsweise wird damit die Abschätzung der Skalierung bei *Simultaneous Localisation and Mapping* (SLAM) [23] ermöglicht. Die Herausforderung bei dieser Sensorkombination ist, dass die unterschiedlichen Sensoren im nahen Infrarot bzw. sichtbaren Spektralbereichen messen [B7] und folglich eine direkte gegenseitige Zuordnung der Beobachtungen nicht gegeben ist. Daher wird eine Methodik gesucht, mit der die relative Orientierung zwischen den Sensoren bestimmt werden kann.



Abbildung 2.15: Mobiler Laserscanner Hokuyo UTM-30LX-EW (links) und RGB-Kamera GoPro 1080p HD HERO2 (rechts).

Zur Bestimmung der relativen Orientierung zwischen den auf einer gemeinsamen Plattform starr fixierten Sensoren werden folgende Verfahrensschritte angegangen. Zuerst werden aus verschiedenen Ansichten synchrone Aufnahme von dem planaren Kalibrierobjekt (Schachbrett) durch den Laserscanner und die Kamera erstellt. Somit ergeben sich paarweise Aufnahmen mit einzelnen Laserscanner-Zeilen und dazu korrespondierenden Kamerabildern. Automatisch wird im Folgenden aus jeder Laserscanner-Zeile das geradlinige Linensegment extrahiert [49], welches den Laserpunkten auf dem planaren Kalibrierobjekt entsprechen. Nach der Methode von Zhang und Pless [104], die als Randbedingung ein planares Kalibrierobjekt voraussetzt, welches durch mehrere geradlinige Linensegmente beschrieben werden kann, wird daraus die relative Orientierung zwischen Laserscanner und Kamera bestimmt [B8]. Die geschätzte relative Orientierung und deren Unsicherheiten für Hebelarm (*Lever-arm*) und Relativausrichtung der Systemkomponenten (*Bore-sight*) [39] ist in Tabelle 2.1 dargestellt. Die geschätzten Werte entsprechen tendenziell dem gemessenen Referenzwert.

	<i>Estimated values</i>	<i>Root mean square error</i>
<i>Lever-arm</i> [mm]	-0.63	±9
	67.10	±47
	-6.45	±12
<i>Bore-sight</i> [deg]	1.08	±0.84
	1.69	±0.37
	-89.30	±0.14

Tabelle 2.1: Geschätzte relative Orientierung und deren Unsicherheiten für Hebelarm (*Lever-arm*) und Relativausrichtung der Systemkomponenten (*Bore-sight*).

Um abschließend eine Projektion der Laserpunkte auf ein verzeichnisfreies Bild vornehmen zu können, werden zusätzlich die Verzeichnungsparameter der Kamera durch eine geometrische Kamerakalibrierung [7, 11, 40] bestimmt und entsprechend berücksichtigt. Exemplarisch ist in Abbildung 2.16 die 3D/2D-Koregistrierung zwischen mobilem Laserscanner und RGB-Kamera dargestellt.



Abbildung 2.16: Visualisierung der 3D/2D-Koregistrierung zwischen mobilem Laserscanner und RGB-Kamera unter Berücksichtigung der geschätzten relativen Orientierung (rote Punkte: detektierte 2D-Punkte von auffälligen Bildbereiche; blaue Punkte, horizontal: Laserpunkte projiziert auf das dazu korrespondierende Bild; grüne Linien, vertikal: skalierte Repräsentation der Entfernungsmessungen für jeden projizierten Laserpunkt).

2.2.2 3D-Koregistrierung

Eine detaillierte Erfassung einer Szene kann durch das Messen von Objektoberflächen durch eine möglichst dichte, genaue und vollständige Punktwolke erreicht werden. Die Punktwolke besteht dann üblicherweise aus mehreren Millionen gemessenen 3D-Punkten. Optional können den 3D-Punkten noch zusätzliche Attribute wie beispielsweise Intensität oder Farbe zugeordnet werden. Um eine gute Flächenabdeckung bei der Erfassung der Szene zu erreichen, werden in der Regel mehrere 2.5D-Messungen, also Entfernungsmessungen, von unterschiedlichen Aufnahmeestandpunkten erfasst und mittels Koregistrierung in ein gemeinsames, übergeordnetes Weltkoordinatensystem überführt. Hierbei ist darauf zu achten, dass für jede der 2.5D-Messungen synchron die absolute Orientierung, also die absolute Position und Lage bekannt ist. Ist eine Erfassung nicht direkt möglich, kann indirekt über eine Verkettung von berechneten relativen Orientierungen zwischen jedem einzelnen Aufnahmestandpunkt eine Vereinheitlichung vorgenommen werden.

Des Weiteren ist eine Unterscheidung hinsichtlich statischer und dynamischer Szenen sinnvoll, da je nach Anforderung unterschiedliche Aufnahmestrategien angewandt werden können. Bei einer statischen Szene kann mit einem stationär positionierten und scannenden aktiven Sensor (z.B. Leica HDS6000, Z+F Imager 5006h, RIEGL VZ-4000; Kapitel 1.1) die Szene sukzessive erfasst werden, da keine Szenenänderungen während der Messung gegeben sind. Anders verhält es sich bei dynamischen Szenen. Erfolgt hier die Erfassung mit einem scannenden aktiven Sensor, kann dies zu ungewollten Bewegungsartefakten führen [91]. Abhilfe wird ermöglicht, indem bei einer dynamischen Szene mit einem bildgebenden aktiven 3D-Sensor (z.B. Microsoft Kinect 1.0, MESA SR-4000 oder PMD CamCube 2.0; Kapitel 1.1) synchrone 2.5D-Messungen durchgeführt werden, wodurch eine momentane Erfassung gewährleistet ist. Bei bewegten 3D-Sensoren ist es zudem sinnvoll, die absolute Position und Lage synchron mit den Beobachtungen zu erfassen, beispielsweise direkt gemessen durch zusätzliche Sensoren in Form von einem Positionsbestimmungssystem gekoppelt mit inertialem Navigationssystem. Nicht immer ist hiermit eine Position- und Lagemessung möglich, beispielsweise ist innerhalb von Gebäuden keine direkte Positionsbestimmung durch ein globales Navigationssatellitensystem (*Global Navigation Satellite System*) möglich, daher werden andere Lösungen angegangen.

Im Folgenden werden neue Methoden vorgestellt, bei denen keine zusätzliche Sensoren zur Bestimmung der Position und der Lage des aktiven 3D-Sensors verwendet werden. Hierfür werden ausschließlich die mit den aktiven 3D-Sensoren erfassten Daten berücksichtigt, um die Aufnahmeestandpunkte anhand der geschätzten relativen Orientierung zu ermitteln, d.h. unter Verwendung von Daten die mit scannenden aktiven 3D-Sensoren (Kapitel 2.2.2.1) bzw. bildgebenden aktiven 3D-Sensoren (Kapitel 2.2.2.2) erfasst werden, wird deren Koregistrierung vorgestellt. Die Koregistrierung zwischen scannenden und bildgebenden aktiven 3D-Sensoren [B15] ist ein Spezialfall der im Rahmen dieser Habilitationsschrift nicht vertieft wird.

2.2.2.1 Koregistrierung bei terrestrischen Laserscanner

Terrestrische Laserscanner (z.B. Leica HDS6000, Z+F Imager 5006h, RIEGL VZ-4000; Kapitel 1.1) ermöglichen eine dichte und sehr genaue 3D-Erfassung. Jedoch reicht in der Regel meist ein einzelner Aufnahmestandpunkt aufgrund von Verdeckungen bei den 2.5D-Messungen nicht aus, um eine Szene vollständig zu erfassen. Die 3D-Messwerte der einzeln erfassten Punktwolken beziehen sich jeweils nur auf das lokale Koordinatensystem des Laserscanners während der Messung und nicht auf ein gemeinsames, übergeordnetes Weltkoordinatensystem. Infolgedessen muss die relative Orientierung des Sensors durch die Koregistrierung der einzeln erfassten Punktwolken bestimmt werden, um eine möglichst vollständige Punktwolke zu erhalten. Die zeitaufwendige Verortung der einzelnen Aufnahmestandpunkt erfolgt bisher meist manuell bzw. halbautomatisch mittels standardisierten Referenzkörpern. Vollautomatische Verfahren zur Koregistrierung sollen daher möglichst eine geringe Berechnungsdauer und zudem eine sehr hohe Genauigkeit aufweisen.

Eine automatische Koregistrierung kann durch das *Iterative Closest Point* (ICP) Verfahren [5, 80] erfolgen. Es handelt sich hierbei um ein zweistufiges Verfahren, bei dem zuerst eine Grobausrichtung und danach eine Feinausrichtung durchgeführt wird. Insbesondere die Grobausrichtung von zwei Punktwolken gestaltet sich schwierig und setzt zudem einschränkende geometrische Nebenbedingungen voraus [1, 10, 19, 20, 21, 94]. Die Feinausrichtung basiert im einfachsten Fall auf einer Zuordnung zwischen jedem Punkt aus der einen Punktwolke zum jeweils nächsten Punkt der anderen Punktwolke. Hierzu ist zu bemerken, dass es sich bei den Punkten um durch die Messung vorgegebene 3D-Positionen handelt, die nicht notwendigerweise einer gemeinsamen Verortung auf der Oberfläche entsprechen müssen. Für die Berechnung wird durch aufwendige iterative Anpassung der Transformationsparameter der Abstand zwischen den zugeordneten Punkten sukzessive minimiert bis ein globales Optimum erreicht ist bzw. die maximale Anzahl von Iterationen durchlaufen ist. Neben der einfachen Zuordnung existieren verbesserte und entsprechend angepasste ICP-Varianten bei denen nicht der Punkt-zu-Punkt-Abstand, sondern der Punkt-zu-Ebene-Abstand optimiert wird.

Der Nachteil von Zuordnungen, die auf Messpunkten basieren, kann umgangen werden, indem eine direkte Zuordnung von identischen 3D-Oberflächenelementen, also gemeinsam verorteten Oberflächenelementen, erfolgt. Wie kann dies effizient erreicht werden? Zweckdienlich hierfür ist eine Bildrepräsentation der sukzessiv erfassten korrespondierenden Entfernungs- bzw. Intensitätsmesswerte. Basierend auf der Nutzung von geometrischer und radiometrischer Bildrepräsentation wurden verschiedene bildbasierte Ansätze zur Registrierung von Punktwolken entwickelt [3, 83]. Aus den einzelnen Intensitätsbildern können charakteristische 2D-Merkmalpunkte (z.B. mit *Scale Invariant Feature Transform* (SIFT) Merkmale [59]) ermittelt werden (Abbildung 2.17), die eindeutig korrespondieren und einer gemeinsamen Verortung entsprechen [6, 48, 98]. Über den Vergleich von 2D-Merkmalpunkten in verschiedenen Intensitätsbildern können 2D/2D- und daraus folglich auch 3D/3D-Punktkorrespondenzen bestimmt werden.

Somit sind verschiedene Lösungsansätze zur Koregistrierung möglich. Ein schnelles und robustes Verfahren, welches das *Efficient Perspective-n-Point* (EPnP) [66] Verfahren mit dem *Random-Sample-Consensus* (RANSAC) [25] Verfahren kombiniert, liefert basierend auf 3D-zu-2D-Punktkorrespondenzen bereits eine sehr gute Grobregistrierung der Punktwolken und ist daher für die Koregistrierung besonders geeignet. Sind große 3D/3D-Distanzen zwischen einzelnen Punkten bei den grob registrierten reduzierten Punktwolken gegeben, werden diese als Ausreißer deklariert und entfernt. Abschließend kann für die verbleibenden Korrespondenzen eine Verfeinerung der Schätzung (Feinregistrierung) durch einen erneuten Durchlauf der vorhergehenden Schritte oder eine ICP-basierte Feinregistrierung erfolgen [B11, B12, B14].

Die entwickelten Methoden wurden an einem frei verfügbaren Datensatz¹, bei welchem Referenzwerte für die 11 Aufnahmestandpunkte vorliegen, getestet und zeigen eine geringe Berechnungsdauer von wenigen Sekunden auf einem Standardrechner, sowie einen absoluten Positionsfehler zwischen dem geschätzten 3D-Aufnahmestandpunkt und der Referenz für die untersuchte Szene in Abbildung 2.19, der im niederen Zentimeter-Bereich für die Grobregistrierung ($e_{abs} < 5$ cm, blau gestrichelt) bzw. Feinregistrierung ($e_{abs} < 3$ cm, rot durchgehend) liegt (Abbildung 2.18). In Abbildung 2.19 sind die geschätzten 11 Aufnahmestandpunkte sowie die Punktwolke nach der Überführung in ein gemeinsames, übergeordnetes Weltkoordinatensystem dargestellt.

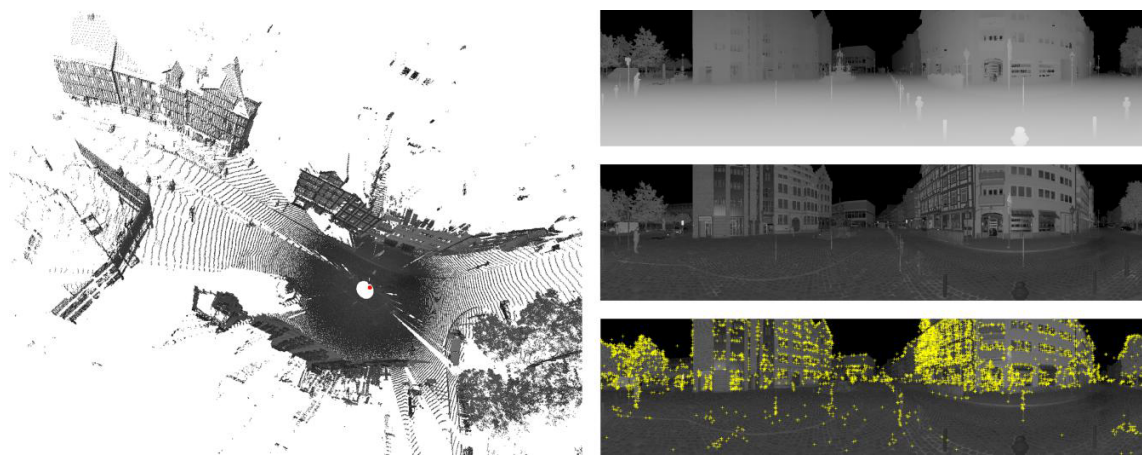


Abbildung 2.17: 3D-Repräsentation einer Punktwolke (links), entsprechende 2D-Repräsentationen für Entfernungswerte und Intensitätswerte sowie Intensitätsbild mit den lokalisierten SIFT-Merkmalen (rechts, von oben nach unten).

¹<http://www.ikg.uni-hannover.de/index.php?id=413&L=de> (Letzter Zugriff April 2015)

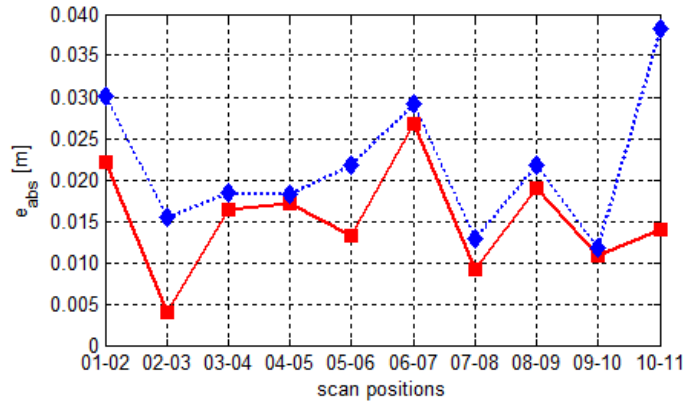


Abbildung 2.18: Absoluter Positionsfehler zwischen dem geschätzten 3D-Aufnahmestandpunkt und der Referenz für die untersuchte Szene in Abbildung 2.19: Grobregistrierung (blau gestrichelt) bzw. Feinregistrierung (rot durchgehend).

2.2.2.2 Koregistrierung bei Entfernungskameras

Entfernungskameras (z.B. Microsoft Kinect 1.0, MESA SR-4000 oder PMD Cam-Cube 2.0; Kapitel 1.1) erfassen flächenhaft und synchron Entfernungs- und Intensitätswerte. Moderne Entfernungskameras mit hohen Bildraten sind daher auch zur Erfassung von dynamischen Szenen geeignet [B2, 79]. Die erfassten Punktwolken beziehen sich auf das lokale Koordinatensystem des aktiven Sensors zum Aufnahmezeitpunkt. Durch die bekannte relative Orientierung zwischen den Aufnahmezeitpunkten der Entfernungskamera können die Punktwolken in ein gemeinsames Koordinatensystem transformiert werden. Infolgedessen muss zur Szenenrekonstruktion eine Registrierung der Punktwolken erfolgen, bei welcher die gesamte, für alle Aufnahmezeitpunkte erfassten 2.5D-Messungen in ein gemeinsames, übergeordnetes Weltkoordinatensystem transformiert wird. Diese Registrierung sollte idealerweise automatisch, schnell und mit hoher Genauigkeit erfolgen.

Um bei der Koregistrierung die Einschränkungen bezüglich des geringen Sichtfeldes der Entfernungskamera und der im Vergleich zu Laserscannern geringeren Messgenauigkeit zu kompensieren, wurde ein modifiziertes Gewichtungsschema basierend auf der Zuverlässigkeit einzelner 3D-Messungen realisiert [B9]. Durch das reguläre Scanraster lassen sich die aufgenommenen Entfernungs- und Intensitätsmesswerte als Entfernungs- und Intensitätsbilder darstellen. In den zu zwei Beobachtungen zugeordneten Intensitätsbildern können über *Scale Invariant Feature Transform* (SIFT) [59] zuverlässig korrespondierende Punkte (2D/2D-Korrespondenzen) detektiert werden. Statt der Verwendung von Intensitätsbildern ist eine erweiterte Kombination aus Entfernungs- und Intensitätsbilder möglich, dann können über *Complex Scale-Invariant Feature Transform* (CSIFT) [B1] die korrespondierende Punkte detektiert werden. Mit Hilfe der entsprechenden Entfernungswerte können daraus wiederum Paare von physikalisch nahezu identischen 3D-Punkten (3D/3D-

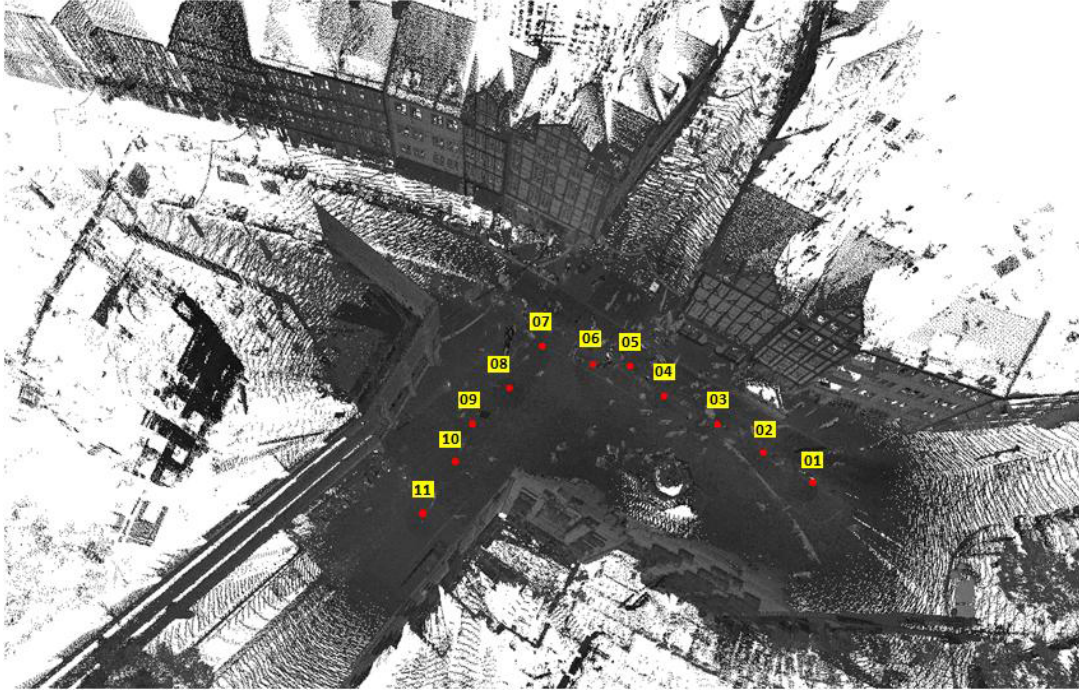


Abbildung 2.19: Punktwolke nach der Überführung in ein gemeinsames, übergeordnetes Weltkoordinatensystem.

Korrespondenzen) abgeleitet werden, aus welchen die Transformationsparameter der relativen Orientierung zwischen den entsprechenden Punktwolken geschätzt werden können (Abbildung 2.20a). Dabei wird der Beitrag jeder 3D/3D-Korrespondenz anhand der Zuverlässigkeit der entsprechenden Entfernungswerte sowie der Zuverlässigkeit der 2D/2D-Korrespondenz, welche die Ähnlichkeit lokaler Intensitätswerte beschreibt, gewichtet.

Eine Bewertung für jede einzelne Beobachtung ist bei diesen Untersuchungen nur eingeschränkt möglich, da kein Referenzmesssystem zur Erfassung der Position und Lage während der Aufnahme verfügbar war. Unter der Annahme, dass eine geradlinige Bewegung entlang einer Seilbahn gegeben ist, der Anfangs- und Endpunkt eingemessen wurde, ergibt sich eine Standardabweichung für die Position quer zur Bewegungsrichtung mit $\sigma_y = 0.0378m$ und ein absoluter Fehler für den Anfangspunkt im Schleifenschluss mit $e_{loop} = 0.0967m$. Es konnte somit gezeigt werden, dass eine automatisierte und rein bildbasierte Schätzung der Position und Lage unter Verwendung von bildgebenden aktiven Sensoren möglich ist. Zusätzliche Untersuchungen für jede einzelne Aufnahme hinsichtlich Position und Lage könnten zukünftig noch angegangen werden, um eine differenzierte Bewertung zu erstellen. Abschließend kann zusätzlich eine Filterung bezüglich der Zuverlässigkeit der gemessenen Entfernungsinformation erfolgen, um unzuverlässige 3D-Messungen aus der Punktwolke zu entfernen (Abbildung 2.20b und c).

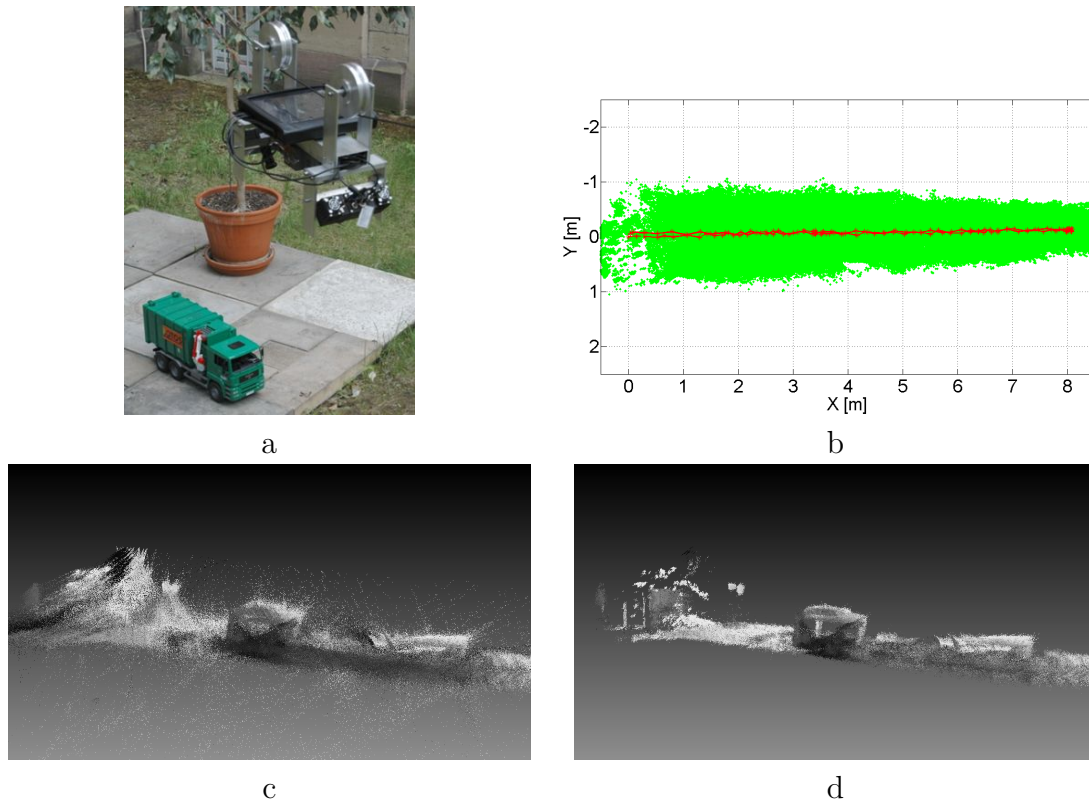


Abbildung 2.20: Koregistrierung bei Entfernungskameras: a) Szeneninhalt, b) auf die Grundebene projizierte registrierte Punktwolke im gemeinsamen, übergeordneten Weltkoordinatensystem (grün) mit verketteter relativer Orientierung (rot), c) ungefilterte Punktwolke und d) gefilterte Punktwolke mit anthropogenem Objekt (mittig).

2.3 Szenenanalyse

Nach der Vereinheitlichung der Daten wird in der dritten Verarbeitungsstufe der Szeneninhalt analysiert. Um den Inhalt einer Szene bezüglich der darin enthaltenen Strukturen möglichst effizient zu analysieren, ist sowohl die schnelle Erfassung durch aktive Sensoren (alle TLS, MLS, ALS; Kapitel 1.1) als auch eine automatisierte Interpretation von Punktwolken von großer Bedeutung. Dabei ist eine wesentliche Herausforderung, jeden 3D-Punkt der Punktwolke automatisch eine semantische Kennzeichnung zuzuweisen (beispielsweise Boden, Gebäudefassade oder Vegetation) [8, 69].

Zur Automatisierung der Szenenanalyse werden vier wesentliche Verfahrensschritte (Abbildung 2.21) vorgeschlagen. Da aktive Sensoren primär der geometrischen Erfassung dienen, werden in diesem Kapitel keine radiometrischen Eigenschaften,

sondern nur geometrische Eigenschaften der Szene in Form von Punktwolken für die Analyse herangezogen. Um einen 3D-Punkt der Punktwolke entsprechend zu charakterisieren, ist die Verwendung der lokalen Nachbarschaft zweckmäßig. Da die Charakterisierung von 3D-Punkten sowohl von der nicht bekannten Szenenstruktur als auch von der zu wählenden Nachbarschaftsanzahl geprägt ist, wird zuerst die Wahl der optimalen lokalen Nachbarschaft zur Merkmalsberechnung angegangen und erläutert (Kapitel 2.3.1). Aus der optimalen lokalen Nachbarschaft werden im nächsten Schritt verschiedene geometrische Merkmale extrahiert (Kapitel 2.3.2) und dem 3D-Punkt zugewiesen. Da die Relevanz der geometrischen Merkmale von den Szeneninhalten beeinflusst wird, ist es sinnvoll für eine performante Klassifikation eine optimale Auswahl von relevanten Merkmalen (Kapitel 2.3.3) zu bestimmen. Abschließend wird basierend auf den bestimmten relevanten Merkmalen eine Klassifikation (Kapitel 2.3.4) durchgeführt.

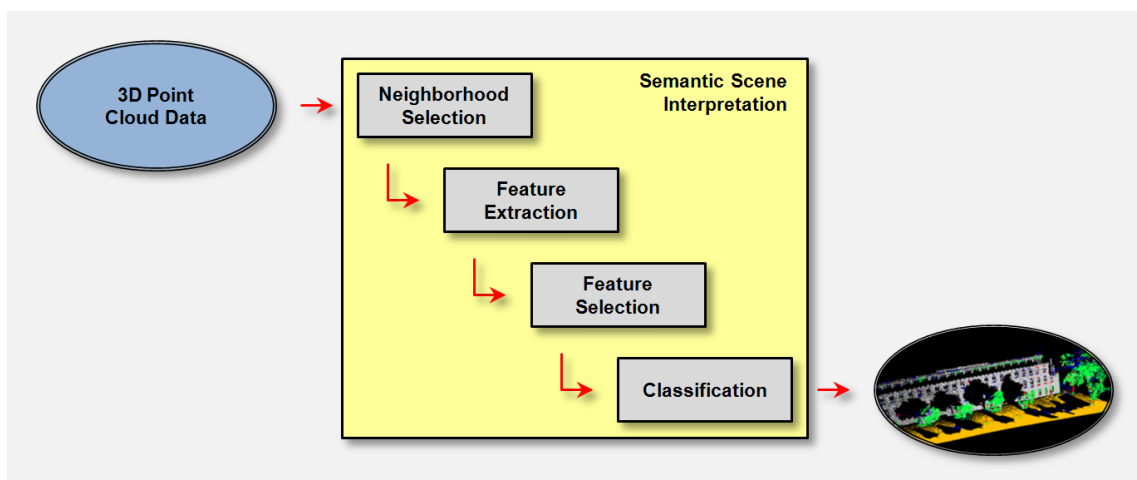


Abbildung 2.21: Szenenanalyse durch semantische Interpretation bei Punktwolken.

2.3.1 Bestimmung der optimalen lokalen Nachbarschaften

Zur Berechnung von Merkmalen für einen 3D-Punkt wird die lokale Nachbarschaft berücksichtigt, da diese den benachbarten Bereich des Punktes und somit indirekt den Punkt selbst charakterisiert. Wesentlich hierbei ist die Größe des benachbarten Bereichs, um den 3D-Punkt bestmöglich semantisch zu charakterisieren, ohne störende Einflüsse von weit entfernten nicht zuträglichen 3D-Punkten einfließen zu lassen.

Üblicherweise wird die lokale Nachbarschaft basierend auf Vorwissen bzgl. der Szene und der Daten empirisch bestimmt, wobei für die lokalen Nachbarschaften aller 3D-Punkte im Datensatz häufig die gleiche Parametrisierung verwendet wird [57, 58,

69, 71]. Da die Größe des benachbarten Bereichs von der lokalen Punktdichte und der lokalen Geometrie abhängt [4, 24], wurde ein neues generisches Verfahren zur automatischen Bestimmung von einer optimalen lokalen Nachbarschaft für jeden individuellen 3D-Punkt der Punktwolke entwickelt [C4, C5, C7].

Um die optimalen lokalen Nachbarschaften (Abbildung 2.21: *Neighborhood Selection*) für jeden 3D-Punkt zu bestimmen wird zuerst die Eigenentropie (*Eigenentropy*) E_λ eingeführt. Die Eigenentropie basiert auf der Kovarianzmatrix, wobei die Elemente der Kovarianzmatrix sich aus den Koordinaten eines 3D-Punktes \mathbf{X} und der dazu k nächsten 3D-Punkte berechnen. Die Kovarianzmatrix ist eine symmetrische positiv definite Matrix, woraus sich drei orthogonale Eigenvektoren mit den Eigenwerten $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ und $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ ableiten lassen. Um einen generischen Ansatz zu erhalten, wird eine Normalisierung über die drei Eigenwerte durch die Summe \sum_λ vorgenommen. Die Eigenentropie E_λ , abgeleitet von der Shannon Entropie [85], wird bestimmt durch

$$E_\lambda = -\epsilon_1 \ln(\epsilon_1) - \epsilon_2 \ln(\epsilon_2) - \epsilon_3 \ln(\epsilon_3), \quad (2.16)$$

wobei ϵ_i durch $\epsilon_i = \lambda_i / \sum_\lambda$ mit $i \in \{1, 2, 3\}$ gegeben ist.

Anhand der durchgeführten Untersuchungen konnte gezeigt werden, dass die Verwendung von optimalen lokalen Nachbarschaften basierend auf der beschriebenen *Eigenentropy-based Scale Selection* zu deskriptiven Merkmalen führt. Im Vergleich mit konstanter Nachbarschaftszahl ($\mathcal{N}_{10}, \mathcal{N}_{50}, \mathcal{N}_{100}$)² bzw. mit variabler Nachbarschaftszahl, beispielsweise bestimmt durch *Dimensionality-based Scale Selection* mit der optimalen Nachbarschaftszahl $\mathcal{N}_{opt,dim}$ ³, konnte das Klassifikationsergebnis durch *Eigenentropy-based Scale Selection* mit der optimalen Nachbarschaftszahl $\mathcal{N}_{opt,\lambda}$ zusätzlich verbessert werden. Das erreichte Klassifikationsergebnis wird durch das Gütemaß \mathcal{F}_1 -Score [92] bewertet, welches die Genauigkeit (*Precision*) und Trefferquote (*Recall*) in einem Gütemaß kombiniert.

In Abbildung 2.22 ist der \mathcal{F}_1 -Score für die Nachbarschaftszahl $\mathcal{N}_{10}, \mathcal{N}_{50}, \mathcal{N}_{100}, \mathcal{N}_{opt,dim}$ und $\mathcal{N}_{opt,\lambda}$ aufgetragen. Als Klassifikation wurde bei dem dargestellten Beispiel exemplarisch die *Random Forest* (RF) [9] Klassifikation verwendet. Die untersuchte Szene (Abbildung 2.26) beinhaltet verschiedene Klassen (Leitung, Pfosten/Stamm, Gebäudefassade, Boden und Vegetation). Für die Nachbarschaftszahl $\mathcal{N}_{opt,\lambda}$ ist der \mathcal{F}_1 -Score generell für jede der fünf Klassen sehr hoch, meist wird damit das beste Ergebnis erreicht.

²Nachbarschaft mit 10, 50, 100 nächsten Nachbarn

³Optimale Nachbarschaft nach [17]

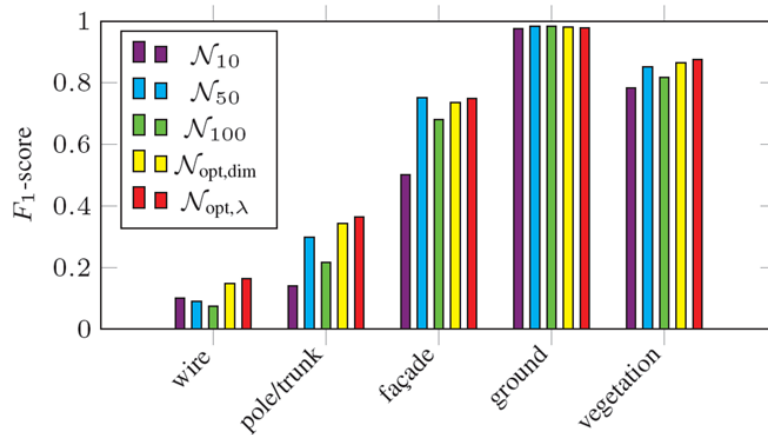


Abbildung 2.22: \mathcal{F}_1 -Score für die Nachbarschaftsanzahl \mathcal{N}_{10} , \mathcal{N}_{50} , \mathcal{N}_{100} , $\mathcal{N}_{opt,dim}$ und $\mathcal{N}_{opt,\lambda}$. Untersuchte Klassen (von links nach rechts) Leitung (*wire*), Pfosten/Stamm (*pole/trunk*), Gebäudefassade (*façade*), Boden (*ground*) und Vegetation (*vegetation*).

2.3.2 Merkmalsextraktion

Um deskriptive Merkmale aus den 3D-Punkten zu extrahieren (Abbildung 2.21: *Feature Extraction*) wird im Folgenden sowohl auf 3D-Merkmale als auch 2D-Merkmale eingegangen. Zur Beschreibung der lokalen Nachbarschaft eines 3D-Punktes \mathbf{X} sind die in Kapitel 2.3.1 bereits erwähnten Eigenwerte $\lambda_1, \lambda_2, \lambda_3$ besonders gut geeignet sind. Aus diesen Eigenwerten lassen sich sowohl 3D-Merkmale (beispielsweise für lineare, planare und volumetrische Strukturen) [71, 101] als auch spezielle geometrische Formprimitive (beispielsweise Punkt bzw. verschiedene Linien und Flächen aus Abbildung 2.23) [C1] bestimmen, die zur Beschreibung von bestimmten Szeneninhalten verwendet werden können, unter anderem für natürliche Objekte (Kapitel 2.4.1) [C2] sowie für anthropogene Objekte (Kapitel 2.4.2) [C3].

In der Literatur werden verschiedene geometrische Eigenschaften zur Berechnung zusätzlicher Merkmale vorgeschlagen, die wesentlichen Eigenschaften sind beispielsweise Winkelstatistiken [67], Höhencharakteristiken und lokale Planaritäten [62], Schiefe und vertikale Profile [33] bzw. Punktabstände und Höhenunterschiede [97]. Neben diesen vorgestellten Eigenwert-basierten Merkmalen, die alleine auf der Geometrie der lokalen Nachbarschaft basieren, werden zusätzliche radiometrische Merkmale (Kapitel 2.1.2 bzw. Kapitel 2.1.3) berücksichtigt [12, 62, 69, 82, 97] um die Klassifikationsleistung zu steigern.

Ohne Anspruch auf Vollständigkeit werden exemplarisch die wesentlichen 3D- und 2D-Merkmale beschrieben, die sich bei der Szenenanalyse bewährt haben [C6, C4, C5, C7] und bei der in Kapitel 2.3.3 folgenden Auswahl relevanter Merkmale verwendet werden.










Primitives	λ_1	λ_2	λ_3
 Isolated point	0	0	0
 End of a line	$\frac{1}{12}$	0	0
 Line	$\frac{1}{3}$	0	0
 Half plane	$\frac{1}{4}$	$\frac{1}{4}\left(1 - \frac{64}{9\pi^2}\right) = 0.07$	0
 Plane	$\frac{1}{4}$	$\frac{1}{4}$	0
 Quarter plane	$\frac{1}{4}\left(1 - \frac{2}{\pi}\right) = 0.09$	$\frac{1}{4} + \frac{1}{2\pi} - \frac{32}{9\pi^2} = 0.05$	0
 Two planes	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8} - \frac{8}{9\pi^2} = 0.03$
 Three planes	$\frac{1}{6}\left(1 - \frac{1}{\pi}\right) = 0.11$	$\frac{1}{6}\left(1 - \frac{1}{\pi}\right) = 0.11$	$\frac{1}{6}\left(1 + \frac{2}{\pi}\right) - \frac{2^6}{3^3\pi^2} = 0.03$
 Two planes 60°	0.25	0.1875	0.01747

Abbildung 2.23: Spezielle geometrische Formprimitive.

3D-Merkmale

Die aus der lokalen Nachbarschaft, respektive 3D-Punkten, abgeleiteten Merkmale werden als 3D-Merkmale bezeichnet. Die 3D-Merkmale *Linearity* L_λ , *Planarity* P_λ und *Scattering* S_λ eignen sich um lineare 1D-Strukturen, planare 2D-Strukturen oder volumetrische 3D-Strukturen zu beschreiben. Zusätzliche 3D-Merkmale sind die *Omnivariance* O_λ , *Anisotropy* A_λ , *Eigenentropy* E_λ und die *Sum of eigenvalues* bezeichnet mit Σ_λ . Ein Indikator für Krümmung ist das Merkmal *Change of curvature* C_λ . Die formale Definition der genannten Merkmale ist gegeben mit

$$\text{Linearity } L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1}, \quad (2.17)$$

$$\text{Planarity } P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1}, \quad (2.18)$$

$$\text{Scattering } S_\lambda = \frac{\lambda_3}{\lambda_1}, \quad (2.19)$$

$$\text{Omnivariance } O_\lambda = \sqrt[3]{\lambda_1 \lambda_2 \lambda_3}, \quad (2.20)$$

$$\text{Anisotropy } A_\lambda = \frac{\lambda_1 - \lambda_3}{\lambda_1}, \quad (2.21)$$

$$\text{Eigenentropy } E_\lambda = - \sum_{i=1}^3 \lambda_i \ln(\lambda_i), \quad (2.22)$$

$$\text{Sum of eigenvalues } \Sigma_\lambda = \lambda_1 + \lambda_2 + \lambda_3, \quad (2.23)$$

$$\text{Change of curvature } C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}. \quad (2.24)$$

Weitere wichtige 3D-Merkmale, jedoch nicht weiter spezifiziert, sind die *Verticality* V [18] und die modifizierte *Local Point Density* D [56].

2D-Merkmale

Anthropogene Objekte sind für die Szenenanalyse relevant, da diese Objekte häufig planare Oberflächen aufweisen und zudem vertikal ausgerichtet sind (beispielsweise Gebäudefassade, Pfosten) ist es konsequent, speziell hierfür Merkmale zu entwickeln. Es ist naheliegend, dass hierfür 3D-Punkte auf eine horizontal orientierte Ebene (Geländeebene) projiziert werden und daraus 2D-Merkmale abgeleitet werden, beispielsweise die *Local Point Density* D_{2D} und der Radius $r_{k-NN,2D}$ [56]. Ein Hinweis auf planare 2D-Strukturen wird durch die Eigenwerte $\lambda_{1,2D}$, $\lambda_{2,2D}$ bzw. deren Verhältnis $R_{\lambda,2D} = \frac{\lambda_{2,2D}}{\lambda_{1,2D}}$ ermöglicht. Ebenso wird die zweidimensionale Variante der *Sum of Eigenvalues* $\Sigma_{\lambda,2D}$ berücksichtigt. Zudem kann über eine *Accumulation Map* [65] die Merkmale *Maximum Height Difference* ΔZ und die *Height Variance* σ_Z bestimmt werden.

2.3.3 Auswahl relevanter Merkmale

Da jedes einzelne Merkmal die Eigenschaften des benachbarten Bereichs beschreibt, sind nur bestimmte Merkmale (Abbildung 2.21: *Feature Selection*) für spezielle Anwendungen geeignet. Normalerweise ist jedoch kein klassenspezifisches Vorwissen hinsichtlich der Merkmalseignung bekannt, daher werden prinzipiell alle verfügbaren Merkmale eingesetzt, um anhand von Trainingsdaten einen Klassifikator zu trainieren. Es ist zu erwarten, dass einige Merkmale irrelevant sind. Durch Reduktion aller verfügbaren Merkmale auf die relevanten Merkmale kann die Klassifikationsleistung gesteigert werden und zusätzlich die Rechenzeit als auch die Anforderungen bezüglich des Speicherbedarfs deutlich reduziert werden [34]. Durch Kombination von willkürlich gewählten Merkmalen bei der Klassifikation und der dazu korrespondierenden Klassifikationsleistung können relevante Merkmale nur mit sehr hohem Rechenaufwand bestimmt werden [12, 50, 62]. Im Gegensatz dazu können klassifikationsunabhängige Abschätzungen hinsichtlich der Eignung des Merkmals anhand von *Filter-based Feature Selection* Methoden besonders einfach und effizient bestimmt werden [C8]. Mit der *Univariate Filter-based Feature Selection* Methode [C6] und der *Multivariate Filter-based Feature Selection* Methode [C7] lässt sich dies weiter spezifizieren.

Univariate Filter-based Feature Selection

Bei der *Univariate Filter-based Feature Selection* Methode impliziert eine Bewertungsfunktion die Relation zwischen Merkmal und Klasse um die Relevanz der Merkmale zu bestimmen. Die hierfür am häufigsten eingesetzten Bewertungsfunktionen sind beispielsweise *Pearson Correlation Coefficient* [72], *Gini Index* [31], *Fisher Score* [26], *Information Gain* [78] oder *Symmetrical Uncertainty* [77].

Multivariate Filter-based Feature Selection

Bei der *Multivariate Filter-based Feature Selection* Methode impliziert eine Bewertungsfunktion sowohl die Relation zwischen Merkmal und Klasse als auch die Relation zwischen den Merkmalen selbst, um die Redundanz der Merkmale zu bestimmen. Das hierfür am häufigsten eingesetzte Bewertungsverfahren ist *ReliefF* [51], weitere Bewertungsverfahren sind *Correlation-based Feature Selection* (CFS) [35], *Fast Correlation-Based Filter* (FCBF) [103] oder *minimal-Redundancy-Maximal-Relevance (mRMR) Criterion* [73].

Hinsichtlich Relevanz (*Maximal Relevance Selection*) bzw. Redundanz (*Minimal Redundancy Selection*) der analysierten Merkmale ergeben sich hierfür Bewertungen die in einer übergeordneten Relevanzmetrik vereinheitlicht werden können. Die Relevanzmetrik basiert auf verschiedenen Strategien zur Analyse der intrinsischen Eigenschaften der gegebenen Trainingsdaten und berücksichtigt verschiedene Aspekte hinsichtlich Maßen bezüglich Abstand, Information, Abhängigkeit oder Konsistenz. Die Betrachtung der Merkmale anhand der Relevanzmetrik ermöglicht eine Einteilung in mehr oder weniger relevante Merkmale (Abbildung 2.24), wobei die für die Klassifikation zu wählende Anzahl von relevanten Merkmalen beispielsweise aus der

Gesamtgenauigkeit der Klassifikation (Abbildung 2.25) abgeleitet werden kann [C6]. Ferner wird durch Abbildung 2.25 verdeutlicht, dass durch alleinige Hinzunahme von zusätzlichen Merkmalen nicht notwendigerweise eine zusätzliche Steigerung der Gesamtgenauigkeit der Klassifikation ermöglicht wird bzw. durch die Berücksichtigung von wenigen relevanten Merkmalen im Vergleich zur Verwendung aller Merkmale die Gesamtgenauigkeit der Klassifikation deutlich gesteigert wird.

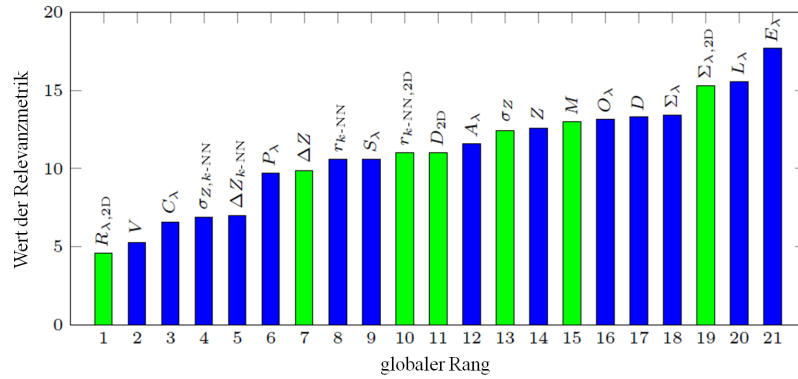


Abbildung 2.24: Rangfolge der 3D-Merkmale (blau) und 2D-Merkmale (grün) entsprechend der Relevanzmetrik aufgetragen, je niedriger der Rang desto relevanter das Merkmal.

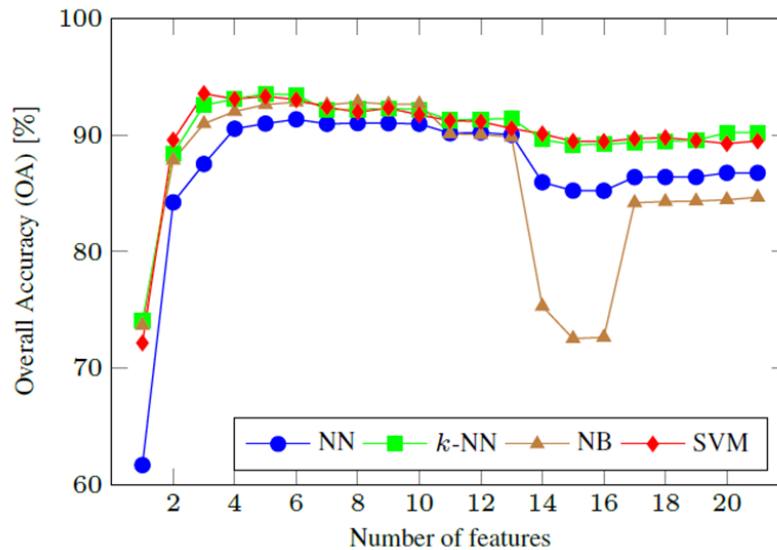


Abbildung 2.25: Gesamtgenauigkeit der Klassifikation (*Overall Accuracy*) von *Nearest Neighbor* (NN), *k-Nearest Neighbor* (k -NN), *Naive Bayesian* (NB) und *Support Vector Machine* (SVM) Klassifikation (Kapitel 2.3.4) aufgetragen über die akkumulierte Rangfolge der Merkmale.

2.3.4 Klassifikation

Die in Kapitel 2.3.3 ermittelten relevanten Merkmale werden in einem Merkmalsvektor zusammengefasst und für jeden 3D-Punkt wird der dazugehörige individuelle Merkmalsvektor zur Klassifikation (Abbildung 2.21: *Classification*) eingesetzt. Für die Untersuchungen sind Verfahren zur überwachten Klassifikation geeignet [100], beispielsweise die *Nearest Neighbor* (NN), *k-Nearest Neighbor* (*k*-NN) [15], *Naive Bayesian* (NB) [44], *Support Vector Machine* (SVM) [14] und *Random Forest* (RF) [9] Klassifikation. Benachbarte 3D-Punkte weisen sehr häufig gleiche Klassenzugehörigkeit auf, durch kontextabhängige Klassifikation (*Contextual Classification*) kann diese Interaktion zwischen benachbarten 3D-Punkten zusätzlich berücksichtigt werden. Insbesondere statistische Modelle sind für kontextabhängige Klassifikation geeignet und werden beispielsweise bei *Conditional Random Field* (CRF) Klassifikation [53] angewandt. Es konnte gezeigt werden, dass durch kontextabhängige Klassifikation die Klassifikationsleistung zusätzlich gesteigert werden kann [C9]. Hierzu ist in Abbildung 2.26 ein Klassifikationsergebnis einer *Random Forest* (RF) Klassifikation verglichen mit einer *Conditional Random Field* (CRF) Klassifikation dargestellt für insgesamt fünf urbane Klassen. Es wird eine bessere Gruppierung mit geringerem Rauschen (Abbildung 2.26b vs. d) durch kontextabhängige Klassifikation erreicht.

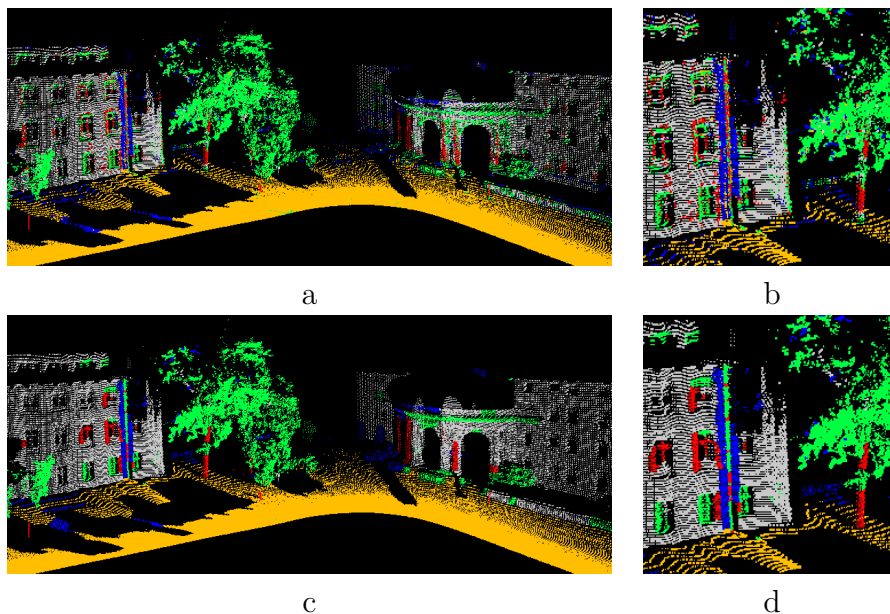


Abbildung 2.26: Klassifikationsergebnis von 3D-Punkten mit optimaler lokaler Nachbarschaft zur Merkmalsberechnung: *Random Forest* (RF) Klassifikation (a) mit Ausschnittsvergrößerung (b), *Conditional Random Field* (CRF) Klassifikation (c) mit Ausschnittsvergrößerung (d). Farbkodierung der Klassen: Leitung (blau), Pfosten/Stamm (rot), Gebäudefassade (grau), Boden (braun), Vegetation (grün).

2.4 Objektrekonstruktion

Nach der sehr allgemein gehaltenen Szenenanalyse (Kapitel 2.3) wird in der vierten Verarbeitungsstufe eine differenzierte Analyse der Szene hinsichtlich einzelner Objekte angegangen. Hierbei sollen einzelne Objekte erkannt bzw. rekonstruiert werden. Im Vergleich zu den vorherigen Verarbeitungsstufen werden in diesem Kapitel hauptsächlich erste Entwicklungen gezeigt die zukünftig weiter vertieft werden können. Bei diesem Kapitel wird wegen der speziellen Formen bestimmter Objekte eine Unterscheidung hinsichtlich natürlicher Objekte (Kapitel 2.4.1) und anthropogener Objekte (Kapitel 2.4.2) vorgenommen.

2.4.1 Natürliche Objekte

Im Folgenden wird ein Ansatz zur Objektrekonstruktion bei natürlichen Objekten vorgestellt. Aus der Objekterkennung sind Verfahren bekannt, die statt einzelner Parameterwerte in Form von 3D-Merkmalen und 2D-Merkmalen (Kapitel 2.3.2) durch eine höher dimensionierte Parametrisierung eine möglichst repräsentative Verteilung aus der Szene bestimmen, die wiederum als Histogramm repräsentiert werden [59, 81, 89, 88]. Zu diesen Verfahren zählt auch die *Shape Distribution* Methode [70], bei der eine Häufigkeitsverteilung für zufällig gewählte 3D-Punkte vorgenommen wird um daraus Histogramme mit spezifischen Eigenschaften zu generieren. In Anlehnung an die ursprüngliche Idee dieses Ansatzes können unter anderem fünf verschiedene Histogramme berechnet werden, deren Parametrisierung (A3, D1, D2, D3, D4) auf geometrischen Maßen basiert, beispielsweise Winkeln, Distanzen, Flächen und Volumen zwischen zufällig gewählten 3D-Punkten:

- Winkel zwischen drei zufällig gewählten 3D-Punkten (A3)
- Distanz eines zufällig gewählten Punktes zum Schwerpunkt aller 3D-Punkte innerhalb der Nachbarschaft (D1)
- Distanz zwischen zwei zufällig gewählten 3D-Punkten (D2)
- Fläche eines Dreiecks zwischen drei zufällig gewählten 3D-Punkten (D3)
- Volumen eines Tetraeders zwischen vier zufällig gewählten 3D-Punkten (D4)

Die Vorteile der spezifischen Eigenschaften der berechneten Histogramme sind deren Rotations- und Translationsinvarianz, die Robustheit gegenüber kleinen Störungen und eine effiziente Berechnung. Aus der Ähnlichkeit zwischen den Häufigkeitsverteilungen der Parametrisierung lassen sich Rückschlüsse auf die Ähnlichkeit von Objekten schließen. Erste Ergebnisse sind vielversprechend und zeigen das Potential hinsichtlich Objektrekonstruktion bei natürlichen Objekten [D1].

2.4.2 Anthropogene Objekte

Anthropogene Objekte sind meist von regulären Formen geprägt, weswegen sich eine Modellierung durch geometrische Formprimitive bzw. einfache Strukturelemente anbietet, beispielsweise durch Punkte, Linien und Flächen. Im Umkehrschluss ist es daher erstrebenswert basierend auf 3D-Punkten innerhalb einer Punktwolke bestimmte Formprimitive zu bestimmen um Hinweise für anthropogene Objekte zu erhalten. Eine gesteigerte Klassifikationsleistung bei der Objektdetektion wurde beispielsweise für Gebäude erreicht, indem mittels *Airborne Laser Scanning* Daten zwischen flachen und geneigten Dachflächen unterschieden wird [D2]. Zudem wird eine verbesserte Datengrundlage bei Gebäudedachflächen erreicht, indem zuerst eine Modellierung der Gebäudedachs als Fläche erfolgt, um dann bei unsicheren Messwerten eine nachträgliche Verdichtung der Daten zu ermöglichen [D4, D5]. Für eine robustere Objektdetektion werden bei Gebäuden stabile Liniensegmente generiert um die Gebäudekanten möglichst vollständig damit zu beschreiben [A8]. Unter der Annahme, dass Gebäudekanten gerade sind, können diese Kanten als Linien modelliert werden. Mit speziellen Einschränkungen ist dann eine subpixel-genaue Lokalisierung von Kanten möglich [D3].

Als Alternative zu dem vorherigen modellbasierten Ansatz kann durch eine merkmalsbasierte Objektdetektion anhand SIFT [59] ohne Hinzunahme von geometrischen Formprimitiven eine Zuordnung zwischen anthropogenem Objekt und Modell erfolgen. In Abbildung 2.27 ist ein Fahrzeug zu sehen, welches trotz perspektivischer Änderung als anthropogenes Objekt detektiert wird.

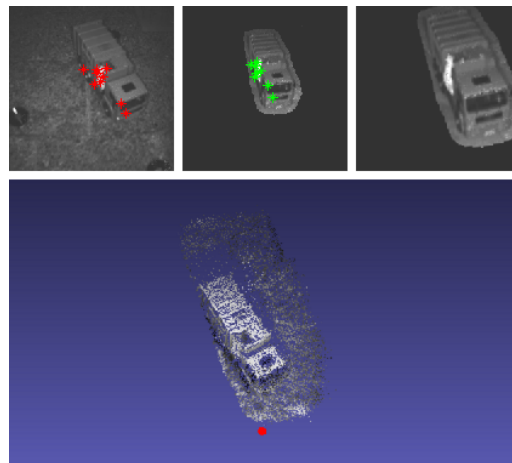


Abbildung 2.27: Detektion und Erkennung von anthropogenen Objekten: Passives Intensitätsbild mit Fahrzeug aus Abbildung 2.20a, *Template* aus einer Datenbank und transformiertes *Template* (oben, v.l.n.r.) sowie zum transformierten *Template* korrespondierende Punktwolke (unten; der rote Punkt entspricht der Sensorposition).

3 Diskussion und Ausblick

In diesem Kapitel werden die zuvor beschriebenen Verarbeitungsstufen diskutiert und es wird davon ausgehend ein Ausblick auf zukünftige Forschungsaktivitäten gegeben. Zudem wird eine Positionierung der dazu entstandenen Schlüsselpublikationen im wissenschaftlichen Umfeld vorgenommen.

Mit der ersten Verarbeitungsstufe **Modellierung und Analyse von Messsignalen** (Kapitel 2.1) werden die mit aktiven Sensoren erfassten Daten für die nachfolgenden Verarbeitungsstufen zweckmäßig aufbereitet. Beim **Phase-Unwrapping bei Entfernungsbildern** (Kapitel 2.1.1) stellen die szenenbedingten Diskontinuitäten eine wesentliche Beschränkung für die Verarbeitung dar, die nicht ohne zusätzlichen Aufwand überwunden werden kann. Eine Weiterentwicklung wird ermöglicht, indem statt einer einzigen Beobachtung mehrere zeitlich aufeinanderfolgende Beobachtungen, erfasst von verschiedenen Standorten, analysiert werden und damit Unzulänglichkeiten beim *Phase-Unwrapping* ausgeschlossen werden, beispielsweise durch Eigenbewegung des aktiven optischen Sensors. Zudem wurde mit der **Signaloptimierung bei Full-Waveform-Laserscanner-Daten** (Kapitel 2.1.2) eine neue Möglichkeit für die Verarbeitung von Laserscanner-Daten vorgestellt. Die hierfür entwickelten Methoden sind nur eingeschränkt bei den aktuell verwendeten Bandbreiten und Empfindlichkeit der Sensorelemente bei aktiven optischen Sensoren einsetzbar. Deswegen sind zusätzliche Weiterentwicklung hinsichtlich der Bandbreite und Empfindlichkeit der Sensorelemente bei aktiven Sensoren erforderlich, um eine noch differenziertere Auswertung der rückgestreuten Strahlung und somit eine verbesserte Analyse der beleuchteten Oberfläche zu ermöglichen. Die **Radiometrische Kalibrierung bei Laserscanner-Daten** (Kapitel 2.1.3) zeigt, dass neben der Erfassung von geometrischen Daten zusätzlich die radiometrischen Daten von Bedeutung für die Auswertung sind. Auch hier ist eine Verbesserung nur dann möglich, wenn die zuvor definierten Randbedingungen eingehalten werden können und zudem rauscharme Signale vorliegen. Dies bedeutet, dass hier ebenfalls eine erhöhte Bandbreite und Empfindlichkeit der Sensorelemente bei aktiven optischen Sensoren relevant ist. Die zu diesem Thema veröffentlichten Schlüsselpublikationen wurden zum einen auf renommierten internationalen Konferenzen der *International Society for Photogrammetry and Remote Sensing (ISPRS)* als *Peer-Reviewed* Konferenzartikel akzeptiert und präsentiert, beispielsweise bei den *Laserscanning* [A5] und *Photogrammetric Image Analysis* [A10] Konferenzen und zudem als begutachtete Artikel in nationalen und internationalen Zeitschriften akzeptiert, beispielsweise *AVN - Allgemeine Vermessungs-Nachrichten* [A4], *Journal of Photogrammetric Engineering & Remote Sensing (PE&RS)* [A7], *Revue Française de Photogrammétrie*

et de Télédétection [A9] und *ISPRS Journal of Photogrammetry & Remote Sensing* [A12]. Der letztgenannte Beitrag [A12] wurde als bester Beitrag des *ISPRS Journal of Photogrammetry & Remote Sensing* (Volume 61) mit dem *U.V. Helava Award* prämiert. Ein Beitrag [A14] wurde als Buchkapitel bei dem internationalen Standardwerk *Topographic Laser Ranging and Scanning: Principles and Processing* berücksichtigt.

Nach der Aufbereitung der Daten werden mit der zweiten Verarbeitungsstufe **Koregistrierung mit 2D- bzw. 3D-Sensoren** (Kapitel 2.2) unterschiedliche Beobachtungen vereinheitlicht. Sowohl bei **Entfernungskamera kombiniert mit thermischer Infrarot-Kamera** (Kapitel 2.2.1.1) als auch bei **Mobiler Laserscanner kombiniert mit RGB-Kamera** (Kapitel 2.2.1.2) werden bistatische Sensoranordnungen eingesetzt, weswegen sich durch unterschiedliche Perspektiven die konstruktionsbedingte Nachteile insbesondere im nahen Entfernungsbereich ergeben. Abhilfe ist prinzipiell durch aufwendigere monostatische Systeme möglich, jedoch erscheint eine zeitnahe technische Umsetzung von kombinierten aktiven und passiven Sensoren nur mit großem Aufwand möglich. Deswegen stellen derzeit die vorgeschlagenen Methoden zur 3D/2D-Koregistrierung mit den perspektivischen Einschränkungen im nahen Entfernungsbereich hierzu eine gute Alternative dar. Bei den entwickelten neuen Methoden zu **Koregistrierung bei terrestrischen Laserscanner** (Kapitel 2.2.2.1) bzw. **Koregistrierung bei Entfernungskameras** (Kapitel 2.2.2.2) wird die merkmalsbasierte Zuordnung texturbedingt meist nur anhand von radiometrischer Bildrepräsentationen durchgeführt, die geometrischen Messungen bleiben hierfür oft ungenutzt. Deswegen erscheint eine Kombination von Radiometrie und Geometrie bzw. die zusätzliche Verwendung von generierten Merkmalsbildern sinnvoll. Erste eigene Ansätze hierzu wurden bereits in der internationalen Open-Access-Zeitschrift *Remote Sensing* publiziert [B1]. Weitere Schlüsselpublikationen sind in den internationalen Zeitschriften *International Journal of Image and Data Fusion* [B8] und *ISPRS Journal of Photogrammetry & Remote Sensing* [B14] erschienen. Zusätzlich wurden auf verschiedenen hochwertigen internationalen Konferenzen begutachtete Beiträge vorgestellt, beispielsweise auf dem *IEEE - Joint Urban Remote Sensing Event (JURSE)* [B11], der *Laserscanning* Konferenz der *ISPRS* [B12] und dem *Technical Commission I Symposium - Sustaining Land Imaging: UAVs to Satellites* der *ISPRS* [B13].

Nach der Vereinheitlichung der Daten wird in der dritten Verarbeitungsstufe durch **Szenenanalyse** (Kapitel 2.3) der Inhalt einer Szene hinsichtlich darin enthaltenen Strukturen analysiert. Die **Bestimmung der optimalen lokalen Nachbarschaften** (Kapitel 2.3.1) eines 3D-Punktes erfolgt in puncto aller 3D-Punkte die innerhalb einer radialsymmetrischen Nachbarschaft als relevant für eine semantische Charakterisierung angesehen werden. Die Prozessierung aller 3D-Punkte ist prinzipiell sehr aufwendig. Bei zukünftigen Untersuchungen könnte anhand von Stichproben eine Untermenge der 3D-Punkte für eine effiziente Schätzung der optimalen lokalen Nachbarschaft berücksichtigt werden. In diesem Zusammenhang radialsymmetrische Nachbarschaften zu verwenden ist naheliegend, jedoch gegebenenfalls nicht optimal,

deswegen wird eine an die örtlichen Gegebenheiten adaptierte Nachbarschaft postuliert, die durchaus asymmetrische Formen annehmen kann. Die **Merkmalsextraktion** (Kapitel 2.3.2) basiert im wesentlichen auf definierten, vielfältigen und deskriptiven Merkmalen. Diese Merkmale sind generisch, jedoch noch nicht anwendungsspezifisch optimiert. Dies bedeutet, dass zukünftig für spezielle Anwendungen gegebenenfalls spezifische Merkmale sowohl konzipiert als auch untersucht werden müssen. Die darauf folgende **Auswahl relevanter Merkmale** (Kapitel 2.3.3) dient zur Steigerung der Klassifikationsleistung. Die hierfür verwendeten Bewertungsfunktionen sind nach den bisherigen Untersuchungen zufolge sehr robust, die Berücksichtigung von zusätzlichen Bewertungsfunktionen ermöglicht gegebenenfalls noch eine verbesserte Optimierung, wobei die mögliche Verbesserung nach derzeitigen Erfahrungen eine sehr große Herausforderung darstellt. Für die **Klassifikation** (Kapitel 2.3.4) wurden bekannte Klassifikationsschemen angewandt um eine möglichst hohe Gesamtgenauigkeit bei der Verarbeitung zu erreichen, methodische Weiterentwicklungen können bei Bedarf zukünftig angegangen werden. Die Schlüsselpublikationen in diesem Kapitel sind erschienen bei den angesehenen internationalen Zeitschriften *ISPRS Journal of Photogrammetry & Remote Sensing* [C4] und *Computers & Graphics Journal* [C5]. Zudem wurde ein *Peer-Reviewed* Konferenzartikel [C7] bei der internationalen Konferenz *ISPRS Technical Commission III Symposium - Photogrammetric Computer Vision and Image Analysis (PCV)* mit dem *PCV 2014 Best Paper - Honourable Mention* prämiert.

Nach der sehr allgemein gehaltenen Szenenanalyse wird in der vierten Verarbeitungsstufe durch **Objektrekonstruktion** (Kapitel 2.4) eine differenzierte Analyse der Szene hinsichtlich einzelner Objekte angegangen. Diese Verarbeitungsstufe ist in der Verfahrensentwicklung noch nicht sehr weit fortgeschritten, da der Schwerpunkt der Forschungsaktivitäten auf den vorherigen Verarbeitungsstufen lag. Zur Rekonstruktion bei **Natürlichen Objekten** (Kapitel 2.4.1) wird eine höher dimensionierte Parametrisierung eingesetzt, um einzelne natürliche Objekte zu bestimmen. Die ersten Ergebnisse hierzu sind vielversprechend. Es besteht weiterhin Forschungsbedarf hinsichtlich der Konzeption von zusätzlichen höher dimensionierten Merkmalen, die optimal an die Objektcharakteristik angepasst sind. Bei **Anthropogenen Objekten** (Kapitel 2.4.2) werden zur Rekonstruktion häufig Verfahren entwickelt die auf Formprimitiven (Strukturelemente) basieren. Diese Formprimitive gilt es möglichst genau und vollständig zu bestimmen, insbesondere die Vollständigkeit kann unter Berücksichtigung von Modellwissen bei verrauschten Messungen verbessert werden, indem schwache bzw. verrauschte Signale zusätzlich extrahiert werden. Eine Schlüsselpublikation [D4] hierzu ist in der internationalen Zeitschrift *ISPRS Journal of Photogrammetry & Remote Sensing* erschienen. Eine weitere Schlüsselpublikation wurde als *Peer-Reviewed* Konferenzartikel bei der *Laserscanning* Konferenz der *ISPRS* [B12] akzeptiert und präsentiert.

Von übergeordneter Bedeutung für zukünftige Forschungsaktivitäten im Umfeld des Habilitationsthemas drängen sich verschiedene Fragestellungen zu den vorgestellten Verarbeitungsstufen auf. Beispielsweise erscheint eine generelle Verbesserung

der Datengrundlage durch die standortoptimierte Erfassung mit Sensoren möglich indem die bestmögliche nächste Mess- bzw. Beobachtungsposition prädiziert wird (*Next Best View Problem*) um durch möglichst wenige Messungen eine vollständige Objekt- bzw. Szenenerfassung zu erreichen. Hierfür kann sequentiell vorgegangen werden bzw. durch die zunehmende Verfügbarkeit von preiswerten Sensoren wird die parallele Erfassung der Szene immer attraktiver, beispielsweise durch den gleichzeitigen Einsatz von mehreren Sensoren. Mehrere mit Sensoren ausgestattete UAS mit kollektiver Intelligenz sind für diese Aufgabe geradezu prädestiniert. Hierbei sind keine Einschränkungen hinsichtlich andersgearteten Sensoren gegeben, denn ein Sensorverbund aus aktiven und passiven Sensoren zur Datenerfassung ist wesentlich geringer anfällig gegenüber sensorspezifischen Limitierungen, wodurch sich zukünftig neue Möglichkeiten zur automatischen Szenencharakterisierung in der Photogrammetrie und Fernerkundung ergeben.

4 Relevante Publikationen als Teil der Habilitationsschrift

Diese kumulative Habilitationsschrift beinhaltet Publikationen des Verfassers zu den in Kapitel 2 vorgestellten Methoden. Den entwickelten Methoden wurden relevante Publikationen zugeordnet und separat in Kapitel A, Kapitel B, Kapitel C und Kapitel D aufgeführt. Insgesamt 22 Schlüsselpublikationen, eine Untermenge der 45 relevanten Publikationen, haben eine spezielle Referenzkennung (Schriftstärke **fett**) und sind als Abdruck vollständig in der Habilitationsschrift enthalten. Nicht referenzierte Publikationen des Verfassers im erweiterten Themengebiet sind in Kapitel Z aufgeführt. Des Weiteren lassen sich relevante Publikationen kapitelweise in begutachtete Zeitschriftenartikel und *Peer-Reviewed* Konferenzartikel (*Full Paper & Double Blind Review*) einteilen und gruppieren. In Tabelle 4.1 sind nur die relevanten Publikationen und Schlüsselpublikationen aufgeführt, die diesen Kriterien entsprechen.

	Kapitel 2.1	Kapitel 2.2	Kapitel 2.3	Kapitel 2.4
	Kapitel A	Kapitel B	Kapitel C	Kapitel D
Zeitschriftenartikel	[A4] [A7] [A9] [A12] [A14]	[B1] [B4] [B8] [B14]	[C4] [C5] [C8]	[D2] [D4]
<i>Peer-Reviewed</i> Konferenzartikel	[A5] [A8] [A10] [A11] [A15]	[B11] [B12] [B13]	[C2] [C6] [C7] [C9]	[D1] [D3] [D5]

Tabelle 4.1: Relevante Publikationen und Schlüsselpublikationen kapitelweise unterteilt in Zeitschriftenartikel und *Peer-Reviewed* Konferenzartikel.

Für die vorliegende Habilitationsschrift wurden diese relevanten Publikationen nach folgenden Gesichtspunkten ausgewählt:

- Zu jedem der in den Kapiteln 2.1 bis 2.4 beschriebenen Forschungsthemen bringt der Verfasser der Habilitationsschrift **mindestens einen begutachteten Zeitschriftenartikel bzw. *Peer-Reviewed* Konferenzartikel** als

Erstautor ein. Diese Artikel wurden in namhaften Zeitschriften bzw. bei namhaften Konferenzen aus der Photogrammetrie, Fernerkundung und Computer Vision publiziert.

- Zu jedem der in den Kapiteln 2.1 bis 2.4 beschriebenen Forschungsthemen wurden **weiterführende Arbeiten** publiziert. Die Autorenschaft dieser Artikel kann variieren, meist erscheint dann der mit dem Thema betraute Doktorand bzw. Kooperationspartner als Erstautor.

An dieser Stelle sei erwähnt, dass ein begutachteter Zeitschriftenartikel [**A12**] mit dem *U.V. Helava Award* und ein *Peer-Reviewed* Konferenzartikel [**C7**] mit dem *PCV 2014 Best Paper - Honourable Mention* prämiert wurde.

Zudem hat der Verfasser bei mehreren dem Habilitationsthema fachnahen Herausgeberschaften [Z7, Z23, Z24, Z25, Z26, Z27] mitgewirkt.

A. Publikationen zu 2.1 Modellierung und Analyse von Messsignalen

- [A1] Bretar F, Chauve A, Mallet C, **Jutzi B** (2008) Managing full waveform LiDAR data: A challenging task for the forthcoming years. In: Chen J, Jiang J, Baudoin A (Eds) XXIth ISPRS Congress: Silk Road for Information from Imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (Part B1): 415-420
- [A2] Gross H, **Jutzi B**, Thoennessen U (2008) Intensity normalization by incidence angle and range of full-waveform LiDAR data. Chen J, Jiang J, Nayak S (Eds) XXIth ISPRS Congress: Silk Road for Information from Imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (Part B4): 405-412
- [A3] **Jutzi B** (2007) Analyse der zeitlichen Signalform von rückgestreuten Laserpulsen. Dissertation, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften (DGK), Reihe C, Nr. 611
- [A4] **Jutzi B** (2010) Extending the range measurement capabilities of modulated range imaging devices by time-frequency-multiplexing. AVN - Allgemeine Vermessungs-Nachrichten, Ausgabe 2/2012: 54-62
- [A5] **Jutzi B** (2009) Investigations on ambiguity unwrapping of range images. In: Bretar F, Pierrot-Deseilligny M, Vosselman G (Eds) Laserscanning 2009. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 3 / W8): 265-270
- [A6] **Jutzi B**, Eberle B, Stilla U (2002) Estimation and measurement of backscattered signals from pulsed laser radar. In: Serpico SB (Ed) Image and signal processing for remote sensing VIII. The International Society for Optics and Photonics (SPIE) Proceedings. Vol. 4885: 256-267
- [A7] **Jutzi B**, Gross H (2010) Investigations on surface reflection models for intensity normalization in airborne laser scanning (ALS) data. In: Heipke C, Jacobsen K, Müller S, Sörgel U (Eds) Journal of Photogrammetric Engineering & Remote Sensing (PE&RS), Vol. 76, No. 9, September 2010: 1051-1060
- [A8] **Jutzi B**, Gross H (2009) Normalization of lidar intensity data based on range and surface incidence angle. In: Bretar F, Pierrot-Deseilligny M, Vosselman G (Eds) Laserscanning 2009. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 3 / W8): 213-218

- [A9] **Jutzi B**, Stilla U (2007) Characteristics of the measurement unit of a full-waveform laser system. *Revue Française de Photogrammétrie et de Télédétection* 182 (2006-2): 17-22
- [A10] **Jutzi B**, Stilla U (2003) Laser pulse analysis for reconstruction and classification of urban objects. In: Ebner H, Heipke C, Mayer H, Pakzad K (Eds) *Photogrammetric Image Analysis PIA03*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3 / W8): 151-156
- [A11] **Jutzi B**, Stilla U (2006) Precise range estimation on known surfaces by analysis of full-waveform laser. In: Förstner W, Steffen R (Eds) *Symposium of ISPRS Commission III: Photogrammetric Computer Vision PCV06*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3): 234-239
- [A12] **Jutzi B**, Stilla U (2006) Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS Journal of Photogrammetry & Remote Sensing* 61 (2): 95-107 doi:10.1016/j.isprsjprs.2006.09.001 U.V. Helava Award 2006
- [A13] **Jutzi B**, Stilla U (2007) Simulation and analysis of full-waveform laser data of urban objects. *Remote sensing and data fusion on urban areas, URBAN 2007*. IEEE 07EX1577 (on CD) [ISBN 1-4244-0712-5]
- [A14] Stilla U, **Jutzi B** (2008) Waveform analysis for small-footprint pulsed laser systems. In: Shan J, Toth CK (Eds) *Topographic Laser Ranging and Scanning: Principles and Processing*. CRC Press, Boca Raton: 215-234
- [A15] Stilla U, **Jutzi B**, Reitberger J, Yao W, Krzystek P (2009) Full Waveform Laserscanning - Auswertemethoden und Anwendungen. *Terrestrisches Laserscanning (TLS2009)*, Schriftenreihe des DVW, Band 60: 49-67 (eingeladener Beitrag)

Extending the Range Measurement Capabilities of Modulated Range Imaging Devices by Time-Frequency-Multiplexing

Erweiterung des Entfernungsmessbereichs bei modulierten Entfernungskameras durch ein Zeit-Frequenz-Multiplexverfahren

Boris Jutzi

In diesem Beitrag wird ein Zeit-Frequenz-Multiplexverfahren zur Auflösung der Phasenmehrdeutigkeiten bei modulierten Entfernungskameras präsentiert, um den Entfernungsmessbereich zu erweitern. Es wird sowohl Phasen-Unwrapping anhand zweier Modulationsfrequenzen als auch ein Vertrauensmaß für die Entfernungsmessung vorgestellt. Für die Untersuchungen werden sowohl Innen- als auch Außenaufnahmen ausgewertet.

Die vielversprechenden Ergebnisse zeigen, dass Entfernungskameras nicht nur im sehr nahen Nahbereich eingesetzt werden können. An einem Beispiel wird gezeigt, dass der vom Hersteller spezifizierte Entfernungsmessbereich um das Vierfache erweitert werden kann, ohne Modifikationen am Sensor vornehmen zu müssen.

Schlüsselbegriffe: Entfernungsbild, Entfernungskamera, Mehrdeutigkeit, Phasen-Unwrapping, Modulationsfrequenz, Nahbereich

In this contribution a time-frequency-multiplexing method for unwrapping the range ambiguity of range imaging device is presented to extend the range measurement capabilities. Beside the phase unwrapping by multiple frequency modulations a confidence measure for the range measurement is proposed. For the investigations an indoor and an outdoor scene were analyzed. The results are promising to utilize range imaging devices not only in very close range. It will be shown that four times of the manufacturers non-ambiguity range specification could be reached without modifying the sensor or improving the illumination unit.

Keywords: Range imaging, RIM, ambiguous, phase-unwrapping, modulation frequency, close range

1 INTRODUCTION

The 3D geometry of the environment is of great interest for a wide variety of applications. In order to obtain a geometric description usually the captured image or range data is analyzed, where in general a high level of automation is desirable.

By utilizing passive imaging sensors the 3D information is gained by textured image data indirectly from several images with costly stereo- or multiple image analysis. These procedures are widely used, but have indispensable claims due to capturing disposition and scene contents. For instance, the illumination conditions should be adequate, the observed materials should be textured and opaque, and the distance between object and camera as well as between the camera viewpoints of stereo images should be sufficient large for gaining a reliable 3D reconstruction.

Beside this, the photogrammetric methods are complemented by active sensor procedures. For instance, a laser scanner captures a sequence of singular range values while accomplishing a time-dependent spatial scanning of the environment. In general spaceborne, airborne as well as terrestrial laser scanner sensors allow a direct and illumination-independent measurement of 3D objects [Shan & Toth, 2008]. For an accurate data acquisition necessarily the scene contents as well as the sensor platform should be static, otherwise a deformation of the environment can appear. In general, with an increasing dynamic of the scene contents respectively sensor platform, the complexity of the analysis increases and the exploitation of 3D information is more and more challenging. To gain 3D information from rapid dynamical processes the capturing of the environment at the same time is essential.

Very recently enhanced types of active imaging sensors have started to meet these requirements, e. g. MESA with the Swiss Ranger series and PMD Vision with the CamCube series. These close range sensors allow to capture a range image and a co-registered monochrome intensity image simultaneously with high repetition rate up to 100 releases per second. The spatial resolution can be up to 204×204 pixels. Beside this the non-ambiguity (sometimes called unique) range is currently about a few meters. In general the measured intensity strongly depends on the used wavelength (usually close infrared) of the laser source and the surface characteristic.

With these new types of sensors for the first time the basic principle to unify advantages of active sensors and the simultaneous capturing of an image for an extended area of dynamical 3D applications is given. Especially the 3D motion or deformation analysis, like autonomous navigation of robots, motion control for game consoles, trajectory tracking of pedestrians for surveying, or maker free 3D measurements of crash tests, are of interest. Beside the hardware and sensor developments [Lange, 2000], nowadays most works focus on geometric (e. g. [Boehm & Pattinson, 2010]; [Kahlmann *et al.*, 2006]) and radiometric calibration (e. g. [Lichti, 2008]) or tracking of objects or automatic extraction of object features.

The terminology for scannerless range imaging systems is multifarious, where the terms Time-of-Flight (TOF) depth camera, 3D range imager, Time-of-Flight Sensors, photonic mixer devices (PMD) [Schwarte *et al.*, 1997] or a combination of the mentioned terms are used. Most of the terms are much more related to the range measurement than to the as well available reflectivity measurement of the observed area. For the procedure the term range imaging with

the abbreviation RIM is more and more established, particularly in Europe.

Especially the relatively large noise influence on the measurement, due to the large amount of ambient radiation in comparison to the emitted radiation, results in a range measurement which is less reliable compared to the performance of airborne laser scanner (ALS) or terrestrial laser scanner (TLS). The major drawbacks of the known RIM devices are:

- an absolute range accuracy of a few centimeters
- a range ambiguity of a few meters

It has to be mentioned that the range ambiguity is closely related to the well-known phase unwrapping problem which is extensively discussed in the radar interferometry community. It is an inverse problem which cannot be solved in general and intensive research is going on this issue until today. To resolve the ambiguity by phase reconstruction various methods are known in literature. A general overview of the existing methods is given in Ghiglia & Pritt [1998], where most of these approaches deal with 2D data sets. By utilizing the Goldstein 2D unwrapping procedure on RIM data an image-based solution was proposed by Jutzi [2009]. However, one large drawback of the methods is the sensitivity of the phase reconstruction to minor measurement errors. Additionally, the reconstruction suffers from multiple integer solutions caused by the unwrapping procedure. Usually the measured environment is unknown and therefore, multiple integer solutions are possible if the topography contains large geometrical discontinuities.

Beside this, from other sensor systems different techniques are known to solve this problem in order to obtain a range non-ambiguity, e. g. by utilizing at least two different modulation frequencies as most continuous-wave (CW) modulated laser scanner and radar systems do or by (pseudo) random modulation. In general, for high modulation frequencies the range measurement shows a high accuracy and the ambiguity range is small, whereas for low modulation frequencies it is vice versa. Therefore, it is always a trade-off to select the best frequency to gain optimal results.

In this paper a method for unwrapping the range ambiguity of range imaging devices is presented to extend the range measurement capabilities. In Section 2 the methodology is proposed by an overview for the measurement principle, the utilized phase unwrapping by multiple frequency modulation, and a confidence measure for the range measurement. Section 3 shows a brief overview of the utilized range imaging sensor with the selected indoor scene and the selected outdoor scene. The detailed experiments and results for both scenes are presented in Section 4. Finally, the derived results are evaluated and discussed, the content of the entire paper is concluded, and an outlook is given.

2 METHODOLOGY

In the following the measurement principle (Section 2.1), the phase unwrapping by multiple frequency modulation (Section 2.2), and a confidence measure for the range measurement (Section 2.3) are specified.

2.1 Measurement principle

The range measurement can be briefly described as follows: A sinusoidal CW modulated signal is transmitted by a LED array in form of monochromatic light (Figure 1a). The emitted light travels to the object, is backscattered by the surface and captured by a receiver unit. The receiver unit is usually a CCD or CMOS array. Single pixels of the array can be subdivided into four collaborating subpixels (Figure 1b).

Concerning a demodulation of the sinusoidal received signal the parameters amplitude A and phase shift $\Delta\varphi$ can be determined. For each measurement per single pixel four neighborhood subpixels are utilized to measure by time gating four intensities with a relative phase shift of 90° , or with other words an absolute phase shift of 0° , 90° , 180° , and 270° (Figure 2a). For each absolute phase shift the corresponding intensity is determined by integration. Then the phase shift $\Delta\varphi$ between the transmitted and received signal can be determined by the intensity values A_0 , A_{90} , A_{180} , and A_{270} (Figure 2b) with

$$\Delta\varphi = \arctan\left(\frac{A_{270} - A_{90}}{A_0 - A_{180}}\right). \quad (1)$$

Based on the phase shift $\Delta\varphi$, the range ΔR to the object is given with respect to the two-way time of flight by

$$\Delta R = \frac{c}{2f_m} \frac{\Delta\varphi}{2\pi}, \quad (2)$$

where f_m is the modulation frequency and c the speed of light.

Unfortunately, the phase shift $\Delta\varphi$ is a wrapped phase and its corresponding range ΔR is ambiguous due to the measurement principle with the utilized modulation frequency. Hence the absolute range R to the object can not be determined directly if the real range is above the modulation range

$$R_m = \frac{c}{2f_m}. \quad (3)$$

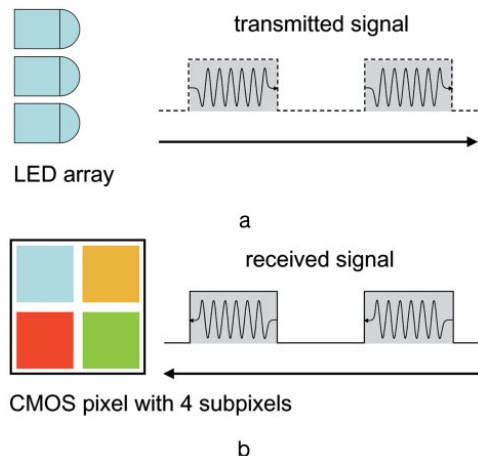


Fig. 1 | Function principle of the RIM: a) transmitter, b) receiver.

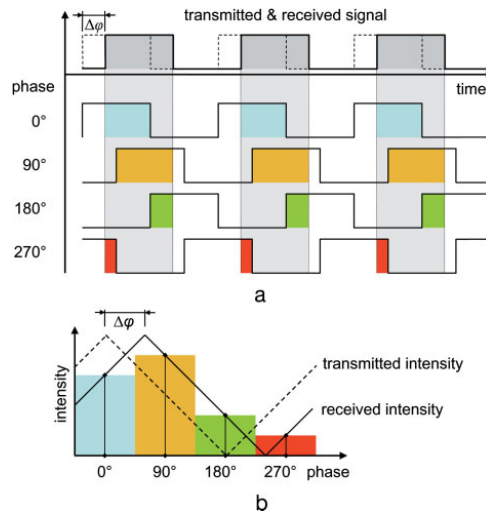


Fig. 2 | Measurement principle of the RIM sensor: a) utilizing time gating to measure the phase shift, b) determining intensity by integration.

Therefore, the unwrapped phase

$$\varphi = 2\pi k + \Delta\varphi \quad (4)$$

has to be known, where the number of periods $k = 0, 1, 2, \dots$ are integer valued. Based on this relationship the absolute range R can be denoted by

$$R = R_m k + \Delta R, \quad (5)$$

with the number of periods k multiplied by the modulation range R_m and added by the measured range ΔR .

2.2 Phase unwrapping by multiple frequency modulation

To resolve the ambiguity of the phase measurement two different modulation frequencies f_{m1} and f_{m2} with $f_{m1} < f_{m2}$ have to be available, which results in two modulation ranges R_{m1} and R_{m2} . Additionally, two conditions have to be satisfied with $k_1 = k_2$ or $k_1 + 1 = k_2$.

Then the range measurement can be extended to a maximum range

$$R_{max} = R_{m1} k_{1,max} = R_{m2} k_{2,max}, \quad (6)$$

with $k_{1,max} = f_1 / (f_2 - f_1)$ and $k_{2,max} = f_2 / (f_2 - f_1)$.

Due to the two conditions two cases A and B have to be considered:

Case A

If the measured ranges are $\Delta R_1 \leq \Delta R_2$ then $k_1 = k_2 = k$ and the absolute range is

$$R = R_{m1}k + \Delta R_1 = R_{m2}k + \Delta R_2, \quad (7)$$

with $k = (\Delta R_2 - \Delta R_1) / (R_{m1} - R_{m2})$.

Case B

If the measured ranges are $\Delta R_1 > \Delta R_2$ then $k_1 + 1 = k_2 = k$ and the absolute range is

$$R = R_{m1}(k - 1) + \Delta R_1 = R_{m2}k + \Delta R_2, \quad (8)$$

with $k = (\Delta R_2 - \Delta R_1 + R_{m1}) / (R_{m1} - R_{m2})$.

In general, for both cases the solution for k should be integer valued.

2.3 Confidence measure for the range measurement

Due to measurement inaccuracies small variations can be expected for the calculated number of periods k_q and therefore, non-integer values will be obtained. If the non-integer value is close to the integer value it can be assumed that the calculated k_q is reliable, if the variation is large the result is not reliable. This is of interest because due to the measurement principle in general for each pixel a range value is captured even if no surface was available. Usually the corresponding active intensity value to this range measurement should be small valued.

In order to avoid unreliable measurements it is obvious to introduce a confidence measure q for the calculated absolute range R . The confidence measure q within the interval $[0, 1]$ can be defined by

$$q = 1 - 2 |k_q - \text{nint}(k_q)|, \quad (9)$$

with $|\cdot|$ for the absolute value and $\text{nint}(\cdot)$ for the nearest integer.

3 CONFIGURATION

A RIM sensor (Section 3.1) was utilized to capture an indoor and an outdoor scene (Section 3.2).

3.1 RIM sensor

For the investigations, a PMD Vision CamCube 2.0 sensor was used. The sensor has a 204×204 pixel array with a pixel size and pitch (spacing) of about $45 \mu\text{m}$. The user can preselect the modulation frequency f_m with 18, 19, 20, and 21 MHz, which results in a modulation range R_m of 8.33, 7.89, 7.5, and 7.14 m. The maximum frame rate is about 25 frames per second and the sensor measures per pixel three features: range, active intensity and passive intensity. Therefore, above three million measurement values per second can be captured.

An example is depicted in *Figure 3*. For the preselected modulation frequencies f_1 18 MHz and f_2 21 MHz the range ambiguity is given by the modulation range $R_{m1}(f_1)$ 8.33m and $R_{m2}(f_2)$ 7.14m. With Formula eq:2.2 the maximum range R_{max} 50 m. Furthermore, the depicted difference between ΔR_2 and ΔR_1 helps to understand

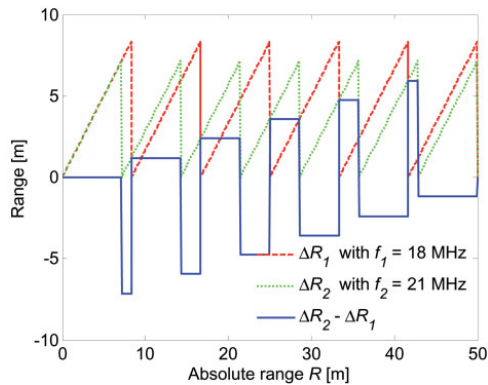


Fig. 3 | Absolute range R to the object compared with measured ranges ΔR for different frequencies (dashed red and dotted green line) and calculated range differences (solid blue line).

the two different cases of interest A and B with $\Delta R_1 \leq \Delta R_2$ as positive values and $\Delta R_1 > \Delta R_2$ as negative values. At the moment, to utilize two different modulation frequencies a temporal sequential capturing of frames by alternating modulation frequencies is realized by time-frequency-multiplexing.

3.2 Scene

A RIM data set of a static indoor and an outdoor scene was recorded by a stationary placed sensor. The photo of the observed scenes are depicted in Figure 4. For the environment no reference data concerning the radiometry or geometry was available.

4 EXPERIMENTS

For the indoor and outdoor experiments the modulation frequencies f_1 18 MHz and f_2 21 MHz for maximum frequency discrimination were selected. The integration time was pushed to the maximum of 40 ms

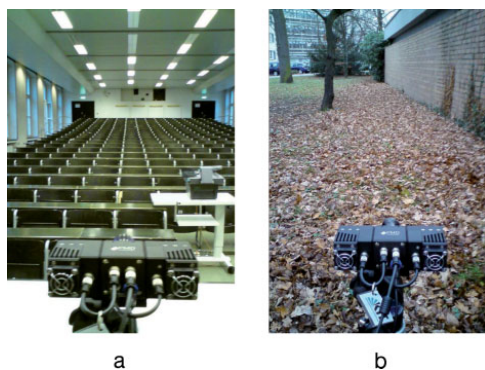


Fig. 4 | Scenes captured with the RIM sensor: a) indoor, b) outdoor.

to gain a high signal-to-noise ratio for the measurement. In this case, saturation could appear in close range or at object surfaces with high reflectivity. All measurement values were captured in raw mode. Only a single image without averaging is depicted in the following figures.

4.1 Indoor

The active range measurement of the indoor scene is only slightly influenced by additional sunlight illumination from outside and artificial lighting from the facility ceiling and therefore, a high signal-to-noise ratio is given. The two range images captured with different modulation frequencies f_1 18 MHz and f_2 21 MHz are depicted in Figure 5. Obviously the depth for modulation range R_{m1} is larger than for R_{m2} and range measurement inaccuracies can be observed, especially at the wrapping discontinuities.

Concerning the formulars in Section 2.2 the numbers of periods k_1 and k_2 can be estimated. The results for the number of periods k_2 are shown in Figure 6, where the estimated number of periods k are encoded by gray values. For the close-by ceiling and benches the estimated parameter k is close to zero. It can be observed that for larger number of periods the variations of the estimated parameter k only slightly increase. Large inconsistencies are visible at the dark colored and polished doors on the left and right side in the back of the room. Unreliable measurement values appear at the polished surfaces in the foreground mainly on the left side where the incidence angle to the surface is steep. These outliers occur due to the low reflectivity or specular surface characteristic which can result in multipath measurements. In general a non-integer value for k is not plausible (Figure 6a), therefore it was rounded to the nearest whole number (Figure 6b). Due to the size of the room for the indoor scene the

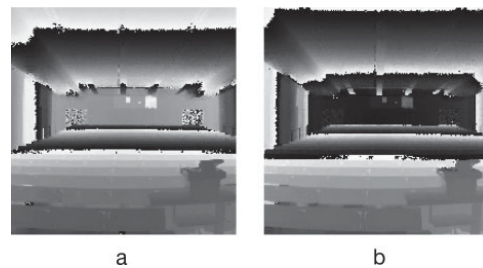


Fig. 5 | Range images captured with different modulation frequencies: a) f_1 18 MHz, b) f_2 21 MHz.

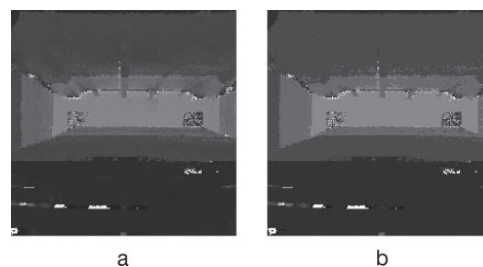


Fig. 6 | Estimated number of periods k_2 : a) non-integer, b) integer.

number of periods is $k_2 = [0, 3]$, which can be as well observed in Figure 6 by the four different gray values.

For the estimated number of periods in Figure 6 up to three different range images can be generated, one for the non-integer and two different ones for the integer case. For each case an example is depicted in Figure 7. Concerning the non-integer value of the range estimation the directly calculated absolute range value is equivalent with averaging the unwrapped range values of the 18 MHz and 21 MHz measurements (Figure 7a). As one would expect, the absolute range image calculated by the integer number of periods appears

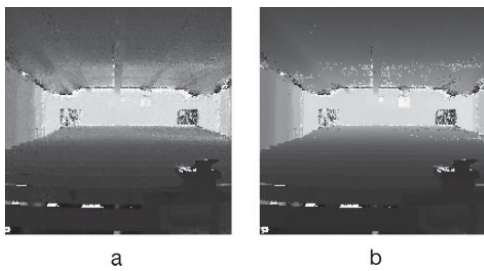


Fig. 7 | Unwrapped range images generated with different estimated number of periods parameters: a) non-integer number of periods, b) integer number of periods and 21 MHz measurements.

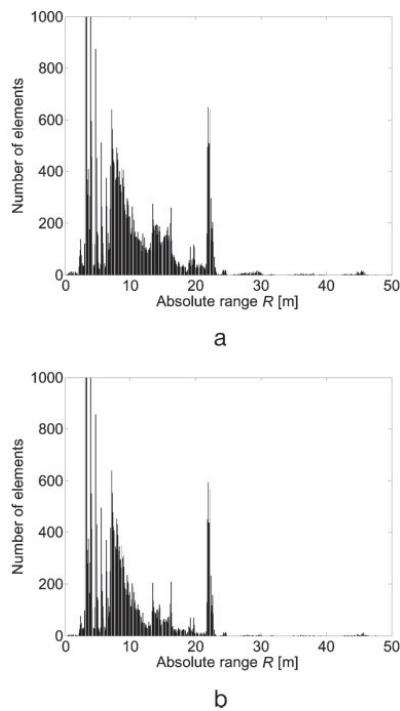


Fig. 8 | Histogram of the absolute range values: a) all estimate values, b) for values with a confidence measure above 0.75.

cleaner with less noise. Two different absolute range images can be calculated by utilizing the 18 MHz and 21 MHz measurements. In Figure 7b the 21 MHz measurement was selected for visualization. Due to the higher modulation frequency a more accurate range measurement can be expected. However it has to be stated that it was not goal of this investigation to validate and compare these two results by a reference measurement.

The unwrapped range values are spread over large distance, where it can be assumed that for large range values the reliability is lower due to the low signal-to-noise ratio. Figure 8a shows a histogram of the estimated absolute range values over the maximum range R_{max} , where most range values are below 23 m. Due to a maximum distance to the central wall at the back of the room of about 23 m, absolute range values above this distance are erroneous.

To evaluate the unwrapping procedure a confidence measure q was introduced with Formula 9 and the results for the scene are visualized as image with the corresponding histogram in Figure 9. Most absolute range values over the entire scene have a high reliability except at the far away ceiling and at the polished surfaces in front, where the incidence angle to the surface is steep. The histogram shows a widely spread distribution with a very high density close to 1, where the highest number of elements is above 0.9, but below 1. For the selected scene 77 % of the absolute range values have a confidence measure above 0.75. A sample of the remaining absolute range values above this empirical preselected threshold for the confidence measure is depicted in Figure 8b.

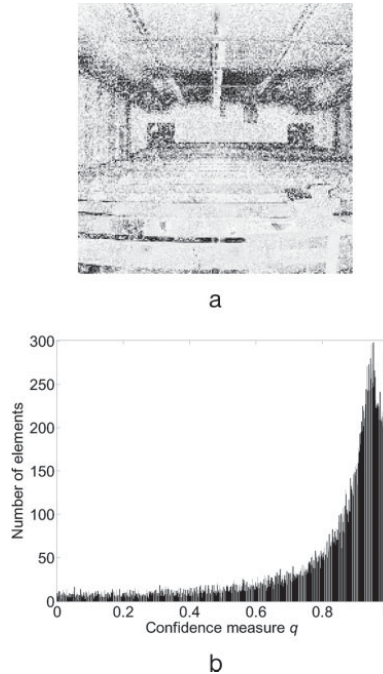


Fig. 9 | Confidence measure q : a) image-based visualization, b) corresponding histogram.

4.2 Outdoor

The outdoor scene provides a challenging measurement environment due to the additional influence of the background illumination by sunlight which decreases the signal-to-noise ratio of the active range measurement. The two range images captured with different modulation frequencies f_1 18 MHz and f_2 21 MHz are depicted in Figure 10, where the depth for modulation range R_{m1} is larger than for R_{m2} .

Again, the numbers of periods k_1 and k_2 are estimated and the results for the number of periods k_1 are shown in Figure 11. For the foreground of the image, which is obviously close-by, the estimated parameter k is close to zero. For larger number of periods the variations of the estimated parameter k increases. Again, the non-integer and integer value for k is depicted in Figure 11 for comparison purposes.

For the estimated number of periods Figure 12 shows the unwrapped range images generated with the non-integer number of periods and integer number of periods utilizing 21 MHz. Again, the range image calculated by the integer number of periods appears cleaner with less noise.

The unwrapped range values are spread over large distance, where it can be assumed that for large range values the reliability decreases. The black bars in Figure 13a show an histogram of the estimated absolute range values, where most range values are below 15 m.

The confidence measure q for the outdoor scene is visualized as image with the corresponding histogram in Figure 14. Again, it can be stated as it would be expected that the most reliable absolute range

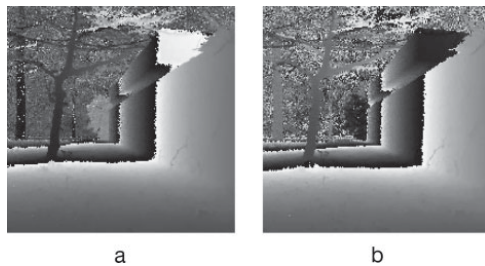


Fig. 10 | Range images captured with different modulation frequencies: a) f_1 18 MHz, b) f_2 21 MHz.

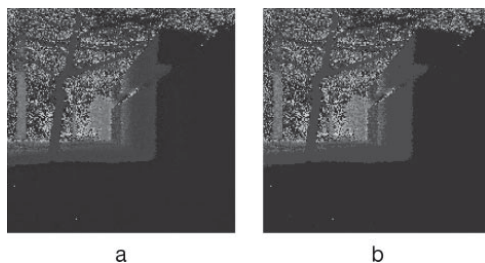


Fig. 11 | Estimated number of periods k_1 : a) non-integer, b) integer.

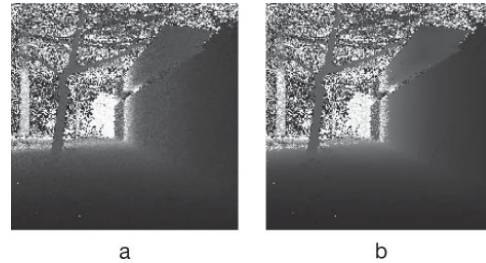


Fig. 12 | Unwrapped range images generated with different estimated number of periods: a) non-integer number of periods, b) integer number of periods and 21 MHz measurements.

values appear for the close range measurements. Behind the tree, approximately at a range of 10 m, the reliability decreases. The histogram shows a widely spread distribution with a high density close to 1, where the highest number of elements is above 0.9, but below 1. For the selected scene 65 % of the absolute range values have a confidence measure above 0.75. A sample of the remaining absolute range values above this empirical preselected threshold for the confidence measure is depicted in Figure 13b.

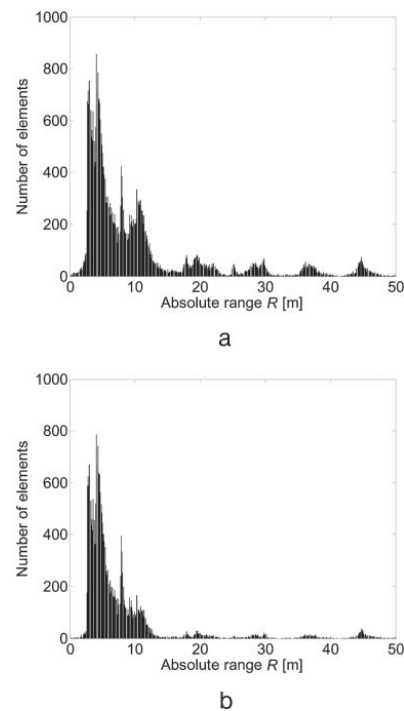


Fig. 13 | Histogram of the absolute range values: a) all estimate values, b) for values with a confidence measure above 0.75.

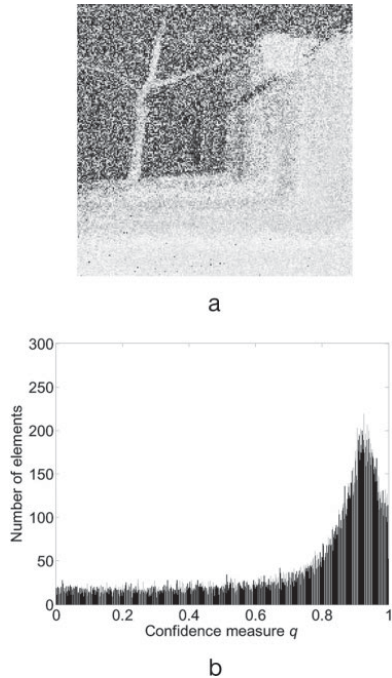


Fig. 14 | Confidence measure q : a) image-based visualization, b) corresponding histogram.

5 CONCLUSION

The goal of these investigations was to extend the measurement capabilities of the range. Therefore, an indoor and an outdoor scene were captured and analyzed. Even without increasing the illumination properties of the system the results are promising to utilize range imaging devices not only in very close range. Furthermore, the outdoor capability could be shown even when the results are not as good as for the indoor scene. However, the results show that four times of the manufacturers non-ambiguity range specification could be reached without modifying the sensor or improving the illumination unit, e. g. by additional illumination modules.

The proposed confidence measure is a useful parameter to evaluate the range measurement and gives evidence about the measurement. It could be shown that a large number of measurements have a high confidence measure, where the highest density is only slightly below 1 (Figure 9b & 14b). On the depicted histograms it looks like a systematic error occurs which might be given by a systematic offset between the 18 MHz and 21 MHz range measurements. Therefore, further investigations have to be done which prove this assumption. However, there are different possibilities to handle this inconsistency, e. g. one is to eliminate these outliers or another is to average the two measurements derived by the different modulation frequencies. This decision can be supported by sub-dividing the confidence measure in different classes.

Usually the range imaging devices have a low range accuracy and the range measurements are sensitive to the signal-to-noise ratio. A simple possibility to increase the measurement reliability is to stretch the integration time or, with other words, to average multiple recordings, but this is of course only helpful for a static scene and a stationary placed sensor.

In future, the extended absolute range values should be validated by resolution and accuracy. Beside the restrictive static scene and a stationary placed sensor, additional

investigations on more dynamic aspects should be focused. However, refined techniques might allow to gain the advantage of an extended range without losing much data capturing speed performance.

Acknowledgement

The author would like to thank Patrick Bradley and Martin Weinmann for many valuable discussions and Christian Teutsch for assistance during the measurement campaign.

References

- Boehm, J.; Pattinson, T. (2010): Accuracy of exterior orientation for a range camera. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 5), pp. 103–108.
- Ghiglia, D. C.; Pritt, M. D. (1998): *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. John Wiley & Sons: New York. Goldstein, R. M.; Zebker, H. A.; Werner, C. L. (1988). Satellite radar interferometry: two-dimensional phase unwrapping. *Radio Science*, 23, pp. 713–720.
- Jutzi, B. (2009): Investigations on ambiguity unwrapping of range images. In: Bretar, F.; Pierrot-Deseilligny, M.; Vosselman, G. (Eds.): *Laserscanning 2009*. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 3/W8), pp. 265–270.
- Kahlmann, T.; Remondino, F.; Ingensand, H. (2006): Calibration for increased accuracy of the range imaging camera SwissRanger. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 5), pp. 136–141.

Lange, R. (2000): 3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. PhD thesis, University of Siegen.

Lichti, D. D. (2008): Self-Calibration of a 3D Range Camera. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences* 37 (Part B5), pp. 927–932.

Schwarte, R.; Xu, Z.; Heinol, H.-G.; Oik, J.; Klein, R.; Buxbaum, B.; Fischer, H.; Schulte, J. (1997): New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). In: Loffeld, O. (Ed.): *3D Sensors and 3D Imaging*, SPIE Proceedings Vol. 3100, pp. 245–254.

Shan, J.; Toth, C. K. (Eds.) (2008): *Topographic Laser Ranging and Scanning: Principles and Processing*. Boca Raton, FL: Taylor & Francis.

Dr.-Ing. Boris Jutzi

INSTITUTE OF PHOTOGRAMMETRY
AND REMOTE SENSING (IPF)
KARLSRUHE INSTITUTE OF TECHNOLOGY
(KIT)

Kaiserstr. 12 | 76128 Karlsruhe
E-Mail: boris.jutzi@kit.edu



Manuskript eingereicht: 22.12.2010 | Im Peer-Review-Verfahren begutachtet

INVESTIGATIONS ON AMBIGUITY UNWRAPPING OF RANGE IMAGES

B. Jutzi

Institute of Photogrammetry and Remote Sensing, Universität Karlsruhe
Kaiserstr. 12, 76128 Karlsruhe, Germany
boris.jutzi@ipf.uni-karlsruhe.de

KEY WORDS: Range imaging, RIM, ambiguous, phase-unwrapping, modulation frequency, close range.

ABSTRACT:

For the first time the basic principle to unify advantages between active sensors and the simultaneous capturing of an image for an extended area of dynamical 3D applications in close range is given by range imaging (RIM) sensors. The drawback of data which is captured with RIM sensors is the absolute range accuracy and the limited non-ambiguity range. From other sensor systems different techniques are known to solve this problem in order to obtain a non-ambiguity range, e.g. by utilizing different modulation frequencies as most continuous-wave (CW) modulated laser scanner and radar systems do or by (pseudo) random modulation. In this paper a post-processing task is presented. The Goldstein 2D unwrapping procedure for unwrapping the range ambiguities of ranging sensors (e.g. RIM sensors or CW-modulated laser scanners) is proposed considering residues, branch cuts and tree estimation strategies and additionally confidence criteria. It could be shown that a range restoration for numerous periods of the ambiguity range is in principle possible with the presented 2D unwrapping procedure.

1. INTRODUCTION

Currently the geometrical 3D capturing and description of the environment are based on image or range data. By utilizing passive imaging sensors the 3D information is gained by textured image data indirectly from several images with costly stereo- or multiple image analysis. These procedures are widely used, but have indispensable claims due to capturing disposition and scene contents. For instance the illumination conditions should be adequate, the observed materials should be textured and opaque, and the distance between object and camera as well as between the camera viewpoints of stereo images should be sufficient large enough for gaining a reliable 3D reconstruction.

Beside this the photogrammetric methods are complemented by laser scanner procedures. These active sensors capture a sequence of singular range values while accomplishing a time dependent spatial scanning of the environment. Beside these basic range measurements the current commercial airborne laser scanner (ALS) developments allow to record the amplitude or the waveform of the backscattered laser pulse (Jutzi & Stilla, 2006). Therefore, laser scanner systems like OPTECH ALTM 3100, TOPEYE MK II, and TOPOSYS HARRIER 56 can be used. The latter system is based on the RIEGL LMS-Q560. More and more waveform capturing scanners are available at the moment, e.g. RIEGL - one of the leading companies for laser scanners - already offers several scanners (LMS-Q560, LMS-Q680, and VQ-480). In general spaceborne, airborne as well as terrestrial laser scanner sensors allow a direct and illumination-independent measurement of 3d objects (Shan & Toth, 2008).

For an accurate data acquisition necessarily the scene contents as well as the sensor platform should be static, otherwise a deformation of the environment can appear. In general with an increasing dynamic of the scene contents respectively sensor platform the complexity of the analysis increases and the exploitation of 3D information is more and more challenging. To gain three-dimensional information from rapid dynamical processes the capturing of the environment at the same time is essential. Very recently enhanced types of active imaging sensors have started to meet these requirements, namely the Swiss Ranger (www.mesa-imaging.ch), the PMD Vision (www.pmdtec.com), and the O3D series (www.ifm.de). These close range sensors (Table 1) allow to capture an range image and a co-registered intensity image simultaneously with high

repetition rate (up to 100 releases per second). The non-ambiguity range is currently below 7.5m and depends on the modulation frequency. The measured intensity strongly depends on the used wavelength (usually close infrared) of the laser source and the surface characteristic.

For the first time the basic principle to unify advantages of active sensors and the simultaneous capturing of an image for an extended area of dynamical 3D applications is given. Especially the 3D monitoring in close range with airborne and terrestrial platforms in problematic weather and illumination conditions or at night is promising with this novel technology. Therefore different applications are building surveillance, traffic monitoring, and driver assistance. Beside this, the 3D motion or deformation analysis, like autonomous navigation of robots, trajectory tracking of pedestrians for surveying, or maker free 3D measurements of crash tests, are of interest.

Another technical advantage is the monostatic sensor configuration, which allows to observe the area of interest from a single point of view, in contrast to the classical stereo observation techniques with passive sensors, which need at least two different viewpoints. The mayor drawbacks are the limited non-ambiguity range and the absolute range accuracy of a few centimeters. Especially the relatively large noise influence on the measurement, due to the large amount of ambient radiation in comparison to the emitted radiation, results in a range measurement which is less reliable compared to the performance of airborne or terrestrial laser scanner (TLS). However concerning the technical progress, most limitations will be overcome soon and in close future systems with expanded operating range and improved image size will be available.

The terminology for scannerless range imaging systems is multifarious, the terms Time-of-Flight (TOF) depth camera, 3D range imager, Time-of-Flight Sensors, photonic mixer devices (PMD; Schwarte, 1997), or often a combination of the mentioned terms are used. Unfortunately, most of the terms are much more related to the range measurement than on the as well available gray value measurement of the observed area. For the procedure the term range imaging with the abbreviation RIM is more and more established, especially in Europe.

Various studies about range imaging have been published in the literature dealing with different interests. Beside the hardware and sensor developments (Lange, 2000), nowadays most works focus on geometric and radiometric calibration:

- Reulke (2006) introduced a geometrical calibration and fused the intensity image derived by the range imaging sensor with high resolution RGB data to improve the texturing of surfaces.
- Kahlmann et al. (2007) focused on the geometric calibration of range imaging sensors and developed a tracking of moving objects (people) approach based on recursive Bayesian filter.
- Lichti (2008) proposed a method for the self-calibration by bundle adjustment of range imaging sensors which allows a simultaneous calibration concerning the spatial distortions and the ranging system.

Other works focused on tracking of objects and automatic extraction of object features:

- For the tracking of human motion and interaction within a range image sequence, Westfeld & Hempel (2008) suggested the combination of complementary radiometric and geometric information to increase accuracy and reliability.
- For a moving range imaging Karel et al. (2007) specified an automatic object segmentation sensor based on a fast minimum covariance determinant approach and evaluated statistically the quality of the data.
- Kim et al. (2008) proposed to utilize more than one synchronized range imaging system to gain multi views for the reconstruction of dynamic scenes.

As briefly mentioned above one drawback of the RIM sensors is the limited non-ambiguity range. From other sensor systems different techniques are known to solve this problem in order to obtain a non-ambiguity range, e.g. by utilizing different modulation frequencies as most continuous-wave (CW) modulated laser scanner and radar systems do or by (pseudo) random modulation.

In this paper a post-processing task is investigated in contrast to the above mentioned and not yet for RIM sensors available technical improvements. It has to be mentioned that the ambiguous range subject is close related to the well known phase unwrapping problem which is extensively discussed in the radar interferometry community. This inverse problem cannot be solved in general and intensive research is going on this issue until today. For instance one large drawback is the sensitivity of the phase reconstruction to minor measurement errors. Additionally, the reconstruction suffers from multiple

integer solutions caused by the unwrapping procedure. Usually the measured environment is unknown and, therefore, multiple integer solutions are possible, if the topography contains large geometrical discontinuities.

In this paper a method for unwrapping the range ambiguities of range imaging sensors is proposed. In Section 2 the measurement principle of range imaging sensors, a 1D straight forward and the Goldstein 2D unwrapping procedure are introduced. In Section 3 the sensor and scene configuration is presented and in Section 4 the data are examined. The analysis by the mentioned unwrapping procedures is described in Section 5 considering residues, branch cuts and tree estimation strategies, and the confidence criteria. Finally, the derived results are evaluated and discussed, the content of the entire paper is concluded, and an outlook is given.

2. METHODOLOGY

2.1 Measurement principle

The range measurement can be briefly described as follows: First a sinusoidal CW modulated signal is transmitted by a LED array in form of monochromatic light. The emitted light travels to the object, is backscattered by the illuminated surface, and captured by a receiver array (usually CCD or CMOS arrays). Concerning a demodulation of the sinusoidal received signal the parameters amplitude A and phase φ can be determined. For each measurement four neighborhood pixels are utilized to measure the four received intensities with a relative phase shift of 90° , or with other words an absolute phase shift of 0° , 90° , 180° , and 270° . Then the phase shift $\Delta\varphi$ between the transmitted and received signal can be determined by the intensity values A_0 , A_{90} , A_{180} , and A_{270} , with

$$\Delta\varphi = \arctan\left(\frac{A_{270} - A_{90}}{A_0 - A_{180}}\right). \quad (1)$$

Based on the phase shift $\Delta\varphi$ the range R to the object is given with respect to the two-way time of flight by

$$R = \frac{c}{2f_m} \frac{\Delta\varphi}{2\pi}, \quad (2)$$

where f_m is the modulation frequency and c the speed of light.

2.2 Data characteristic & feature convention

The range ambiguity ΔR can be denoted by

Type	MESA Swiss Ranger		PMD [Vision]		
	SR-3000	SR-4000	O3 ^a	S3 ^a	CamCube 2.0
URL	www.mesa-imaging.ch		www.pmdtec.com		
Image size	176x144	176x144	64x48	64x48	204x204
Focal length [mm]	8	10	TBD	-	12.8
Field of View (FoV) [°]	47.5x39.6	43.6x34.6	40x30	40x30	40x40
Pixel size [µm]	40x40	40x40	100x100	100x100	TBD
Wavelength [nm]	850	850	850	850	870
Power (optical) [W]	≤1	≤1	ca. 1	ca. 4	TBD
Frame rate [1/s]	max. 15-20	max. 54	max. 25	max. 20	max. 25
Modulation frequency [MHz]	20	29, 30, 31	20	20	variable
Non-ambiguity range [m]	7.5	5	7.5	7.5	7.5
Size [mm]	60x50x65	65x65x68	60x42x54	120x75x95	60x187x60
Outdoor feasibility	no	yes ^b	yes ^b	yes ^b	yes ^b

Table 1. Specification overview of selected range imaging sensors: MESA Swiss Ranger series (www.mesa-imaging.ch) and PMD Vision series (www.pmdtec.com). URLs accessed on June 2009.

^aPMD [Vision] O3/S3 are equivalent to IFM (www.ifm.de) O3D100/ O3D200, ^bSuppression of background illumination.

$$\Delta R = \frac{c}{2f_m} . \quad (3)$$

To estimate the absolute range R for the ambiguity an integer k is multiplied with the range ambiguity ΔR

$$R = \Delta R k . \quad (4)$$

To resolve the range ambiguity various methods are known in literature. A general overview of the existing methods is given in Ghiglia & Pritt (1998). Most of these approaches deal with 2D data sets. In contrast to this, the capturing of a scene with a RIM system delivers a 3D data set composed out of voxels with two spatial coordinates x, y and one time coordinate t

$$Q(x, y, t) , \quad (5)$$

where for each voxel different features are measured, e.g. *range* R , *intensity* I , and *confidence-of-the-measurement* C . Then for each feature a single data cube is given by Q_R , Q_I , and Q_C .

In the following, a stationary sensor setup is assumed for observing a high dynamical temporal and spatial scene. Two methods were examined and a brief overview is given in the following Sections.

2.3 1D straight forward unwrapping

With this straight forward approach a separability of the data set is assumed based on a sequential accomplished 1D unwrapping. Additionally, a mask, which is available from the data cube Q_C with the feature *confidence-of-the-measurement*, can be used to mask out unreliable voxels.

The drawback of this approach is that only a single voxel in the neighborhood of the six connected voxels (joint faces connection) is considered and the result depends on the processing direction and order. Therefore the topology is principally ignored and due to this, the 1D processing causes an erroneous unwrapping which results in a striped pattern. This inadequate approach was implemented mainly for comparison purposes and to visualize the problematic of ambiguity range unwrapping.

2.4 Goldstein 2D unwrapping

The Goldstein approach is described in detail in various publications, e.g. Goldstein et al. 1988. Originally it was developed for phase unwrapping in radar interferometry. The suggested solutions to reconstruct the unknown phase can be analogue transferred to the ambiguous range problematic. A brief description of unwrapping the ambiguous range will be given in this section.

The goal of unwrapping is to find integers k which can be added to the measured values to gain a continuous representation. The measured values are within a cycle of zero and the non-ambiguity value. In general phase unwrapping approaches are based on processing the changes between the pixels or respective voxels in the direct neighborhood by gradient calculation. Then the values are integrated by predefined rules and finally, if a discontinuity is detected, the most likelihood integer solution for unwrapping is added. To get a reasonable solution it is essential to find an optimal integration path for the gradient.

The unwrapping procedure is highly over determined. Therefore, different constraints have to be assumed. The key assumption is moderate changes within the neighborhood with relative changes below the ambiguity value. Values above are

called discontinuities and have to be bypassed by the restoration procedure. The discontinuities can be reduced to inconsistencies within the range values, so-called residues. Residues are given if the sum of four neighborhood pixels calculated in circular direction is unequal to zero. This procedure is path dependent and further it is very sensitive to noise.

The identified residues are connected to generate so-called branch cuts. Usually the length of the branch cuts (distance between the residues) should be as short as possible. The idea behind the branch cut is to find close by negative and positive residues (sometimes called dipoles) which can be compensate by each other if the total charge along the branch cut is zero. If the total charge is nonzero, the search continues for additional close by residues. Each associated residue is connected to the tree by means of a branch cut and the total charge is calculated. If the total charge is zero the tree is considered. The disadvantage of this procedure is that coordinates (position) of the branch cuts are arbitrarily chosen leaving out important context information. A better solution might be to utilize a more expensive approach, like for instance the Mask-cut-algorithm, which take into account the quality concerning the position of the selected cuts.

For the continuative search of associated residues, regardless if the residues have been previously assigned, they are added to a new tree. This results in a dendritic form of the branch cuts. Finally, the derived tree has to be bypassed for the integration calculation to utilize the unwrapping procedure.

This procedure can be additionally supported by the *confidence-of-the-measurement* to mask out unreliable range values.

3. CONFIGURATION

3.1 Sensor

For the investigations a Swiss Ranger SR-4000 sensor was used with the specifications listed in Table 1. The sensor has a 176 x 144 pixel array with a pixel size and pitch (spacing) of about 40 μm . The user can preselect the modulation frequency with 29, 30, and 31 MHz, which results in a maximum non-ambiguity range of 5.17, 5.00, and 4.84 m. The maximum frame rate is about 50 frames per second. Therefore, the number of 3D points measured by a range imaging system is above one million points per second which is equivalent to the current point capturing rate of the fastest close range laser scanners.

3.2 Scene

A range image sequence of an indoor scene was recorded by a stationary placed sensor. 100 frames were captured with a frame rate of 12 frames per second while the person was moving in direction to the sensor within a furnished room. A single RGB image of the observed scene is depicted in Figure 1. For the environment no reference data concerning the radiometry or geometry were available.



Figure 1. RGB image of the observed indoor scene.

4. DATA EXAMINATION

To depict the neighborhood relations of the feature *range*, the data cube was sliced in different directions: equivalent to the captured frames in y - x slices, vertically in t - y slices, and horizontally in x - t slices. The same procedure was done for the feature *intensity*. Figure 2 shows a set of images for the different slices in the space-space and space-time domain with the corresponding range and intensity images. The images have been normalized for visualization purposes and range and intensity values are depicted as gray values.

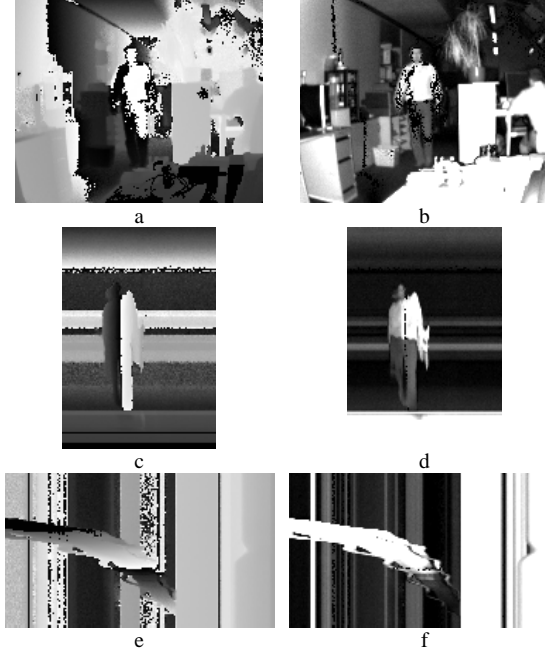


Figure 2. Corresponding *range* and *intensity* images differently orientated: a&b) x - y , c&d) t - y , and e&f) x - t slice.

For the measurements a modulation frequency of 29 MHz was selected which results in a range ambiguity of $\Delta R=5.17$ m. The range ambiguity is below the extension of the room. Figure 2a obviously shows several range ambiguity crossings. The discontinuities of the gray values can be seen by comparing them with the continuous appearance of the gray values of the intensity in Figure 2b, e.g. on the plan wall on left side. Further the range and slope dependent measured intensity values, which decrease with increasing range, are noticeable on the wall and on the ceiling. The intensity I can be normalized by the corresponding range r with

$$I \propto \frac{1}{r^2}. \quad (6)$$

Due to the stationary sensor setup the t - y and x - t slices contain mainly a stripy pattern, which is typical for a static scene. This pattern is interrupted by a specific representation of the dynamic procedures within the scene, which can be recognized as motion area. In Figure 2c-f the moving person is visible, but the representation is obviously deformed. In Figure 2c&e within the motion area the black to white crossover of the regions shows an object (person) crossing the ambiguity range during the measurements.

Beside the feature *range* and *intensity* the *confidence-of-the-measurement* is available. In Figure 3a the quality of the measurement is coded by gray values, dark values for a low confidence (bad) and bright values for a high confidence (good). Obviously the measurements close to the ambiguity range appear with a bad signal-to-noise ratio and, therefore, the confidence is low. Furthermore, the quality of the measurement is range dependent, measurements in far range are less reliable than in close range. This can be observed for instance at the wall on the left side of the room. Further if the measured intensity is above the dynamic range, the receiver saturates and the measurement is unusable. The statistic of the quality for the investigated slice is depicted in Figure 3b. It can be seen that most of the measurements are reliable, but about 11% of the measurements are unreliable, where the *confidence-of-the-measurement* C is equivalent to 0.

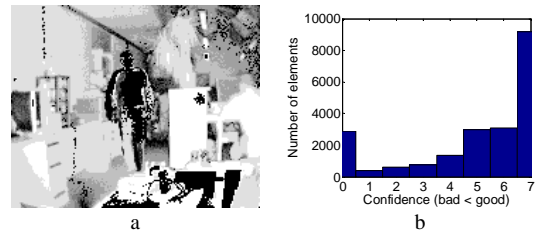


Figure 3. *Confidence-of-the-measurement* image (a) and corresponding histogram (b) for x - y slice.

5. ANALYSIS RESULTS

5.1 1D straight forward unwrapping

The 1D straight forward unwrapping does not take into account all neighborhood relations. An example for the unwrapping procedure is depicted, where in Figure 4a the original ambiguous range (dashed red line), the ambiguity range (dotted green line), and the 1D unwrapped range (solid blue line) are shown. Furthermore, the corresponding intensity values are presented in Figure 4b. Comparing the characteristic of the ambiguous range with the intensity values, the unreliability of the measured values is obvious (e.g. at pixel coordinate 25). At this range the intensity values are noisy compared to the nearby values. After unwrapping the dynamic range interval implies about four periods of the range ambiguity ΔR in the depicted example.

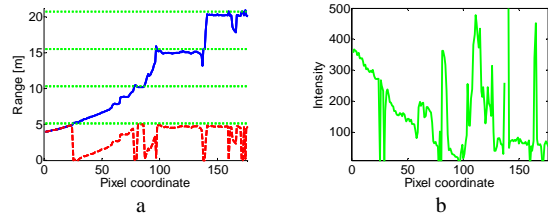


Figure 4. 1D examples for the characteristic of corresponding range and intensity values of a single row: a) Original ambiguous range (dashed red line), ambiguity range (dotted green line), and 1D unwrapped range (solid blue line), b) intensity (solid green line).

Utilizing this approach on the x - y slices it can be shown that the derived results depend on the unwrapping direction due to the discontinuities of the range values. In Figure 5 the results for different processing directions are presented. Most of the

failures are induced by unreliable measurements, which will be demonstrated in the following Section.

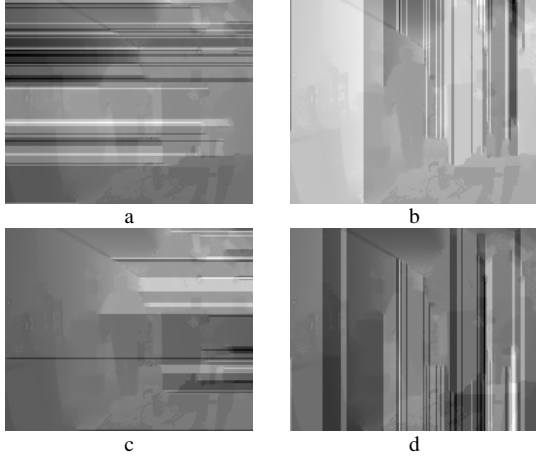


Figure 5. 1D straight forward unwrapping results for different processing directions: a) Right to left, b) bottom to top, c) left to right, d) top to bottom.

5.2 Residues, branch cuts and tree estimation

First, all residues are calculated for all x - y slices of the data cube Q_R . In Figure 6a the negative and positive residues (black and white colored pixels) for the x - y slice in Figure 2a are depicted. All in all 64 residues were determined. To proof the reliability of the residues for each residue the corresponding *confidence-of-the-measurement* voxel is extracted. A histogram with the statistic for the reliability of all residues is depicted in Figure 6b. It clearly shows that most of the determined residues base on an unreliable measurement, as 26 residues have the *confidence-of-the-measurement* 0 and, therefore, they are unreliable. This coincides with the already mentioned statement that the procedure to calculate residues is very sensitive to noise.

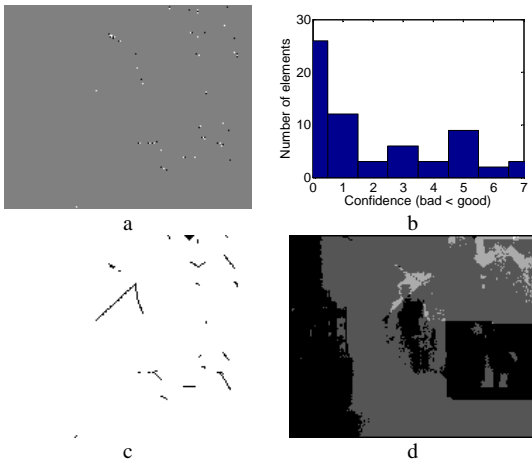


Figure 6. Intermediate results: a) Extracted residues, b) reliability of the residues, c) branch cuts and trees, d) number of gained range offsets ΔR .

In Figure 6c all connected residues in form of branch cuts are shown, where all branch cuts are linked together to the final trees. For the following integration procedure to unwrap the

range areas it is interdicted to cross the branch cuts. The number of gained range offsets ΔR coded by integer values are depicted in Figure 6d. For the black areas the number is zero, this denotes the original range is already non-ambiguous. The gray areas show that ambiguous range areas have been detected and solutions up to a range ambiguity of four periods (bright gray areas) could be determined.

Concerning the residues on the afore mentioned 1D straight forward unwrapping approach in Section 5.1 it can be easily shown with Figure 7 that the remaining discontinuities derived by the erroneous range unwrapping have their origin at the coordinates of the residues (black pixel).

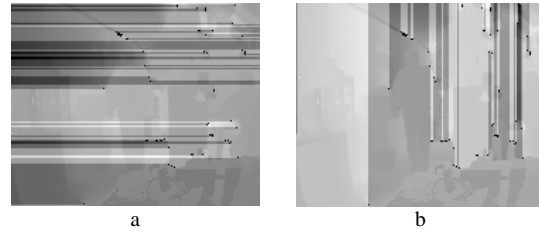


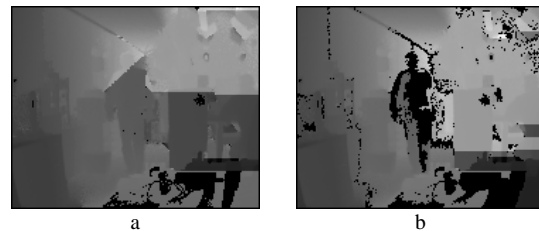
Figure 7. Visualization of residues (black pixel) and 1D straight forward unwrapping results of Figure 5a&b.

5.3 Goldman 2D unwrapping concerning the confidence criteria

The given data cube Q_C (*confidence-of-the-measurement*) provides information about the quality of the measurement, where the ranking of the quality is within the interval 0-7 and the values start from unreliable (value 0) up to excellent (value 7). A histogram for a single x - y slice is shown in Figure 3.

In the following the quality of the measurement on the reconstruction of the absolute range is investigated. Therefore, different tests were carried out to verify the influence of the quality of the measurement on the reconstruction of the absolute range. Overall seven tests were carried out by utilizing the confidence masks to verify the influence of the quality of the measurement on the reconstruction for the absolute range by the 2D unwrapping process. Only the reliable ambiguity range values above a given threshold are considered for processing. According to this the low quality measurements have been rejected.

Selected results are depicted in Figure 8. In close range, below two periods of the ambiguity range and with modest discontinuities the range unwrapping is reliable (e.g. the wall on the left side). Erroneous range unwrapping can be observed in all four results, where most of the wrong unwrapping range values appear at far range, above two periods of the ambiguity range. At this range the data quality is poor and the range values might be noisy. In addition to this, numerous discontinuities are at this range. Furthermore, it has to be mention that the results mainly depend on the seed point initialization and on the connectivity of the segments.



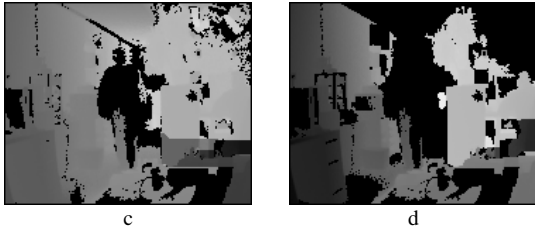


Figure 8. Unwrapping results considering confidence masks: a) Interval 0-7 (equivalent to all data processing), b) interval 2-7, c) interval 4-7, and d) interval 6-7.

5.4 Range dependent intensity normalization

Finally, the range dependent correction of the measured intensity is calculated by utilizing Formula 4. The measured intensity (Figure 2b) can be compared with the range corrected intensity (Figure 9a). The derived intensity for the same material (e.g. wall on the left) is equalized over the complete range area. Of a selected 1D example (dotted white line in Figure 9a) the corresponding intensity values (solid blue line) are depicted in Figure 9b. For comparison purposes the measured intensity (solid green line) and the unwrapped range corrected intensity (dashed red line) is shown. Still some artifacts are remaining from unreliable pixels at the 1D example and within the intensity image (e.g. at the person).

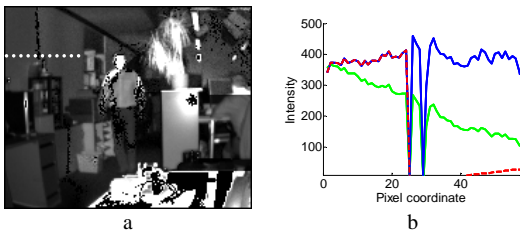


Figure 9. Range corrected intensity. a) Image, b) 1D example.

6. EVALUATION AND DISCUSSION

The evaluation of the results is difficult because geometric and radiometric reference data are not available. Therefore, the evaluation was performed by visual criteria. It could be observed that the range unwrapping sometimes fails, where most of the wrong range values appear at far range, above two periods of the ambiguity range, because at this range the data quality is poor and the range values might be noisy. However, in general an improvement could be gained.

Obviously the main disadvantage of the Goldman 2D unwrapping approach is the way the branch cuts are determined because they were selected by the criteria to be as short as possible (Figure 6c) and they do not rely on topographical aspects. This should be improved by more expensive approaches, like e.g. the Mask-cut-algorithm, which take into account the quality concerning the position of the selected cuts. In general the reconstruction suffers from multiple integer solutions if the topography contains large geometrical discontinuities.

By considering the confidence-of-the-measurement the result could be further improved, but in this case it is always a trade of between incomplete and erroneous results.

7. CONCLUSION AND OUTLOOK

It could be shown that a range restoration for numerous periods of the ambiguity range is in principle possible with the presented 2D unwrapping procedures. For future work the confidence-of-the-measurement might be a reliable basis for a quality guided unwrapping approach. Furthermore, due to the availability of a 3D data set, a 3D unwrapping procedure might be promising. Beside these approaches which mainly base on a single sensor system, the utilization of more than one synchronized range imaging system to gain multi views might be of interest to solve the range unwrapping problem.

ACKNOWLEDGEMENT

The author would like to thank the members of MESA Imaging for assistance during the measurements.

REFERENCES

- Ghiglia, D. C., Pritt, M. D., 1998. Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software. John Wiley & Sons: New York.
- Goldstein, R. M., Zebker, H. A. Werner, C. L., 1988. Satellite radar interferometry: two-dimensional phase unwrapping. *Radio Science*, 23, pp. 713-720.
- Jutzi, B., Stilla, U., 2006. Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS Journal of Photogrammetry & Remote Sensing* 61 (2): pp. 95-107. [doi:10.1016/j.isprsjprs.2006.09.001]
- Kahlmann, T., Remondino, F., Guillaume, S., 2007. Range imaging technology: new developments and applications for people identification and tracking. In: Beraldin, J.-A., Remondino, F., Shortis, M. R. (Eds.) *Videometrics IX*, SPIE Proceedings Vol. 6491, 64910C.
- Karel, W., Dorninger, P., Pfeifer, N., 2007. In Situ Determination Of Range Camera Quality Parameters By Segmentation. In: Gruen A, Kahmen H (Eds.) *Optical 3-D Measurement Techniques VIII*, pp. 109-116.
- Kim, Y. M., Chan, D., Theobalt, C., Thrun, S., 2008. Design and calibration of a multi-view TOF sensor fusion system. In *Proceedings of IEEE CVPR Workshop on Time-of-flight Computer Vision* 23-28 June 2008, pp. 1-7.
- Lange, R., 2000. 3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. PhD thesis, University of Siegen.
- Lichti, D. D., 2008. Self-Calibration of a 3D Range Camera. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences* 37 (Part B5), pp. 927-932.
- Reulke, R., 2006. Combination of distance data with high resolution images. In: Maas, H.-G., Schneider, D. (Eds.) *ISPRS Commission V Symposium: Image Engineering and Vision Metrology*, International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences 36 (Part B).
- Schwarte, R., Xu, Z., Heinol, H.-G., Olk, J., Klein, R., Buxbaum, B., Fischer, H., Schulte, J., 1997. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). In: Loffeld, O. (Ed.) *3D Sensors and 3D Imaging*, SPIE Proceedings Vol. 3100, pp. 245-254.
- Shan, J., Toth, C.K., (Eds.) 2008. *Topographic Laser Ranging and Scanning: Principles and Processing*. Boca Raton, FL: Taylor & Francis.
- Westfeld, P., Hempel, R., 2008. Range Image Sequence Analysis by 2.5-D Least Squares Tracking with Variance Component Estimation and Robust Variance Covariance Matrix Estimation. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences* 37 (Part B5), pp. 933-938.

Investigations on Surface Reflection Models for Intensity Normalization in Airborne Laser Scanning (ALS) Data

Boris Jutzi and Hermann Gross

Abstract

The analysis of laser scanner data is of great interest for gaining geospatial information. Especially for segmentation, classification, or visualization purposes, the intensity measured with a laser scanner device can be helpful. For automatic intensity normalization, various aspects are of concern, like beam divergence and atmospheric attenuation, both depending on the range. Additionally, the intensity is influenced by the incidence angle between beam propagation direction and surface orientation. To gain the surface orientation, the eigenvectors of the covariance matrix for object points within a nearby environment are determined. After normalization the intensity does no longer depend on the incidence angle and is influenced by the material of the surface only. For surface reflection modeling, (a) the Lambertian, (b) the extended Lambertian, and (c) the Phong reflection model are introduced, to consider diffuse and specular backscattering characteristics of the surface. An airborne measurement campaign was carried out to investigate the influences of the incidence angle on the measured intensity. For investigations, 17 urban areas, such as traffic, building, and vegetation regions were studied and the derived improvements are depicted. The investigation shows that large intensity variation caused by the object surface orientation and the distance between sensor and object can be normalized by utilizing the standard Lambertian reflection model.

Introduction

The processing of laser scanner data is of great interest to gain geospatial information. Spaceborne, airborne as well as terrestrial laser scanners allow a direct and illumination independent measurement of 3D objects in a fast, contact free, and accurate way (Shan and Toth, 2008). Besides the classical range measurements, the current generation of airborne laser scanners allows to record the amplitude or the complete waveform of the backscattered laser pulse. Various laser scanners and systems provide nowadays these capabilities, such as the OPTECH ALTM 3100, TOPEYE MK II, and RIEGL LMS-Q560. More and more waveform capturing scanners are available and tend to be state-of-the-art, e.g., RIEGL, one of

the leading companies for laser scanners, already offers several small-footprint full-waveform laser scanners like the LMS-Q560, LMS-Q680, and VQ-480.

The received waveform of a backscattered laser pulse depends on the absolute range, range variation, and reflectance of the illuminated surface. To interpret the received waveform, a fundamental understanding of the physical background of pulse propagation and surface interaction is important. The waveform of each laser pulse can be described by a distributed series of range values combined with amplitude values. For a Gaussian shaped waveform of the emitted laser pulse, the received waveform can be approximated by one or more parameterized Gaussian curves (Hofton *et al.*, 2000; Persson *et al.*, 2005; Wagner *et al.*, 2006). Due to this approximation, the temporal position, width and amplitude caused by the object surfaces can be estimated (Jutzi and Stilla, 2006). With these parameters, the geometry and reflectance of the illuminated surface can be investigated, whereby the material reflectance features from the measured data mainly depend on the incidence angle of the beam on the surface, the surface properties and the laser wavelength (Jelalian, 1992).

Related Works

In the terminology of laser scanning, the reflectance is widely used as a synonym for the amplitude or energy, where the energy of each pulse is the integral over its waveform. For a Gaussian pulse, this can be simplified and approximated by the product of amplitude and width. Besides *amplitude* or *energy*, the term *intensity* is used. Various studies about surface reflectance and intensity calibration have been published in the literature in this context:

- Briese *et al.* (2008) proposed to use natural surfaces with known backscattering characteristic measured by a reflectometer for radiometric calibration of full-waveform data.
- Höfle and Pfeifer (2007) showed a data and a model-driven method for correcting the intensity for specific influences. The corrected intensity is successfully used to generate intensity images with lower systematic errors.
- Kaasalainen *et al.* (2007 and 2009) suggested using laboratory-measured reference targets for calibrating the intensity values of airborne laser scanner sensors.
- Katzenbeisser (2003) observed for flat surfaces that the measured intensity provide a reasonable mean for the

Boris Jutzi is with the Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76128 Karlsruhe, Germany (boris.jutzi@kit.edu).

Hermann Gross is with the Fraunhofer Institute of Optics, System Technologies and Image Exploitation, Gutleuthausstr. 1, 76275 Ettlingen, Germany.

Photogrammetric Engineering & Remote Sensing
Vol. 76, No. 9, September 2010, pp. 000–000.

0099-1112/10/7609-0000/\$3.00/0
© 2010 American Society for Photogrammetry
and Remote Sensing

reflectance, if the measured intensity is corrected by the known distance.

- Kukko *et al.* (2007) measured for various urban materials the dependency of the intensity on the incidence angle.
- Pfeifer *et al.* (2007) studied the influence on the intensity for surfaces with varying incidence angles, known reflectance and scattering characteristics. It is shown that the range dependent inverse-square model might be insufficient to estimate the accurate intensity.
- Pfeifer *et al.* (2008) proposed the potential of retrieving material properties from natural surfaces after investigating radiometric reference targets with terrestrial laser scanners.
- Reshetyuk (2006) investigated the surface reflectance of various materials and its influences on the measured range and intensity.
- Wagner *et al.* (2006) gives a review on their proposed calibration procedure and scattering model concerning the cross section and backscattering parameter of an object with diffuse (Lambertian) surface characteristics.

Especially for registration, segmentation, classification, or visualization purposes an intensity normalization of the measurements derived by an airborne laser scanner is of great interest. Obviously, the variation of the incidence angle increases if data from several flights with different paths (flight stripes) are fused. Further, the atmospheric conditions can change slightly while the measurement is carried out or change considerably if multi-temporal data is acquired. For these reasons, the range, the incidence angle and the atmospheric attenuation has to be taken into account to normalize the intensity.

To give an example for the dependency of the intensity on the incidence angle, an airborne image together with the corresponding intensity values from two different flights are visualized in Figure 1. The viewing direction of the sensor system is depicted by a white arrow. The area of interest is the gabled roof. Obviously, roof areas oriented towards the sensor system show higher intensity values, while roof areas on the opposite side show significantly smaller intensity values.

These intensity values, which are estimated from the laser data, strongly correlated to the incidence angle of the laser beam on the surface. Therefore, as likewise Höfle and Pfeifer (2007) showed, the intensity can be normalized by considering the incidence angle derived by the sensor and object position as well as its surface orientation. The estimation of the surface orientation is derived by utilizing the available data concerning the nearby neighborhood of each measured laser point. Based on the estimated surface orientation diffuse and

specular material reflectance is investigated. Further influences on the measurement are not studied in this work, like the non-linear effects on the measured intensity induced by electronic receiver components or inhomogeneous intensity distribution of transmitted laser in the footprint or inhomogeneous reflectance characteristic within the illuminated area.

The paper is organized as follows. In Section *Methodology*, the physical conditions of the Lambertian and Phong reflection models, a data-driven parameter estimation approach, and the methodology for the calculation of the incidence angle are introduced. In the following section, the *Evaluation of the Intensity Calibration* is described. The Section *Test Environment* gives a description of the data set, the regions of interest and the influence of the flight paths on the incidence angle. The *Results* for the data-driven parameter estimation are shown in the following section, concerning parameter estimation and the normalization results for all investigated regions. The derived results are discussed in general and in detail for selected regions in the subsequent Section *Discussion*. Finally, the gained experiences are given in the Section *Conclusion*.

Methodology

In general, each echo from full-waveform laser scanning data has attributes such as the relative 3D coordinate, signal amplitude a_G , and signal width w_G at full-width-at-half-maximum derived from the Gaussian approximation. Additionally to these attributes of the measurement, the corresponding absolute 3D coordinate of the sensor position by the flight trajectory is available.

The shape and size of the received waveform mainly depends on the material characteristic, the reflectance characteristic of the surface, and the inclination angle between the surface normal and the laser beam direction. The typical surface attributes which can be extracted from a waveform are the absolute range, range variation, and reflectance corresponding to the waveform features: time, width, and amplitude. A representative reflectance of the surface can only be extracted for a unique surface area. This is usually given for single echoes on objects larger than the laser footprint, for multiple echoes, e.g., derived from a vegetated environment, it is indistinguishable.

The intensity is estimated with a Gaussian approximation by multiplying the width with the amplitude, and

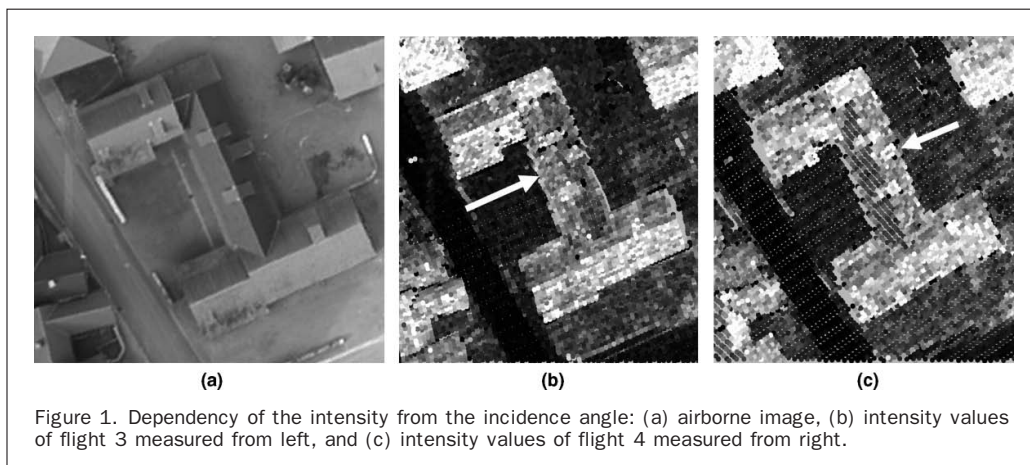


Figure 1. Dependency of the intensity from the incidence angle: (a) airborne image, (b) intensity values of flight 3 measured from left, and (c) intensity values of flight 4 measured from right.

further it is adapted by the range between sensor and object with respect to the extinction by the atmosphere and the divergence of the laser beam. It describes the reflectance influenced by geometry and material of the object at this point. For each particular echo caused by partially illuminated object surfaces, an individual intensity value influenced by the part of the footprint is received.

Laser Beam, Transmission, and Reflection Model

The received intensity I_r of a monostatic laser scanner system can be calculated by the amplitude a_G and width w_G of the received signal approximation with $I_r = a_G \cdot w_G$. Considering an energy balance, the received intensity I_r depends on the transmitted intensity I_t , the distance R to the object surface, and for a Lambertian reflection model, the incidence angle ϑ , which is given by the angle between the transmitter direction and the surface normal vector with:

$$I_r = I_t \cdot C_t \cdot C_r \cdot R^{-2} \cdot e^{-2\alpha R} \cdot \cos(\vartheta) \cdot f(c_s) \quad (1)$$

where C_t and C_r are constant terms of the transmitter and the receiver (Kamerman, 1993; Pfeifer *et al.*, 2007). The atmospheric attenuation along the way from the transmitter to the illuminated object surface and return to the receiver is $e^{-2\alpha R}$. The function $f(c_s)$ depends on c_s , which encloses all additional influences, e.g., the surface material or roughness.

Equation 1 is valid for object surface sizes larger than the footprint of the laser beam. Because the scaling is not relevant, all constant terms can be ignored. Using the received amplitude a_G and width w_G of the signal, the range corrected intensity can be calculated with:

$$I_R = a_G \cdot w_G \cdot R^2 \cdot e^{2\alpha R} \quad (2)$$

The intensity I_R does not depend on the distance R anymore. I_R is only influenced by the material properties and the incidence angle. Non-linear effects of the signal capturing unit (e.g., photodiode, amplifier) are not considered by this assumption.

For all points with high planarity concerning the close neighborhood, the measured intensity is normalized by:

$$I_{Rc} = I_R / \cos(\vartheta) \quad (3)$$

with respect to the incidence angle. The illumination direction \bar{e}_i is calculated from the sensor to the object position. For each measured point, the covariance matrix is calculated by including all points in a predefined environment. Its eigenvalues are used to estimate the planarity of an object surface. The eigenvector \bar{e}_3 with respect to the smallest eigenvalue λ_3 is the normal vector of the object surface. Further details about estimating the planarity are given in Sub-section *Surface Orientation*. With the normalized vectors the required divisor can be calculated with the inner product $\cos(\vartheta) = |\bar{e}_i \cdot \bar{e}_3|$.

With the proposed framework, the radiometric calibration of the intensity due to atmospheric influences and surface orientation is feasible. Finally, the normalized intensity I_{Rc} depends on the wavelength of the laser and the material properties only. The influence of speckle effects on the measurement is neglected in this framework.

Data-driven Parameter Estimation

In order to adapt the dependency of the intensity on the object distance, atmospheric attenuation, and incidence angle, an extended Lambertian reflection model is introduced. Therefore, Equations 2 and 3 are extended for each single measured point i by:

$$I_{A_i} = I_t R_i^a e^{2bR_i} \cos^c(\vartheta_i) f_k(i) \quad (4)$$

with $I_i = a_{G,t} \cdot w_{G,t}$, R_i the distance between the sensor and the object, ϑ_i the incidence angle, and a, b, c constant parameters. Herein a describes the beam divergence. The exponent $2bR$ concerns the attenuation by the two-way propagation of the laser beam. The term c models the type of reflectivity and f_k is supposed to be constant inside each homogeneous region, and normalizes the term to be equal with 1. Inside a homogeneous region the adapted intensity should be nearly constant. Set $T_i = I_i R_i^a e^{2bR_i} \cos^c(\vartheta_i)$, and let Nr be the set of the region numbers. The following minimization problem has to solved:

$$\sum_{\substack{i \in \bigcup_{k \in Nr} \text{region}(k)}}} \varepsilon_i^2 = \sum_{k \in Nr} \sum_{i \in \text{region}(k)} \left(T_i f_{k(i)} - 1 \right)^2 \rightarrow \text{Min} \quad (5)$$

where ε_i marks the error of the i^{th} point of the point cloud with the considered region. For known parameters a, b, c the region weighting factor is:

$$f_k = \frac{\sum_{i \in \text{region}(k)} T_i}{\sum_{i \in \text{region}(k)} T_i^2} \quad (6)$$

Extended Reflection Model for Diffuse and Specular Material Reflectance

To enhance the above mentioned Lambertian reflection model and its extended version, the empirical Phong reflection model (Phong, 1975) is introduced. Besides the diffuse surface scattering k_d the proposed Phong model can handle also specular surface characteristics k_s . The general Equation is given by:

$$I_{out} = I_{amb} k_a + I_{in} \left[k_d \cos(\vartheta) + k_s \cos^n(2\vartheta) \right] \quad (7)$$

where I_{in} is the incoming intensity and I_{out} the outgoing intensity. Ignoring the ambient intensity I_{amb} by setting ($k_a = 0$) and considering $k_a + k_d + k_s = 1$ for the diffuse reflectance parameter $k_d = 1 - k$, is derived. The remaining free parameters k_s and n are the weighting factors for the specular part of the reflection. The adaptation of the Equation yields

$$I_{AP} = I R^a e^{2bR} \left/ \left((1 - k_s) \cos(\vartheta) + k_s \cos^n(2\vartheta) \right) \right. \quad (8)$$

with the specular reflectance parameter k_s and the degree of the specular reflectance n , which can be iteratively optimized within a homogeneous region.

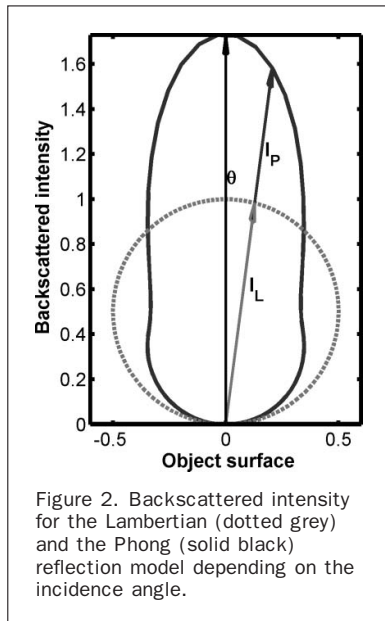
The dependency of the backscattered intensity from the incidence angle is shown in Figure 2 for the Lambert and the Phong reflection model with the parameters $k_s = 0.6$ and $n = 4$. For small incidence angles, the Phong model gives higher intensity values than the Lambert model and lower values for greater incidence angles.

Determining the Surface Orientation

The orientation of the illuminated surface has to be estimated to accomplish the radiometric calibration of the intensity. For each measured point in the data set, all points in a small spherical neighborhood are considered to calculate the covariance matrix and the corresponding eigenvalue and eigenvector (Gross and Thoennessen, 2006). With the determined eigenvalues plane surface areas can be segmented and the orientation of the surface can be estimated. To decide whether a point belongs to a planar surface or not, the planarity $p = (\lambda_2 - \lambda_3) / \lambda_1$ based on descend-sorted eigenvalues of the covariance matrix (West *et al.*, 2004) is used.

Evaluation of the Intensity Calibration

For evaluating the adapted intensity typical object types (e.g., roof, road, grassland, field areas) of a laser scanning



large variety of surface slopes. The large variety of surface slopes enables to evaluate the utilized reflection models.

For the selected regions, the given intensity is normalized by the optimized cosine exponent (reflectance adaptation parameter c) and the division by the cosine of the incidence angle ϑ . By this division, the normalized intensity value increases compared to the original one. For the intensity sample x the mean value $\mu(x)$ and the standard deviation $\sigma(x)$ is used for the calculation of the variation coefficient $V_c(x) = \sigma(x)/\mu(x)$. This term is scale invariant and takes into account the dependency of the standard deviation of the intensity as presented by Pfeifer *et al.* (2007).

The variance of the incidence angle for each region increases in general if data from more than one flight are used. For an assessment, the ratio of the variation parameter $R_V(\text{region}) = V_{c,\text{after}}(\text{region})/V_{c,\text{before}}(\text{region})$ for all selected regions after versus regions before normalization are calculated. If the ratio is smaller than 1, the intensity could be improved.

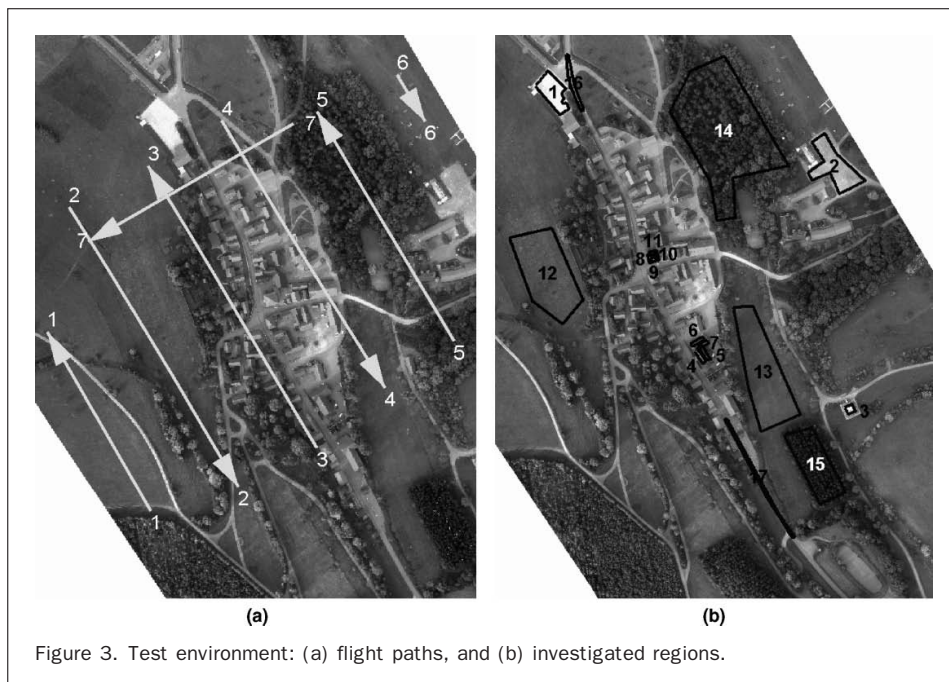
Test Environment

A measurement campaign on a scene containing buildings, streets, grassland, and trees in an urban environment was carried out to investigate the influences of the incidence angle on the measured intensity. The data was captured with a RIEGL LMS-Q560 full-waveform laser scanner. Several flight paths with different trajectories to gain overlapping stripes were performed.

Data Set

The data set is fused by all points gathered within seven flights. Six flight paths are parallel. Flight path seven crosses the others (Figure 3a). A square region in the middle

earth observation were selected and different orientated planar regions with homogeneous surface reflectance have been investigated. The regions were manually selected by the criteria to consist of the same material and to have a



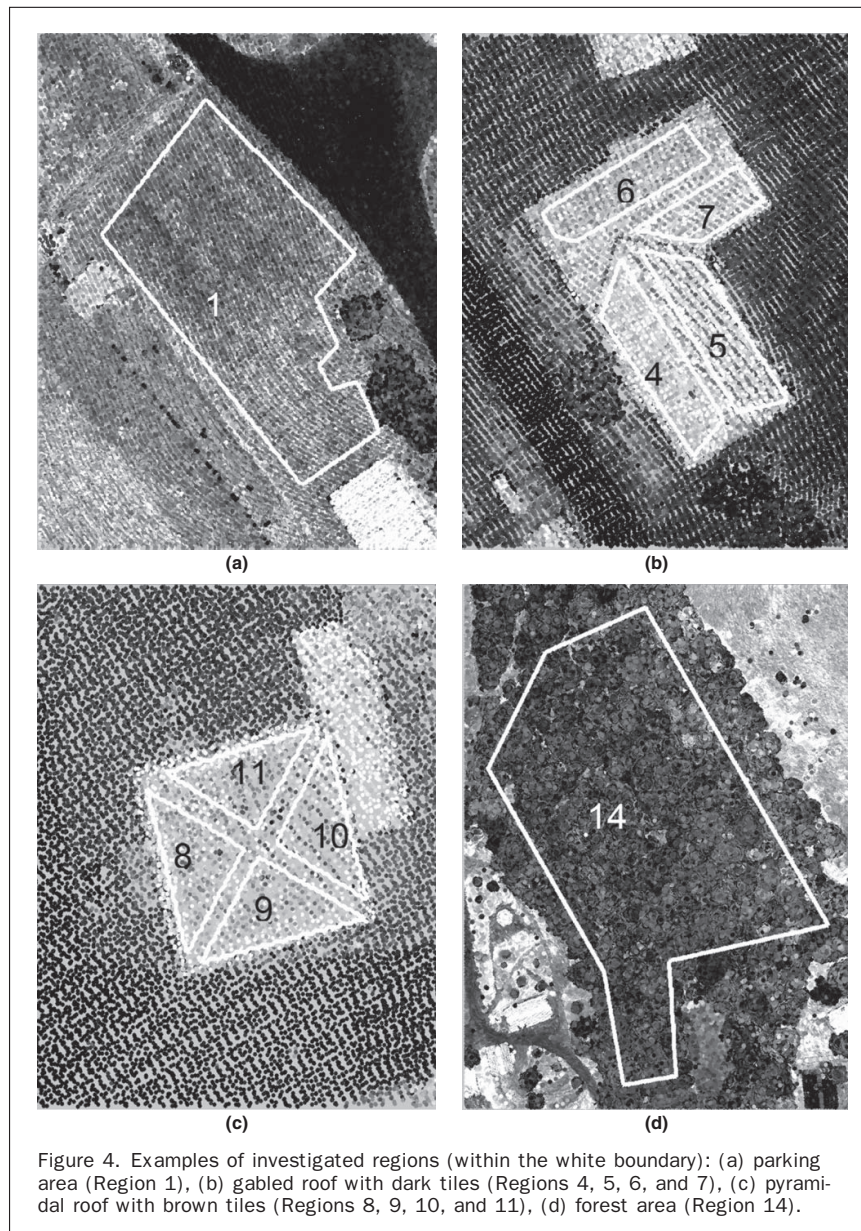
of the scene with the size of $200\text{ m} \times 200\text{ m}$ is covered by an average density of about 11 points per square meter, where flights 1, 2, 3, 4, 5 have a density of 0.6, 3.1, 3.6, 2.9, and 0.9 points/m^2 , respectively.

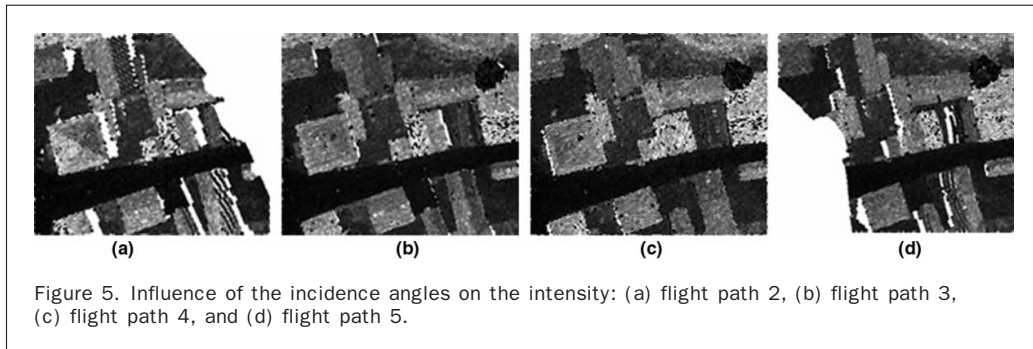
The calculation of the incidence angle and the planarity is based on the determination of the covariance for each point by including all neighbor points inside a sphere with predefined radius. For a radius of 1 m the average of 25 points are considered if the data of all flight paths are utilized.

Regions of Interest

For investigations, 17 specific Regions (Figure 3b and Figure 4) were manually selected from the data set to prove the concept and usage of the normalization method for different orientated object regions and different kinds of materials:

- Region 1 and 2: parking areas made of different materials (Figure 4a),
- Region 3: one building with a flat bright roof,
- Regions 4, 5, 6, and 7: one L-shaped building with different orientated parts of a gabled roof made of dark tiles (Figure 4b),





- Regions 8, 9, 10, and 11: one building with different orientated parts of a pyramidal roof made of brown tiles (Figure 4c),
- Regions 12 and 13: two great grassland areas,
- Regions 14 and 15: two forest areas (Figure 4d), and
- Regions 16 and 17: two street areas.

Influence of the Flight Paths on the Incidence Angle

The influences of the flight path respectively of the local incidence angle on the measured intensity values is shown in Figure 5. The flight paths 2 to 5 have a distance of about 330 m. Already a small offset (116 m) between the two flight paths 2 and 3, yields to significant different intensity values as depicted in Figure 5a and 5b.

The building with four different orientated parts of a pyramidal roof (Regions 8, 9, 10, and 11) out of the same material on the left border of the image shows that large intensity values are given if the plane normal vector points to the sensor. Smaller intensity values are given for the same area if the normal vector is directed away from the sensor. The variation of the intensity in general is depicted in Figure 5.

Results

Parameter Estimation of the Lambertian Models for the Investigated Data Set

For both Lambertian models, the beam divergence of the laser beam is set to $a = 2$ due to geometric consideration and the atmospheric extinction coefficient $b = 0.001725$ [1/m] was provided with the measured data. Then, the estimated extinction parameter is equal to the attenuation of 7.5 dB/km. Typically the atmospheric attenuation versus wavelength (0.7 to 10.6 μm) for lasers is within the range of 0.2 [dB/km] for extremely clear weather conditions up to 9 [dB/km] for light fog or rain (Jelalian, 1992). For the standard Lambertian model $c = -1.0$ is utilized and by optimization for the extended Lambertian model $c = -0.79$ is estimated. For the standard Lambertian model according to Equation 3, a slightly greater mean error than for the optimized extended version can be observed (Table 1). The enhancement is not essential.

Normalization Results for the Investigated Regions

To study the influence of the incidence angle on the intensity, the selected regions, which are shown in Figure 3b, were processed. Table 2 shows the relevant information and derived results in the following order: (a) each selected Region, (b) number of points of the point cloud

TABLE 1. MEAN ERRORS FOR THE LAMBERTIAN MODELS

	a	b[1/m]	c	mean(e_r^2)
Standard	2.0	0.001725	-1.00	0.1365
Extended	2.0	0.001725	-0.79	0.1358

per region by Σ Points, (c) mean value $mean(I_r)$ and (d) standard deviation $std(I_r)$ of the original intensity, (e) mean value of the normalized intensity derived by utilizing the Lambertian I_{Rc} reflection model by $mean(I_{Rc})$, (f) corresponding standard deviation $std(I_{Rc})$, (g) coefficient before $V_C(I_r) = std(I_r)/mean(I_r)$ and (h) after the normalization $V_C(I_{Rc})$, and (i) ratio of the variation coefficients $V_C(I_{Rc})/V_C(I_r)$. Ratio values below 1 are improvements gained for the intensity by considering the extinction by the atmosphere and the laser beam divergence in dependence on the distance between sensor and object surface as well as the incidence angle given by the normal vector of the surface and the beam direction.

In Table 3 for each region, the derived variation coefficients of the investigated reflection models are listed for direct comparison. First, the already given variation coefficients of the Lambertian with $\cos(\vartheta)^{-1}$, second, the extended Lambertian considering the empirically optimized surface reflectance adaptation parameter $c = -0.79$ with $\cos(\vartheta)^{-0.79}$, and third, the Phong reflection model considering the empirically determined parameters $k_s = 0.2$ and $n = 4$. The rates of the variation coefficients after versus before the normalization of the intensity are for the extended Lambertian model for most of the regions better and for the Phong model in the most of the regions worse than for the standard version.

Discussion

In the following, a discussion on the intermediate and final results is given.

Normalization Results for the Investigated Regions

The results derived by the Lambertian reflection model (Equation 3) in Table 2 have to be discussed. Comparing the traffic related areas of the given data set, it can be observed that the parking areas (Region 1 and 2) have three to five times higher intensities than the street areas (Region 16 and 17) but nearly the same variation coefficient before and after the normalization. Due to the small variation of the incidence angle and the low intensity

TABLE 2. RESULTS DERIVED BY UTILIZING THE LAMBERTIAN REFLECTION MODEL FOR THE SELECTED REGIONS

a	b	c	d	e	f	g	h	i
Region	SPoints	mean(I_r)	std(I_r)	mean(I_{Rc})	std(I_{Rc})	$V_C(I_r)$	$V_C(I_{Rc})$	$V_C(I_{Rc})/V_C(I_r)$
1	18667	2988	705	1421	243	0.24	0.17	0.72
2	19294	4725	1152	1778	423	0.24	0.24	0.98
3	1398	2907	778	1288	314	0.27	0.24	0.91
4	1016	5287	1086	2885	397	0.21	0.14	0.67
5	943	4836	1316	2724	419	0.27	0.15	0.57
6	798	4354	782	2614	392	0.18	0.15	0.83
7	537	5195	1071	3074	455	0.21	0.15	0.72
8	454	4568	1000	2500	411	0.22	0.16	0.75
9	429	4651	1015	2577	430	0.22	0.17	0.76
10	333	4792	2100	2476	978	0.44	0.39	0.90
11	423	4422	1063	2479	397	0.24	0.16	0.67
12	100261	3180	569	1300	209	0.18	0.16	0.90
13	72065	3021	773	1410	348	0.26	0.25	0.96
14	216597	1758	915	841	446	0.52	0.53	1.02
15	24074	1579	1167	831	558	0.74	0.67	0.91
16	5068	1089	257	489	105	0.24	0.22	0.91
17	5793	1199	269	599	110	0.22	0.18	0.82

TABLE 3. RATIOS OF THE VARIATION COEFFICIENTS FOR THE LAMBERTIAN AND THE PHONG REFLECTION MODELS

Region	$V_C(I_{Rc})/V_C(I_r)$	$V_C(I_A)/V_C(I_r)$	$V_C(I_{AP})/V_C(I_r)$
	Lambert ($c = -1$)	Lambert ($c = -0.79$)	Phong ($k_s = 0.2$ and $n = 4$)
1	0.72	0.72	0.80
2	0.98	0.98	1.01
3	0.91	0.90	0.99
4	0.67	0.70	0.67
5	0.57	0.53	0.59
6	0.83	0.82	0.84
7	0.72	0.71	0.72
8	0.75	0.78	0.75
9	0.76	0.75	0.79
10	0.90	0.92	0.91
11	0.67	0.67	0.68
12	0.90	0.88	1.13
13	0.96	0.96	1.01
14	1.02	1.01	1.03
15	0.91	0.94	0.85
16	0.91	0.91	0.97
17	0.82	0.81	0.86

values, improvements from 2 percent to 28 percent could be derived.

In general for the roof areas, significant improvements up to 43 percent could be gained, especially if the roof parts were differently orientated and a large incidence angle on these parts was given, e.g., gabled roof with grey tiles with Regions 4 to 7, and pyramidal roof with brown tiles with Regions 8 to 11. It could be observed that the intensity for all roof parts was nearly in the same range and independent from the visible colors of the tiles. A lower improvement of 9 percent was given for flat bright roof with Region 3, due to the small variation of the incidence angle.

For the vegetation environment, no significant improvements were expected, due to the large variation of natural objects and therefore, the low planarity in close neighborhood. The results for the grassland areas (Region 12 and 13) and forest area (Region 15) do only show very small improvements up to 10 percent or no relevant changes for the forest area in Region 14.

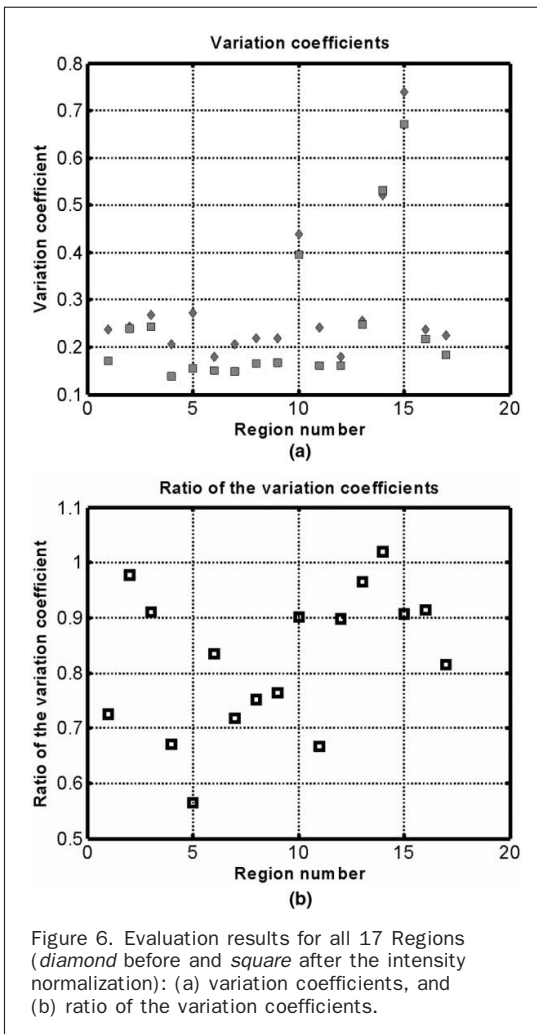
Concerning the extended Lambertian (Equation 4) and Phong (Equation 7) reflection model, (Table 3) the investigations show that only for a few cases the extended Lambertian reflection model gives insignificantly better results than the standard model. With the Phong reflection model, the results are worse for all regions except for the forest Region 15. Various internal tests showed that other parameter values do not provide essentially better results for this data set. Due to the large variation of the intensity measurements in this data set unfortunately no reliable improvements could be gained. Therefore, these two reflection models are not further discussed in this paper.

Detailed Comparison of the Intensity Normalization Results

In Figure 6a, the variation coefficients for all regions are plotted; *diamond* symbols mark the values before and *square* symbols after the intensity normalization (Table 2, column $V_C(I_r)$ and $V_C(I_{Rc})$). All regions except Region 14 give better results. The ratios of the variation coefficients $V_C(I_{Rc})/V_C(I_r)$ are drawn in Figure 6b. All regions containing man-made objects show an improvement between 2 percent to 43 percent due to the intensity normalization. For the street areas (Region 16 and 17), the original intensity is very low, but nevertheless a small improvement could be gained.

The following figures show the intensity distribution depending on the incidence angle before and after the normalization for a parking area (Figure 7) and the pyramidal roof regions (Figure 8) colored in accordance with the flight identification. The influence of the incidence angle on the intensity is obvious, where the theoretical trend of the intensity is depicted by the black curves. The proposed normalization can compensate the mean intensity but not the high variation of the intensity for constant incidence angle. It can be assumed that the high variation of the intensity is strongly influenced by the measurement procedure.

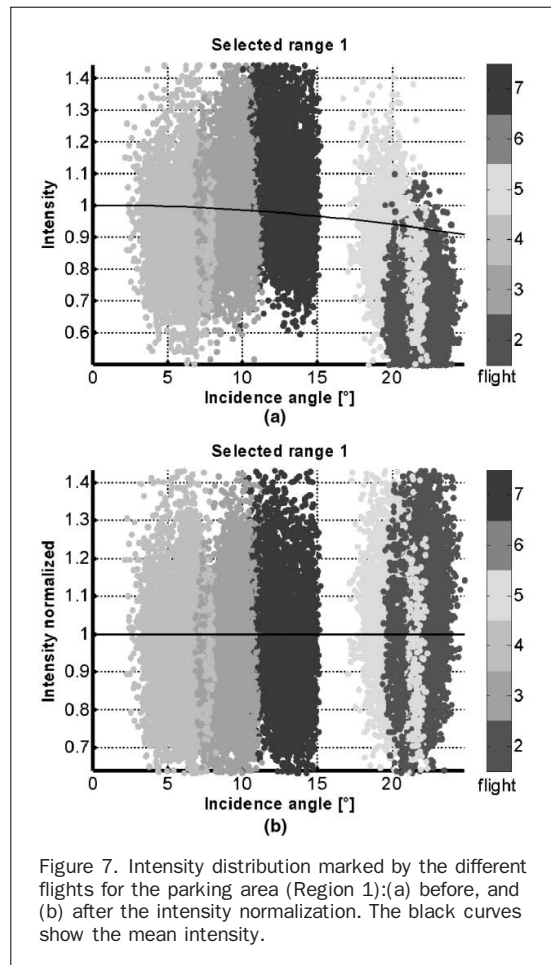
Figure 9 shows the distribution of the intensity values of a parking area (Region 1, Figure 4a) with incidence angles from 0° up to 10° before and after the intensity normalization. It could be shown that in general the intensity variation in the area is improved with the normalization. In Figure 10 the high standard deviation for each incidence angle is depicted. The variation can not be corrected by using the incidence angle only, it is influenced by the measurement techniques itself and also some local disturbances.



Conclusion

Investigations on airborne laser scanning data for intensity normalization on man-made and natural areas were done. The intensity was adapted concerning the influence of the atmospheric extinction, the laser beam divergence as well as the incidence angle between laser beam, and normal vector of the object surface under the assumption that the material has a Lambertian reflectivity. To refine the surface model beside the standard Lambertian (Equation 3), the extended Lambertian (Equation 4) and the Phong (Equation 7) reflection model was introduced. For evaluation, nearly homogeneous regions were selected interactively. Due to the dependency of the standard deviation from the mean value of the intensity, the variation coefficient as measure for the assessment is recommended.

A significant improvement could be presented in this paper by utilizing the standard Lambertian reflection model. Concerning the extended Lambertian and Phong reflection models, no reliable improvements could be gained even after parameter optimization. However, for future investigations



data captured in close range might be of interest, e.g., by a terrestrial laser scanning (TLS) device, where less intensity variation can be expected.

Our investigation demonstrates that large intensity variation with respect to the surface orientation and beam propagation direction can be reduced and the intensity can be already improved within the standard Lambertian reflection model. The limitation of this approach is given for intensity variations depending on the sensor technique (e.g., low signal-to-noise ratio at the receiver, variation of the emitted pulse intensity), measurement technique (e.g., speckle effects), object materials (e.g., inhomogeneous material reflectance), and local geometry (e.g., roughness of the surface, partly illuminated surface areas). These effects cannot be compensated by the proposed approach.

References

- Briese, C., B. Höfle, H. Lehner, W. Wagner, M. Pfennigbauer, and A. Ullrich, 2008. Calibration of full-waveform airborne laser scanning data for object classification, *Laser Radar Technology and Applications XIII* (M.D. Turner and G.W. Kamerman, editors), SPIE Proceedings Vol. 6950, pp. 69500H.

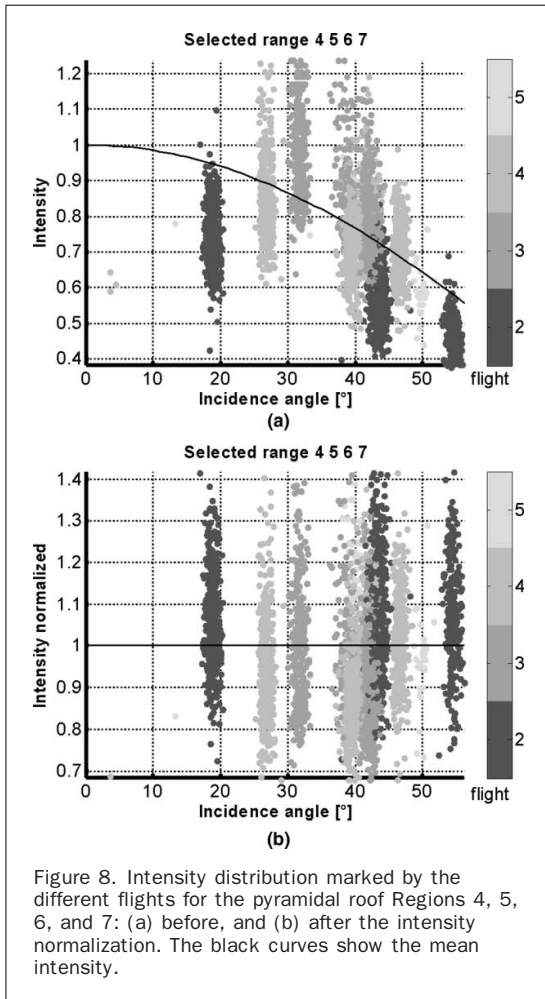


Figure 8. Intensity distribution marked by the different flights for the pyramidal roof Regions 4, 5, 6, and 7: (a) before, and (b) after the intensity normalization. The black curves show the mean intensity.

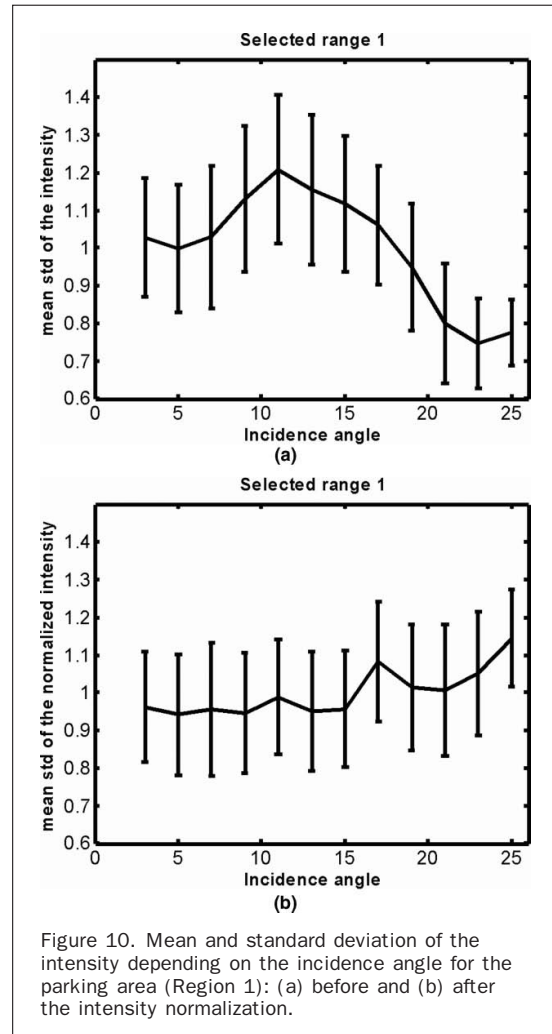


Figure 10. Mean and standard deviation of the intensity depending on the incidence angle for the parking area (Region 1): (a) before and (b) after the intensity normalization.

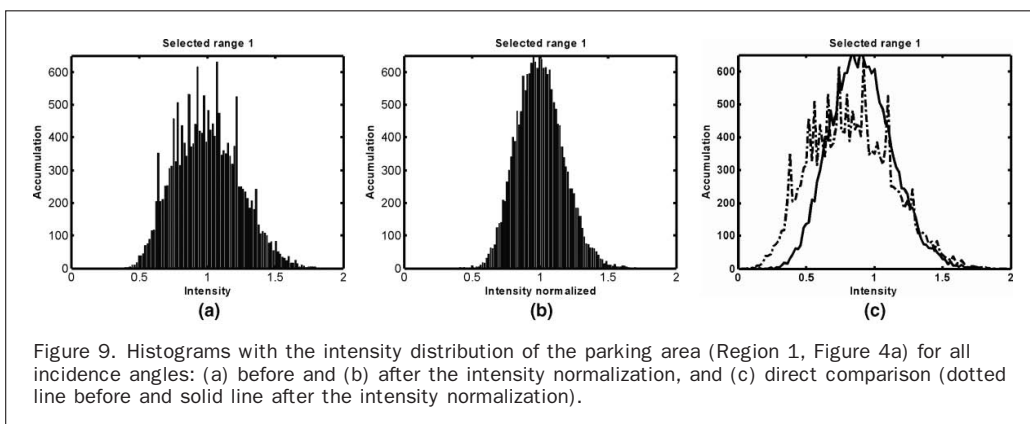


Figure 9. Histograms with the intensity distribution of the parking area (Region 1, Figure 4a) for all incidence angles: (a) before and (b) after the intensity normalization, and (c) direct comparison (dotted line before and solid line after the intensity normalization).

- Gross, H., and U. Thoennessen, 2006. Extraction of lines from laser point clouds, *Proceedings of the Symposium of ISPRS Commission III: Photogrammetric Computer Vision PCV06* (W. Förstner and R. Steffen, editors), *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (Part 3), pp. 86–91.
- 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(6):415–433.
- Hofton, M.A., J.B. Minste, and J.B. Blair, 2000. Decomposition of laser altimeter waveforms, *IEEE Transactions on Geoscience and Remote Sensing*, 38(4):1989–1996.
- Jelalian, A.W., 1992. *Laser Radar Systems*, Artech House, Boston, Massachusetts.
- 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., J. Hyypä, P. Litkey, H. Hyypä, E. Ahokas, A. Kukko, and H. Kaartinen, 2007. Radiometric calibration of ALS intensity, *Proceedings of Laserscanning 2007* (P. Rönnholm, H. Hyypä, and J. Hyypä, editors), *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 36(Part 3-W52):201–205.
- Kaasalainen, S., A. Krooks, A. Kukko, and H. Kaartinen, 2009. Radiometric calibration of terrestrial laser scanners with external reference targets, *Remote Sensing 2009*, 1(3):144–158, doi:10.3390/rs1030144.
- Kamerman, G.W., 1993. Laser radar, *Active Electro-Optical Systems, The Infrared & Electro-Optical Systems Handbook* (C.S. Fox, editor), SPIE Optical Engineering Press, Michigan.
- Katzenbeisser, R., 2003. Technical Note on Echo Detection, URL: <http://www.toposys.de/pdf-ext/Engl/echo-detec3.pdf> (last date accessed: 23 June 2010).
- Kukko, A., S. Kaasalainen, and P. Litkey, 2007. Effect of incidence angle on laser scanner intensity and surface data, *Applied Optics*, March 2008, 47(7):986–992.
- Persson, Å., U. Söderman, J. Töpel, and S. Ahlberg, 2005. Visualization and analysis of full-waveform airborne laser scanner data, *Proceedings of Laserscanning 2005* (G. Vosselman and C. Brenner, editors), *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (Part 3/W19):109–114.
- Pfeifer, N., P. Dorninger, A. Haring, and H. Fan, 2007. Investigating terrestrial laser scanning intensity data: Quality and functional relations, *Proceedings of the International Conference on Optical 3-D Measurement Techniques VIII* (A. Gruen and H. Kahmen, editors), Zürich, Switzerland, ISBN 3-906467-67-8, pp. 328–337.
- Pfeifer, N., B. Höfle, C. Briese, M. Rutzinger, and A. Haring, 2008. Analysis of the backscattered energy in terrestrial laser scanning data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(Part B5):1045–1052.
- Phong, B.T., 1975. Illumination for computer generated pictures, *Communications of the ACM*, 18(6):311–317.
- Reshetyuk, Y., 2006. Investigation of the influence of surface reflectance on the measurements with the terrestrial laser scanner Leica HDS 3000, *Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, 131(2):96–103.
- Shan, J., and C.K. Toth (editors), 2008. *Topographic Laser Ranging and Scanning: Principles and Processing*, Taylor & Francis, Boca Raton, Florida.
- 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(2):100–112.
- West, K.F., B.N. Webb, J.R. Lersch, S. Pothier, J.M. Triscari, and A.E. Iverson, 2004. Context-driven automated target detection in 3D data, *Proceedings of Automatic Target Recognition XIV* (F.A. Sadjadi, editor), SPIE, 5426:133–143.

CHARACTERISTICS OF THE MEASUREMENT UNIT OF A FULL-WAVEFORM LASER SYSTEM

B. Jutzi ^a, U. Stilla ^b

^a FGAN-FOM Research Institute for Optronics and Pattern Recognition, 76275 Ettlingen, Germany - jutzi@fom.fgan.de

^b Photogrammetry and Remote Sensing, Technische Universität München, 80290 München, Germany - stilla@bv.tum.de

Commission I, WG I/2

KEY WORDS: Accuracy, Analysis, Laser scanning, LIDAR, Measurement, Full-waveform.

ABSTRACT:

The analysis of data derived by full-waveform laser scanning systems is of great interest. In this study, we estimate the impulse response of a laser scanning system capturing the waveform. Considering the impulse response of the system is important for precise waveform analysis. The characteristic of the system has to be mentioned to interpret the pulse properties in an accurate way. We determined by experiments the impulse responses of the laser scanning system for different bandwidths, namely 250 MHz, 750 MHz, 1 GHz, and 6 GHz. By considering the impulse response the measured transmitted waveform of the emitted laser pulse can be adapted. The similarity of the adapted transmitted and received waveform is compared and discussed.

1. INTRODUCTION

The automatic generation of 3-d models for a description of man-made objects, like buildings, is of great interest in photogrammetric research. Laser scanner systems allow a direct and illumination-independent measurement of the range. Laser scanners capture the range of 3-d objects in a fast, contact free and accurate way. Overviews for laser scanning systems are given in (Huising & Pereira, 1998; Wehr & Lohr, 1999; Baltsavias, 1999). A general overview how to develop and design laser systems can be found in textbooks (Jelalian, 1992; Kamermann, 1993).

Current pulsed laser scanner systems for topographic mapping are based on time-of-flight techniques to determine the range of the illuminated object. The elapsed time between the emitted and backscattered laser pulses is typically determined by a threshold detection with analogous electronics. Some systems capture multiple reflections caused by objects which are smaller than the laser beam footprint located in different ranges. Such systems usually record the first and the last backscattered laser pulse.

First pulse as well as last pulse exploitation is used for different applications like urban planning or forestry surveying. While first pulse registration is the optimum choice to measure the hull of partially penetrable objects (e.g. canopy of trees), last pulse registration should be chosen to measure non-penetrable surfaces (e.g. ground surface below vegetation).

Beside the first or last pulse exploitation the complete waveform in between is of interest, because it includes the backscattering characteristic of the illuminated field. Investigations on the waveform analysis were done to explore the vegetation concerning the bio mass, foliage or density (e.g. trees, bushes, and ground). NASA has developed a prototype of the Laser Vegetation Imaging Sensor (LVIS) recording the waveform to determine the vertical density profiles in forests (Blair *et al.*, 1999). This experimental airborne system operates in altitudes up to 10 km and provides a large footprint diameter (up to 80 m) to study different land cover classes.

The spaceborne Geoscience Laser Altimeter System (GLAS) on the Ice, Cloud and Land Elevation Satellite (ICESat) determines changes in range through time, height profiles of clouds and aerosols, ice sheet and land elevations, and vegetation (Brenner *et al.*, 2003; Zwally *et al.*, 2002). It operates with a large footprint diameter (70 m) on Earth and measures elevation changes with decimeter accuracy (Hoften *et al.*, 2000). Table 1 gives an overview of additional specifications.

Beside large footprint systems first developments of small footprint systems were done for monitoring the nearshore bathymetric environments with the Scanning Hydrographic Operational Airborne Lidar Survey system (SHOALS). SHOALS has been in full operation since 1994 (Irish & Lillycrop, 1999; Irish *et al.*, 2000). Recent developments of commercial laser scanner systems led to systems that allow capturing the waveform: RIEGL LMS-Q560, OPTTECH ALTM 3100, TOPEYE II. The systems mentioned above are specified to operate with a transmitted pulse width of 5 ns and allow digitization and capturing the waveform with approximately 1 GSample/s. Additional specifications of the small footprint laser systems mentioned above are shown in Table 2.

To interpret the received waveform of the backscattered pulse, a fundamental understanding of the physical background of pulse propagation and surface interaction is important (Jutzi *et al.*, 2002; Wagner *et al.*, 2003). The influence of the surface on the transmitted waveform is discussed by Steinvall (2000) for objects with different shapes taking into account different reflection characteristics. Gardner (1982) and Bufton (1989) investigated the pulse spreading by the impact of the surface structure, e.g. surface slope and vertical roughness within the laser footprint.

The recording of the received waveform offers the possibility to use different methods for the range determination, e.g. peak detection, leading edge detection, average time value detection, constant fraction detection. This topic was investigated by different authors, e.g. Der *et al.*, 1997; Steinvall & Carlsson, 2001; Jutzi & Stilla, 2003; Thiel & Wehr, 2004; Wagner *et al.*, 2004; Vandapel *et al.*, 2004. The analysis of the pulse shape increases the reliability, accuracy, and resolution.

Large footprint laser systems	LVIS	GLAS
Wavelength [nm]	1064	1064
Pulse length at FWHM ^a [ns]	10	6
Pulse repetition rate [Hz]	100-500	40
Laser output energy [mJ]	5	75
Laser beam divergence [mrad]	~5	0.11
Operating altitude [km]	10	600
Scan angle [°]	± 7	0
Detector bandwidth [MHz]	90	0.16 & 1
Digitizer sampling rate [GSamples]	0.5	1

Table 1. Specifications of large footprint laser systems capturing the waveform: LVIS (Blair *et al.*, 1999), and GLAS (Zwally *et al.*, 2002).

^a Full-width-at-half-maximum

The range estimation is further improved by the comparison between the transmitted and the received waveform. This can be done by signal processing methods (e.g. cross-correlation, inverse filtering), if the sampling of the waveform is done with a high sampling rate. The maximum of the cross-correlation between the transmitted and received signal estimates the range value with a higher reliability and accuracy than considering the received waveform only (Hofton & Blair, 2001; Jutzi & Stilla, 2005; Thiel *et al.*, 2005).

Beside the range determination further surface features can be studied by waveform analysis, namely *reflectance*, *slope* and *roughness*. This specific surface features have an influence on the *amplitude* and *width* of the received waveform (Brenner *et al.*, 2003; Jutzi & Stilla 2002; Steinvall *et al.*, 2004; Wagner *et al.*, 2006). For a parametric description of the pulse properties a Gaussian decomposition method on the waveform can be used (Hofton *et al.*, 2000; Jutzi & Stilla 2005; Persson *et al.*, 2005; Söderman *et al.*, 2005). Nowadays, waveform analysis is more and more established for remote sensing applications especially in forestry (Hug *et al.*, 2004; Reitberger *et al.*, 2006).

Depending on the application different surfaces have to be analyzed, e.g. for urban objects we have to deal with different elevated objects. In rural environment we have to deal with statistically distributed natural objects. The impact of the scene on the received waveform will be discussed using some standard examples (Figure 1). Different elevated object surfaces within the beam corridor lead to a mixture of different range values. A simple situation is given by a horizontal plane surface which will lead to a small pulse (Figure 1b). A plane which is slanted in relation to the viewing direction shows different range values within the footprint. This range interval which is given by the size of the footprint and the orientation of the plane leads to a spread of the pulse width (Figure 1c). A deformation of the pulse form can also be caused by perpendicularly oriented plane surfaces shifted by a small step

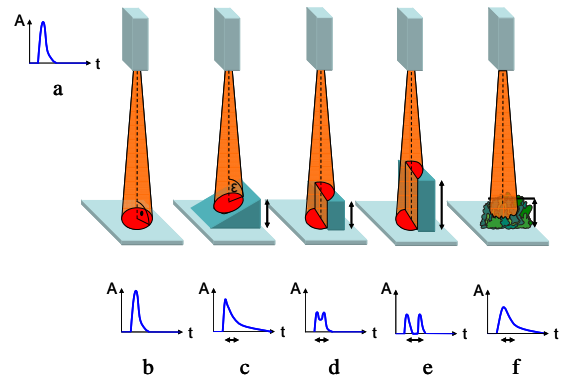


Figure 1. Effects of the surface on the received waveform.

- a) transmitted waveform,
- b) plane surface,
- c) sloped surface,
- d) two slightly different elevated areas,
- e) two significantly different elevated areas,
- f) randomly distributed small objects.

in viewing direction (Figure 1d). A large step leads to two separate pulses (Figure 1e). Several surfaces with different range within the beam can result in multiple pulses. Randomly distributed small objects (e.g. by vegetation) spread over different range values within the beam leads as well to a spread of the pulse width (Figure 1f). These examples show the influence on the waveform by standard surface situations.

Beside the influence of the surface on the waveform the waveform is additionally affected by the used system, especially by the measurement unit itself. For precise waveform analysis it is relevant to consider the system characteristic. This system characteristic can be described by the impulse response of the used laser scanning system. To investigate the impulse response of laser scanning systems experiments are carried out by using receivers with various bandwidths.

In Section 2 an overview on the experimental setup is given. We show a method to estimate the impulse response of the laser scanning system in Section 3. The adaptation of the transmitted waveform by the impulse response and an evaluation for the similarity of the transmitted and received waveform is presented in Section 4. In Section 5 the method is proofed by experiments and results are depicted.

2. EXPERIMENTAL SETUP

An experimental setup was built up for exploring the capabilities of a laser scanning system, which allows capturing the waveform.

Small footprint laser systems	RIEGL LMS-Q560	OPTECH ALTM 3100	TOPEYE II
Wavelength [nm]	1550	1064	1064
Pulse length at FWHM ^a [ns]	4	-	5
Pulse repetition rate [kHz]	Up to 100	50	50
Laser output energy [mJ]	-	-	-
Laser beam divergence [mrad]	≤ 0.5	0.3 or 0.8	1
Operating altitude [m]	< 1500	< 2500	< 1000
Scan angle [°]	± 22.5	± 25	± 20
Detector bandwidth [MHz]	-	-	-
Digitizer sampling rate [GSamples]	1	1	1

Table 2. Specifications of small footprint laser systems capturing the waveform: RIEGL LMS-Q560 (<http://www.riegl.com>), Optech ALTM 3100 with Intelligent Waveform Digitiser (<http://www.optech.on.ca>), and TopEye Mark II (<http://www.topeye.com>).

^a Full-width-at-half-maximum

2.1 Laser system

The laser system has three main components: an emitter unit, a motion control unit, and receiver unit.

2.1.1 Emitter unit

We use a laser pulse system with a pulse duration of 5 ns at full-width-at-half-maximum (FWHM) and a high repetition rate (42 kHz). The average power of the laser is up to 10 kW. The multi-mode Erbium fiber laser operates at a wavelength of 1550 nm with a beam divergence of 1 mrad. The transmitted waveform of the emitted pulse shows strong random modulation for each emitted pulse. In Figure 2 two examples of the transmitted waveform are depicted. The shape of the waveform depends on the design of the laser system, where the system uses a photodiode to pump the multi mode fiber cavity and a fiber amplifier.

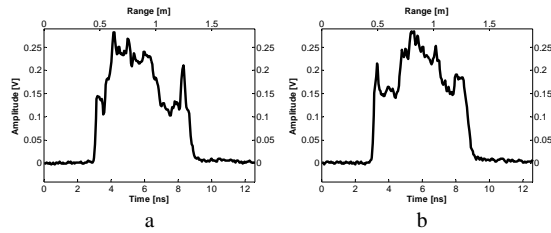


Figure 2. Two samples of the transmitted waveform.

2.1.2 Motion control unit

For the 2-d scanning process a moving mirror is used for an elevation scan with ± 15 degrees in vertical direction (320 raster steps) and a rotation stage for an azimuth scan with 360 degree rotation in horizontal direction (variable number and spacing of the raster steps).

2.1.3 Receiver unit

The receiver unit to capture the waveform usually contains PIN^d or APD^e photodiodes. For our investigations we used receivers with various bandwidths (250 MHz, 750 MHz, 1 GHz, 7 GHz, and 12 GHz) containing photodiodes sensitive at wavelengths of around 900 to 1700 nm. An overview for the specifications of the receivers which were used in the experiment is given by Table 3. Furthermore, we use an A/D converter with 20 GSamples/s. The A/D conversion and digital recording is accomplished by using a digital memory oscilloscope (Le Croy - Wavemaster 8600A), where the bandwidth of the oscilloscope is limited to 6 GHz.

Receivers	TTI-TIA 950 ^a	TTI-TIA 950 ^b	New Focus 1611	TTI-TIA 4000 ^c	New Focus 1544 ^c
3dB bandwidth [MHz]	250	750	1000	7000@6000	12000@6000
Detector Material/Type	InGaAs/PIN ^d	InGaAs/PIN	InGaAs/PIN	InGaAs/APD ^e	InGaAs/PIN
Wavelength [nm]	900-1700	900-1700	900-1700	950-1650	950-1650
Minimum NEP [pW/ \sqrt{Hz}]	3	3	20	16	33
Detector Diameter [μ m]	100	100	100	30	25

Table 3. Specifications of the receivers used in the experiment

^a Gain of 10.0

^b Gain of 1.0

^c Bandwidth is limited by the digital oscilloscope to 6 GHz

^d Positive intrinsic negative diode

^e Avalanche photodiode

3. ESTIMATING THE IMPULSE RESPONSE OF THE SYSTEM

The measured received waveform of the backscattered laser pulse depends on the transmitted waveform $s[t]$ of the emitted laser pulse, the impulse response $h_R[t]$ of the measurement unit for the received waveform, the spatial beam distribution of the used laser $P[x,y]$, and the illuminated surface $S[x,y,z]$. The received waveform $r_M[x,y,z,t]$ can be expressed by a convolution of the relevant terms mentioned above and we get

$$r_M[x, y, z, t] = s[t] * h_R[t] * P[x, y] * S[x, y, z] \quad (1)$$

where (*) denotes the convolution operation. The spatial beam distribution is typically Gaussian, and the surface characteristic can be described by its geometry and its reflectance properties (usually a mixture of diffuse and specular). If the transmitted waveform $s[t]$ is as well measured it can be described by

$$s_M[t] = s[t] * h_S[t], \quad (2)$$

where $h_S[t]$ is the impulse response of the measurement unit of the transmitted waveform. Both impulse responses are mainly affected by the used photodiode(s) and amplifier(s).

Assuming a perfectly flat surface perpendicular to the propagation direction of the laser beam we derive with Equation 1

$$r_M[t] = s[t] * h_R[t]. \quad (3)$$

The measured waveforms $s_M[t]$ and $r_M[t]$ may not necessarily represent the real waveforms of the laser pulses. Depending on the bandwidth of the used photodiodes and amplifiers the waveforms of the used laser system show more or less detailed information. The depicted transmitted waveforms in Figure 2 are recorded with a high bandwidth of 6 GHz and are sampled with 20 GSamples/s. However, we want to characterize the system by its impulse response $h[t]$ independent of the bandwidth of the measurement unit. But, how can we determine the impulse response of the used system?

For estimating the impulse response $h[t]$ of the measurement unit a deconvolution is necessary. The deconvolution can be easily formulated by transforming Eq. 2 and 3 in frequency domain

$$\underline{S}[f] = \frac{\underline{S}_M[f]}{\underline{H}_S[f]} = \frac{\underline{R}_M[f]}{\underline{H}_R[f]}. \quad (4)$$

Terms $\underline{S}[f]$, $\underline{S}_M[f]$, $\underline{R}_M[f]$, $\underline{H}_S[f]$, and $\underline{H}_R[f]$ are the Fourier transforms of the corresponding terms $s[t]$, $s_M[t]$, $r_M[t]$,

$h_s[t]$, and $h_R[t]$.

This can be written as

$$\underline{R}_M[f] = \underline{S}_M[f] \cdot \frac{\underline{H}_R[f]}{\underline{H}_S[f]} = \underline{S}_M[f] \cdot \underline{H}[f], \quad (5)$$

where $\underline{H}[f]$ is the transfer function of the system.

To estimate the transfer function $\underline{H}[f]$ a frequency-domain division of $\underline{R}_M[f]/\underline{S}_M[f]$ has to be carried out. By the inverse Fourier transformation of $\underline{H}[f]$ we obtain $h[t]$.

Usually the measurement of the transmitted and received waveforms is affected by noise from photodiode and amplifier. Then a straightforward direct division operation leads to a noisy transfer function and as well to a noisy impulse response (Figure 3a) without any significant information.

To avoid this, a sample of N single impulse responses is averaged and we receive the averaged impulse response

$$\bar{h}[t] = \frac{1}{N} \sum_{n=1}^N h_n[t], \quad (6)$$

which describes the characteristic of the system. An averaged impulse response ($N=1000$) is depicted in Figure 3b.

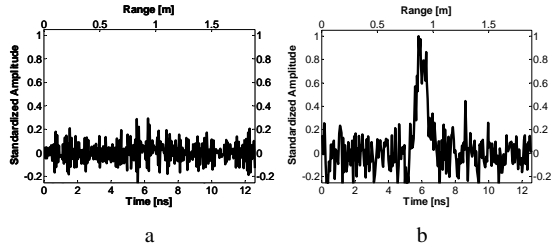


Figure 3. Impulse response of the system:
a) Example of a single impulse response $h_1[t]$,
b) Averaged impulse response $\bar{h}[t]$.

4. ADAPTATION OF THE TRANSMITTED WAVEFORM AND EVALUATION

If the impulse response is determined the transmitted waveform $s_M[t]$ can be adapted to consider the characteristic of the measurement system. Therefore the adaptation of the transmitted waveform is given by

$$s_{\bar{h}}[t] = s_M[t] * \bar{h}[t], \quad (7)$$

where $s_{\bar{h}}[t]$ describes the *adapted transmitted waveform*.

Assuming a perfectly flat surface the shape of the adapted transmitted waveform should be closer to the shape of the received waveform of the backscattered pulse than the transmitted waveform of the emitted pulse. The similarity of the adapted transmitted waveform and the received waveform should be generally increased with this adaptation.

To proof the similarity of the adapted transmitted waveform $s_{\bar{h}}[t]$ and the received waveform $r_M[t]$ the correlation is the favorite method. Therefore the maximum coefficient of the normalized cross-correlation function has to be determined. The normalized cross-correlation function is defined with

$$k'[\tau] = R_{sr}[t + \tau] = \frac{\sum_{t=0}^{M-1} s_{\bar{h}}[t] \cdot r_M[t + \tau]}{\sqrt{\sum_{t=0}^{M-1} s_{\bar{h}}[t]^2 \cdot \sum_{t=0}^{M-1} r_M[t]^2}}, \quad (8)$$

where M is the length of the correlation function $k'[\tau]$ and the maximum correlation coefficient R' is derived by

$$R' = \max(k'[\tau]). \quad (9)$$

To compare the transmitted waveform $s_M[t]$ and the received waveform $r_M[t]$ the correlation coefficient is calculated in the same manner as shown in Equation 8 and 9, where the maximum correlation coefficient is denoted with R . Generally the correlation coefficient is close to 1 for high similarity of the shape of the two waveforms. The result received for R' and R can be compared with each other and evaluated.

5. EXPERIMENTS

To study the impulse response of laser scanning systems experiments are carried out by using receivers with various bandwidths. Figure 4 depicts a schematic description of the processing chain for the experiments.

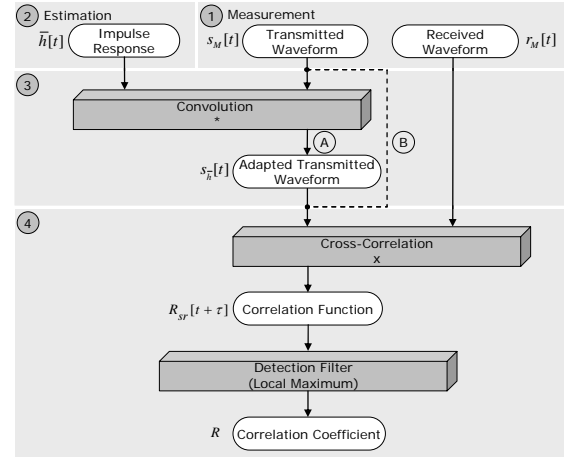


Figure 4. Processing chain.

First the transmitted and the received waveform are measured with the laser scanning system (Figure 4-1) for a sample of $N=1000$ emitted pulses, where the transmitted and the received waveforms are recorded by separate receivers of the same type and with the same bandwidth.

It has to be mentioned that the receiver TTI-TIA 950 operates at bandwidths of 250 and 750 MHz depending on the pre-selected gain factor, which can be set manually. Furthermore, for the experiments there wasn't an equal pair of receivers with the supported 6 GHz bandwidth of the digital memory oscilloscope available. Therefore we selected two different receivers with a higher bandwidth, where the TTI-TIA 4000 (7 GHz) was used to capture the transmitted waveform and the New Focus 1544 (12 GHz) to capture the received waveform of the laser pulse.

The illuminated target is a flat surface with an orientation perpendicular to the laser beam propagation direction and high material reflectance positioned at a range of approximately 80 m. The high material reflectance is necessary to capture the

Receivers	TTI-TIA 950	TTI-TIA 950	New Focus 1611	TTI-TIA 3000 ^a New Focus 1544 ^a
3dB bandwidth [MHz]	250	750	1000	@6000
Ⓐ Average value \bar{R} of the correlation coefficient	0.9734	0.9756	0.9737	0.9688
Ⓑ Average value \bar{R}' of the correlation coefficient	0.9782	0.9817	0.9776	0.9745
Ⓐ Standard deviation σ_R of the correlation coefficient	0.0085	0.0585	0.0072	0.0321
Ⓑ Standard deviation $\sigma_{R'}$ of the correlation coefficient	0.0075	0.0588	0.0068	0.0320

Table 4. Evaluation of the experiments.

^aBandwidth is limited by the digital oscilloscope to 6 GHz

waveform with receivers of low sensitivity. Especially the receivers with bandwidths greater than 1 GHz have a very low sensitivity.

To estimate the impulse response of the system (Figure 4-2) the measured sample is processed as described in Section 3. The average of the single impulse responses derived by the various receiver pairs is depicted in Figure 5. Measurements are carried out for the following bandwidths: 250 MHz, 750 MHz, 1 GHz, and 6 GHz. With increasing bandwidth the width of the impulse response decreases. Caused by the small number of the measured samples ($N=1000$) the estimated impulse response shows still some noise characteristics, especially the impulse responses derived by the receivers with low bandwidths. For further processing only the values greater than zero of the impulse response itself are considered and the background noise is set to zero. Beside this straightforward estimation of the impulse response further investigations in frequency-domain for noise reduction (e.g. low pass filtering) might be of interest for performance optimization, but this was not further investigated in this paper.

However the estimated impulse response is convolved with the transmitted waveform for each emitted pulse to determine the adapted transmitted waveform $s_{\bar{h}}[t]$ (Figure 4-3).

To evaluate the transmitted waveforms $s_M[t]$ and $s_{\bar{h}}[t]$ the correlation coefficient of the transmitted and the received

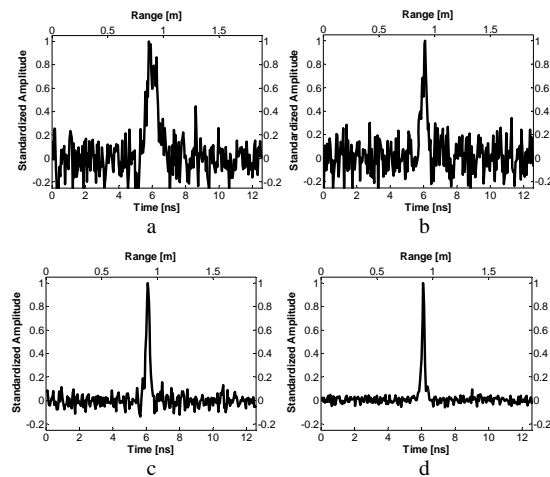


Figure 5. Impulse responses of the laser scanning system for various receivers:
a) 250 MHz bandwidth,
b) 750 MHz bandwidth,
c) 1 GHz bandwidth,
d) 6 GHz bandwidth.

waveform is calculated for each single emitted laser pulse (Figure 4-4). The correlation coefficient is extracted from the correlation function by detecting the local maximum and determines the maximum value. To compare the adapted (Figure 4-A) and the corresponding measured transmitted waveform (Figure 4-B, dashed line), the correlation coefficients are determined and compared with each other. Then the average values \bar{R} (\bar{R}') and the standard deviation values σ_R ($\sigma_{R'}$) of the correlation coefficient R (R') for 1000 samples is calculated for each measurement. The results are presented in Table 4.

For all investigated bandwidths the average value \bar{R}' of the correlation coefficient derived by the adapted transmitted waveform $s_{\bar{h}}[t]$ is higher than the average value \bar{R} derived by the transmitted waveform $s_M[t]$. The standard deviation $\sigma_{R'}$ of the correlation coefficient derived by the adapted transmitted waveform is lower for three of four measurements.

6. DISCUSSION

All experiments are carried out in the same manner, only receivers with various bandwidths are used for each measurement. The similarity of the shape of the transmitted waveform and the received waveform is generally very high, but could be increased with the *adapted transmitted waveform*. It has to be mentioned, that the significance is not as high as we expected. In relation to the amplitude of the estimated impulse response large noise influences could be observed at low bandwidths. With increasing the bandwidth the width of the impulse response decreases. This might depend on the manufacturing accuracy of the used photodiodes. The width of the adapted impulse responses $\bar{h}[t]$ is smaller than the width of the transmitted waveform $s_M[t]$ of the emitted laser pulse (compare Figure 3 and 5). The convolution of the impulse response with the transmitted waveform has a low pass effect on the transmitted waveform.

7. CONCLUSION

In this work it was shown that the waveform is affected by the measurement unit. The influence of the measurement unit on capturing the transmitted and received waveforms of the laser pulse can be described by the impulse response which characterizes the system. To study the measurement unit we determined the impulse response of the laser scanning system for different bandwidths, namely 250 MHz, 750 MHz, 1 GHz, and 6 GHz. Considering the impulse response by calculating the adapted transmitted waveform increases the similarity of the shape of the transmitted and the received waveform. For precise waveform analysis the similarity of both measured waveforms is important to determine the influence of the surface in an accurate way. This is helpful to extract surface features with

high accuracy, e.g. range determination by cross-correlation. The experiments we carried out are general investigations for a laser scanning system which records the full-waveform of laser pulses.

REFERENCES

- Baltsavias EP (1999) Airborne laser scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 164-198.
- Blair JB, Rabine DL, Hofton MA (1999) The Laser Vegetation Imaging Sensor (LVIS): A medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry & Remote Sensing* 56: 112-122.
- Brenner AC, Zwally HJ, Bentley CR, Csatho BM, Harding DJ, Hofton MA, Minster JB, Roberts LA, Saba JL, Thomas RH, Yi D (2003) Geoscience Laser Altimeter System (GLAS) - Derivation of Range and Range Distributions From Laser Pulse Waveform Analysis for Surface Elevations, Roughness, Slope, and Vegetation Heights. Algorithm Theoretical Basis Document - Version 4.1.
- Bufton JL (1989) Laser Altimetry Measurements from Aircraft and Spacecraft. *Proceedings of the IEEE*, Vol. 77, No. 3, 463-477.
- Der S, Redman B, Chellappa R (1997) Simulation of error in optical radar measurements. *Applied Optics* 36 (27), 6869-6874.
- Gardner CS (1982) Target signatures for laser altimeters: an analysis. *Applied Optics*, Volume 21, Issue 3, 448-453.
- Hofton MA, Blair JB (2001) Laser altimeter return pulse correlation: A method for detecting surface topographic change. *Journal of Geodynamics special issue on laser altimetry* 34, 491-502.
- Hofton MA, Minster JB, Blair JB (2000) Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience & Remote Sensing* 38 (4), 1989-1996.
- Hug C, Ullrich A, Grimm A (2004) LITEMAPPER-5600 - a waveform digitising lidar terrain and vegetation mapping system. *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences* 36, Part 8/W2, 24-29.
- Huising EJ, Gomes Pereira LM (1998) Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of Photogrammetry & Remote Sensing* 53: 245-261.
- Irish JL, Lillycrop WJ (1999) Scanning laser mapping of the coastal zone: the SHOALS system. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 123-129.
- Irish JL, McClung JK, and Lillycrop WJ (2000) Airborne lidar bathymetry: the SHOALS system. *PIANC Bulletin*. 2000 (103): 43-53.
- Jelalian AW (1992) *Laser Radar systems*. Norwood, MA, Boston: Artech House.
- Jutzi B, Eberle B, Stilla U (2002) Estimation and measurement of backscattered signals from pulsed laser radar. In: Serpico SB (ed) (2003) *Image and Signal Processing for Remote Sensing VIII*, SPIE Proc. Vol. 4885: 256-267.
- Jutzi B, Stilla U (2003) Laser pulse analysis for reconstruction and classification of urban objects. In: Ebner H, Heipke C, Mayer H, Pakzad K (eds) *Photogrammetric Image Analysis PIA'03*. *International Archives of Photogrammetry & Remote Sensing*. Vol. 34, Part 3/W8, 151-156.
- Jutzi B, Stilla U (2005) Measuring and processing the waveform of laser pulses. In: Gruen A, Kahmen H (eds) *Optical 3-D Measurement Techniques VII*. Vol. I, 194-203.
- Kamermann GW (1993) *Laser Radar*. In: Fox CS (ed) *Active Electro-Optical Systems, The Infrared & Electro-Optical Systems Handbook*. Michigan: SPIE Optical Engineering Press.
- Papoulis A (1984) *Probability, Random Variables, and Stochastic Processes*. Tokyo: McGraw-Hill.
- Persson Å, Söderman U, Töpel J, Ahlberg S (2005) Visualization and analysis of full-waveform airborne laser scanner data. In: Vosselman G, Brenner C (Eds) *Laser scanning 2005*. *International Archives of Photogrammetry & Remote Sensing* 36, Part 3/W19, 109-114.
- Reitberger J, Krzystek P, Heurich M (2006) Full-Waveform analysis of small footprint airborne laser scanning data in the Bavarian forest national park for tree species classification. In: Koukal T, Schneider W (Eds) *3D Remote Sensing in Forestry*. 218-227.
- Skolnik MI (1980) *Introduction to radar systems*. McGraw-Hill International Editions, Second Edition.
- Söderman U, Persson Å, Töpel J, Ahlberg S (2005) On analysis and visualization of full-waveform airborne laser scanner data. *Laser Radar Technology and Applications X*. In: Kamerman W (ed) *SPIE Proc. Vol. 5791: 184-192*.
- Steinval O (2000) Effects of target shape and reflection on laser radar cross sections. *Applied Optics* 39 (24), 4381-4391.
- Steinval O, Carlsson T (2001) Three-dimensional laser radar modeling. In: Kamerman GW (Ed) *Laser Radar Technology and Application VI*, *SPIE Proc. Vol. 4377, 23-34*.
- Steinval O, Larsson H, Gustavsson F, Chevalier T, Persson Å, Klasén L (2004) Characterizing targets and backgrounds for 3D laser radars. *Military Remote Sensing*. In: Kamerman W, Willetts DV (eds) *SPIE Proc. Vol. 5613: 51-66*.
- Thiel KH, Wehr A (2004) Performance Capabilities of Laser-Scanners - An Overview and Measurement Principle Analysis. *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences* 36, Part 8/W2, 14-18.
- Thiel KH, Wehr A, Hug C (2005) A New Algorithm for Processing Fullwave Laser Scanner Data. *EARSel 3D-Remote Sensing Workshop*, CDROM.
- Vandapel N, Amidi O, Miller JR (2004) Toward Laser Pulse Waveform Analysis for Scene Interpretation. *IEEE International Conference on Robotics and Automation (ICRA 2004)*.
- Wagner W, Ullrich A, Briese C (2003) Der Laserstrahl und seine Interaktion mit der Erdoberfläche. *Österreichische Zeitschrift für Vermessung & Geoinformation, VGI* 4/2003, 223-235.
- Wagner W, Ullrich A, Ducic V, Melzer T, Studnicka N (2006) Gaussian Decomposition and Calibration of a Novel Small-Footprint Full-Waveform Digitising Airborne Laser Scanner. *ISPRS Journal of Photogrammetry & Remote Sensing*, 60 (2), 100-112.
- Wagner W, Ullrich A, Melzer T, Briese C, Kraus K (2004) From single-pulse to full-waveform airborne laser scanners: Potential and practical challenges. In: Altan MO (ed) *International Archives of Photogrammetry & Remote Sensing*. Vol. 35, Part B3, 201-206.
- Wehr A, Lohr U (1999) Airborne laser scanning - an introduction and overview. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 68-82.
- Zwally HJ, Schutz B, Abdalati W, Abshire J, Bentley C, Brenner A, Bufton J, Dezio J, Hancock D, Harding D, Herring T, Minster B, Quinn K, Palm S, Spinhirne J, Thomas R (2002) ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics* 34 (3-4), 405-445.

LASER PULSE ANALYSIS FOR RECONSTRUCTION AND CLASSIFICATION OF URBAN OBJECTS

B. Jutzi, U. Stilla

FGAN-FOM Research Institute for Optronics and Pattern Recognition
76275 Ettlingen, Germany
{jutzi,stilla}@fom.fgan.de

Commission III

KEY WORDS: Object recognition, Urban areas, Waveform analysis, Pulsed laser radar.

ABSTRACT:

Current pulsed laser radar systems for topographic applications are based on time-of-flight ranging techniques to determine the range of the illuminated object. The signal analysis to determine the time-of-flight typically operates with analogous threshold detection. In this paper we describe investigations for digital recording of received laser pulses and a detailed analysis of the pulse waveform. Recording the complete signal is advantageous because it allows processing depending on different tasks, respecting intermediate results, and considering neighbourhood relations. Two different techniques for measurement of time resolved laser pulses are presented. An experimental system for a fast recording of signals was constructed. For principal investigations a test board with urban materials was measured by single photon detection technique and visualized by a data cube. Based on these spatio-temporal data, different features are extracted and depicted by grey value images to describe macro, meso, and micro structures. Measurements were carried out with partial occlusion of the test board by vegetation. Multiple reflections are counted by processing the data cube using a spatio-temporal filter and peak detector. Initial results of object segmentation are shown.

1. INTRODUCTION

The automatic generation of 3-d models for a description of man-made objects, like buildings, is of great interest. In photogrammetry a spatial surface is classically determined by triangulation of corresponding image points from two or more pictures of the surface. The points are manually searched or automatically detected by analyzing image structures (e.g. stereo correlation techniques). Besides this indirect measurement using object characteristics which depend on natural illumination, active laser radar systems allow a direct and illumination-independent measurement of the range. The laser radar captures the range information of 3-d objects in a fast, contactless and accurate way. In future applications the reliability of measurements and the capability for object recognition could be increased by analyzing object features.

Current pulsed laser radar systems for topographic applications are based on time-of-flight ranging techniques to determine the range of the illuminated object. Overviews are given in Baltsavias (1999), Huising & Pereira (1998) and Wehr & Lohr (1999). The time-of-flight is measured by the elapsed time between the emitted and backscattered laser pulses. The signal analysis to determine the elapsed time typically operates with analogous threshold detection (e.g. peak detection, leading edge detection, constant fraction detection). Currently first pulse as well as last pulse exploitation are used for different applications (urban planning, forestry surveying, power line monitoring) to capture a three-dimensional scene (digital terrain model, digital surface model, city model).

While first pulse registration is the optimum choice when surveying the top of objects (e.g. canopy), last pulse registration should be chosen if the final elevation model should describe the ground surface. Figure 1a shows a section of an image taken

in first pulse mode in October. The foliage of the trees is visible. Figure 1b was taken in last pulse mode in January. The branches and foliage are not visible and the building areas are smaller than in Figure 1a. Due to multiple pulse reflection at the boundary of the buildings and the processing by first or last pulse mode, building areas dilate or erode. Similar results can be achieved by corresponding rank value filters. In some areas no response could be received due to the weak reflectivity of snow and ice at the operating wavelength of 1,55 μm .

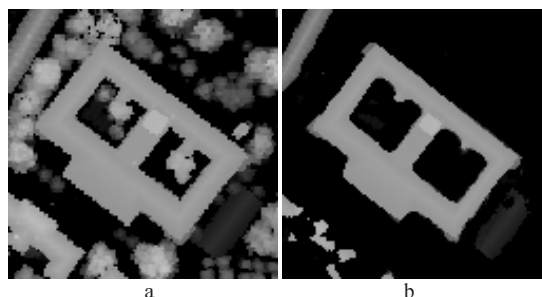


Figure 1. Sections of elevation images of an urban scene (Test area Karlsruhe, Germany). a) first pulse mode, b) last pulse mode.

Beside the first or last pulse exploitation the complete signal waveform in between might be of interest, because it includes the backscattering characteristic of the illuminated field. This enhanced information is helpful to explore the vegetation concerning the bio mass, foliage or density (e.g. trees, bushes, ground). Several groups are conducting research on natural objects. NASA has developed a prototype of a Laser Vegetation Imaging Sensor (LVIS) for recording the waveform to determine the vertical density profiles in forests (Blair *et al.*, 1999). The spaceborne Geoscience Laser Altimeter System

(GLAS) determines distance to the Earth's surface, and a profile of the vertical distribution of clouds and aerosols. In some applications (e.g. observation of climate changes), clouds are objects of interest. In other application clouds are considered as obstacles, which limit the visibility of the illuminated object.

In addition to range, some systems deliver a reflectance value derived from the intensity of the backscattered laser light. This reflectance value can be used for separating segments of artificial objects from vegetation (Hug & Wehr, 1997; Maas, 2001).

In this paper we describe investigations for a detailed analysis of laser pulses. In Section 2 different techniques for measurement of time resolved laser pulses are discussed. The experimental system for a fast recording of signals is described in Section 3. In Section 4 the performed experiments are explained and the measured data are depicted. The determined features of the data are presented in Section 5. First steps for a segmentation of objects are shown in Section 6.

2. MEASUREMENT TECHNIQUES

2.1 Laser radar systems

Depending on the application, laser radar systems can be designed in different ways (Kamermann, 1993). They may differ in techniques concerning e.g. the modulation, detection, or measurement.

Concerning the modulation techniques laser systems can be split in continuous wave (cw) laser or pulsed laser. For applications in remote sensing the pulsed laser with the higher power density compared to cw laser is of advantage, because it allows for operating at long ranges. In this work we focus on pulsed laser systems.

Detection techniques can be divided in coherent detection and direct detection (Jelalian, 1992). Coherent detection considers the wave front of the received signal compared to a reference signal emitted from a cw laser. In direct detection laser systems the received optical energy is focused onto a photosensitive element that generates a value that depends on the optical power. The possibility to detect multiple or single photons is described in the next section.

Measurement techniques for range determination can be distinguished by the exploited signal properties such as phase, amplitude, frequency, polarization, time, or any combination of them. We are interested in measuring and analyzing the received pulse waveform, i.e. the dependence of the intensity over time.

2.2 Recording the signal waveform

Two detection techniques seem to be appropriate to record the temporal characteristic of the backscattered signal: multi photon detection (Figure 2) and single photon detection (Figure 3).

2.2.1 Multi photon detection: The classical measurement technique for direct detection operates with a photodiode (usually PIN- or avalanche photodiode). The photodiode generates an electrical signal (voltage or current) that is directly proportional to the optical power of the incident light (multi photon). For a detailed analysis of the signal waveform a digitizing receiver unit is essential. Analyzing the signal

waveform of the emitted short duration laser pulse with only a few nanoseconds pulse width requires a receiver unit processing the data with a bandwidth of some GHz and an appropriate sampling rate. Increasing bandwidth results in decreasing sensitivity of the photodiode which can be compensated by increasing power of the emitting laser source. Figure 4a shows a pulse digitized by 5 GSample/s.

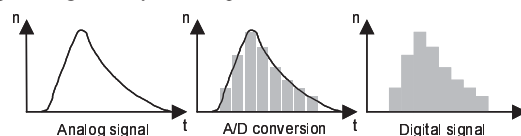


Figure 2. Digital recording of the signal waveform with multi photon detection

2.2.2 Single photon detection: The integration of the incoming radiation by multi photon detection can be seen as collecting a countable number of photons for a limited observation time of one pulse. Let us assume a stationary scene and a stationary sensor platform. In this case the statistical properties of the laser radiation do not change with the time and time-average quantities are equal to the ensemble quantities. Under these assumptions the radiation ensembles are stationary and ergodic (Papoulis, 1984; Troup, 1972). This means that counting single photon events with their time-of-flight into time bins of a histogram is closely related to integration of multi photons over the time (Alexander, 1997; Gagliardi & Karp, 1976; Loudon, 2000). For receiving the temporal waveform of the pulse, a histogram of the photon arrivals over time has to be measured.

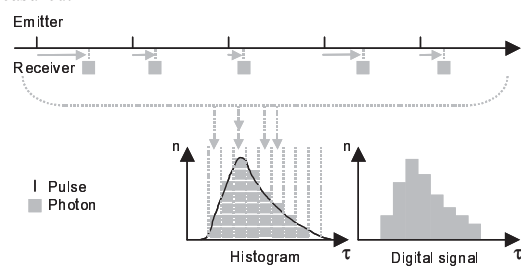


Figure 3. Digital recording of the signal waveform with single photon detection

Many pulses are necessary to obtain a signal waveform with single photon detection. The quality of the sampled signal waveform depends on the number of photon counts. Figure 4b shows a pulse plotted from a histogram containing the time-of-flight measurements from 6135 photons distributed in 25 bins (bin width $\hat{=} 40$ ps).

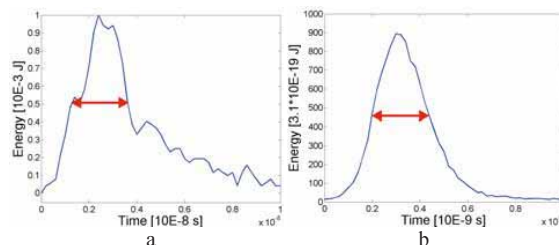


Figure 4. Examples for pulses backscattered from a diffuse surface. a) multi photon detection (FWHM = 2,3 ns), b) single photon detection (FWHM = 0,3 ns)

The full-width-half-maximum (FWHM) of the pulse in Figure 4a is about eight times of the pulse in Figure 4b. For experimental investigations on the signal waveform a short duration laser pulse is helpful. In contrast to previous work (Jutzi *et al.*, 2003), this paper focusses on single photon detection.

3. EXPERIMENTAL SETUP

An experimental setup was constructed for exploring the capabilities of recognizing urban objects using a laser system. For principle investigations of the influence of different object properties (e.g. geometry, material) on the signal waveform a pulsed laser system with single photon detection is used. In contrast to single photon approaches which exploit only a few photons for a range measurement (Ho *et al.*, 1999; Priedhorsky *et al.*, 1996), we are interested in receiving more photons to obtain a detailed distribution for analyzing the signal waveform.

The measurements are carried out indoors by an experimental setup which consists of a laser system and a target with different urban materials.

3.1 Laser system

The laser system consists of three main components: (i) an emitter unit, (ii) a receiver unit, and (iii) a motion control unit.

We use a short duration laser pulse system with a high repetition rate. The pulsed diode laser operates at a wavelength of 640 nm. The average power of the laser is up to 30 mW and pulse duration with less than 70 ps FWHM. The user-selectable repetition rate is from 2.5 MHz to 80 MHz and allows an unambiguous range measurement up to 60 m. This maximum unambiguous range is given by the slowest repetition rate and the time-of-flight of the laser pulse. The receiver unit consists of a photomultiplier tube which detects up to 3 million single photon events per second and a device which exploits the received data and computes a histogram. The receiver unit limits the temporal resolution to a minimum pulse width of 300 ps (FWHM) and the minimum recording time is 1 ms per measurement of the signal waveform. For the 2-d scan, a mirror, two linear actuators, and a motion controller are used to redirect the laser beam. The angle accuracy of the motion control unit is approximately 20 μ rad. All units are synchronized and controlled with a PC.

3.2 Targets

Depending on the application different properties of urban surfaces can be sensed by a laser system. According to the size of the focused surface geometry in relation to the beam (footprint and wavelength) we divide surface structures into macro, meso, and micro structures (Figure 5).

3.2.1 Macro Structure: We interpret macro structures as structures which are much more extended than the beam footprint d . Laser range measurements taken by a scan allow a reconstruction of large object structures like different roof shapes (e.g. flat roof, gabled roof, hip roof, etc.). The scan is typically spaced greater than the spatial beam width for aerial survey purposes.

3.2.2 Meso Structure: We interpret meso structures as structures which are much less extended than the beam footprint d and much greater than the wavelength λ . Different elevated

object surfaces within the beam corridor lead to a mixture of different range values. This may be caused by small elevated objects (e.g. a chimney), a slanted plane, or vegetation like branches or leaves of a tree.

3.2.3 Micro Structure: We interpret micro structures as structures which are less extended than the wavelength λ . Depending on the roughness the irradiance is more or less reflected. A single measurement of the backscattered beam intensity (amplitude) gives information about the roughness of the surface and the material. Small roughness results in specular reflectance and large roughness in diffuse reflectance.

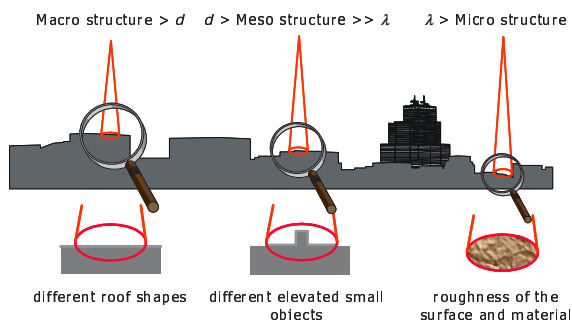


Figure 5. Different levels of details sensed by a laser system

4. EXPERIMENTS

4.1 Test board

According to the focused structure, class relevant signal features have to be examined. For this purpose a test board was designed which consist of different urban materials (Figure 6): corrugated iron (left above), pebbles (background), slanted slate plate (right above), and flat roof tile (right below). The materials have different backscattering characteristics and are mounted in different ranges.

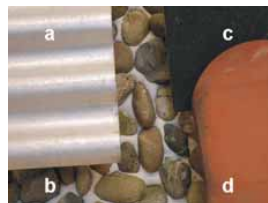


Figure 6. Test board with different urban materials.

a) corrugated iron, b) pebbles, c) slanted slate plate, d) flat roof tile

4.2 Scan

The test board was illuminated by a pulsed laser with a wavelength of approximately 640 nm and a footprint of approximately 5 mm (beam divergence 1 mrad). The time-of-flight distribution of the received photon events (cf. Figure 4a) was recorded with a temporal resolution of 40 ps and 4096 bins which results in a range interval of approximately 25 m and a range resolution of 6 mm. A section of the test board (approximately 0.4 m x 0.3 m) was scanned by 128 columns steps and 128 rows steps.

4.3 Visualization

Displaying the time-of-flight as distance the data set can be seen as a spatial distribution. Assuming Cartesian coordinates the spatial distribution may be visualized by a cube (x,y,t) . Figure 7 shows a section of the cube ($\Delta t=32$ bins), where the grey values correspond to the intensity of the signal (number of photons). The corrugated iron is on the left side and the roof tile on the right side of the cross range plane. The pebbles are inside the selected cube. The temporal dependency of the uppermost row and the right column on the signal is depicted on top plane and right plane. For visualization purposes the dynamic level was adjusted.

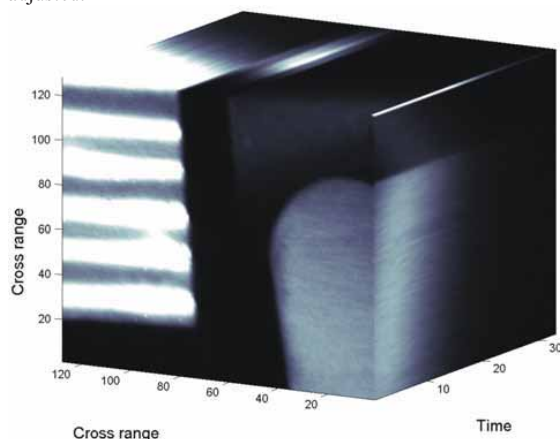


Figure 7. Visualization of time-dependent intensity values of the scanned test board

5. ANALYSIS

5.1 Pulse analysis for macro structures

For measuring macro structures with extended surfaces we assume to receive single pulses, i.e. we neglect the boundaries (discontinuities) within the footprint. Storing the data cube opens the possibility of using different ranging techniques to exploit the received pulses $s(t)$, e.g. peak detection, leading edge detection, constant fraction detection, average time value detection (Figure 8).

5.1.1 Peak detection: The range value is determined by the maximum pulse amplitude, where the highest reflectance is expected (Figure 8a). Regional spikes on the pulse waveform strongly effect the range detection. We processed the range values without temporal filter. For noisy signals, a filter is recommended to determine the global maxima.

5.1.2 Leading edge detection: The threshold crossing by the exceeding pulse waveform determines the range (Figure 8b). The threshold value can be a predefined fixed value, but then the ranging detection strongly depends on the pulse waveform, amplitude, and width of the pulse. We used half of the maximum amplitude of the pulse for a threshold with linear interpolation for range determination. This ranging detection gives the shortest range values.

5.1.3 Constant fraction detection (CFD): The pulse waveform $s(t)$ is inverted and delayed by a fixed time τ and added to the original pulse (Figure 8c). The combined signal $c(t)$ give a constant fraction signal with a zero crossing point,

which is insensitive to the pulse amplitude, but depends on the pulse waveform and width (Kamermann, 1993). We used 300 ps (FWHM) for delay time τ .

$$t_{CFD} \Leftrightarrow c(t) = s(t) - s(t + \tau) = \begin{cases} t_{CFD} & \text{if } c(t) = 0 \\ 0 & \text{if } c(t) \neq 0 \end{cases} \quad (1)$$

5.1.4 Average time value detection: The temporal centre of gravity of the pulse waveform is determined (Figure 8d). It delivers good results for small noise, various pulse amplitudes and pulse widths. The asymmetric pulse shape results in range detection, which is slightly behind the range value detected with the peak detection.

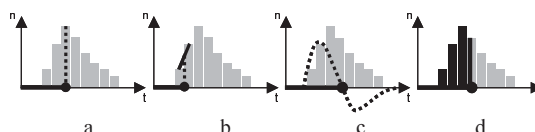


Figure 8. Ranging techniques. a) peak detection, b) leading edge detection, c) constant fraction detection, d) average time value detection

The range images processed with the different ranging techniques are depicted in Figure 9. The range is visualized by grey values whereas the slanted slate plate (right above) shows increasing values and the corrugated iron (left) shows wavy grey values. For visualization purposes the temporal jitter of the different techniques is eliminated by adjusting the average value of the complete range images to each other. The range resolution with peak detection in Figure 9a appears coarser than in the other Figures. The smoothest range resolution is processed with average time value detection with Figure 9d.

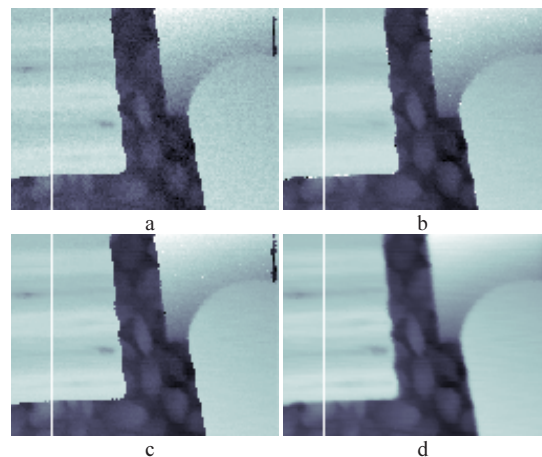


Figure 9. Range images. a) peak detection, b) leading edge detection, c) constant fraction detection, d) average time value detection

In Figure 10a range profiles taken from range images in Figure 9 (see vertical line) are depicted. The leading edge and the constant fraction detection show small systematic ranging errors for different pulse widths. Peak detection appears noisy. Discontinuities in the waveform are a general problem for all ranging techniques. Depending on the signal quality, a spatio-temporal filter might be helpful to increase the precision of range detection.

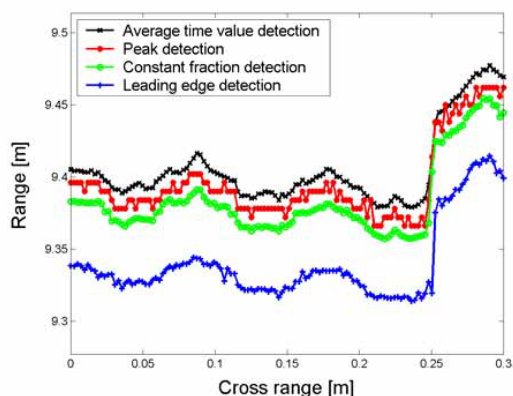


Figure 10. Range profiles

5.2 Pulse analysis for meso structures

Different elevated object surfaces within the beam corridor lead to a mixture of different range values. A plane which is slanted in viewing direction shows different range values within the footprint. This range interval which is given by the size of the footprint and the orientation of the plane leads to a temporal spread of the pulse (Figure 11b). A deformation of the pulse waveform can also be caused by perpendicularly oriented plane surfaces shifted by a small step in viewing direction (Figure 11c). A large step leads to two separate pulses (Figure 11d). Several surfaces with different range within the beam result in multiple pulses. These examples are simplified and not exhaustive.

The increased pulse width indicates an uneven, slanted or small stepped surface. A peak detector is used to separate a single pulse from multiple pulses in the data cube. The pulse width is calculated as standard deviation from the time-of-flight values of the photons. In Figure 12 the pulse widths of maximum pulses are depicted by different grey values. The slanted slate plate (top right) and the contours of the pebbles (centre) appear brighter than the surroundings.

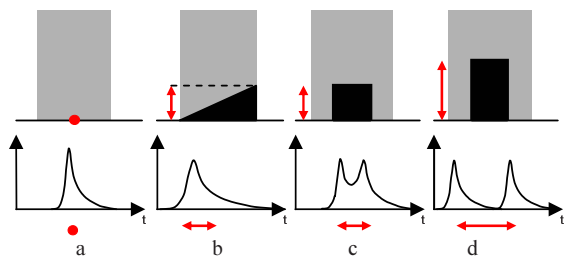


Figure 11. Surface and pulse waveform. a) flat surface, b) slanted surface, c) small step, d) large step



Figure 12. Pulse width image (Angle of the slate plate: 33°).

5.3 Pulse analysis for micro structures

Depending on the material and the surface roughness (micro structure) objects show different reflectance properties. The amount of backscattered photons varies with these properties. The received maximal pulse amplitude and pulse power P

$$P = \frac{1}{2T} \int_{-T}^T s^2(t) dt \quad (2)$$

can be used to discriminate different materials. Figure 13 depicts the pulse power visualized by different grey values. The horizontally oriented areas of the corrugated iron (left) shine bright caused by specular reflectance. The slate plate (top right) appears black caused by the strong absorption.



Figure 13. Pulse power image (Gamma corrected)

6. MULTIPLE PULSES

Both first pulse and last pulse detection is used for photogrammetric applications which allows to consider or neglect the presence of vegetation (see Figure 1). In general, from vegetation a various number of pulse reflections can be expected depending on the density and structure. The captured data cube opens the possibility to analyze the data concerning multiple pulses. This might be of interest for discriminating different types of vegetation. Multiple reflection can also be observed for urban structures which are smaller than the footprint (e.g. power lines).

A scene containing vegetation which partly occludes the test board is shown in Figure 14a. Figure 14b depicts the temporal signal waveform of a single measurement. The multiple pulses were caused by branches in different ranges within the beam. Multiple reflections are counted by a peak detector. Figure 14c shows the number of counted pulses by different grey values. The dark areas result from the test board which reflects background as a single pulse. The branches of the foreground cause additional pulses and appear brighter. More than two pulses can be observed at the crossing of branches.

Due to the noise the measured signals were filtered in the data cube by a spatio-temporal Gaussian implement by a 3x3x3 filter. The result of a peak detection is visualized by a 3-d view in Figure 15. A simple segmentation was performed by a threshold operation on the range to separate the branches against the test board.

7. CONCLUSION

The recording of the signal waveform has different advantages for the analysis:

- The recorded data cube of the received signals allows exploiting different features without selecting a special feature in advance.

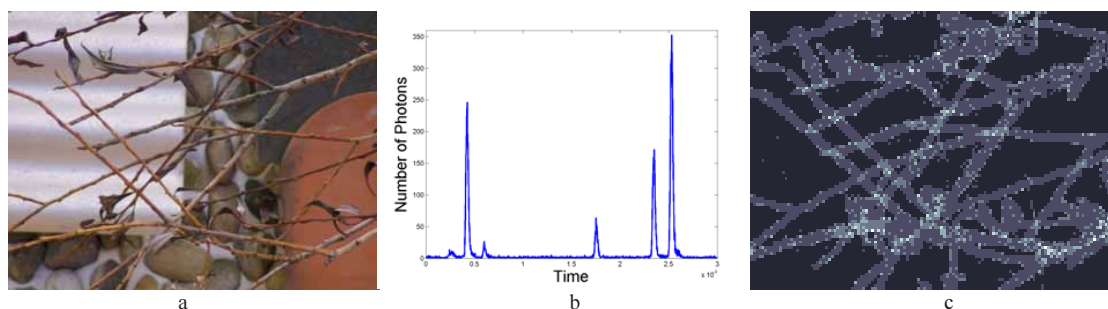


Figure 14. Multiple pulses. a) vegetation (branches) in front of the test board, b) signal waveform of a single measurement (Number of pulses: 6), c) number of pulses visualized by different grey values

- For processing the data cube an adaptive threshold operation is possible.
- Object recognition from laser data often requires further image processing steps like filtering or segmentation. Instead of exploiting neighbourhood relations within the images, the data cube opens the possibility of a spatio-temporal filtering or segmentation.

For single photon detection we assumed a stationary scene and sensor platform. In our experiments the acquisition time for recording a received pulse by single photon counting and by using a low power laser was about 100 ms. Other authors, e.g. (Pellegrini et al., 2000), carried out a range measurement in 10 ms. The experimental system we used is limited to an acquisition time of 1 ms. Applications using an airborne system require a shorter acquisition time.

The experiments we carried out are general investigations for a laser system which operates with multi photon detection. Future work focuses on a fast digitizing acquisition of single laser pulses with multi photon detection and the analysis of the received signal.

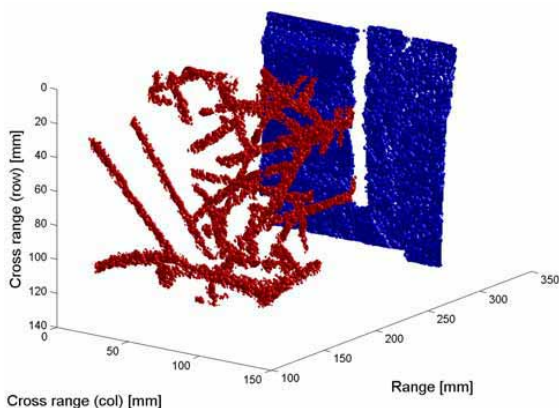


Figure 15. 3-d visualization of segmented objects (branches and test board)

REFERENCES

- Alexander SB (1997) Optical Communication Receiver Design. In: SPIE Tutorial Texts in Optical Engineering, vol. TT22, Washington, Bellingham: SPIE
- Baltsavias EP (1999) Airborne laser scanning: existing systems and firms and other resources. ISPRS Journal of Photogrammetry & Remote Sensing 54: 164-198
- Blair JB, Rabine DL, Hofton MA (1999) The Laser Vegetation Imaging Sensor (LVIS): A medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography. ISPRS Journal of Photogrammetry & Remote Sensing 56: 112-122
- Gagliardi RM, Karp S (1976) Noncoherent (Direct) Detection. Optical Communications. New York: John Wiley & Sons
- Ho C, Albright KL, Bird AW, Bradley J, Casperson DE, Hindman M, Priedhorsky WC, Scarlett WR, Smith RC, Theiler J, Wilson KS, (1999) Demonstration of Literal Three-Dimensional Imaging. Applied Optics 38: 1833-1840
- Hug C, Wehr A (1997) Detecting and identifying topographic objects in laser altimeter data. ISPRS, International Archives of Photogrammetry & Remote Sensing, Vol. 32, Part 3-4W2: 19-26.
- Huising EJ, Gomes Pereira LM (1998) Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. ISPRS Journal of Photogrammetry & Remote Sensing 53: 245-261
- Jelalian AW (1992) Laser Radar systems. Norwood, MA, Boston: Artech House
- Jutzi B, Eberle B, Stilla U (2003) Estimation and measurement of backscattered signals from pulsed laser radar. In: Serpico SB (ed) Image and Signal Processing for Remote Sensing VIII, SPIE Proc. Vol. 4885: 256-267
- Kamermann GW (1993) Laser Radar. In: Fox CS (ed) Active Electro-Optical Systems, The Infrared & Electro-Optical Systems Handbook. Michigan: SPIE Optical Engineering Press
- Loudon R (2000) The Quantum Theory of Light. Oxford: Oxford University Press
- Maas HG (2001) The suitability of airborne laser scanner data for automatic 3D object reconstruction. In: Baltsavias EP, Gruen A, Van Gool L (eds) Automatic Extraction of Man-Made Objects From Aerial and Space Images (III), Lisse: Balkema
- Papoulis A (1984) Probability, Random Variables, and Stochastic Processes. Tokyo: McGraw-Hill
- Pellegrini S, Buller GS, Smith JM, Wallace AM, Cova S (2000) Laser-based distance measurement using picosecond resolution time-correlated single-photon counting. Measurement Science and Technology 11: 712-716
- Priedhorsky WC, Smith RC, Ho C (1996) Laser ranging and mapping with a photon-counting detector. Applied Optics 35: 441-452
- Troup GJ (1972) Photon Counting and Photon Statistics. In: Sanders JH, Stenholm S (eds) Progress in Quantum Electronics, vol. 2 part 1, Oxford: Pergamon
- Wehr A, Lohr U (1999) Airborne laser scanning – an introduction and overview. ISPRS Journal of Photogrammetry & Remote Sensing 54: 68-82



ELSEVIER

Available online at www.sciencedirect.com

 ScienceDirect

PHOTOGRAMMETRY
& REMOTE SENSING

ISPRS Journal of Photogrammetry & Remote Sensing 61 (2006) 95–107

www.elsevier.com/locate/isprsjprs

Range determination with waveform recording laser systems using a Wiener Filter

Boris Jutzi ^{a,*}, Uwe Stilla ^b

^a *FGAN-FOM Research Institute for Optronics and Pattern Recognition, 76275 Ettlingen, Germany*

^b *Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, 80290 Muenchen, Germany*

Received 28 December 2005; received in revised form 31 August 2006; accepted 1 September 2006

Abstract

Current pulsed laser scanning systems determine the range to an object surface by a time-of-flight measurement. Critical measurement situations occur in discriminating the ranges of surfaces close to their edges or of small objects within the beam footprint which are closely located in range. Capturing the complete waveform of the laser pulse allows discriminating differences in a range smaller than the length of the laser pulse. The capabilities of this technique can be predicted by modeling the emitted pulse, the surface, and the backscattered pulse. Due to the varying waveforms of the emitted pulses each individual emitted pulse is recorded and considered for the determination of the surface features. A deconvolution is used to remove the characteristic of the transmitted waveform from the received waveform to obtain a surface response. A Wiener Filter reduces the noise of the determined surface response. For extraction of temporal position, length, and amplitude the corresponding surface features are approximated by Gaussians using the Levenberg–Marquardt Method. Experiments have shown that a stepped surface within the beam with a step smaller than ten times of the pulse length can be distinguished.

© 2006 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Keywords: Laser scanning; Waveform analysis; Signal processing; Feature extraction

1. Introduction

The automatic generation of 3-d models describing man-made objects like buildings is of great interest in photogrammetric research (Stilla et al., 2005). In photogrammetry the distance to a surface is classically derived from a triangulation of corresponding image points from two or more pictures of the surface. The points are chosen manually or detected automatically by analyzing image structures. Besides this indirect mea-

surement using object characteristics, which depends on natural illumination, active laser scanning systems allow a direct and illumination-independent measurement of the range. Laser scanners capture the range of 3-d objects in a fast, contactless, and accurate way. An overview of airborne laser scanning systems is given in Huising and Gomes Pereira (1998), Wehr and Lohr (1999), Baltsavias (1999).

Current pulsed laser scanning systems for topographic mapping are based on time-of-flight ranging techniques to determine the distance to the illuminated object. The time-of-flight is derived by the elapsed time between the emitted and backscattered laser pulses. The signal analysis to determine this time typically operates with analog threshold

* Corresponding author. Tel.: +49 7243 992 337; fax: +49 7243 992 299.

E-mail address: boris.jutzi@fom.fgan.de (B. Jutzi).

0924-2716/\$ - see front matter © 2006 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

doi:10.1016/j.isprsjprs.2006.09.001

detection (e.g. peak detection, leading edge detection, constant fraction detection) (Jutzi and Stilla, 2003). A few systems capture multiple reflections caused by objects at different ranges with surfaces smaller than the footprint of the beam. Most systems capture the first and the last backscattered laser pulse. First pulse as well as last pulse exploitation is used for different applications like urban planning or forestry surveying. While first pulse registration is the optimal choice to measure the hull of partially penetrable objects (e.g. canopy of trees), last pulse registration should be chosen to measure non-penetrable surfaces (e.g. ground surface). Fig. 1a shows a section of an image taken in first pulse mode. The foliage of the trees is visible. Fig. 1b was taken in last pulse mode. The branches and foliage are not visible anymore.

Critical measurement situations can occur if a single pulse is strongly deformed or more than one pulse is

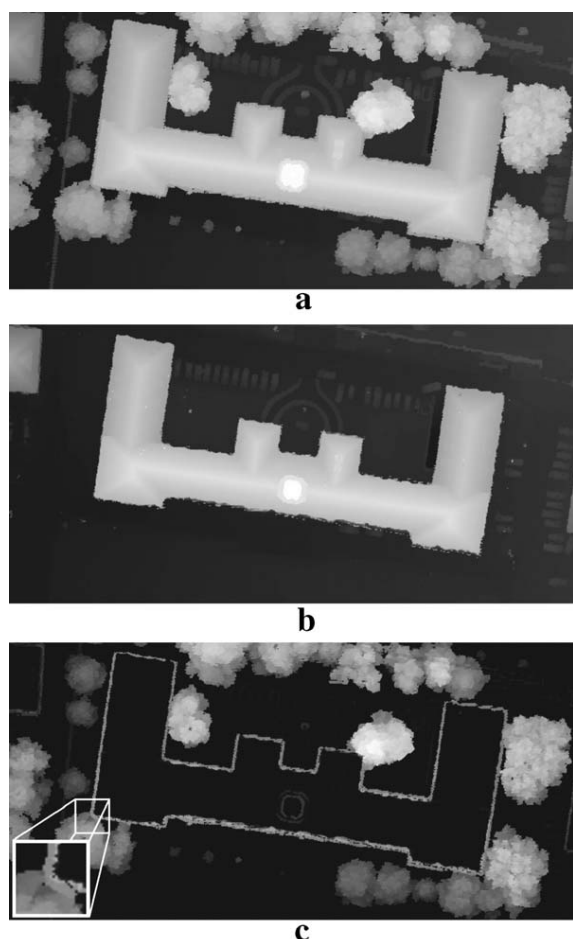


Fig. 1. Section of an urban scene. a) Elevation images captured by first pulse mode, b) elevation images captured by last pulse mode, c) difference image of first and last pulse mode.

backscattered (Fig. 2). The following examples give a selection of these situations:

- i. A sloped surface covers a range interval within the footprint of the beam and leads to a deformation (widening) of the backscattered pulse depending on the slope and the size of the footprint. Analog detectors, which are for example measuring the leading edge of the pulse, only find a single range value which is typically shifted from the mean distance of the surface towards the sensor.
- ii. Two different elevated areas within the footprint of the beam, as occurring at building walls (ground and roof) lead to two pulses with a temporal offset. In this case, analog detectors can measure two different range values, but the coverage of the areas within the footprint cannot be determined without any further information (Vosselman, 2002). Depending on the processing of first or last pulse data, building areas dilate or erode. To visualize the various sizes of the building footprints in first and last pulse image, a difference image was calculated (Fig. 1c). The ambiguous pixels of the building are visible as a bright stripe along the building contours. A section of this ambiguous area was enlarged and is depicted in Fig. 1c (bottom left).
- iii. Two areas at slightly different elevations within the footprint of the beam will lead to two superimposed backscattered pulses (Wagner et al., 2006) if the range difference of the areas is smaller than the length of the laser pulse (Katzenbeisser, 2003). In general an analog detector measures a single range value.
- iv. Small objects within the footprint distributed randomly around a mean elevation value (e.g. average height of crop plants on a field) lead to a widening of the backscattered pulse depending on this distribution. This results in a widened pulse with a low intensity which is difficult to detect with a fixed threshold of an analog system.

These examples suggest that the complete waveform in between the first pulse and last pulse might be of interest, because it includes the backscattering characteristic of the illuminated field.

Prior work on analyzing the full waveform was carried out by NASA to examine vegetation with respect to bio mass, foliage, or density (e.g. trees, bushes, and ground). The Laser Vegetation Imaging Sensor (LVIS) was used to record the waveform and to determine the vertical density profiles in forests (Blair et al., 1999). The spaceborne Geoscience Laser Altimeter System (GLAS)

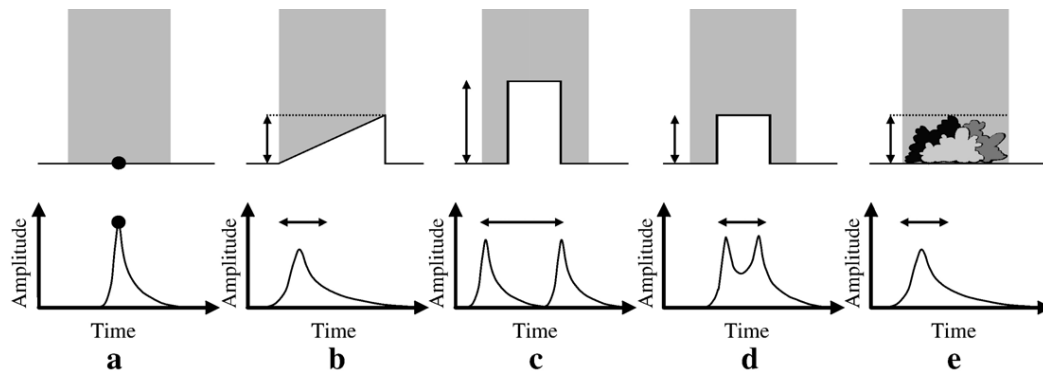


Fig. 2. Surface characteristic and pulse form. a) Plane surface, b) sloped surface, c) two significantly different elevated areas, d) two slightly different elevated areas, e) randomly distributed small objects.

determines the distance to the Earth's surface, and a profile of the vertical distribution of clouds and aerosols (Brenner et al., 2003). In the work of both systems the surface characteristics were determined by comparing a parametric description of the transmitted and received waveform. The shape of the surface response was not determined.

Apart from the range measurement of laser scanning systems, some systems additionally deliver a single reflectance value derived from the amplitude or the power of the backscattered laser light. The amplitude is defined as the signal maximum and the power by signal integration of the measured laser light. These values give radiometric information about the surveyed area. This amplitude value or power value can be used for separating segments of artificial objects from vegetation (Hug and Wehr, 1997; Maas, 2001), classify individual trees or forest stands according to species (Moffiet et al., 2005) or to perfectly texture 3-d scene models (Sequeira et al., 1999). Vosselman (2002) suggested to use the intensity of the laser beam response in order to estimate and improve the position of the edge in between areas with different reflectance properties. Examining the power of the backscattered pulse allows increasing the position accuracy of an object edge, too (Jutzi et al., 2005).

To enable an interpretation of the measured waveform of a backscattered pulse, understanding the physical background of pulse propagation and surface interaction is necessary (Der et al., 1997; Jutzi et al., 2002; Steinvall and Carlsson, 2001; Wagner et al., 2004, 2006). Especially in the case of multiple backscatters, the received waveform may be complex. For analysis of the waveform and to separate fine structures, models for pulse propagation and surface interaction have to be introduced.

In this paper, we propose a method for a detailed analysis of the full waveform of laser pulses. The

presented approach improves the range resolution and allows discriminating multiple surface responses. In Section 2, the waveform of the emitted and backscattered pulse, the laser beam, and the illuminated surface are modeled for capturing the measurement situation. The algorithms for gaining and discriminating surface responses are described in Section 3. In Section 4 outdoor experiments with different surface configurations are presented. Finally in Section 5, the received discrimination results of the surface responses are discussed concerning accuracy and reliability.

2. Modeling

In this section, a model for the waveform of the backscattered laser pulse is derived. This waveform depends on the transmitted waveform of the emitted laser pulse, the spatial energy distribution of the beam, and the material and geometric reflectance properties of the surface.

2.1. Modeling the transmitted waveform of the emitted pulse

Depending on the laser system, the waveform of a pulse may appear in different shapes. Different models of the waveform are known from the literature. Brenner et al. (2003) proposed a simple temporal symmetric Gaussian distribution for modeling the waveform of the spaceborne Geoscience Laser Altimeter System (GLAS). A waveform with an exponential distribution is applied by Steinvall (2000), while Wagner et al. (2004) uses a rectangular distribution. The laser source's modulation effect was not captured in the mentioned models. Some laser systems, like the used multi mode Erbium fiber laser, typically show variations of the

waveform, which cannot be neglected for a detailed analysis. Fig. 3a–c gives examples of emitted pulses of our laser system.

Generally, the basic waveform depends on the pulse generation process of the laser source. A strong random intensity fluctuation is caused by the mode-beating of the multi mode laser light, which is generated by interference of neighboring longitudinal modes within the laser resonator. Furthermore, external disturbances, fluctuations of relaxation oscillation, and inharmonic oscillation (spiking) randomly modulate the waveform.

The basic waveform $s(t)$ of the used laser system can be described by a time delayed Gaussian with the amplitude a and the pulse length w :

$$s(t) = \frac{2a}{w} \cdot \sqrt{\frac{\ln 2}{\pi}} \cdot \exp\left\{-4 \cdot \ln 2 \cdot \frac{(t-\tau)^2}{w^2}\right\}. \quad (1)$$

The length of a pulse is defined by one-half of the pulse's maximum amplitude, known as Full-width-at-half-maximum (FWHM).

The random modulation of the basic waveform, caused by intensity fluctuations, is modeled by noise following a Gaussian distribution $m(t)$. This multiplicative component is defined by the parameters μ for the modulation offset and σ representing the modulation standard deviation. Finally, the modulated waveform $s_m(t)$ is

$$s_m(t) = s(t) \cdot m(t) \text{ with } m(t) \sim N(\mu, \sigma^2). \quad (2)$$

Other laser systems may generate waveforms which appear with different shapes, e.g. a Q-switched laser with an exponential shape. The exponential waveform can be described by $s(t)=t^2 \cdot \exp\{-t/w\}$. If the shape of the transmitted waveform differs from the presented the model has to be adapted.

2.2. Modeling the spatial energy distribution of the emitted pulse

The spatial energy distribution of a laser pulse (beam profile) depends on the used pump source, the optical resonator, and the laser medium. Often, these profiles are modeled by a cylindrical distribution (top-hat form) or by a 2d-symmetric Gaussian distribution (Kamer-mann, 1993). Measurements of the beam profile in the near field have shown that a cylindrical distribution fits our data best (Fig. 4). For the treatment of monostatic systems it is convenient to use a spherical coordinate system with origin in the detector and emitter lens. We take the optical axis as the polar axis of the coordinate

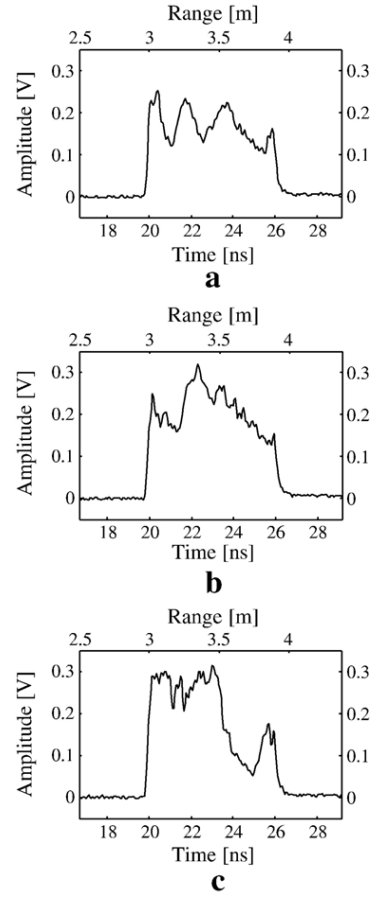


Fig. 3. Examples of the transmitted waveform.

system and call the range r , the zenith angle ε , and the azimuth angle α . In this case the spatial irradiance distribution K of the laser beam over the angle ε is described by

$$K(\varepsilon) = \begin{cases} E \approx \varepsilon_0^2 \pi & \text{if } 0 \leq \varepsilon \leq \varepsilon_0 = \text{beam radius,} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

2.3. Modeling of the surface

In this section, we describe the influence of the surface on the laser beam by geometrical shape and material properties. Capturing the interaction of the emitted pulse with the illuminated surface is relevant to analyze the received waveform. The effects on the observed data of the surface structure are dependent on their scale. Structures with scales below the laser wavelength may result in interference effects (speckle). If the receiver aperture is large compared to the correlation size of the speckle pattern, the intensity fluctuations due to speckle will be smoothed by aperture

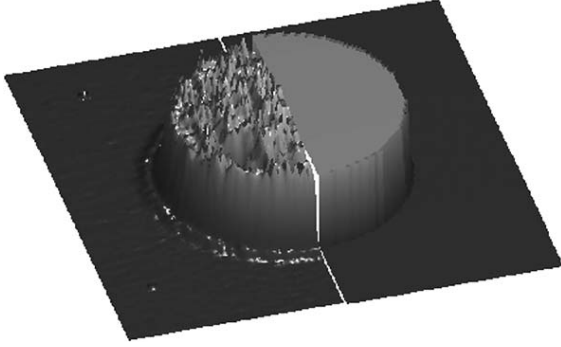


Fig. 4. Spatial energy distribution of the system: measured profile (left) and modeled profile (right).

averaging. In this case only insignificant fluctuations of the waveform with slightly lower amplitude are measured. Because these conditions are often met, speckle effects will be neglected in this paper.

2.3.1. Geometrical reflectance

For handling the imaging properties of extended surfaces, we assume the following: (i) object and sensor are stationary, (ii) the illuminated part of the object is locally plane, (iii) the distance from the sensor is very large compared to the beam footprint. In this case we assume for the area of the footprint that the plane surface differs from the corresponding sphere only marginally. Let us denote the range of a unbounded surface A along the optical axis by r_{0A} , the angle between the surface normal and the optical axis by φ , the slope direction by α_A , and the beam radius by b . The observation geometry is depicted in Fig. 5a and b. The range $r_A(\varepsilon, \alpha)$ of an illuminated surface point depends on zenith angle ε and azimuth angle α and is given by

$$r_A(\varepsilon, \alpha) = r_{0A}(1 + \varepsilon \cdot \tan(\varphi) \cdot \cos(\alpha - \alpha_A)). \quad (4)$$

2.3.2. Material reflectance

Using a monostatic system, the angles of the incident and reflected light coincides, but the beam attenuation is still related to the angle of incidence φ_m . A special case is given for specular surfaces with material reflectance $\rho_m = 1$:

$$\rho_{\text{specular}}(\varphi) = \begin{cases} 1 & \text{if } \varphi = 0, \\ 0 & \text{if } \varphi \neq 0. \end{cases} \quad (5)$$

For many surfaces the reflectance is assumed to be uniform and isotropic, also known as Lambertian

diffuser. In this case the observed reflectance depends on material reflectance ρ_m and the angle of incidence φ :

$$\rho_{\text{diffuse}}(\varphi) = \rho_m \cdot \cos^2(\varphi). \quad (6)$$

A discussion of reflectance measurements for different materials is shown in Jelalian (1992). There are strong variations within the reflectance ρ_m for the same material, caused by the measurement situation. These

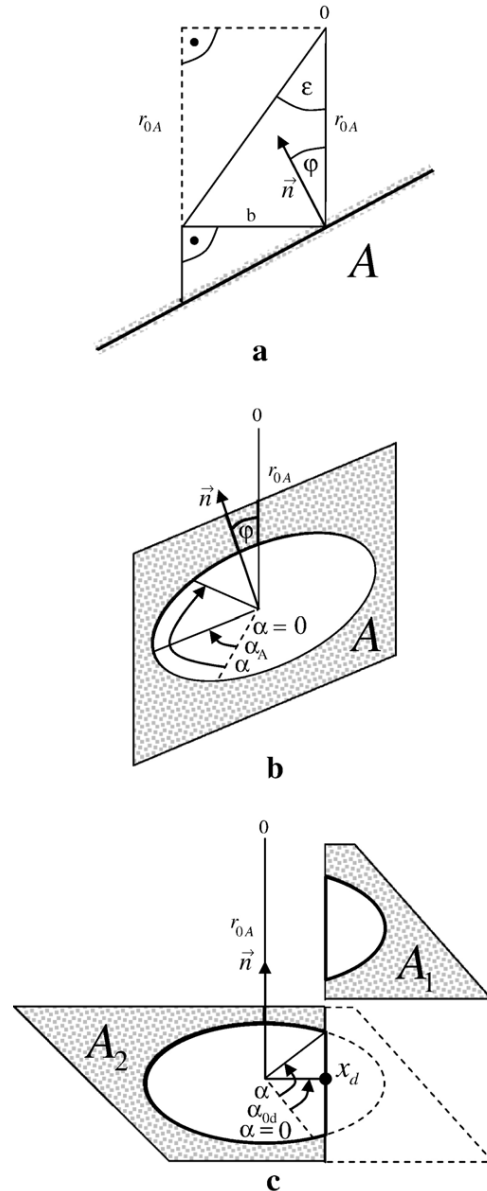


Fig. 5. Schematic illustration of the surface geometry. a) Side view of a single surface, b) oblique view of a single surface, c) oblique view of two differently elevated surfaces.

values should be treated with some caution if they are to be used for material discrimination.

Combining the information presented in Sections 2.3.1 and 2.3.2 we finally get the general expression for the reflectance density function ρ_A of a diffuse or specular surface with respect to the geometry and material:

$$\begin{aligned}\rho_A(r, \varepsilon, \alpha) &= \rho_{diffuse}(\varphi) \cdot \delta(r_A(\varepsilon, \alpha), r) \\ &= \rho_m \cdot \cos^2(\varphi) \cdot \delta(r_A(\varepsilon, \alpha), r), \text{ or} \\ \rho_A(r, \varepsilon, \alpha) &= \rho_{specular}(\varphi) \cdot \delta(r_A(\varepsilon, \alpha), r),\end{aligned}\quad (7)$$

where δ represents the delta function with $\delta(r_A, r) = \begin{cases} 1 & \text{if } r = r_A, \\ 0 & \text{if } r \neq r_A. \end{cases}$ The delta function $\delta(r_A, r)$ is used to describe the surface in range r_A .

2.3.3. Multiple surface reflections

We have assumed that just one plane object surface occurs within the laser beam. If differently elevated object surfaces are partly illuminated by the laser beam, we will receive a mixture of reflections at different ranges. Multiple reflections of the surfaces A_1, A_2, \dots, A_n within the beam corridor are calculated by:

$$\begin{aligned}\rho_{multiple}(r, \varepsilon, \alpha) &= \sum_{i=1}^n \rho_{A_i}(r, \varepsilon, \alpha) \\ &\quad \times \delta(r_A_i(\varepsilon, \alpha), \min(\{r_{A_i}(\varepsilon, \alpha) | i = 1 \dots n\})).\end{aligned}\quad (8)$$

The delta function δ is used to describe each single surface A_i in range r_{A_i} . Illuminated surface areas are described by $\min(\{r_{A_i}(\varepsilon, \alpha) | i = 1 \dots n\})$ which considers the line of sight and disregards the occluded areas. The complete reflection $\rho_{multiple}$ is given by the sum of all partial surface reflections.

2.4. Atmospheric transmission

The one-way propagation path for the atmospheric transmission η_a with the atmospheric extinction coefficient σ_a (Kamermann, 1993) is given by

$$\eta_a(r) = \exp\{-\sigma_a \cdot r\}.\quad (9)$$

This formula is seldom used, but might be of importance for measurements under bad weather conditions.

2.5. Receiver efficiency

The receiver attenuates the perceived radiant intensity which is described by S and depends on the aperture

diameter D and object range r . Assuming $D \ll r$, the receiver characteristic is given by

$$S(D, r) = \frac{\pi D^2}{4\pi r^2} \cdot \eta_r,\quad (10)$$

with η_r being the receiver efficiency.

2.6. Calculating the received waveform of the back-scattered pulse

The received power $P(t)$ depends on the modulated waveform of the transmitted laser pulse (Eq. (2)), the spatial energy distribution of the laser beam (Eq. (3)), the atmospheric transmission on the way from the sensor to the surface (Eq. (9)), the reflectance density function (geometrical and material) (Eq. (7)), the atmospheric transmission on the way from the surface to the sensor (Eq. (9)), the receiver attenuation (Eq. (10)), and the illuminated surface area over the infinitesimal volume element $r^2 \sin(\varepsilon) d\alpha d\varepsilon dr$:

$$\begin{aligned}P(t) &= \int_{r=0}^{\infty} \int_{\varepsilon=0}^{\pi/2} \int_{\alpha=0}^{2\pi} s_m(t - \frac{2r}{c}) \\ &\quad \times K(\varepsilon) \cdot \eta_a(r) \cdot \rho_A(r, \varepsilon, \alpha) \cdot \eta_a(r) \cdot S \\ &\quad \times (D, r) \cdot r^2 \sin(\varepsilon) d\alpha d\varepsilon dr,\end{aligned}\quad (11)$$

where c is the speed of light. This can be written in the form

$$\begin{aligned}P(t) &= \int_{r=0}^{\infty} s_m(t - \frac{2r}{c}) \cdot H(r) dr, \text{ with} \\ H(r) &= \int_{\varepsilon=0}^{\pi/2} \int_{\alpha=0}^{2\pi} K(\varepsilon) \cdot \eta_a(r) \cdot \rho_A(r, \varepsilon, \alpha) \\ &\quad \times \eta_a(r) \cdot S(D, r) \cdot r^2 \sin(\varepsilon) d\alpha d\varepsilon.\end{aligned}\quad (12)$$

Under the assumption (iii) of Section 2.3.1, that the general range r_{0A} of the system origin to the surface is much greater than the geometrical reflectance properties inside the range interval $[r_{0A} - \Delta r, r_{0A} + \Delta r]$ the formula $H(r)$ can be simplified

$$H(r) \approx C_{r_{0A}} \int_{\varepsilon=0}^{\pi/2} \int_{\alpha=0}^{2\pi} K(\varepsilon) \cdot \rho_A(r, \varepsilon, \alpha) \cdot r^2 \sin(\varepsilon) d\alpha d\varepsilon, \quad (13)$$

$$\begin{aligned}\text{with } C_{r_{0A}} &= \eta_a^2(r_{0A}) \cdot S(D, r_{0A}) \\ &= \exp\{-\sigma_a \cdot 2r_{0A}\} \cdot \frac{\pi D^2}{4\pi r_{0A}^2} \eta_r.\end{aligned}$$

With this approximation, $P(t)$ depends on the transmitted waveform $s_m(t)$, the spatial energy distribution of

the laser beam $K(\varepsilon)$ and the reflectance density function $\rho_A(r, \varepsilon, \alpha)$ integrated over the illuminated surface area, only.

Assume the geometry according to Fig. 5c the two differently elevated specular plane plates A_1 and A_2 are partly illuminated by the laser beam. Now let the surface A_2 be bounded by a straight line and $x_d = [r_{0d}, \varepsilon_{0d}, \alpha_{0d}]$ be the point on the line with the lowest distance to the optical axis we obtain

$$H(r) = C_{r_{0d}} \int_{\varepsilon=0}^{\pi/2} \int_{\alpha=0}^{2\pi} K(\varepsilon) \times [\delta(r_{A_1}(\varepsilon, \alpha), r) \cdot \delta(r_{A_1}(\varepsilon, \alpha), \min(\{r_{0A_2}, r_{A_1}(\varepsilon, \alpha)\})) + \delta(r_{0A_2}, r) \cdot \delta(r_{0A_2}, \min(\{r_{0A_2}, r_{A_1}(\varepsilon, \alpha)\}))] \times r^2 \sin(\varepsilon) d\alpha d\varepsilon$$

$$r_{A_1}(\varepsilon, \alpha) = \begin{cases} r_{0A_1} & \text{if } \varepsilon > \text{asin}\left(\frac{\sin(\varepsilon_{0d})}{\cos(\alpha - \alpha_{0d}) \cdot \cos(\varepsilon_{0d})}\right) \\ \approx \frac{\varepsilon_{0d}}{\cos(\alpha - \alpha_{0d})}, & \\ \infty & \text{otherwise.} \end{cases} \quad (14)$$

3. Analyzing the waveform

Analyzing the waveform of the backscattered pulse is useful to obtain information about the illuminated surface. For this analysis, very short pulses compared to the illuminated structures in range are advantageous, e.g. for resolving fine structures of branches the pulse length has to be shorter than some nanoseconds. In general, short pulses have a lower pulse power and are more difficult to detect. In the case of a bad signal to noise ratio, a matched filter approach can be used (Jutzi and Stilla, 2004). However, this results in only a single estimate of the travel time of the pulse without further features. In this paper we focus on pulses wider than a nanosecond and a moderate pulse power (10 kW).

For analyzing the surface response it has to be extracted from the measured signal. The following steps have to be carried out: (i) detection of the backscattered pulse, (ii) deconvolution of the transmitted with the received waveform, and (iii) Wiener filtering to estimate the surface function.

3.1. Detection of the backscattered pulses

First, in order to analyze the waveform, the backscattered pulse of interest has to be detected and extracted.

For pulse detection, a noise dependent threshold is estimated to discriminate a single pulse from the background noise. A signal interval without pulses is processed to estimate the noise and to characterize it by its mean and its standard deviation. In practice, this can also be done from a signal including pulses, as the pulse duration is negligible compared to the signal duration. If the signal is higher than three times of the noise's standard deviation for at least 5 ns, a pulse will be assumed to have been found and a waveform interval including the pulse will be accepted for further processing.

3.2. Deconvolution

The received waveform $p(t)$ corresponds to a convolution of the transmitted waveform $s_m(t)$ and the surface response $h(t)$:

$$p(t) = s_m(t) * h(t). \quad (15)$$

By transforming $p(t)$ into the Fourier domain and solving the resulting equation for the spectral surface function $\underline{H}(f)$ we obtain

$$\underline{H}(f) = \frac{\underline{P}(f)}{\underline{S}_m(f)}. \quad (16)$$

For calculating $\underline{H}(f)$, the functions $\underline{P}(f)$ and $\underline{S}_m(f)$ have to be known, which means that $p(t)$ and $s_m(t)$ have to be measured. Measurements of the waveform are always associated with a receiver noise term $n(t)$, which is added to the signal

$$s_{m,n}(t) = s_m(t) + n(t) = s(t) \cdot m(t) + n(t) \quad (17)$$

and

$$p_n(t) = p(t) + n(t). \quad (18)$$

Depending on the receiver noise, large numerical errors may appear for $\underline{H}(f)$. Therefore, we need a filter that reduces the noise, without smearing the surface function.

3.3. Wiener Filter

For the estimation of the surface function, we use the Wiener Filter (Wiener, 1949). This Optimal Filter minimizes the mean squared error between the uncorrupted

surface function $\underline{H}(f)$ and the estimated surface function $\hat{H}(f)$:

$$\int_{-\infty}^{+\infty} |h(t) - \hat{h}(t)|^2 dt = \int_{-\infty}^{+\infty} |\underline{H}(f) - \hat{H}(f)|^2 df \quad (19)$$

is minimized.

A solution has to be found where the Wiener Filter $W(f)$ is a real function

$$W(f) = \frac{|\underline{P}(f)|^2}{|\underline{P}(f)|^2 + |\underline{N}(f)|^2} \quad (20)$$

$\underline{N}(f)$ can be easily estimated from the background noise. $\underline{P}(f)$ depends on the received waveform and the surface function. For designing a Wiener Filter, $|\underline{P}(f)|^2$ has to be estimated.

Assuming that a plane surface is perpendicular to the pulse propagation direction and the surface is illuminated by an infinitesimal footprint, then $h(t) = \delta(t)$ (Dirac delta function) and $\underline{P}(f) = \underline{S}_m(f)$. Instead of $s_m(t)$ (Eq. (2)), we measure $s_m(t) + n(t)$. For the estimation of $s_m(t)$, we low-pass-filter the received signal in time domain using a linear binominal filter. In this case, we receive for the Wiener Filter

$$W(f) = \frac{|\underline{S}_m(f)|^2}{|\underline{S}_m(f)|^2 + |\underline{N}(f)|^2} \quad (21)$$

The estimation of the surface function is then given by

$$\hat{H}(f) = \frac{\underline{P}_n(f)}{\underline{S}_m(f)} \cdot W(f) \quad (22)$$

and the estimated surface response $\hat{h}(t)$ is obtained by transforming $\hat{H}(f)$ into time domain.

3.4. Approximation of the surface response

By determining the temporal position, length, and amplitude of the waveform, surface features like the range, elevation variations, and reflectivity of the surface can be derived from the extracted surface response. These surface features are extracted by fitting a Gaussian $\tilde{h}(t)$ to the estimated surface response

$$\begin{aligned} \hat{h}(t) \rightarrow \tilde{h}(t) &= \sum_{i=1}^n \tilde{h}_{A_i}(t) \\ &= a_{A_i} \cdot \exp\left\{-4 \cdot \ln 2 \cdot \frac{(t-t_{A_i})^2}{w_{A_i}^2}\right\}. \end{aligned} \quad (23)$$

For the estimation of the three parameters (i) time value t_{A_i} , (ii) temporal pulse length w_{A_i} , and (iii)

maximum amplitude a_{A_i} of the Gaussian, the iterative Levenberg–Marquardt Method is used (Marquardt, 1963). For n total number of responses, i presents the current response to process. The generalized iteration rule to estimate the parameters $\tilde{q}_{A_i} = [t_{A_i}, w_{A_i}, a_{A_i}]$ is described by

$$\begin{aligned} &\{\tilde{q}_{A_i,k} - \tilde{q}_{A_i,k+1}\} \\ &= \{J^T \Sigma^{-1} J\}^{-1} J^T \Sigma^{-1} \{\hat{h}(t) - \tilde{h}_{A_i,k}(t)\}, \end{aligned} \quad (24)$$

where $\tilde{q}_{A_i,k}$ are the pulse parameters $t_{A_i,k}$, $w_{A_i,k}$, and $a_{A_i,k}$ of the current iteration step k , J is the Jacobian matrix of the estimated Gaussian $\tilde{h}(t)$, and Σ is the covariance matrix of the estimated pulse parameters.

The iteration starts with $t_{A_i,1}$ as the temporal position of the surface response maximum, $w_{A_i,1}$ as the length of the surface response, and $a_{A_i,1}$ as the surface response maximum. The iteration is repeated until the change in $\tilde{h}_{A_i,k}(t)$ is below a specified tolerance. From the estimated position t_{A_i} the range value r_{A_i} to the object is determined by evaluating $r_{A_i} = t_{A_i} c / 2$, where c is the speed of light. From the estimated w_{A_i} the varying elevations of the object can be derived by $d_{A_i} = w_{A_i} c / 2$, where the determined elevation variations due to vegetation and slopes depend on the lasers footprint size. Skew waveforms of the surface response are not considered in this approach.

A measure describing the quality of the estimated parameters t_{A_i} , w_{A_i} , and a_{A_i} is given by their variances $\sigma_{t_{A_i}}^2$, $\sigma_{w_{A_i}}^2$, and $\sigma_{a_{A_i}}^2$. These variances can be determined directly from the main diagonal of the covariance matrix. If the variances are all below a given threshold, the surface response will be accepted. Then the estimated Gaussian $\tilde{h}(t)$ will be subtracted from the estimated surface response $\hat{h}(t)$ and the remaining waveform will be processed again in the same manner. This kind of processing is repeated until all responses with high quality (low values for the variances $\sigma_{t_{A_i}}^2$, $\sigma_{w_{A_i}}^2$, and $\sigma_{a_{A_i}}^2$) are detected.

If the temporal distance ΔT in between the single responses $\hat{h}_{A_i}(t)$ and $\hat{h}_{A_{i+1}}(t)$ is close to $0.85 w_c$, then the estimated parameters for each single surface will have to be revised, because of the overlapping single responses. w_c is derived from the system and will be explained in the next paragraph. Fig. 6 shows a simulation of two overlapping single responses (dotted lines) for different distances together with the received surface response (solid line). In Fig. 6a the temporal distance ΔT in between the single responses is $2 w_c$ ($w_c = 7.5$ cm), then the maxima position of the surface response and the single responses are very close to each other. In Fig. 6b the temporal distance ΔT of the single responses is w_c .

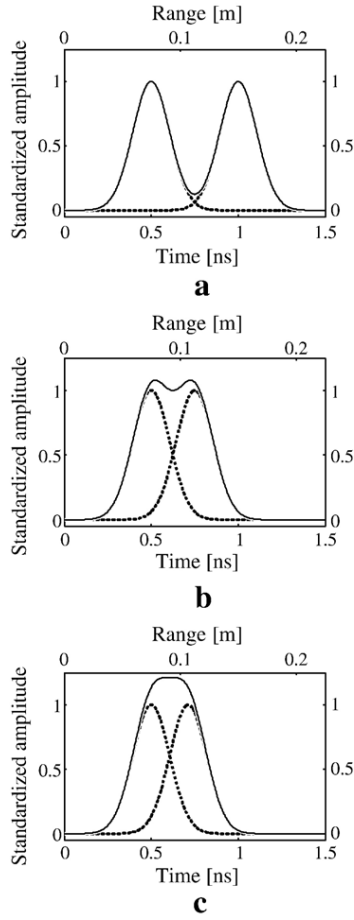


Fig. 6. Overlapping responses (dotted lines) and received surface response (solid line) with different temporal distances ΔT between single responses. a) $\Delta T = 2 w_c$, b) $\Delta T = w_c$, c) $\Delta T = 0.85 w_c$.

The surface response has two small peaks, whereas the maxima positions of the peaks are closer to each other than the maxima positions of the single responses. For a temporal distance $0.85 w_c$ the surface response shows only a single maximum (Fig. 6c). In the case of two close located responses, the estimation is performed by two Gaussians with 6 parameters ($t_{A_1}, w_{A_1}, a_{A_1}, t_{A_2}, w_{A_2}, a_{A_2}$); for three responses, 9 parameters ($t_{A_1}, w_{A_1}, a_{A_1}, t_{A_2}, w_{A_2}, a_{A_2}, t_{A_3}, w_{A_3}, a_{A_3}$) have to be estimated; and so on.

3.5. Impulse response of the measurement unit

If data from a real system is used the impulse response of the measurement unit has to be taken into account in the parameter estimation, as the properties of the measurement unit affect the detected signal. This means that the measured waveform, which is originally defined by a convolution of the transmitted waveform and the surface response is additionally convolved with

the impulse response. The length of the impulse response is influenced by the bandwidth of the used receiver unit. A given bandwidth for a Gaussian frequency function with a cutoff frequency f_c leads to an impulse response length w_c :

$$w_c = \frac{2}{\pi} \sqrt{\ln \sqrt{2} \cdot \ln 2} \cdot \frac{1}{f_c}. \quad (25)$$

w_c is the impulse response length of the system and defines the length of the signal measured by a given system when the surface response is an ideal Dirac delta function $\delta(t)$.

4. Experiments

An experimental setup with a pulsed Erbium fiber laser (wavelength: $1.55 \mu\text{m}$, beam divergence: 1 mrad, pulse length: 5 ns at FWHM) was built up for exploring the capabilities of waveform analysis for discriminating objects close in range. For capturing the transmitted and the received waveforms, two InGaAs detectors and two amplifiers with an overall bandwidth $f_c = 1 \text{ GHz}$ were used. Both analog signals were sampled with a rate of 20 GSamples/s.

The used bandwidth results in an impulse response length of $w_c = 0.3 \text{ ns}$ (Eq. (25)). An example of an impulse response derived by averaging 1000 samples of a signal

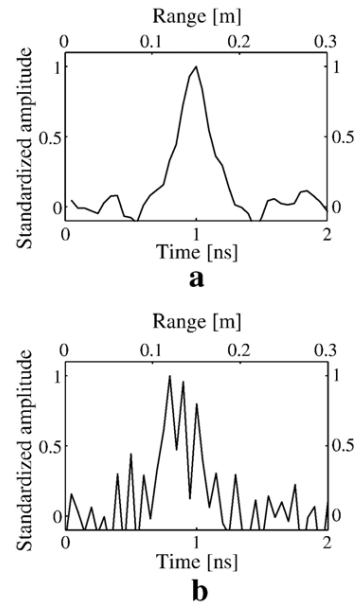


Fig. 7. Impulse response. a) Averaged impulse response estimated by 1000 samples, b) single impulse response with noise.

scattered by a specular surface (plane metal plate) is shown in Fig. 7a.

For reducing the noise of a single impulse response (Fig. 7b), a low pass filter is used. As the low pass filter reduces the signal bandwidth, the length of the shown impulse response increases from $w_c=0.3$ ns to $\bar{w}_c=0.5$ ns. The length \bar{w}_c of the low pass filtered impulse response is

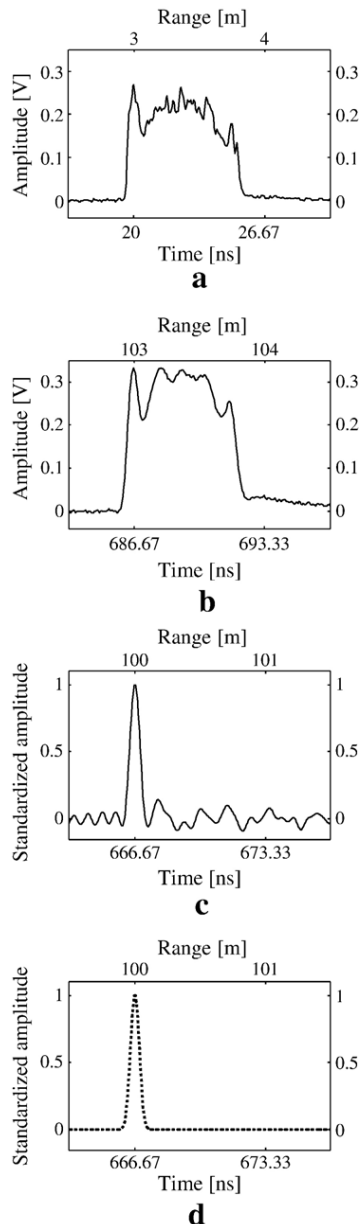


Fig. 8. Example of the measurements and results for the response of a single surface. a) Transmitted waveform of the emitted pulse, b) received waveform of the backscattered pulse, c) estimated surface response, d) approximated surface response.

the lower bound of the length of the surface response to resolve a single surface. This length is relevant to discriminate multiple surfaces which are located close to each other.

The investigations explore the reliability of a range measurement for a single surface (Section 4.1) and the range discrimination of two surfaces (Section 4.2). For both experiments, 500 samples of the transmitted and received waveform from a plane plate with specular scattering characteristic are captured. The range accuracy is measured by the standard deviation σ_{r_A} of the range estimation. The ability of discriminating two nearby surfaces is tested by using two plates arranged behind each other in propagation direction of the laser beam (Eq. (14)) with a distance below the pulse length. Each surface was illuminated partly by half of the beam footprint. The range of the plates to the laser was about 100 m and the distance in between them 0.15 m, which corresponds to $2\bar{w}_c$ of the example in Fig. 7a.

4.1. Single surface

For estimating the range accuracy to a plane surface at the range of 100 m a single surface response is estimated. Fig. 8a and b show an example of a single measurement illustrating the emitted and the backscattered pulse. In this figure, the depicted area of interest has a range interval of 2 m. Fig. 8b shows the shape of the received waveform which is similar to the transmitted waveform in Fig. 8a, but different in amplitude. The estimated surface response obtained by the Wiener Filter is depicted in Fig. 8c. It shows a single strong peak with a maximum at the range of 100 m surrounded by small signal ripples. The result of Fig. 8c is further processed by the Levenberg–Marquardt Method for gaining the parameters of the surface response. From the estimated parameters, the waveform of the surface response is calculated. An example of this waveform is shown in Fig. 8d.

The distribution of the range values is depicted in Fig. 10a by a histogram. All of the 500 emitted pulses which were backscattered at the plane surface located approximately at 100 m in range were detected and the surface responses were calculated. The range distribution shows a standard deviation of $\sigma_{r_A}=6.2$ mm.

4.2. Two surfaces

Two plane plates are positioned at 100 m and 100.15 m, and the transmitted and the received waveform are captured. In this case, the goal of processing the waveform is to detect, discriminate and estimate the range

of the two surface responses ($i=1,2$) as accurate as possible.

Fig. 9a and b shows examples of a single measurement with the emitted and the backscattered pulse. The small range shift of the plates generates an overlap of the returned waveforms. This mixed signal is slightly longer than the transmitted one with a dissimilar shape. The result of the Wiener Filter is depicted in Fig. 9c. It shows

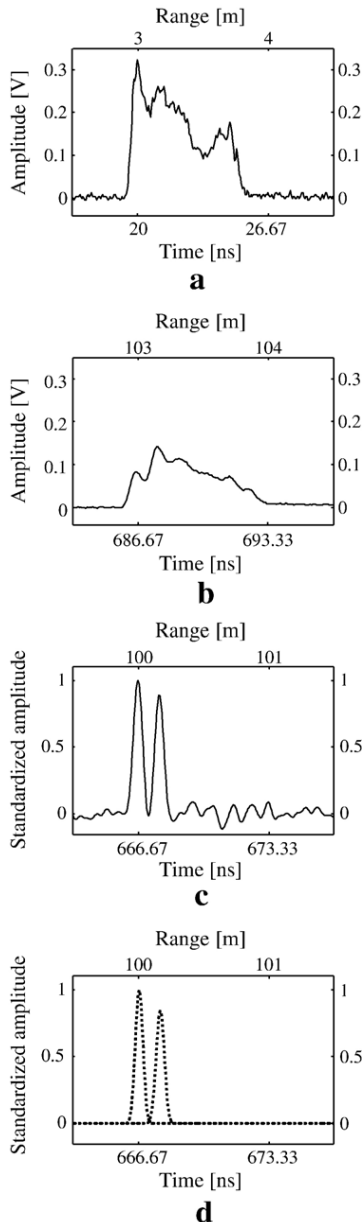


Fig. 9. Example of the measurements and results for the response of two surfaces (distance 0.15 m). a) Transmitted waveform of the emitted pulse, b) received waveform of the backscattered pulse, c) estimated surface response, d) approximated surface responses.

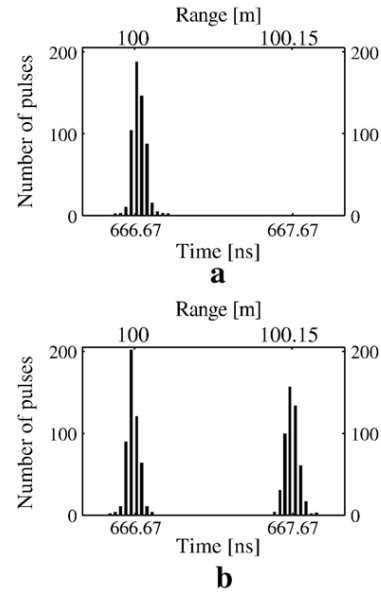


Fig. 10. Histograms of the estimated range values. a) Single surface, b) two surfaces (distance 0.15 m).

two strong peaks at the range of 100 m and 100.15 m. The two estimated surface responses are visualized in Fig. 9d. The curves of the estimated surface responses get in touch at the bottom, and the single responses overlap slightly in between each other. The overlapping of the single surface responses is a problem for accurate range estimation. Because the overlap of the responses generates an overlay of the surface responses, the original maxima position of the single responses are shifted apart from each other. Thus the maximum position which is usually used for range estimation gives inaccurate range values. By using an improved estimation (Section 3.4) of the single surface responses, this inaccuracy can be eliminated.

The amount of 500 pulses was measured from two plane surfaces with a range distance of 0.15 m located at 100 m in range. The histogram of the range values shows a bimodal distribution with two significant peaks (Fig. 10b). The left range distribution shows a standard deviation of $\sigma_{r_{A_1}} = 5.5$ mm and the right range distribution shows a standard deviation of $\sigma_{r_{A_2}} = 6.6$ mm. The average distance for 500 measurements is $\bar{\Delta r} = 0.149$ m with the standard deviation of $\sigma_{\Delta r} = 5.9$ mm.

5. Discussion

In general, analog techniques for range determination by measuring the time-of-flight of laser pulses have a strong limitation regarding their capability of discriminating surfaces which are very close located in range.

Katzenbeisser (2003) has shown that the discrimination of nearby pulses will only be possible if the distance of two backscattering object surfaces is at least greater than half of the pulse length. By recording the waveform with a high bandwidth of the receiver unit, an adequate sampling rate of the analog-to-digital converter, and the proposed algorithms for discriminating multiple surface responses, the range resolution can be significantly improved. Therefore, the received as well as the transmitted waveform have to be recorded. We have shown that surfaces with a distance corresponding to a tenth of the emitted pulse length can be resolved. The used laser system allows discriminating two surface responses with a distance of 0.15 m. It has to be mentioned that in the experiments specular surfaces were used. Only a few natural surfaces are specular. But, for investigating the limitations of the laser system, e.g. discriminating stepped surfaces (objects with small distances) specular surfaces seems to be the best choice due to the proper signal-to-noise-ratio.

Furthermore, recording of the transmitted waveform enables to isolate the surface response from the received waveform. By applying the Levenberg–Marquardt Method to the surface response, three properties of the response can be estimated, namely time value, temporal pulse length, and maximum amplitude. From these properties we derive the corresponding surface features, namely range, elevation variations, and reflectivity.

The spatial energy distribution was modeled and measured in the near field. For the far field different profiles for modeling are used (Section 2.2). The transition from near-field to far-field is given by the Rayleigh range $z_{\text{Ray}} = \frac{\pi \omega_0^2}{\lambda}$, where the wavelength is λ and the beam waist is ω_0 (Siegman, 1986). Assuming a wavelength of $\lambda = 1.55 \mu\text{m}$ and a beam-waist diameter of $D = 2\omega_0 \approx 60 \text{ mm}$ we receive a Rayleigh range of approximately 1800 m.

For investigating the limits in resolving small distances in range within the beam we have chosen plane metal plates with specular surfaces. These specific surfaces with high reflectivity and minimum surface slope were used for the experiments to capture a waveform with high quality. Diffuse surfaces with high material reflectance might deliver similar results but were not investigated in this paper. For the range measurements in 100 m, the results of the experiments show a high reliability ($\sigma_{r_A} \leq 6.6 \text{ mm}$). The plates which were shifted by 0.15 m were supposed to reflect half of the pulse energy, each. The configuration was not adjusted in an optimal way so that the maxima amplitudes of the surface responses differ (Fig. 9c). The measurement of the average distance $\Delta r = 0.149 \text{ m}$ was obtained with a high reliability ($\sigma_{\Delta r} = 5.9 \text{ mm}$).

The limits to resolve small distances in range generally depend on the length of the impulse response w_c and the corresponding bandwidth f_c of the used measurement unit. Discriminating two responses by peak detection will be possible if the temporal distance in between single responses is greater than $0.85 w_c$. Otherwise, the shape of the waveform of the overlapping responses resembles a single widened peak. If the temporal distance in between single responses is close to $0.85 w_c$, range estimation by the maxima positions of the surface response will be inaccurate, because of the overlapping single responses (Section 3.4). The range estimation can be revised with the Levenberg–Marquardt Method considering multiple surfaces to increase the accuracy.

The temporal pulse length and amplitude of the waveform were not investigated further in this paper. The temporal pulse length seems to be of special interest for distinguishing natural from man-made surfaces.

References

- Baltsavias, E.P., 1999. Airborne laser scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 164–198.
- Blair, J.B., Rabine, D.L., Hofton, M.A., 1999. The Laser Vegetation Imaging Sensor (LVIS): a medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 112–122.
- Brenner, A.C., Zwally, H.J., Bentley, C.R., Csatho, B.M., Harding, D.J., Hofton, M.A., Minster, J.B., Roberts, L.A., Saba, J.L., Thomas, R.H., Yi, Y., 2003. Geoscience Laser Altimeter System (GLAS) — derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights. Algorithm Theoretical Basis Document — Version 4.1. http://www.csr.utexas.edu/glas/pdf/Atbd_20031224.pdf (Accessed September 1, 2006).
- Der, S., Redman, B., Chellappa, R., 1997. Simulation of error in optical radar measurements. *Applied Optics* 36 (27), 6869–6874.
- Hug, C., Wehr, A., 1997. Detecting and identifying topographic objects in laser altimeter data. *International Archives of Photogrammetry and Remote Sensing* 32 (Part 3-4W2), 19–26.
- Huising, E.J., Gomes Pereira, L.M., 1998. Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (5), 245–261.
- Jelalian, A.W., 1992. *Laser Radar Systems*. Artech House, Norwood, MA, Boston.
- Jutzi, B., Stilla, U., 2003. Laser pulse analysis for reconstruction and classification of urban objects. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (Part 3-W8), 151–156.
- Jutzi, B., Stilla, U., 2004. Extraction of features from objects in urban areas using space-time analysis of recorded laser pulses. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B2), 1–6.
- Jutzi, B., Eberle, B., Stilla, U., 2002. Estimation and measurement of backscattered signals from pulsed laser radar. In: Serpico, S.B.

- (Ed.), (2003) *Image and Signal Processing for Remote Sensing VIII*. SPIE Proceedings, vol. 4885, pp. 256–267.
- Jutzi, B., Neulist, J., Stilla, U., 2005. Sub-pixel edge localization based on laser waveform analysis. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3-W19), 109–114.
- Kamermann, G.W., 1993. Laser radar. In: Fox, C.S. (Ed.), *Active Electro-Optical Systems, the Infrared and Electro-Optical Systems Handbook*. SPIE Optical Engineering Press, Michigan, pp. 1–76.
- Katzenbeisser, R., 2003. Technical note on echo detection. <http://www.toposys.de/pdfext/Engl/echo-detec3.pdf>. (Accessed September 1, 2006).
- Maas, H.G., 2001. The suitability of airborne laser scanner data for automatic 3D object reconstruction. In: Baltsavias, E.P., Gruen, A., Van Gool, L. (Eds.), *Automatic Extraction of Man-Made Objects From Aerial and Space Images (III)*. Balkema, Lisse, pp. 291–296.
- Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 11 (2), 431–441.
- Moffiet, T., Mengersen, K., Witte, C., King, R., Denham, R., 2005. Airborne laser scanning: exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing* 59 (5), 289–309.
- Sequeira, V., Ng, K.C., Wolfart, E., Gonçalves, J.G.M., Hogg, D.C., 1999. Automated reconstruction of 3D models from real environment. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (1), 1–22.
- Siegman, A.E., 1986. *Lasers*. University Science Books, Sausalito, CA.
- Steinvall, O., 2000. Effects of target shape and reflection on laser radar cross sections. *Applied Optics* 39 (24), 4381–4391.
- Steinvall, O., Carlsson, T., 2001. Three-dimensional laser radar modeling. In: Kamerman, G.W. (Ed.), *Laser Radar Technology and Application VI*. SPIE Proceedings, vol. 4377, pp. 23–34.
- Stilla, U., Rottensteiner, F., Hinz, S. (Eds.), 2005. Object Extraction for 3D City Models, Road Databases, and Traffic Monitoring — concepts, algorithms, and evaluation (CMRT05). *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36 (Part 3-W24).
- Vosselman, G., 2002. On estimation of planimetric offsets in laser altimetry data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (Part 3A), 375–380.
- Wagner, W., Ullrich, A., Melzer, T., Briese, C., Kraus, K., 2004. From single-pulse to full-waveform airborne laser scanners: potential and practical challenges. In: Altan, M.O. (Ed.), *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 35 (Part B3), pp. 201–206.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T., Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (2), 100–112.
- Wehr, A., Lohr, U., 1999. Airborne laser scanning — an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 68–82.
- Wiener, N., 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, Cambridge, MA.

7

Waveform Analysis for Small-Footprint Pulsed Laser Systems

Uwe Stilla and Boris Jutzi

CONTENTS

7.1	Introduction.....	215
7.2	Characterization of a Laser System.....	217
7.2.1	Laser Type.....	217
7.2.2	Modulation Technique.....	218
7.2.3	Measurement Technique.....	218
7.2.4	Detection Technique.....	219
7.2.5	Multiphoton Detection.....	219
7.2.6	Single-Photon Detection.....	220
7.2.7	Construction.....	221
7.3	Modeling.....	221
7.3.1	Waveform of the Laser Pulse.....	221
7.3.2	Spatial Energy Distribution.....	222
7.4	Analyzing the Waveform.....	223
7.4.1	Peak Detection.....	224
7.4.2	Leading Edge Detection.....	224
7.4.3	Constant Fraction Detection.....	225
7.4.4	Center of Gravity Detection.....	226
7.4.5	Gaussian Decomposition.....	227
7.4.6	Deconvolution.....	227
7.5	Attribute Extraction.....	228
7.6	Summary and Outlook.....	232
	References.....	233

7.1 Introduction

Aerial photogrammetry and airborne laser scanning (ALS) are the two most widely used methods for generating digital elevation models (DEMs), including digital terrain models (DTMs) that depict ground topography and digital surface models (DSMs) that depict the height of the ground, structures, and vegetation cover. In photogrammetry, the distance to a spatial surface is classically derived from a triangulation of corresponding image points from two or more overlapping images of the surface. These points are chosen manually or

detected automatically by analyzing image structures. In contrast to photogrammetry, active laser scanner systems allow a direct and illumination-independent measurement of the distance to a surface, otherwise known as the range.

The interrelationship between aerial photogrammetry and ALS has been intensely discussed within the aerial surveying community in the last decade. Different comparison factors concerning data acquisition (e.g., coverage, weather conditions, costs, etc.) and surface reconstruction (e.g., accuracy, redundancy, post-processing time, etc.) have to be taken into account to choose the optimal method for a certain mapping campaign. An example of a study comparing photogrammetric image matching versus laser scanning for generation of high-quality DEMs for glacier monitoring (Lenhart et al., 2006) is given in Wuerlaender et al. (2004). In contrast to a decision to use one or the other technique, in some fields of applications a combined processing of laser data and stereo images is advantageous as shown in the generation of the extraterrestrial DTMs of Mars (Albertz et al., 2005; Spiegel et al., 2006) or DSMs for building characterization.

Conventional pulsed laser scanner systems for topographic mapping are based on time-of-flight ranging techniques to determine the range to the illuminated object. The time-of-flight is measured by the elapsed time between the emitted and backscattered laser pulses. The signal analysis to determine this time typically operates with analog threshold detection. For targets that have surfaces at different ranges illuminated by a single laser pulse, more than one backscattered pulse may be detected per emitted pulse. Most ALS systems are able to capture, at a minimum, the range for the first- and last-detected backscattered pulses. Some systems acquire ranges up to as many as five per emitted pulse for multiple backscattered pulses. First-pulse detection is the optimal choice to measure the hull of partially penetrable objects or the so-called volume scattering targets (e.g., canopy of trees). Last-pulse detection should be chosen to measure nonpenetrable surfaces (e.g., ground surfaces).

Currently, some commercial ALS systems not only capture the range for multiple pulse reflections but also digitize and record the received signal of the reflected laser energy, which allows for the so-called full-waveform analysis. This offers the possibility of analyzing the waveform off-line using digital signal processing methods in order to extract different surface attributes from the received signal based on the shape of the return pulses.

In the last decade, some waveform analysis investigations were carried out to explore the structure of vegetation and estimate aboveground biomass. For example, NASA developed the waveform-recording laser vegetation imaging sensor (LVIS) to measure vertical density profiles in forests (Blair et al., 1999). This experimental airborne system operates at altitudes up to 10 km and acquires waveforms for large diameter laser footprints (nominally 20 m) acquired across a wide swath. Another NASA system operating with a large footprint is the spaceborne geoscience laser altimeter system (GLAS) mounted on the ice, cloud, and land elevation satellite (ICESat). GLAS measures height distributions of atmospheric clouds and aerosols, and surface elevations of ice sheets, land topography, and vegetation (Brenner et al., 2003; Zwally et al., 2002). It is a profiling system that operates with a footprint diameter of 70 m and measures elevation changes with decimeter accuracy (Abshire et al., 2005; Schutz et al., 2005). In the analysis of data from both systems, surface characteristics are determined by comparing a parametric description of the transmitted and received waveforms (Hofton et al., 2000; Harding and Carabajal, 2005). Because the laser footprint is large and illuminates multiple surfaces, the resulting return waveform is an integrated, spatially nonexplicit representation of the range to illuminated surfaces separated both vertically and horizontally. The geometric organization of surfaces within a single footprint can therefore not be determined.

In contrast to large-footprint systems, small-footprint ALS systems illuminate only one or a few surfaces within the footprint, yielding waveforms with distinct return pulses corresponding to specific surfaces. One of the first developed small-footprint waveform-recording systems is the scanning hydrographic operational airborne lidar survey (SHOALS) instrument used for monitoring nearshore bathymetric environments. SHOALS has been in full operation since 1994 (Irish and Lillycrop, 1999; Irish et al., 2000). More recently commercial full-waveform ALS systems for terrestrial mapping have been developed (Jutzi and Stilla, 2006b; Wagner et al., 2006), which operate with a transmitted pulse width of 4–10 ns and allow digitization and acquisition of the waveform with approximately 0.5–1 GSample/s. Reitberger et al. (2006, 2007) have recently reported results that show clearly the potential of airborne, small-footprint, full-waveform data for the comprehensive analysis of tree structure and species classification. A set of key attributes have been defined and extracted based on the 3D distribution of the returns in combination with their characteristics in the full-waveform signal, providing information about tree microstructure such as the organization of the trunk and branches.

In this chapter, we focus not on a given application in the context of a data set from a given sensor rather than on general principles. Specifically, we describe the different approaches for designing a laser system, modeling the spatial and temporal properties of the emitted laser pulses, detecting return pulses, and deriving attributes from the waveform. We emphasize aspects of the received waveform that are especially relevant for the newly available small-footprint, full-waveform commercial systems that yield distinct return pulses when multiple surfaces are illuminated by a laser pulse.

The design of a laser system impacts its measurement capabilities and the manner in which the signal has to be modeled and analyzed. Section 7.2 gives a brief overview of the features that characterize the design of laser ranging systems. Section 7.3 focuses on modeling of the temporal waveform and the spatial beam distribution. Different strategies for pulse detection are explained in Section 7.4. Section 7.5 describes the attributes that can be extracted from a single laser shot and presents an analysis of an entire scene that was recorded with an experimental small-footprint, full-waveform laser ranging system. A summary and outlook is given in Section 7.6.

7.2 Characterization of a Laser System

Depending on the application, laser systems can be designed in different ways. They may differ in techniques concerning the type of laser used, the modulation, type of measured features, detection method, or design of beam paths (construction). Figure 7.1 sketches a simplified overview of features characterizing a laser system. More detailed descriptions can be found in Kamermann (1993).

7.2.1 Laser Type

A laser works as an oscillator and an amplifier for monochromatic radiation (infrared, visible light, or ultraviolet). The operative wavelength of available lasers is located between 0.1 μm and 3 mm. For comparison, it should be mentioned that the visible domain is from 0.37 to 0.75 μm . To achieve a good signal-to-noise ratio (SNR) over long ranges, conventional scanning laser systems emit radiation with high energy. However, this could endanger the health of humans due to the focusing of laser radiation on the retina, which is most

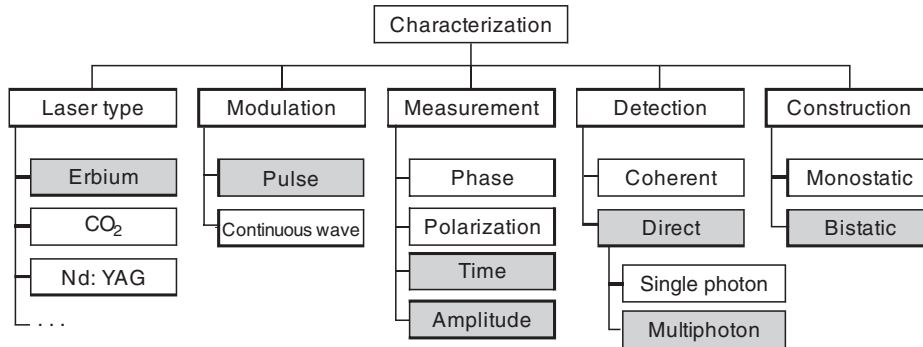


FIGURE 7.1

Features characterizing a laser system (with shaded boxes indicating the characteristics that are the focus of this chapter).

susceptible to damage at visible wavelengths. For this reason, most eye-safe laser systems used for mapping purposes operate with a wavelength outside the visible spectrum. This allows working with an emitted energy that is many times higher (up to 10^6) compared to the visible domain without the potential for retinal damage. Eye-safe lasers of greatest interest for long-range laser scanning are erbium-fiber lasers, carbon dioxide (CO_2) lasers, and neodymium:YAG (Nd:YAG) lasers.

Erbium-fiber lasers are optically pumped by a semiconductor diode and the active medium is an erbium-doped fiber. Their construction can be compact while still achieving a high output power. CO_2 lasers use carbon dioxide in gas form as the active medium. Although their construction is simple, their large size and mass are significant disadvantages. Solid-state Nd:YAG lasers can be pumped by various sources that define the characteristics of the emitted laser radiation. In this contribution we focus on erbium-fiber lasers (Figure 7.1).

7.2.2 Modulation Technique

Concerning modulation techniques, laser systems can be divided into two groups: continuous wave (cw) and pulsed lasers. A cw laser continuously emits electromagnetic radiation. The temporal energy distribution of the transmitted signal is influenced by amplitude modulation (AM) or frequency modulation (FM). Depending on the applied modulation technique, specific measurement techniques (Section 7.2.3) are required. A pulsed laser emits electromagnetic radiation in pulses of short duration. For laser ranging it is desirable to emit a pulse as short as possible and with a pulse energy as high as possible in order to obtain a precise range with a high probability of detection. However, design limitations on maximum peak power introduce a trade-off that requires a compromise between the length and the energy of the pulse. The length of the pulse (full width at half maximum, FWHM) is typically between 2 and 10 ns. For applications in remote sensing with long ranges, pulsed lasers with higher power density as compared to cw lasers are advantageous. In this contribution we focus on pulsed laser systems (Figure 7.1).

7.2.3 Measurement Technique

Measurement techniques using laser systems can be distinguished by the exploited signal properties such as phase, amplitude, frequency, polarization, time, or any combination

of them. An amplitude modulated cw laser system is used to measure the range by exploiting the phase of a sinusoidal modulated signal. A phase difference Φ_d can be determined from a given phase of the transmitted signal and the measured phase of the received signal. With wavelength λ_m of the modulated signal, a corresponding range r can be calculated by $r = \lambda_m \cdot \Phi_d / 4\pi$. If the measurement of the phase difference Φ_d cannot be distinguished from $\Phi_d + n\pi$, the unambiguity interval of the range measurement will be limited to a maximum range $r_{\max} = \lambda_m / 2$. Assuming that the system is able to resolve an angle difference $\Delta\Phi_d$, the range resolution Δr corresponds according to $\Delta r = \lambda_m \cdot \Delta\Phi_d / 4\pi$. To increase the range resolution for a given $\Delta\Phi_d$, the modulation wavelength λ_m has to be decreased. However, this results in a reduction of the unambiguity interval of the range determination.

The problem of ambiguity can be solved by using multiple simultaneous offset sinusoidal modulation frequencies (multiple-tone sinusoidal modulation). In this case the maximum modulation wavelength defines the unambiguity and the minimum modulation wavelength defines the range resolution. In addition to this, partially illuminated surfaces with different ranges within the beam corridor result in a superimposed signal depending on the range and the reflectance of the surface. Because only a single phase value can be determined at the receiver, the ambiguities caused by the partially illuminated surfaces cannot be resolved (Thiel and Wehr, 2004). An incorrect intermediate value is measured.

The measurement of the amplitude value is feasible for cw lasers as well as for pulsed lasers. The amplitude is influenced by background radiation, the range of the object to the laser system, and the size, reflectance, slope, and roughness of the illuminated surface. In this chapter, we are interested in measuring and analyzing the received pulse waveform, i.e., the dependence of the intensity over time (Figure 7.1).

7.2.4 Detection Technique

Detection techniques can be divided into coherent detection and direct detection (Jelalian, 1992). Coherent detection is based on signal amplification due to constructive interference of the wave front of the received signal with that of the reference signal emitted from a cw laser. In direct detection laser systems, the received optical energy is focused onto a photosensitive element that generates an output signal that depends on the received optical power. Two direct detection techniques are appropriate for recording the temporal characteristics of the backscattered signal: multiphoton detection and single-photon detection.

7.2.5 Multiphoton Detection

The classical measurement technique for direct detection operates with a photodiode. For optical detectors, a positive intrinsic negative diode (PIN) or the more sensitive avalanche photodiode (APD) is used. The photodiode generates an electrical signal (voltage or current) that is directly proportional to the optical power of incident light composed of multiple photons. Figure 7.2 sketches a pulse resulting from a varying number of photons n over time t . For a detailed analysis of the analog signal a digitizing receiver unit is essential. To analyze the signal of the emitted short duration laser pulse with only a few nanoseconds pulse width, a high bandwidth receiver that resolves the signal at gigahertz rates and a correspondingly high digitizer sampling rate is required. Increasing bandwidth results in decreasing sensitivity of the photodiode, which can be compensated by increasing the power of the emitting laser source. An example of an Nd:YAG laser pulse sampled with 5 GSample/s is given in Figure 7.4a.

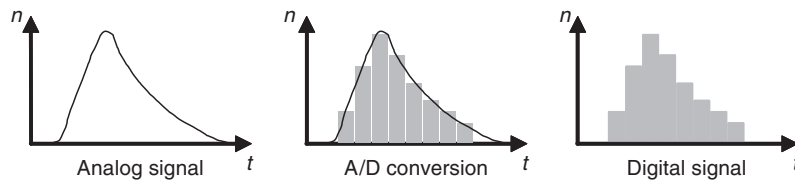


FIGURE 7.2
Digital recording of the signal with multiphoton detection.

7.2.6 Single-Photon Detection

The principle of single-photon detection is depicted in Figure 7.3. A short duration pulse is emitted by the laser source. A single photon of the backscattered pulse is detected by the receiver after time interval τ_1 . This event blocks the receiver for a certain period of time during which no further photons are able to trigger the receiver. The time-of-flight of this single event is collected into a corresponding time bin of a histogram. After the period of blocking, the receiver is open to detect a new single-photon event. Multiple measurements are repeated and the time-of-flight of each single event (τ_2, τ_3, \dots) is registered into the corresponding time bin of the histogram.

Let us assume a stationary scene and a stationary sensor platform. In this case, the statistical properties of the laser radiation do not change with the time and time-average quantities are equal to the ensemble quantities. Under these assumptions, the radiation ensembles are stationary and ergodic (Papoulis, 1984; Troup, 1972). The counting of single-photon events with assignment of their time-of-flight into time bins of a histogram is closely related to the integration of multiphotons over time (Alexander, 1997; Gagliardi and Karp, 1976; Loudon, 1973). In other words, the temporal waveform of the pulse can be reconstructed from a histogram of single-photon arrivals over time.

Many transmit pulses are necessary to obtain the waveform with single-photon detection. The quality of the sampled waveform depends on the number of photon counts. Various optical detectors can be used for this purpose, namely, PMT (photomultiplier tubes), MCP (microchannel plate), or APD detectors. Figure 7.4b shows a pulse plotted from a histogram containing the time-of-flight measurements from 16252 photons distributed in 50 bins, where the bin width is 40 ps. Note that the FWHM of the pulse

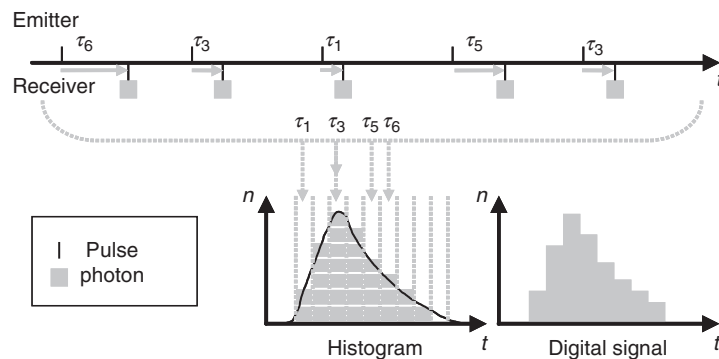
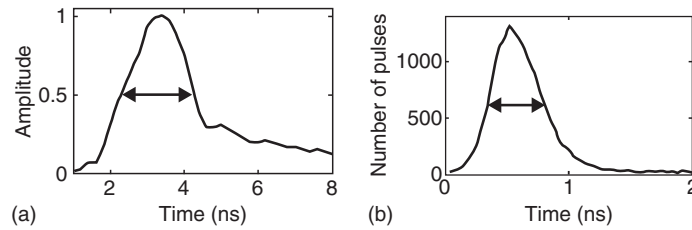


FIGURE 7.3
Digital recording of the signal with single-photon detection.

**FIGURE 7.4**

Examples for pulses backscattered from a diffuse surface. (a) Multiphoton detection (FWHM = 2.1 ns) and (b) single-photon detection (FWHM = 0.4 ns).

in Figure 7.4a is about five times of the pulse in Figure 7.4b. In this chapter, we focus on multiphoton detection (Figure 7.1).

7.2.7 Construction

Depending on the construction of the transmitter and receiver optics, monostatic and bistatic laser systems can be distinguished.

Monostatic laser systems have transmitter and receiver optics collocated on the same optical axis. A disadvantage of this construction is the higher number of components compared to a bistatic laser system, increasing the effort needed to optimally align the components. Advantages of this construction are the isogonal measurement of angles and the exact measurement of ranges, because the illuminated surface area and the observed field of view of the receiver are coincident for all ranges.

Bistatic laser systems have a transmitter and receiver optics that are spatially separated and thus the illumination and view angles are divergent. In general, both optics are close together and oriented in nearly the same direction. Objects are illuminated via a lens from the transmitter optic and the backscattered radiation is transferred via a separate lens to the receiver optic. An advantage of this measurement system is that the design can be easily constructed. A disadvantage of this design is that depending on the range to the illuminated surface the angle between the transmitter and the receiver optic varies. Furthermore, depending on the range, only a partial overlap is obtained between the illuminated surface and the observed field of view.

7.3 Modeling

The received waveform depends on the transmitted waveform of the emitted laser pulse, the spatial energy distribution of the beam, and the geometric and reflectance properties of the surface. In order to describe the temporal and spatial properties of a pulse, appropriate models that parameterize the pulse attributes have to be introduced.

7.3.1 Waveform of the Laser Pulse

Mathematical functions can be used to approximate the shape of laser pulse waveforms. Depending on the system, the shape may be best represented by a rectangular, exponential, or Gaussian distribution. A simple model is given by a rectangular distribution $s(t)$ with an amplitude a , pulse width w , and time delay τ

$$s(t) = \text{arect} \left\{ -\frac{(t-\tau)}{w} \right\} = \begin{cases} a & \text{for } -\tau \leq t \leq w-\tau \\ 0 & \text{else.} \end{cases} \quad (7.1)$$

Especially for short laser pulses a rectangular model often differs from the measured shape. A waveform with an exponential distribution (e.g., for a Q-switched laser) is applied by Steinvall (2000)

$$s(t) = t^2 \exp \left\{ -\frac{(t-\tau)}{w} \right\} \quad (7.2)$$

A temporally symmetric Gaussian distribution for modeling the waveform of the spaceborne GLAS is proposed by Brenner et al. (2003). The basic waveform $s(t)$ of the used laser system can be described by

$$s(t) = \frac{2a}{w} \sqrt{\frac{\ln(2)}{\pi}} \exp \left\{ -4\ln(2) \frac{(t-\tau)^2}{2w^2} \right\} \quad (7.3)$$

The width of a pulse w is commonly defined as one-half of the pulse's maximum amplitude known as FWHM.

7.3.2 Spatial Energy Distribution

The spatial energy distribution of a laser (also known as the beam profile) depends on the pump source, the optical resonator, and the laser medium. In general, beam profiles are modeled by a cylindrical distribution (top-hat form) or by a 2D-symmetric Gaussian distribution (Kamermann, 1993). The measured cylindrical beam distribution of a pulsed

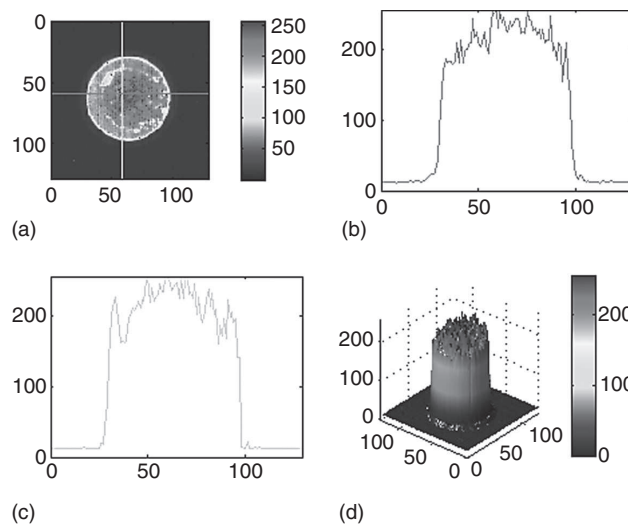


FIGURE 7.5 (See color insert following page XXX.)

Measured cylindrical beam distribution (top-hat form). (a) 2D visualization, with the row (red) and column (green) of the maximum intensity indicated, (b) profile of the indicated row, (c) profile of the indicated column, and (d) 3D visualization.

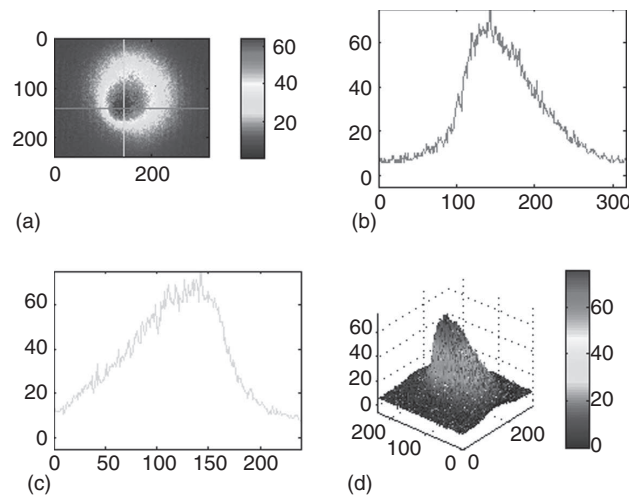


FIGURE 7.6 (See color insert following page XXX.)

Measured Gaussian beam distribution. (a) 2D visualization, with the row (red) and column (green) of the maximum intensity indicated, (b) profile of the indicated row, (c) profile of the indicated column, and (d) 3D visualization.

erbium fiber laser that operates at a wavelength of 1550nm is depicted in Figure 7.5. A Gaussian beam distribution of a Raman shifted Nd:YAG laser that operates at a wavelength of 1543nm is depicted in Figure 7.6. Both measurements differ more or less from these idealized distributions.

7.4 Analyzing the Waveform

Various detection methods are used to extract attributes of the reflecting surface from the waveform. To obtain the surface attributes, each waveform $s(t)$ is analyzed. The surface within the beam corridor generates a return pulse. To detect and separate this pulse from the noise, a signal-dependent threshold is estimated using the signal background noise. For example, in one particular implementation if the intensity of the waveform is above three times the noise standard deviation ($3\sigma_n$) for a duration of at least the full-width-half-maximum of the pulse, a pulse is assumed to be found. The section of the waveform including the detected pulse is passed onto the subsequent processing steps.

Typical surface attributes to extract from a waveform are range, elevation variation, and reflectance corresponding to the waveform attributes of time, width, and amplitude.

A rough surface, i.e., a surface of a certain vertical extent, will widen the laser pulse upon reflection. Therefore, the width of the pulse is a measure of the elevation variation of the surface. In addition, the widening of the pulse causes the reflected photons to be spread over a greater amount of time, thus reducing the peak amplitude. Therefore, to estimate the elevation variation or reflectance attributes of a surface, the pulse width and amplitude have to be known. Estimating just the amplitude of a pulse without considering this dependency will lead to inaccurate and noisy reflectance values.

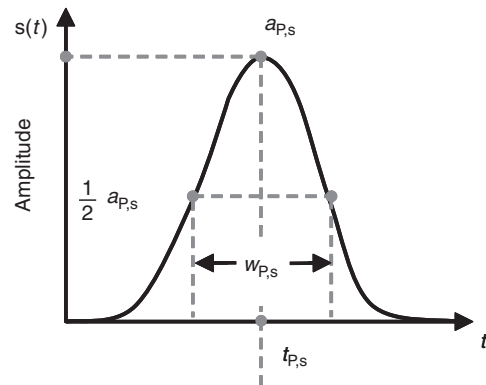


FIGURE 7.7
Attribute extraction with peak detection.

Determination of the range to a surface can be accomplished with different schemes that include peak detection, leading edge ranging, constant fraction detection, center of gravity detection, and Gaussian decomposition and deconvolution. A discussion of key elements of each approach follows.

7.4.1 Peak Detection

The values of range $r_{P,s}$ and amplitude are determined at the maximum pulse amplitude $a_{P,s}$ (Figure 7.7), and the width $w_{P,s}$ is estimated at full-width-half-maximum of the pulse. Local spikes on the pulse waveform strongly effect the attribute extraction. Therefore, for noisy signals, a smoothing filter is recommended to determine the global maxima.

7.4.2 Leading Edge Detection

A threshold crossing of the pulse waveform leading edge determines the range value $r_{LE,s}$ (Figure 7.8). The threshold value can be a predefined fixed value, but then the ranging detection strongly depends on the amplitude and width of the pulse waveform, introducing a ranging bias dependent on pulse shape referred to as range walk. The half of the maximum amplitude $a_{LE,s}$ of the pulse for a threshold is used for range determination.

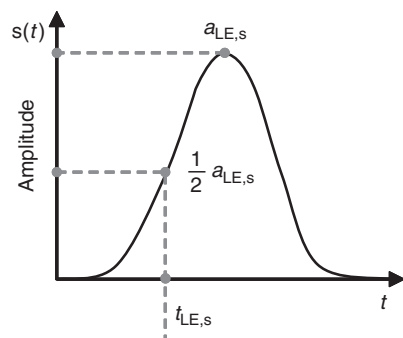


FIGURE 7.8
Attribute extraction with leading edge detection.

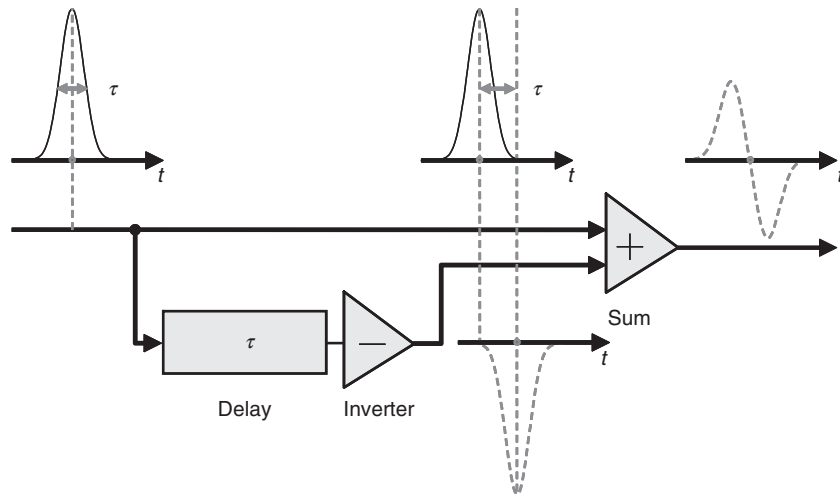


FIGURE 7.9
Simplified schematic visualization of the processing steps for the constant fraction detection.

7.4.3 Constant Fraction Detection

A ranging implementation designed to be insensitive to amplitude-dependent biases applies a constant fraction detection circuit in which the pulse waveform $s(t)$ is inverted and delayed by a fixed time τ and added to the original pulse (Figure 7.9). The combined signal $s_{CFD}(t)$ gives a constant fraction signal with a zero crossing point at t_{CFD} (see Figure 7.10).

$$s_{CFD}(t_{CFD}) = 0 \quad \text{with} \quad s_{CFD}(t) = s(t) - s(t + \tau) \quad (7.4)$$

The determined t_{CFD} is insensitive to the pulse amplitude, but depends on the pulse waveform and width (Kamermann, 1993). A suitable value for the delay time τ is the FWHM of the waveform.

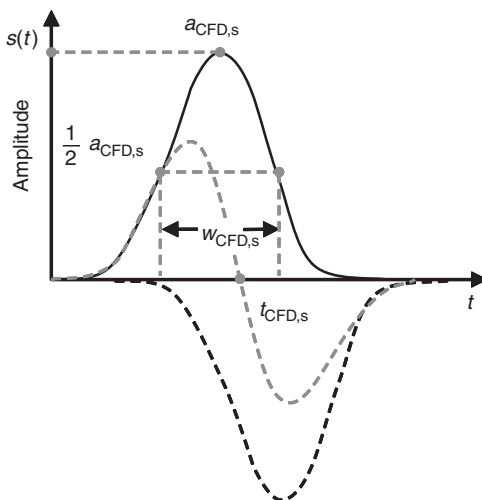


FIGURE 7.10
Attribute extraction with constant fraction detection.

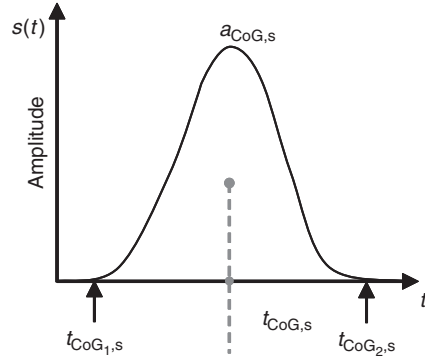


FIGURE 7.11
Attribute extraction with center of gravity detection.

For symmetric waveforms, the traditional constant fraction algorithm delivers unbiased ranging results. However, for an asymmetric noisy waveform the delayed signal should be reversed in time as well, to avoid ambiguities of the zero crossing point.

7.4.4 Center of Gravity Detection

The temporal center of gravity of the pulse waveform is $t_{CoG,s}$ (Figure 7.11). The time value (range) is determined by integrating the pulse waveform $s(t)$

$$t_{CoG,s} = \frac{\int_{t=t_{CoG1,s}}^{t_{CoG2,s}} t s(t) dt}{\int_{t=t_{CoG1,s}}^{t_{CoG2,s}} s(t) dt} \quad (7.5)$$

It delivers good results for returns with various pulse amplitudes and pulse widths that have low noise. For returns with an asymmetric pulse shape skewed to longer ranges, this method results in a detected range that is slightly longer than the range value obtained with the peak detection.

The following methods to further process the pulse properties are not part of the center of gravity algorithm, but are well suited to complement it. Generally, integration over a section of the signal has the advantage of reducing the noise dependence compared to the aforementioned methods relying on single samples. We call the integral of the waveform $s(t)$ shown in the denominator of Equation 7.4 as the pulse strength. From this, the value a_{CoG} can be calculated assuming a Gaussian and using the Inverse error function (erf^{-1}) and the width w

$$a_{CoG} = \frac{2 \text{erf}^{-1}(0.5)}{\sqrt{\pi} w} \int_{t=t_{CoG1}}^{t_{CoG2}} s(t) dt \quad (7.6)$$

Furthermore, the width $W_{CoG,s}$ is approximated by the width of the central pulse area contributing 0.76 of this pulse strength with

$$\int_{t=t_{CoG}-\frac{w_0}{2}}^{t_{CoG}+\frac{w_0}{2}} s(t) dt = \text{erf}(\sqrt{\ln 2}) \int_{t=t_{CoG1}}^{t_{CoG2}} s(t) dt \quad (7.7)$$

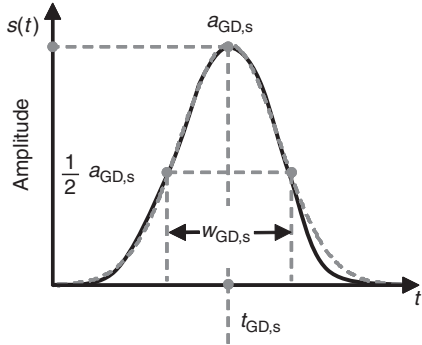


FIGURE 7.12 Attribute extraction with Gaussian decomposition algorithm.

7.4.5 Gaussian Decomposition

Assuming a Gaussian function for the waveform (Equation 7.3), the surface attributes can be extracted by estimating the parameters of the adapted function:

$$s(t) = a_{GD,s} \exp \left\{ -4 \ln(2) \frac{(t - t_{GD,s})^2}{(w_{GD,s})^2} \right\} \quad (7.8)$$

For a parametric description of the pulse properties, a Gaussian decomposition on the waveform can be used. Different methods are known, for example, Expectation Maximation (EM) algorithm (Persson et al., 2005) and Gauss–Newton or Levenberg–Marquardt algorithm (Hofton et al., 2000; Jutzi and Stilla, 2005; Reitberger et al., 2006). In Figure 7.12, the estimated attributes of the received waveform are depicted.

7.4.6 Deconvolution

Analysis of a received waveform in order to extract the attributes of the illuminated surface is a difficult task because different processes impact the shape of the waveform. The received waveform $s(t)$ of a laser pulse depends on the transmitted waveform $r(t)$, the impulse response of the receiver, the spatial beam distribution of the laser pulse, and the geometric and reflectance properties of the illuminated surface. The impulse response of the receiver is mainly affected by the photodiode and amplifier, and the spatial beam distribution typically has a Gaussian distribution. Let us assume that a receiver consists of an ideal photodiode and that the amplifier has an infinite bandwidth with a linear response. In that case the received waveform depends mainly on the transmitted waveform $r(t)$ and the properties of the illuminated surface. The 3D characteristics of the surface can be captured by a time-dependent surface representation, referred to as the surface response $h(t)$. In this case the received waveform $s(t)$ can be expressed as

$$s(t) = h(t) * r(t) \quad (7.9)$$

where $(*)$ denotes the convolution operation. By transforming $s(t)$ into the Fourier domain and solving the resulting equation for the spectral surface function $\underline{H}(f)$, we obtain

$$\underline{H}(f) = \underline{S}(f) / \underline{R}(f) \quad (7.10)$$

The surface response $h(t)$ is obtained by transforming $H(f)$ into the time domain. By applying a Gaussian decomposition method to the surface response, surface attributes can be extracted. The deconvolution removes the characteristics of the transmitted waveform from the received waveform and enables a description of the observed surface.

For a reliable deconvolution a high SNR for the received waveform is essential. In addition to this, the waveform has to be captured with a high bandwidth receiver and with an adequate sampling rate of the analog-to-digital converter. Furthermore, it has to be mentioned that large numerical errors may appear depending on the receiver noise. A Wiener filter used for deconvolution reduces the noise of the determined surface response (Jutzi and Stilla, 2006a). This method allows discriminating differences in range, e.g., given by a stepped surface within the beam, which are smaller than the length of the laser pulse. Experiments have shown that a step smaller than 10 times of the pulse length can be distinguished.

7.5 Attribute Extraction

An example of a signal profile applied to multiple pulses is depicted in Figure 7.13. The waveform parameters for each detected pulse of this signal profile are estimated by a Gaussian decomposition method using the Levenberg–Marquardt algorithm. The extracted attributes are described in the table. The estimated waveform is shown below the original waveform in Figure 7.13.

By comparing the range values in the table, we can observe that the distance between the first and second pulse is about 10 m and between the third and fourth pulse about 2.5 m. The third pulse shows the highest maximum amplitude and the pulse width of the first and

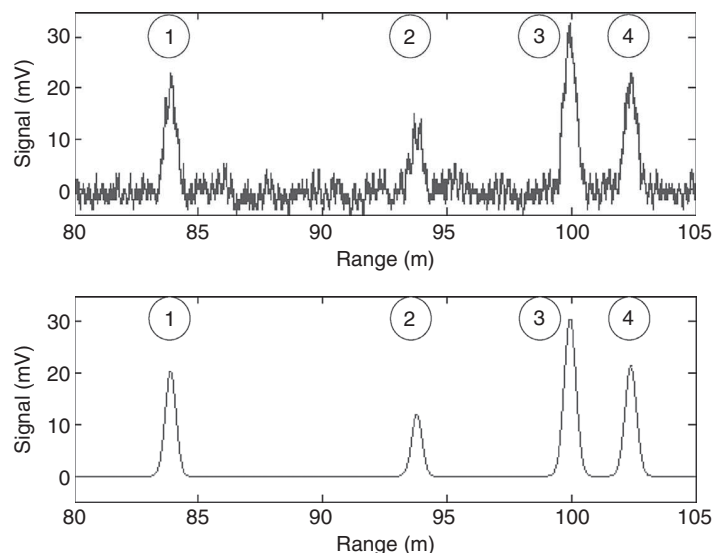


FIGURE 7.13
Signal profile with multiple reflections.

second pulse is slightly lower than the third and fourth pulse. It is not possible to classify the type of surfaces illuminated within the single-beam corridor using the amplitude, pulse width, and range properties alone. For assigning each return pulse to a specific surface type additional information is required, which can result from the 3D geometrical relationships of the returns within a point cloud.

As an example of information retrieval achieved by combining return pulse properties and the spatial relationship of the returns, a full-waveform data set of a test scene was captured by scanning along the azimuth and elevation and recording the return intensity sampled over time t . Neglecting angular variations of the scan, the measured intensities as a function of time t sampled over the azimuth and elevation can be interpreted as a 3D data set forming a cuboid with Cartesian coordinates x , y , and t . The sampling along the time axis can be recalculated into corresponding range values z . These data cuboid can be analyzed in several different ways.

Figure 7.14 shows a set of image slices ($y-t$ planes). The second slice from the left ($x = 4$) shows vegetation in the center (near range) and building structures on the right side (far range). The grey values correspond to the intensity of the signal. The intensity values along the marked solid line are the intensity values of the waveform shown in Figure 7.13.

Note that although this way of displaying the data suggests that a full 3D representation of the scene has been obtained, this is in fact not the case. Just as with point clouds measured by conventional laser scanners, the data cube represents only 2.5d information. This is because of occlusion effects that are dependent on the target size in relation to the beam footprint size. It is possible that the laser pulse is mostly intercepted by and backscattered from the first illuminated surface along the propagation direction and that the following surfaces along the laser vector are hidden, giving weak or no reflections. For instance, a tree with dense foliage may return only a single reflection response per laser pulse even though multiple surfaces are present after the first detected return along the path of the laser vector.

In the following, we use the Wiener filter method to extract attributes from received waveforms. The extraction is carried out without considering spatial neighborhood relations. The results of the extracted surface attributes from the data cuboid are shown in xy plane by

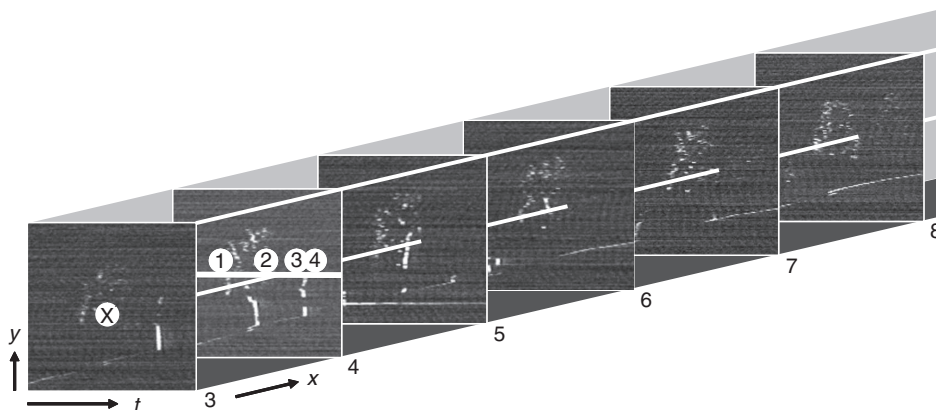


FIGURE 7.14
Vertical image slices with ground, vegetation, and building structures.

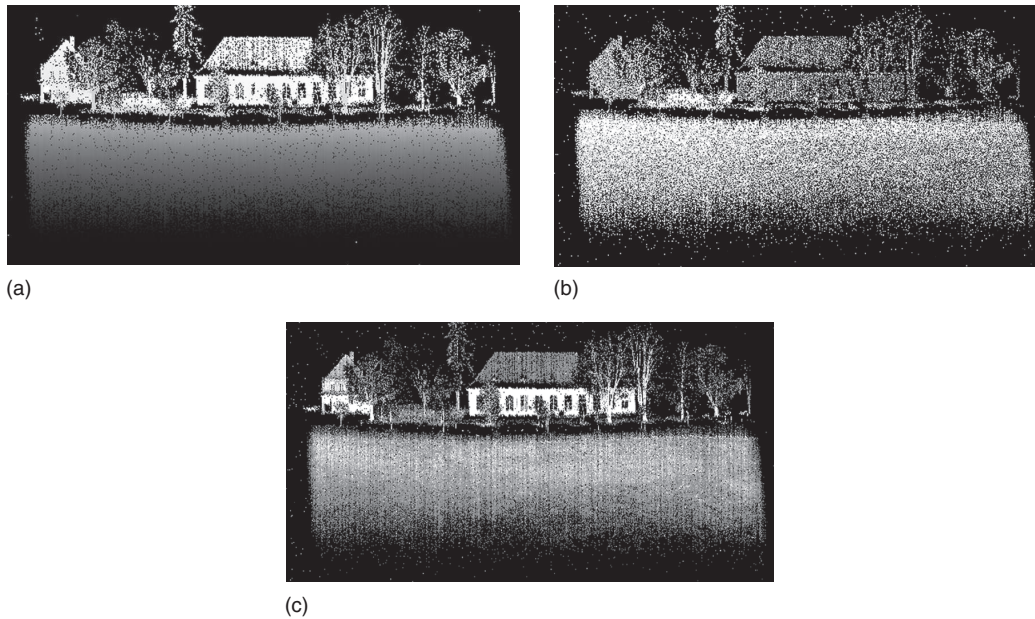


FIGURE 7.15
Extracted attributes of pulses: (a) range, (b) pulse width, and (c) intensity.

images of 320×600 pixels (Figure 7.15). Figure 7.15a shows the range image, Figure 7.15b shows the width of the pulse, and Figure 7.15c shows the intensity of the pulse. Larger values for an attribute are displayed by brighter pixels. Due to the fact that only a single value can be shown in the 2D images, only the first reflection is considered in cases where multiple reflections are present for a laser pulse.

The attributes maximum amplitude and pulse width were extracted using the Wiener filter and examined for their ability to discriminate different object classes. The entire scene and three object classes, namely, buildings, trees, and meadow, are shown in Figure 7.16. Column A of the figure depicts range images of the entire scene (Figure 7.16a, column A) and of the selected object classes (Figure 7.16b through d (column A)). Column B depicts scatter plots of maximum amplitude versus pulse width for the entire scene and the object classes (Figure 7.16, column B).

In Figure 7.16a, it can be observed that for the entire scene small values of the maximum amplitude have a large spreading of the pulse width. By decomposing the scene into the three object classes, it is apparent that the vegetation (trees and meadow) in most cases produces the signal returns with small maximum amplitudes but high values for the pulse width. The building in Figure 7.16b shows small values for the pulse width with little variation and a large range in the maximum amplitude values. Furthermore, the maximum amplitude of the building shows a cluster of higher values than that from the trees and meadow. These high values may result from high reflectance by the white façade. The trees are depicted in Figure 7.16c, where the pulse width shows high variation with mostly small maximum amplitude values. The meadow (Figure 7.16d) induces large variations of the pulse width and small maximum amplitude values. In general, the trees and the meadow produce larger pulse widths than the building.

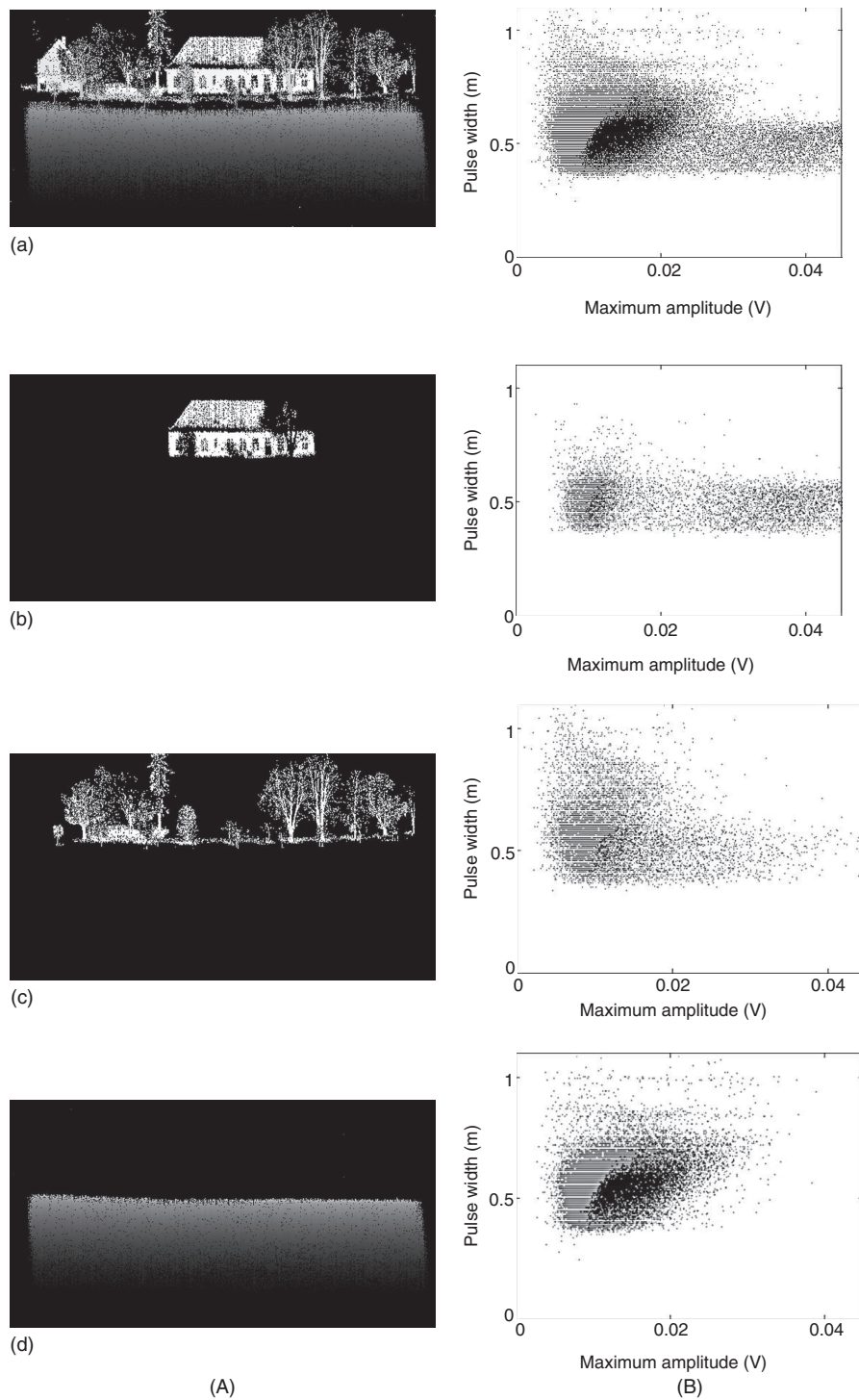


FIGURE 7.16 Comparison of maximum amplitude versus pulse width for selected objects of the measured scene. (a) Scene, (b) building, (c) trees, and (d) meadow. Column (A) denotes range images, and column (B) denotes maximum amplitude versus pulse width.

7.6 Summary and Outlook

It has been shown that full-waveform analysis enables extraction of more information compared to classical analogous pulse detection methods. First, recording of the received waveform offers the possibility for the end-user to select different methods to extract range information. The shape of the pulse and the entire signal can be considered for determining the range more accurately. Further improvements on reliability and accuracy can be derived by signal-processing methods based on the transmitted and the received waveform, e.g., deconvolution. Additionally, attributes of the surface can be derived from a parametric description of the waveform. The attributes maximum amplitude and pulse width may support the discrimination between volume scatterer (vegetation) and hard targets (man-made objects) (Kirchhof et al., 2007).

In the preceding description laser pulses were analyzed without considering the information of neighboring measurements. For reconstructing man-made objects, the introduction and test of hypotheses about the shape of the surface (e.g., plane, sphere) may efficiently support the analysis of single waveforms. Two different strategies assuming a planar shape in the local neighborhood of the surface and introducing this assumption into the signal analysis should be addressed. Both strategies combine information from top-down (surface primitives) and bottom-up (signal processing) for an extended analysis of full waveform laser data.

The first strategy (Kirchhof et al., 2007) uses an iterative processing of waveforms considering a predicted shape of the waveform from the local neighborhood. A presegmentation based on surface attributes is carried out to distinguish between partly penetrable objects (e.g., trees, bushes) and impenetrable surfaces (e.g., roof, wall). Derived range values from presegmentation of the impenetrable surfaces are used to automatically generate surface primitives (e.g., planes).

This allows a refinement of each range value, considering the surface geometry in a close neighborhood. Furthermore, partly occluded surface areas are extended by prediction of the expected range values. This prediction is further improved by considering the surface slope for the estimated received waveform. Expected pulses are simulated and correlated with the received waveforms. Accepted points that were missed in the first processing step due to weak signal response are associated to the point cloud. The procedure is repeated several times until all appropriate range values are considered to estimate the surface.

The second strategy (Stilla et al., 2007) uses a slope compensated stacking of waveforms. Weak pulses with a low SNR are discarded by classic threshold methods and get lost. In signal and image processing, different stacking techniques are used to improve the SNR.

For detection of weak laser pulses, hypotheses for planes of different slopes (e.g., angle difference 5°) are generated. According to the slope of the hypothesis, the waveforms in the local neighborhood are shifted in range. A superimposed signal is calculated from the stack of shifted waveforms. The maxima of superimposed signals from all hypotheses are compared to verify a hypothesis. Each signal is assessed by a likelihood value with respect to its contribution to the accepted hypothesis. Finally, signals are classified according to the likelihood values obtained using two thresholds and visualized by the traffic-light paradigm. The results contain detected pulses reflected from objects, which cannot be predicted by the previously detected point cloud.

Both strategies show promising results that encourage the continuity of work on the analysis of full-waveform laser pulses.

References

- Abshire, J.B., Sun, X.L., Riris, H., Sirota, J.M., McGarry, J.F., Palm, S., Yi, D.H., and Liiva, P., 2005. Geoscience Laser Altimeter System (GLAS) on the ICESat mission: On-orbit measurement performance. *Geophysical Research Letters*, 32:L21S02, doi:10.1029/2005GL024028.
- Albertz, J., Attwenger, M., Barrett, J., Casley, S., Dorninger, P., Dorrer, E., Ebner, H., Gehrke, S., Giese, B., Gwinner, K., Heipke, C., Howington-Kraus, E., Kirk, R.L., Lehmann, H., Mayer, H., Muller, J.-P., Oberst, J., Ostrovskiy, A., Renter, J., Reznik, S., Schmidt, R., Scholten, F., Spiegel, M., Stilla, U., Waehlich, M., and Neukum, G., 2005. HRSC on Mars Express—Photogrammetric and Cartographic Research. *Photogrammetric Engineering and Remote Sensing*, 71(10):1153–1166.
- Alexander, S.B., 1997. Optical communication receiver design. In: *SPIE Tutorial Texts in Optical Engineering*, TT22, SPIE Press, Bellingham, WA.
- Blair, J.B., Rabine, D.L., and Hofton, M.A., 1999. The Laser Vegetation Imaging Sensor (LVIS): A medium-altitude, digitization-only, Airborne Laser Altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2–3):112–122.
- Brenner, A.C., Zwally, H.J., Bentley, C.R., Csatho, B.M., Harding, D.J., Hofton, M.A., Minster, J.B., Roberts, L.A., Saba, J.L., Thomas, R.H., and Yi, D., 2003. Geoscience Laser Altimeter System (GLAS)—Derivation of range and range distributions from Laser Pulse Waveform Analysis for surface elevations, roughness, slope, and vegetation heights. Algorithm Theoretical Basis Document—Version 4.1. Available at http://www.csr.utexas.edu/glas/pdf/Atbd_20031224.pdf (accessed September 1, 2006).
- Gagliardi, R.M. and Karp, S., 1976. *Noncoherent (Direct) Detection: Optical Communications*. John Wiley & Sons, New York.
- Harding, D.J. and Carabajal, C.C., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32:L21S10, doi:10.1029/2005GL023471.
- Hofton, M.A., Minster, J.B., and Blair, J.B., 2000. Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4):1989–1996.
- Irish, J.L. and Lillycrop, W.J., 1999. Scanning laser mapping of the coastal zone: The SHOALS System. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2–3):123–129.
- Jelalian, A.W., 1992. *Laser Radar Systems*. Artech House, Boston, MA.
- Jutzi, B. and Stilla, U., 2005. Measuring and processing the waveform of laser pulses. In: Gruen, A., Kahmen, H., Eds., *Optical 3-D Measurement Techniques VII*, Vol. 1, pp. 194–203. Vienna University of Technology.
- Jutzi, B. and Stilla, U., 2006a. Range determination with waveform recording laser systems using a Wiener filter. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(2):95–107 [doi:10.1016/j.isprsjprs.2006.09.001].
- Jutzi, B. and Stilla, U., 2006b. Characteristics of the measurement unit of a full-waveform laser system. Symposium of ISPRS Commission I: From Sensors to Imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 1/A).
- Kamermann, G.W., 1993. Laser radar. In: Fox, C. S., Ed., *Active Electro-Optical Systems, The Infrared and Electro-Optical Systems Handbook*. SPIE Optical Engineering Press, Ann Arbor, MI.
- Kirchhof, M., Jutzi, B., and Stilla, U., 2008. Iterative processing of laser scanning data by full waveform analysis in close neighbourhood. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(1), pp. 99–114 [doi:10.1016/j.isprsjprs.2007.08.006].
- Lenhart, D., Kager, H., Eder, K., Hinz, S., and Stilla, U., 2006. Hochgenaue Generierung des DGM vom vergletscherten Hochgebirge—Potential von Airborne Laserscanning. Arbeitsgruppe Automation in der Kartographie: Tagung 2005, Mitteilungen des Bundesamtes für Kartographie und Geodäsie, Band 36:65–78 (in German).
- Loudon, R., 1973. *The Quantum Theory of Light*. Clarendon Press, Oxford.
- Papoulis, A., 1984. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Tokyo.

- Persson, Å., Söderman, U., Töpel, J., and Ahlberg, S., 2005. Visualization and analysis of full-waveform airborne laser scanner data. In: Vosselman, G., Brenner, C., Eds., *Laserscanning 2005. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Enschede, 36(Part 3/W19):109–114.
- Reitberger, J., Krzystek, P., and Stilla, U., 2006. Analysis of full waveform lidar data for tree species classification. In: Förstner, W., Steffen, R., Eds., Symposium of ISPRS Commission III: Photogrammetric Computer Vision PCV06. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Bonn, 36(Part 3):228–233.
- Reitberger, J., Krzystek, P., and Stilla, U., 2007. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing*, 29(5):1407–1431.
- Schutz, B.E., Zwally, H.J., Shuman, C.A., Hancock, D., and DiMarzio, J.P., 2005. Overview of the ICESat Mission. *Geophysical Research Letters*, 32, L21S01, [doi:10.1029/2005GL024009].
- Spiegel, M., Stilla, U., and Neukum, G., 2006. Improving the exterior orientation of Mars Express regarding different imaging cases. *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, 36(4) (on CD).
- Stilla, U., Yao, W., and Jutzi, B., 2007. Detection of weak laser pulses by full waveform stacking. In: Stilla, U. et al., Eds., PIA07: Photogrammetric Image Analysis 2007. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Munich, 36(3/W49A):25–30.
- Steinvall, O., 2000. Effects of target shape and reflection on laser radar cross sections. *Applied Optics*, 39(24):4381–4391.
- Thiel, K.H. and Wehr, A., 2004. Performance capabilities of laser-scanners—An overview and measurement principle analysis. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 8/W2):14–18.
- Troup, G.J., 1972. Photon Counting and Photon Statistics. In: Sanders, J.H., Stenholm, S., Eds., *Progress in Quantum Electronics*, Vol. 2 (Part 1), Oxford, Pergamon.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T., and Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2):100–112.
- Wuerlaender, R., Eder, K., and Geist, T., 2004. High quality DEMs for glacier monitoring: Image matching versus laser scanning. In: Altan, M.O., Ed., *International Archives of Photogrammetry and Remote Sensing*, 35(Part B7):753–758.
- Zwally, H.J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., and Thomas, R., 2002. ICESat's Laser Measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34(3–4):405–445.

B. Publikationen zu 2.2 Koregistrierung mit 2D- bzw. 3D-Sensoren

- [B1] Bradley PE, **Jutzi B** (2011) Improved feature detection in fused intensity-range images with Complex SIFT (CSIFT). Remote Sensing - Open Access Journal 2011, 3 (9): 2076-2088 doi:10.3390/rs3092076
- [B2] Hinz S, Weinmann M, Runge P, **Jutzi B** (2011) Potentials of image based active ranging to capture dynamic scenes. In: Heipke C, Jacobsen K, Rottensteiner F, Müller S, Sörgel U (Eds) High-resolution earth imaging for geospatial information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 4 / W19) (on CD)
- [B3] Hoegner L, Hanel A, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2014) Towards People Detection from Fused Time-Of-Flight and Thermal Infrared Images. In: Schindler K, Paparoditis N (Eds) ISPRS Technical Commission III Symposium. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-3, 2014: 121-126
- [B4] Hoegner L, Roth L, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2013) Fusion von Time-of-Flight-Entfernungsdaten und thermalen IR-Bildern. AVN - Allgemeine Vermessungs-Nachrichten, Ausgabe 5/2014: 192-197
- [B5] Hoegner L, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2013) Co-registration of Time-of-Flight (TOF) camera generated 3d point clouds and thermal infrared images (IR). In: Seyfert E (Ed) 33. Wissenschaftlich-Technische Jahrestagung der DGPF, 2013 (22): 481-488
- [B6] Hoegner L, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2012) Synchrone Koregistrierung von 3d Punktwolken und thermischen Infrarotbildern. In: Luhmann T (Hrsg) Optische 3D-Messtechnik - Photogrammetrie - Laser-Scanning. Beiträge der 12. Oldenburger 3D-Tage 2013. Herbert Wichmann Verlag, Heidelberg
- [B7] **Jutzi B**, Weinmann M, Meidow J (2013) Improved UAV-borne 3D mapping by fusing optical and laserscanner data. In: Grenzdörffer G, Bill R (Eds) UAV-g2013. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W2, 2013: 223-228
- [B8] **Jutzi B**, Weinmann M, Meidow J (2013) Weighted Data Fusion for UAV-borne 3D Mapping with Camera and Line Laserscanner. International Journal of Image and Data Fusion 5 (3): 226-243 doi:10.1080/19479832.2014.889228

- [B9] Weinmann M, Dittrich A, Hinz S, **Jutzi B** (2013) Automatic feature-based point cloud registration for a moving sensor platform. In: Heipke C, Jacobsen K, Rottensteiner F, Sörgel U (Eds) High-resolution earth imaging for geospatial information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W1, 2013: 373-378
- [B10] Weinmann M, Hoegner L, Leitloff J, Stilla U, Hinz S, **Jutzi B** (2012) Fusing passive and active sensed images to gain infrared-textured 3D models. In: Shortis M, El-Sheimy N (Eds) XXII ISPRS Congress: Imaging a sustainable future. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 39 (B1): 71-76 doi:10.5194/isprsarchives-XXXIX-B1-71-2012
- [B11] Weinmann M, **Jutzi B** (2013) Fast and Accurate Point Cloud Registration by Exploiting Inverse Cumulative Histograms (ICHs). Joint Urban Remote Sensing Event, JURSE 2013. IEEE CFP13RSD (on USB) [ISBN 978-1-4799-0212-5]: 218-221
- [B12] Weinmann M, **Jutzi B** (2011) Fully automatic image-based registration of unorganized TLS data. In: Lichti DD, Habib AF (Eds) Laserscanning 2011. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 5 / W12) (on CD)
- [B13] Weinmann M, Leitloff J, Hoegner L, **Jutzi B**, Stilla U, Hinz S (2014) Thermal 3D Mapping for Object Detection in Dynamic Scenes. In: Toth CK, Jutzi, B (Eds) ISPRS Technical Commission I Symposium. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-1, 2014: 53-60
- [B14] Weinmann M(artin), Weinmann M(ichael), Hinz S, **Jutzi B** (2011) Fast and automatic image-based registration of TLS data. In: Bretar F, Wagner W, Paparoditis N (Eds) ISPRS Journal of Photogrammetry & Remote Sensing 66 (6): 62-70 doi:10.1016/j.isprsjprs.2011.09.010
- [B15] Weinmann M, Wursthorn S, **Jutzi B** (2011) Semi-automatic image-based co-registration of range imaging data with different characteristics. In: Stilla U, Rottensteiner F, Mayer H, Jutzi B, Butenuth M (Eds) Photogrammetric Image Analysis PIA11. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W22): 119-124

Article

Improved Feature Detection in Fused Intensity-Range Images with Complex SIFT (CSIFT)

Patrick Erik Bradley * and Boris Jutzi

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Kaiserstr. 12, D-76128 Karlsruhe, Germany; E-Mail: boris.jutzi@kit.edu

* Author to whom correspondence should be addressed; E-Mail: bradley@kit.edu;
Tel.: +49-721-608-47301; Fax: +49-721-608-48450.

Received: 1 July 2011; in revised form: 28 July 2011 / Accepted: 23 August 2011 /
Published: 16 September 2011

Abstract: The real and imaginary parts are proposed as an alternative to the usual Polar representation of complex-valued images. It is proven that the transformation from Polar to Cartesian representation contributes to decreased mutual information, and hence to greater distinctiveness. The Complex Scale-Invariant Feature Transform (CSIFT) detects distinctive features in complex-valued images. An evaluation method for estimating the uniformity of feature distributions in complex-valued images derived from intensity-range images is proposed. In order to experimentally evaluate the proposed methodology on intensity-range images, three different kinds of active sensing systems were used: Range Imaging, Laser Scanning, and Structured Light Projection devices (PMD CamCube 2.0, Z+F IMAGER 5003, Microsoft Kinect).

Keywords: image-based registration; SIFT; complex-valued image; mutual information; active sensor; range imaging; laser scanning; structured light projection

1. Introduction

The detection of local features in data is of general interest in several disciplines, e.g., Photogrammetry, Remote Sensing, and Computer Vision. According to [1], good features should have the following properties: repeatability, distinctiveness/informativeness, locality, quantity, accuracy, and efficiency. A general overview of the performance of some important algorithms and resulting descriptors at points of

interest is given by [2]. If points of interest are detected they can be utilised to locate correspondences between images. Various applications are known that are based on such points of interest, e.g., image-based registration, object recognition and segmentation, image stitching, self localisation, egomotion and trajectory estimation, as well as 3D reconstruction.

Typically, image-based registration methods focus on intensity images, which gives rise to the question of how to treat combined intensity-range data that can be obtained from particular sensors. For instance, existing registration methods for such data use either the intensity information or the range information, which is treated either as a point cloud, e.g., by applying the costly Iterative Closest Point (ICP) algorithm [3], or as an image, e.g., by applying the Scale-Invariant Feature Transform (SIFT) [4,5].

One might suggest that a separate treatment of the intensity and range data might be a sufficiently combined method. Indeed, some authors use the classical SIFT on range and intensity images separately [6–10]. This combined concept can in fact be viewed as representing the data with complex numbers which are close to the nature of the data itself: range measurements are often in fact phase measurements and intensity is obtained by measuring the amplitude. A possible dependence between range and intensity due to significant mutual information cannot be excluded. Depending on the application, greater mutual information can be desirable [11]. However, in feature detection low mutual information is important, and also fulfills the requirement “*Distinctiveness/informativeness: The intensity patterns underlying the detected features should show a lot of variation, such that features can be distinguished and matched.*”, as outlined by [1]. Therefore, considering the other traditional representation of complex numbers by the real and imaginary parts becomes important for fused intensity-range images. The fusion of intensity-range data asks for their holistic treatment. In the case of the Polar or Cartesian representation of such images, the Complex Scale-Invariant Feature Transform (CSIFT) is a natural generalisation of SIFT. Any particular interest detector, e.g., SURF, MSER, Harris, can be generalised to complex-valued images, SIFT has been chosen for reasons of example only.

Traditionally, the data consist of radiometric images captured with a passive sensor, e.g., a digital camera. Most active sensors, e.g., range imaging, laser scanning, or structured light projection devices provide additional intensity information beside the range. The measured intensity of active sensors can generally be separated in an active and passive intensity. The active intensity is often described as an amplitude and depends just on the measured scattering received by the active illumination with the sensor, e.g., a laser or diode. The passive intensity measured with an active sensor is often called background illumination, and depends on the illumination given by available extraneous light, e.g., sun light. The passive illumination captured with an active sensor might usually have low spectral information, due to the spectral bandpass filters which are in general used. Further, the range is measured which is for most users of main interest. Sometimes only a phase measurement is utilised to determine the range, where a limited uniqueness range is given by the lowest modulation frequency. These data can be described in a unified manner using complex numbers. This has the advantage of providing a general framework for passive and active sensors, without restrictions.

The aim of this article is to provide a method for obtaining in complex-valued images more independent real-valued representations by transformations which decrease mutual information. At the same time the method is aimed at increasing the number of features as well as the uniformity of their distribution.

2. Methodology

The complex-valued image description is introduced. Different representations of complex-valued images are compared with respect to mutual information. An evaluation method for estimating the uniformity of feature distributions is proposed.

2.1. Complex-Valued Image Description

The data provided by active sensors consists of the active and passive intensities together with the range information. In this article, it is assumed that the latter is given by a phase measurement. In this wide-spread method, the phase is usually interpreted as actual distance. However, this approach causes problems if the observed object is at a distance beyond the uniqueness range. For this reason, we will always interpret that information as what it is: namely a *phase* value (Note: The measured phase difference (a physical quantity) can be represented by a phase value (a mathematical quantity).). Hence, a description of an active-passive image using complex numbers becomes quite natural.

Throughout this article, x, y are image coordinates, $r = r(x, y)$ is the range image, $I_a(x, y, r)$ the active intensity image, and $I_p(x, y)$ the passive intensity image. The latter does not depend on the range r . The complex-valued image function is now defined as:

$$f(x, y, r) = I_p(x, y) + I_a(x, y, r)e^{i\phi(x, y, r(x, y))} \quad (1)$$

where the phase $\phi = \phi(x, y, r) \in [0, 2\pi)$ is defined via the range

$$r = n \cdot 2\pi\ell + \phi\ell \quad (2)$$

with $n \in \mathbb{N}$. Notice that passive intensity is treated here as an offset. In this setting, $2\pi\ell$ is the *uniqueness range* of the camera with $\ell \in \mathbb{N}$. The natural number ℓ is a multiple of some unit of length, and n is the “wrapping number”.

The two standard ways of representing complex numbers yield two different image representations: the Polar representation

$$f = |f| e^{i \arg(f)} \quad (3)$$

where

$$|f| = \sqrt{I_p^2 + 2I_p I_a \cos \phi + I_a^2} \quad (4)$$

$$\arg(f) = \arctan \frac{I_a \sin \phi}{I_p + I_a \cos \phi} \quad (5)$$

and the Cartesian representation

$$f = \operatorname{Re}(f) + i\operatorname{Im}(f) \quad (6)$$

where

$$\operatorname{Re}(f) = I_p + I_a \cos \phi \quad (7)$$

$$\operatorname{Im}(f) = I_a \sin \phi \quad (8)$$

Throughout the article, it is assumed that

$$\max |f| = 1 \quad (9)$$

for all complex images. This normalisation can be achieved through division by the maximal value of $|f|$. The remainder of this article will discuss these two different representations of complex images coming from different types of sensors from the entropy perspective.

2.2. Mutual Information in Complex-Valued Images

If a representation of complex-valued images f with real values is given, the image-value dimension is at least two. However, the information content of data is known to depend on their representation. For complex-valued images, this means that some real-valued representations could be more preferred than others from the viewpoint of information theory. For this purpose, the differential entropy is defined as

$$E_q = - \int_R \rho(q) \log \rho(q) dq \quad (10)$$

where R is the range of quantity q , dq is a probability measure and $\rho(q)$ is the distribution function of q .

If $q = (A, \omega)$, then $E_q = E_{A,\omega}$ becomes the joint entropy of amplitude $A = |f|$ and phase $\omega = \arg(f)$. Likewise, $E_{X,Y}$ is the joint entropy of the real and imaginary parts $X = \text{Re}(f)$, $Y = \text{Im}(f)$ of the complex-valued image:

$$f = A \cos \omega + iA \sin \omega \quad (11)$$

It is a fact that the entropy of a system depends on the choice of coordinates, the change in entropy being dependent on the Jacobian of the transformation (cf. e.g., [12]). In the case of complex-valued images, this general result specialises to a preference of Cartesian over Polar coordinates:

Theorem 2.1. *The transformation from the Polar to the Cartesian image representation increases the entropy. More precisely, it holds true that*

$$E_{A,\omega} = E_{X,Y} + \langle \log A \rangle \quad (12)$$

where

$$\langle \log A \rangle = \int \rho(X, Y) \log A(X, Y) dXdY < 0 \quad (13)$$

and $\rho(X, Y)$ is the joint distribution function of $X = \text{Re}(f)$ and $Y = \text{Im}(f)$.

Proof. The statement follows from the well-known transformation rule of the distribution function:

$$\rho(X, Y) = \rho(A, \omega) \cdot |J| \quad (14)$$

where J is the Jacobian of $(A, \omega) \rightarrow (X, Y)$, the transformation from the Polar to the Cartesian image representation. In this case, $J = A(X, Y) = \sqrt{X^2 + Y^2} \leq 1$, since A is normalised amplitude by Equation (9). It follows that the mean $\langle \log A \rangle$ is negative. \square

As a consequence, Theorem 2.1 allows to compute the difference in mutual information

$$MI(a, b) = E_a + E_b - E_{a,b} \tag{15}$$

for the pairs (X, Y) and (A, ω) from the individual entropies:

$$\begin{aligned} \mu &:= MI(X, Y) - MI(A, \omega) \\ &= (E_X + E_Y) - (E_A + E_\omega) + \langle \log A \rangle \end{aligned} \tag{16}$$

and the quantity μ becomes a measure for the gain of independence by transforming from Polar to Cartesian image representation. Namely, $MI(a, b) = 0$ if and only if the quantities a and b are independent, and $MI(a, b)$ can be interpreted as the degree of dependence between a and b . This allows to formulate:

Conjecture 2.2. *For almost all complex-valued images, there is a gain of independence by transforming from the Polar to the Cartesian image representation. In other words: $\mu < 0$ for almost all complex-valued images.*

In fact, the experiments of Section 3 indicate that

$$E_X + E_Y \leq E_A + E_\omega \tag{17}$$

which means that $\leq \langle \log A \rangle \leq 0$.

2.3. Naive Approach

For range measurements within the uniqueness range, the well-known inverse-square law of active intensity implies the approximation:

$$I_a \approx \frac{I_p}{\phi^2} \tag{18}$$

where the phase ϕ is identified with the range r (w.l.o.g. $\ell = 1$ in Equation (2)). This means that it does make sense to consider I_a and I_p as correlated and detect features only in the pair (I_a, ϕ) . This is called the *naive approach*. Hence, there are two successive transformations leading to our complex image in Polar representation:

$$I_a e^{i\phi} \rightarrow |f| e^{i \arg(f)} \rightarrow \text{Re}(f) + i \text{Im}(f) \tag{19}$$

where

$$|f|^2 = I_a^2 (\phi^4 + 2\phi^2 \cos \phi + 1) \tag{20}$$

$$\arg(f) = \arctan \frac{\sin \phi}{\phi^2 + \cos \phi} \tag{21}$$

$$\text{Re}(f) = I_a (\phi^2 + \cos \phi) \tag{22}$$

$$\text{Im}(f) = I_a \sin \phi \tag{23}$$

with Jacobian of the composite map being

$$J'' = I_a (\phi^2 \cos \phi - 2\phi \sin \phi + 1) \tag{24}$$

We wish to exclude the possibility that the benefit from $\langle \log |f| \rangle \leq 0$ of the second transformation (Polar to Cartesian image representation with Jacobian $J = |f|$) is jeopardised by the first transformation with Jacobian J' .

The relation between J , J' and J'' for general composed transformations $(a, b) \rightarrow (a', b') \rightarrow (a'', b'')$ is known to be

$$J'' = J \cdot J' \quad (25)$$

Hence,

$$E_{a,b} = E_{a'',b''} + \langle \log |J| \rangle + \langle \log |J'| \rangle \quad (26)$$

and it follows that

$$\langle \log |J''| \rangle = \langle \log |J| \rangle + \langle \log |J'| \rangle \quad (27)$$

where the means are each taken over the corresponding probability distribution. Hence, we would like to exclude large positive values of $\langle \log |J'| \rangle$. From Equations (20), (24) and (25), it follows that

$$J' = \frac{\phi^2 \cos \phi - 2\phi \sin \phi + 1}{\sqrt{\phi^4 + 2\phi^2 \cos \phi + 1}} \quad (28)$$

which depends only on ϕ . Notice that the denominator is strictly positive, and a closer look reveals that $\langle \log |J| \rangle < 0$ if ϕ is not concentrated in some specific small neighbourhood of π .

Notice, that the inverse-square law in Equation (18) can be used to estimate missing values of I_a or I_p in order to obtain our complex image representation. Use of this will be made in the following section.

2.4. Feature Distribution in Complex-Valued Images

Scale-space feature detection usually involves finding extrema in real-valued functions, and these are obtained from the image through filtering. In the case of complex-valued images f , it makes sense to detect features individually in the components of a representation over the real numbers. This means, for the Polar representation, the detection of features in $|f|$ and in $\arg(f)$, and for the Cartesian representation in $\text{Re}(f)$ and $\text{Im}(f)$. The classical SIFT can be applied to any kind of real-valued images. In particular, applying SIFT to the pairs $(|f|, \arg(f))$ or $(\text{Re}(f), \text{Im}(f))$ componentwise defines CSIFT. If the complex-valued image is represented by the pair (u, v) of real values, a *feature* for CSIFT is defined as a point which is a classical SIFT-feature for u or v .

The preferred representation usually has the desired property that it contains more features, and these are also more uniformly distributed over the image grid than in other representations. More texture in an image can be obtained by increasing the entropy. Hence, a transformation whose Jacobian has absolute value less than one yields more texture by Equation (14), and Theorem 2.1 then says that the Cartesian representation yields more structure than the Polar representation. On the other hand, using the scale-space equation $\frac{\partial}{\partial t} f = \Delta f$ aims at finding texture which is sufficiently persistent through the filtering cascade. Hence, increasing entropy of the image derivative leads to more persistent texture. And also from this persistence point of view, the Cartesian representation turns out more advantageous than the Polar representation:

Theorem 2.3. Transforming from Polar to Cartesian image representation increases the entropy of the scale-space derivatives:

$$E_{\dot{A}, \dot{\omega}} = E_{\dot{X}, \dot{Y}} + \langle A \cdot |\cos \dot{\omega} \sin \omega - \sin \dot{\omega} \cos \omega| \rangle \tag{29}$$

with

$$\langle A \cdot |\cos \dot{\omega} \sin \omega - \sin \dot{\omega} \cos \omega| \rangle < 0 \tag{30}$$

where the expectation value is taken over the joint probability distribution of \dot{X} and \dot{Y} .

Proof. In the light of Theorem 2.1, the statement follows from the Jacobian

$$J = A \cdot (\cos \dot{\omega} \sin \omega - \sin \dot{\omega} \cos \omega) \tag{31}$$

of the transformation of derivatives. □

A Cartesian feature in a complex-valued image is defined to be a scale-space feature for $\text{Re}(f)$ or $\text{Im}(f)$, and, similarly, a Polar feature of f is a scale-space feature for $|f|$ or $\arg(f)$. Consequently, one can formulate:

Conjecture 2.4. The expected number of Cartesian features is larger than the expected number of Polar features for almost all complex-valued images f .

It is natural that a mere increase in the number of features is not sufficient for many applications, e.g., the more the points of interest are concentrated in one small portion of the image, the less relevant their number becomes for estimating the relative camera pose. Hence, an important issue is the distribution of features on the image grid. In fact, it is often desired to know that they are sampled from a uniform distribution.

For n independent, identically distributed random variables X_i , the empirical distribution function $F_n(x)$ is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{(-\infty, x]}(X_i) \tag{32}$$

where $\delta_{(-\infty, x]}$ is the indicator function

$$\delta_{(-\infty, x]}(Y) = \begin{cases} 1, & Y \leq x \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

Then, by the Glivenko-Cantelli Theorem [13,14], the F_n converge uniformly to their common cumulative distribution function F :

$$\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \longrightarrow 0 \quad a.s. \tag{34}$$

with increasing number n of observations. For arbitrary cumulative distribution functions F , the expression $\|F_n - F\|_\infty$ is known as the *Kolmogorov-Smirnov statistic*. It has the general properties of a distance between cumulative distribution functions. Therefore, it will be called here the *KS-distance*.

In the case of a snapshot taken from a scene, one can assume that the observed n features are produced by the scene independently from another. By viewing the scene as the single source of features, one can further assume that the features are identically distributed. In other words, they can be assumed taken from a common cumulative distribution function F .

However, there seems to be no straightforward generalisation of the KS-distance to the multivariate case, as indicated by [15], in particular the proposed generalisations seem to lack robustness. Therefore, we simply propose the Euclidean norm of the two coordinate-wise KS-distances:

$$d(S, \lambda) := \sqrt{\|F_n^x - \lambda_x\|^2 + \|F_n^y - \lambda_y\|^2} \quad (35)$$

where S is a sample of n points in the plane, F_n^i is the empirical distribution function, λ_i and λ are the cumulative density functions of the uniform distribution on the i -coordinate axis and on the plane, respectively. This will be called the *Euclidean KS-distance* to uniformity. Conjecturally, there will be more uniformity in the detected scale-space features for the Cartesian representation than in those for the Polar representation. Let $S_{\text{Cartesian}}$ be the sample of Cartesian features, and S_{Polar} the sample of Polar features of a given complex-valued image f . Then:

Conjecture 2.5. *For almost all complex images f , the transformation from Cartesian to Polar image decreases the Euclidean KS-distance to uniformity:*

$$d(S_{\text{Cartesian}}, \lambda) < d(S_{\text{Polar}}, \lambda) \quad (36)$$

where λ is the uniform distribution on the image plane.

Conjecture 2.2 says that the pair $S_{\text{Re}(f)}, S_{\text{Im}(f)}$ will be more independent than $S_{|f|}, S_{\text{arg}(f)}$. Conjecture 2.4 says that there will be more Cartesian than Polar features. Intuitively, these conjectures together support Conjecture 2.5.

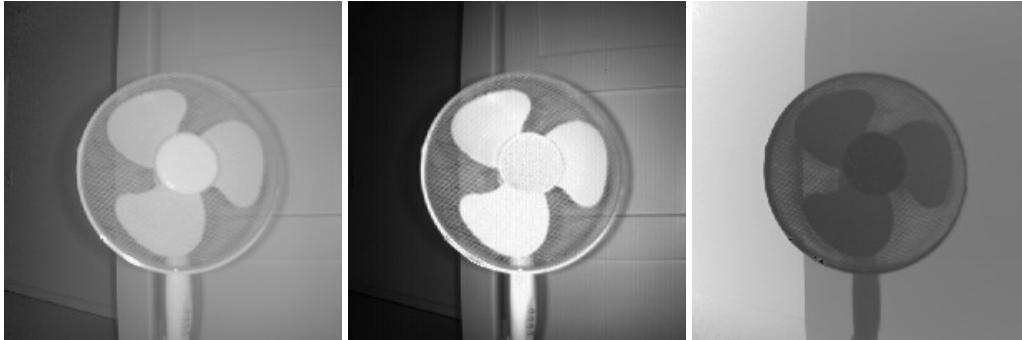
3. Experiments

In order to experimentally evaluate the proposed methodology, three different kinds of active sensing systems were used to capture data from three different scenes: Range Imaging (RIM), Terrestrial Laser Scanning (TLS), and Structured Light Projection (SLP) devices.

3.1. Range Imaging (RIM)

The PMD [Vision] CamCube 2.0 captures range and intensity simultaneously. The former is obtained through phase measurements, and the latter is differentiated between active and passive intensity. The measured active intensity depends on the illumination by the sensor, whereas the passive intensity depends on the background illumination (e.g., the sun or other light sources). The uniqueness range of 7.5 m is defined by the utilized modulation frequency of 20 MHz.

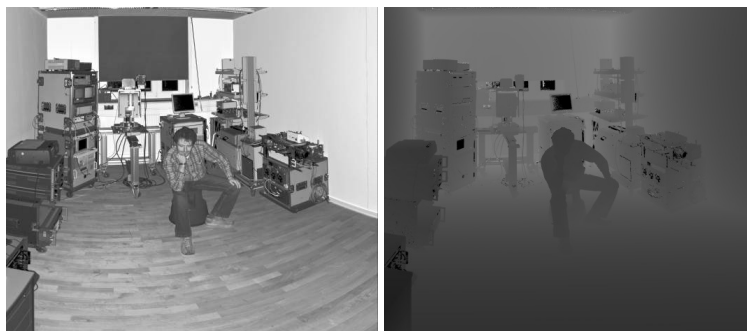
The data meet exactly the requirements of the methodology. The image size for all data is 204×204 pixels. Figure 1 shows the values of I_p , I_a , and ϕ , respectively.

Figure 1. RIM. I_p , I_a , ϕ (from left to right).

3.2. Terrestrial Laser Scanning (TLS)

The Z+F IMAGER 5003 is a standard phase-based laser scanner with survey-grade accuracy within the mm range. While scanning the scene a phase is measured and by using the modulation frequency the range between scanner and object surface is calculated. Again the uniqueness range of 79 m is given by the lowest utilized modulation frequency. Also measured is the intensity I_{raw} which is a mixture of active and passive intensity, because the different intensities are not distinguished.

To adapt the measured data to the methodology, the range image is converted to a phase image ϕ , and the intensity is separated to active and passive intensity using the inverse-square law in Equation (18). The indoor scene in Figure 2 guarantees that all range values lie within the uniqueness range. The selected image area has 532×468 pixels. The I_{raw} and ϕ are shown.

Figure 2. TLS. Left: I_{raw} , Right: ϕ .

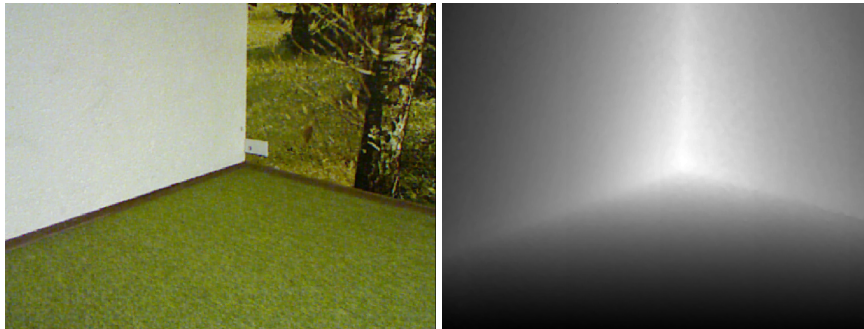
3.3. Structured Light Projection (SLP)

In Microsoft's Kinect for the Xbox 360, active range measurement is based on continuously-projected infrared structured light. Additionally an RGB image is synchronously captured. Due to the multi-static sensor design where the sensors are at different locations, the data are captured with a slight parallax. Through a calibration, the images are properly aligned.

Finally, in order to adapt the measured data to the methodology, the measured range image is interpreted as a phase image with uniqueness range given by the measured maximal range value. The RGB image,

after conversion to gray values, is interpreted as the passive intensity image I_p . In the same way as with the TLS data, Equation (18) is used to estimate I_a . The images have 480×640 pixels. Figure 3 shows I_p and ϕ .

Figure 3. SLP. Left: I_p , right: ϕ .



4. Results

Table 1 shows the entropy results for the various images. A first observation is that the real parts $\text{Re}(f)$ have similar entropies as the absolute values $|f|$ and are relatively high, whereas the entropies of the imaginary parts $\text{Im}(f)$ are relatively low and similar to those of the angle values $\arg(f)$. The last two columns show firstly that the main contribution to the gain of independence μ comes from the Jacobian of the transformation. Secondly, there is an extra gain due to the observed validity of inequality (17). In all examples, the Cartesian images are more independent than the Polar images.

Table 1. Entropy, mean log-Jacobian, and mutual information.

	$E_{I_{\text{raw}}}$	Polar		Cartesian		$-\langle \log f \rangle$	$-\mu$
		$E_{ f }$	$E_{\arg(f)}$	$E_{\text{Re}(f)}$	$E_{\text{Im}(f)}$		
RIM	7.4184	6.6761	1.0976	7.0763	1.2991	0.9494	1.2929
TLS	7.2282	7.213	4.9165	6.6763	4.6484	1.2193	1.4772
SLP	7.8106	7.1925	1.0543	7.2595	0.9879	0.8303	0.8309

For the feature detection with CSIFT, Vedaldi's Matlab implementation of SIFT [16] was used. CSIFT was applied on two snapshots of the same scene. A pair of homologous matched feature points is given by a successful CSIFT matching of candidate keypoints at most one pixel apart. Table 2 reveals that complex images contain more homologous matched feature points than the sole intensity image, and the Cartesian representation contains more than the Polar representation.

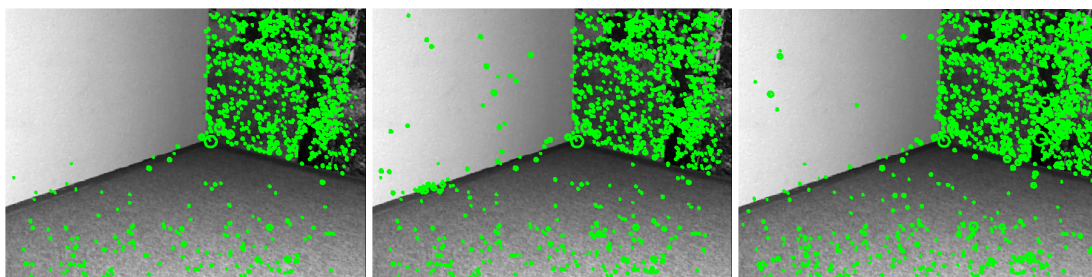
Table 2. Number of homologous matched feature points.

	I_{raw}	$ f $	$\arg(f)$	Polar	$\text{Re}(f)$	$\text{Im}(f)$	Cartesian
RIM	159	159	59	218	156	119	275
TLS	1335	1335	222	1557	567	1133	1700
SLP	1004	1004	81	1085	967	1084	2051

Table 3 indicates that the complex feature points are more uniformly distributed than those in the corresponding intensity image. In all examples, the Cartesian representation has the smallest Euclidean KS-distance to uniformity. Figure 4 depicts the locations of the homologous matched feature points in the SLP case. It can be observed that the raw intensity has no valid detections in the left sector consisting of a homogeneous surface. Both complex image representations have more texture in that sector, therefore valid features are detected. The Polar image contains a cluster in the left border between the left and the bottom sector which is likely responsible for the relatively large horizontal KS-distance to uniformity. The right sector has a large amount of texture in all cases. However, in contrast to the others, the Cartesian image contains valid features very near the right boundary of that sector.

Table 3. KS-distances to uniformity of homologous matched feature points.

	horizontal			vertical			Euclidean		
	S_{raw}	S_{Polar}	$S_{\text{Cartesian}}$	S_{raw}	S_{Polar}	$S_{\text{Cartesian}}$	S_{raw}	S_{Polar}	$S_{\text{Cartesian}}$
RIM	0.18294	0.17035	0.16402	0.12487	0.11194	0.12745	0.22149	0.2039	0.20754
TLS	0.1699	0.18024	0.1815	0.17369	0.16188	0.15585	0.24297	0.24264	0.23921
SLP	0.04855	0.064263	0.059843	0.051329	0.033411	0.036508	0.070652	0.072429	0.0701

Figure 4. Homologous matched feature points in SLP image: S_{raw} , S_{Polar} , $S_{\text{Cartesian}}$ (from left to right).

5. Conclusions

High mutual information of variables a , b means that they are redundant. Therefore, it is of interest to compare the two standard representations of complex-valued images: the Polar and the Cartesian representation. We have deduced through theoretical considerations the general conjecture that the

mutual information of the real and imaginary parts of an information is lower than the amplitude and phase parts. We have verified this experimentally by applying CSIFT, and have found that not only does the number of valid detected features increase through transforming from the Polar to the Cartesian image representation, but also the uniformity of their distribution.

The implication is that in feature detection within fused intensity-range images, as e.g., obtained by Range Imaging, Laser Scanning or Structured Light Projection devices, the Cartesian representation is to be preferred over the Polar representation in order to achieve more distinctiveness and informativeness, one of the requirements for local features from [1].

Acknowledgements

The authors would like to thank Sven Wursthorn and Martin Weinmann for fruitful discussions.

References

1. Tuytelaars, T.; Mikolajczyk, K. Local invariant feature detectors: A survey. *Found. Trends Comput. Graph. Vis.* **2008**, *3*, 177–280.
2. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *10*, 1615–1630.
3. Rusinkiewicz, S.; Levoy, M. Efficient Variants of the ICP Algorithm. In *Proceedings of the 3rd International Conference on 3D Digital Imaging and Modeling (3DIM)*, Quebec City, QC, Canada, 28 May–1 June 2001; pp. 145–152.
4. Lowe, D. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the International Conference on Computer Vision (ICCV 1999)*, Corfu, Greece, 20–23 September 1999; pp. 1150–1157.
5. Lowe, D. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91C110.
6. Haker, M.; Böhme, M.; Martinetz, T.; Barth, E. Scale-Invariant Range Features for Time-of-Flight Camera Applications. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW08*, Anchorage, AK, USA, 23–28 June 2008.
7. Barnea, S.; Filin, S. Keypoint based autonomous registration of terrestrial laser point-clouds. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 19–35.
8. Böhm, J.; Becker, S. Automatic Marker-Free Registration of Terrestrial Laser Scans Using Reflectance Features. In *Proceedings of the 8th Conference on Optical 3D measurement Techniques*, Zurich, Switzerland, 9–12 July 2007; pp. 338–344.
9. Wang, Z.; Brenner, C. Point Based Registration of Terrestrial Laser Data Using Intensity and Geometry Features. In *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, Beijing, China, 3–11 July 2008; Volume XXXVII, pp. 583–590.
10. Weinmann, M.; Weinmann, M.; Hinz, S.; Jutzi, B. Fast and automatic image-based registration of TLS data. *ISPRS J. Photogramm. Remote Sens.* **2011**, in press.

11. Sawada, H.; Mukai, R.; Araki, S.; Makino, S. Polar coordinate based nonlinear function for frequency-domain blind source separation. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2003**, *E86-A*, 590–596.
12. Hnizdo, V.; Gilson, M. Thermodynamic and differential entropy under a change of variables. *Entropy* **2010**, *12*, 578–590.
13. Glivenko, V. Sulla determinazione empirica della legge di probabilita. *Giorn. Ist. Ital. Attuari* **1933**, *4*, 92–99.
14. Cantelli, F. Sulla determinazione empirica della legge di probabilita. *Giorn. Ist. Ital. Attuari* **1933**, *4*, 221–424.
15. Lopes, R.; Reid, I.; Hobson, P. The Two-Dimensional Kolmogorov-Smirnov Test. In *Proceedings of the XI International Workshop on Advanced Computing and Analysis in Physics Research*, Amsterdam, The Netherlands, 23–27 April 2007.
16. Vedaldi, A. SIFT for Matlab. Available online: <http://www.vlfeat.org/~vedaldi/code/sift.html> (accessed on 1 September 2011).

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>.)

POTENTIALS OF IMAGE BASED ACTIVE RANGING TO CAPTURE DYNAMIC SCENES

S. Hinz^a, Ma. Weinmann^a, P. Runge^b, B. Jutzi^a

^a Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Englerstr. 7, 76131 Karlsruhe, Germany

^b Geodetic Institute, Karlsruhe Institute of Technology (KIT), Englerstr. 7, 76131 Karlsruhe, Germany

KEY WORDS: Point cloud, range imaging, multi-view, airborne sensing, dynamic observations.

ABSTRACT:

Obtaining a 3D description of man-made and natural environments is a basic task in Computer Vision and Remote Sensing. To this end, laser scanning is currently one of the dominating techniques to gather reliable 3D information. The scanning principle inherently needs a certain time interval to acquire the 3D point cloud. On the other hand, new active sensors provide the possibility of capturing range information by images with a *single* measurement. With this new technique image-based active ranging is possible which allows capturing dynamic scenes, e.g. like walking pedestrians in a yard or moving vehicles. Unfortunately most of these range imaging sensors have strong technical limitations and are not yet sufficient for airborne data acquisition. It can be seen from the recent development of highly specialized (far-)range imaging sensors – so called flash-light lasers – that most of the limitations could be alleviated soon, so that future systems will be equipped with improved image size and potentially expanded operating range. The presented work is a first step towards the development of methods capable for application of range images in outdoor environments. To this end, an experimental setup was set up for investigating these proposed possibilities. With the experimental setup a measurement campaign was carried out and first results will be presented within this paper.

1. INTRODUCTION AND MOTIVATION

Currently the 3D geometrical capturing and description of the environment is based on (multi-view) image or range data. By utilizing passive imaging sensors the 3D information is gained indirectly from several images with stereo- or multiple image analysis. These procedures are widely used but, for certain kinds of applications, they have indispensable limitations due to the constrained camera set-up, the scene contents, and last but not least because of the inherently ill-posed problem of 3D reconstruction from 2D images. For instance the illumination conditions should be adequate, the observed materials need to be textured and opaque, and the distance between object and camera as well as between the camera observation points of stereo images should be sufficiently large enough for gaining a reliable 3D reconstruction.

The photogrammetric methods are complemented by direct measurement procedures like laser scanning. These active sensors capture a sequence of single range values while conducting a time dependent spatial scanning of the environment. In general space-borne, airborne (ALS) as well as terrestrial laser scanners (TLS) provide a direct and illumination-independent measurement of 3D objects (Shan & Toth, 2008; Vosselman & Maas, 2009). For continuous-wave (CW) modulated laser scanning devices the measuring rate is currently between 150000 and 700000 measurements per second and the operating distance is up to 100m. For pulse modulated laser scanning devices the measuring rate is currently between 10000 and 300000 measurements per second and the operating distance is up to 3000m. However, it must be considered that the time-dependent acquisition of the 3D laser points can cause significant artefacts in the point cloud in case the captured scene contains moving objects.

2. STATE OF THE ART AND CHALLENGES

For deriving accurate 3D world coordinates from range measurements, scene as well as the sensor platform must be static or their relative motions must be known precisely.

Otherwise deformation artefacts of the environment will appear and have to be considered before transferring the measured data in a 3D model. In general with an increasing dynamic of the scene contents, respectively sensor platform, the complexity of the analysis increases and the exploitation of three-dimensional information is more and more challenging, especially for laser scanning systems (Toth & Grejner-Brzezinska, 2006; Yao et al., 2010).

Very recently, enhanced types of active imaging sensors are available, namely Swiss Ranger (www.mesa-imaging.ch) and PMD Vision (www.pmdtec.com). These close-range sensors allow to capturing a range image and a co-registered intensity image simultaneously with a high frame rate up to 100 frames per second, so that not just one (or few) points are captured at the same time but a whole frame. The use of both active and passive illumination provides furthermore information of the ambient light, yet allows also to controlling and adjusting the measurement signal – most prominently regarding frame rate, integration time and modulation frequency – to accommodate for the current acquisition conditions in the best manner possible. Another technical advantage is the monostatic sensor configuration, which allows for observing the area of interest from a single point of view, in contrast to the classical stereo observation techniques with passive sensors, which need at least two different observation points. Henceforth, the advantages of active 3D measurement sensors over images and the simultaneous acquisition of areal data have been unified. This concept thus contains much potential for the automatic analysis dynamic scenes in fully 3D. Especially the 3D monitoring with terrestrial or even airborne platforms in challenging weather and illumination conditions is promising with this novel technology.

The major drawbacks are the limited absolute range accuracy of a few centimeters and the limited unambiguous range: Especially the relatively large noise influence on the measurement – which stems from the large amount of ambient radiation in comparison to the emitted radiation – causes significant inaccuracies of the range measurement. Regarding this aspect, the performance of range imaging (RIM) is usually less reliable than airborne or terrestrial laser scanners.

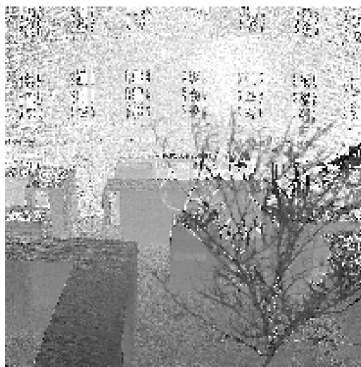
The unique range of the most commercial systems is currently less than 10m and depends on the user-specified modulation frequency. This range measurement restriction can however be significantly relaxed by image- or hardware-based unwrapping procedures, which operate also in far range (Jutzi, 2009; Jutzi, 2011). It could be shown that a range maximum of more than four times of the manufacturers non-ambiguity range specification could be reached without modifying the sensor or improving the illumination unit, e.g. by additional illumination modules. An outdoor example is given by Figure 1: the range images are captured with different modulation frequencies (18 & 21 MHz) and the unwrapped range image is depicted below. The maximum distance within the scene is about 30m.



Range image captured with 18MHz



Range image captured with 21MHz



Unwrapped range image

Figure 1. Range images captured with different modulation frequencies (top: 18MHz; center: 21 MHz) and corresponding unwrapped range image (bottom).

It can be seen from Figure 1 that, although the unwrapped range image is quite noisy concerning the absolute range measurement, the artificial gray value edges of back-folded range measurements have disappeared. With the current technical progress, it can be assumed that the mentioned limitations will be alleviated soon and future systems might be featured by expanded operating range and improved image size.

Beside this the registration procedure is challenging. Usually additional sensor components, e.g. like INS (Inertial Navigation System) and GPS (Global Positioning System), to gain orientation and position of the sensor. Direct measurement of position and attitude of the sensor might still contain systematic errors as it is for instance well-known from strip-adjustment of ALS data. Hence, image-based registration techniques like shown in our previous work (Weinmann et al., 2011; Weinmann & Jutzi, 2011), possibly combined with a bundle approach, should be included to improve accuracy.

In following, some conceptional perspectives regarding the acquisition of dynamic scenes with RIM sensors are described (Section 3). In Section 4, the constructed multi-view range imaging device is introduced, while a “toy scene” for the investigations is shown in Section 5. First results of the measurement campaign and recommendations are given in Section 6. The paper closes with a brief conclusion and outlook.

3. CONCEPTIONAL PERSPECTIVES

While active range scanning devices are more and more established in close-range photogrammetry and computer vision, first – still experimental – developments such as flash-light lasers already show the potential that range imaging devices can be applied to capture larger scenes in the near future. Hence, also the basis for monitoring of highly dynamic scenes can be envisioned. In contrast to the 3D geometry derived by passive sensing techniques (e.g. photogrammetry) the range information is available directly without processing delays. Yet it should be noticed that the range information captured with a single static device is not fully 3D, as only range information corresponding to the well-known bundle of viewing rays can be measured. Still, when using a multi-view camera set-up the observed object or monitored scene can be captured from different directions so that also real 3D descriptions can be derived, even with fewer restrictions than in photogrammetry. In general the multi-view active range imaging can strongly support navigation, (co-)registration, and observing temporal scene changes if a reliable matching procedure is available.

To simulate a future operation of RIM sensors in airborne scene monitoring fairly realistically, a scaled test scenario has been set up. Instead of mounting RIM sensors at unmanned aerial vehicles (UAVs), which involves much efforts and expenses due to the large payload of several kilograms for both the RIM sensors and the data recording system, a sort of cable-car has been constructed (see Figures 3 and 4), on which two RIM sensors and the recording unit have been mounted. This allows (quasi-)airborne monitoring in low altitudes, as they also appear in UAV videos, for capturing dynamic 3D observations like walking pedestrians in a yard or other moving objects.

The RIM sensors can be turned into different pointing directions. For the current tests mainly the over-head option was of used, whereby two general constellations were of main

interest: the convergent (a) and the parallel (b) acquisition geometries (see Figure 2):

a) The convergent acquisition geometry (in over-head pointing direction) provides a wider range of viewing directions due to the oblique optical axes, which can be very helpful for 3D object reconstruction and characterization. On the other side, this concept is very challenging for image-based registration due to the different perspectives and the different object appearance in the two images.

b) The parallel acquisition geometry adapts the so-called normal case of photogrammetry and eases many tasks such as co-registration and mosaiking. Especially for image based registration, this constellation is more cooperative due to the similar viewing geometry and similar object appearance in both images. The point density in scenes with steep relief reduces however.

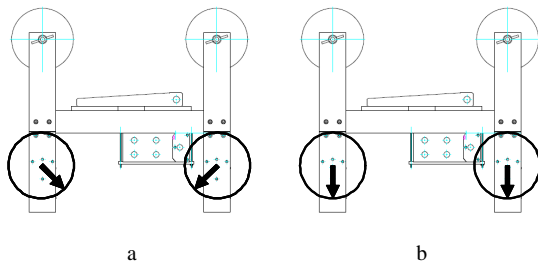


Figure 2. Image acquisition geometry: a) convergent, b) parallel.

4. MULTI-VIEW ACTIVE RANGE IMAGING SYSTEM

To investigate the potentials of multi-view range imaging systems, an experimental setup based on the above-mentioned cable car concept was developed. The system includes various components of the main sensor rack (Figures 3 and 4):

- two RIM sensors (PMD Vision CamCube 2.0)
- unit for variable multi-view options (viewing possibilities are approximately $\pm 90^\circ$ wrt. nadir direction)
- data recording unit for both sensors (notebook with solid state hard disk)
- independent power supply (12V battery with 6.5Ah),
- cable car wheels,
- ropes (100m length).

The measurement staying power is at least 60 minutes and can easily be extended to several hours by utilizing a battery with a larger capacity. For instance, the power consumption of a single PMD Vision CamCube 2.0 is typically between 17W (@2.5ms integration time) and 35W (@10ms integration time). A navigation system to record the absolute position and viewing direction is not on board, as image-based navigation is of our main interest for further investigation. Therefore more focus was put on accurately synchronized image acquisition. To measure the position of the cameras externally, an in-house laser-tracker system could be optionally used.

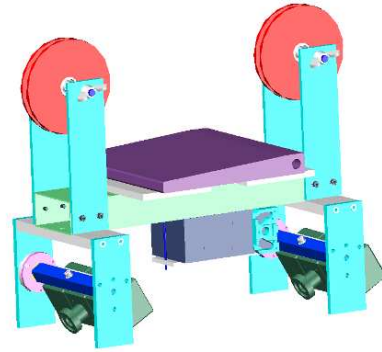


Figure 3. Visualized CAD model of the experimental device: sensor rack carrying two RIM sensors, capturing unit, and power supply.

The RIM sensors are two PMD Vision CamCube 2.0. The sensors have a 204 x 204 pixel array with a pixel size and pitch (spacing) of about 45 μm . The field of view is 40x40°. The maximum frame rate is about 25 frames per second and the sensor measures three features per pixel: range, active intensity and passive intensity. Therefore, above three million measurement values per second can be captured.

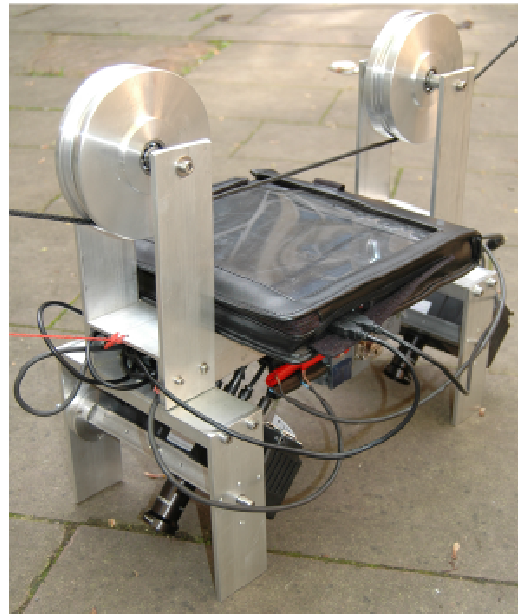


Figure 4. Experimental device ready for measurement.

5. EXPERIMENTS

To obtain first tests and assessments, an outdoor “toy scene” has been set up. It contains bare soil, concrete, a small movable model vehicle and a plant (see Figure 6), where the cable-car with the experimental device could pass by. The scene has been captured with an integration time of 10ms to gain a reasonable signal-to-noise ratio, which is important especially for outdoor measurements. Furthermore various acquisition geometries have been tested. Therefore the rope for the cable-car was mounted at

two points of approximately 3 to 5m height for recording the scene (Figure 5). Please note, that in case of using a convergent viewing geometry as shown in Figure 2a, this height allows already to exceeding the ambiguity range of the PMD sensors.

All tests proved the functionality of the multi-view measurement system.



Figure 5. Experimental device while recording a scene.



Figure 6. Experimental device while recording with scene content.

6. CAPTURED DATA

The captured data can be interpreted in different ways. The measured intensity of active sensors can be generally separated

in an active and passive intensity. The active intensity is often described as amplitude and depends just on the measured scattering received by the active illumination with the sensor, e.g. a laser or diode. The passive intensity measured with an active sensor is often called background illumination, and depends on the illumination given by available extraneous light, e.g. sun light. The passive illumination captured with an active sensor might usually have low spectral information, due to the spectral bandpass filters which are in general used. Further, the range is measured which is for most users of main interest. Sometimes only a phase measurement is utilized to determine the range, where a limited uniqueness range is given by the lowest modulation frequency.

However the captured data for the two acquisition geometries is shown in Section 6.1 (convergent) and Section 6.2 (parallel). For each acquisition geometry, two nearly aligned scene images are depicted. A reliable synchronization of the data is currently an open task which has to be investigated in the future.

6.1 Convergent acquisition geometry

The front sensor is backward looking and the back sensor is forward looking. Obviously with this acquisition geometry more data from the objects side is captured. Active intensity, passive intensity, and phase are depicted in Figure 7. In this case the intensity images look similar except of the different selected dynamic range.

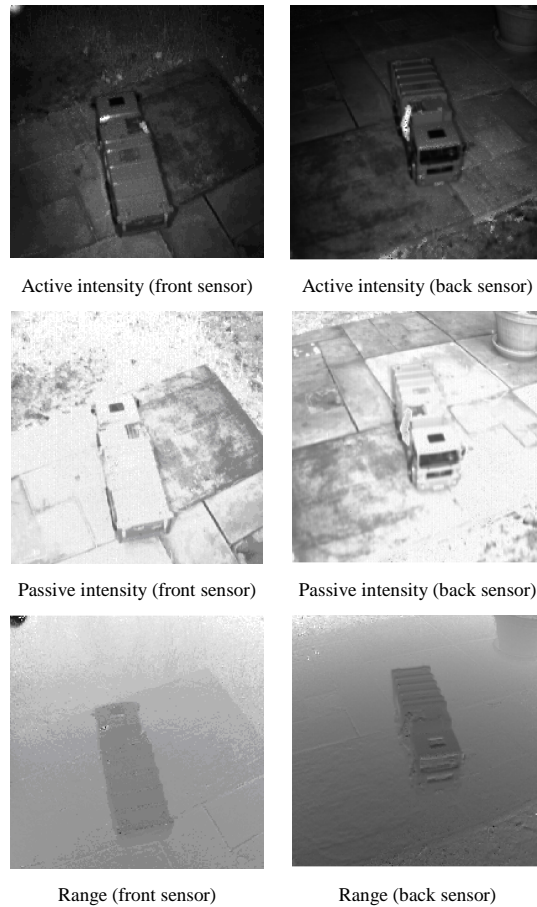


Figure 7. Image results of the convergent acquisition geometry.

With the captured images a point cloud can be generated and visualized (Figure 8). Obviously the point cloud is affected by some noise so that – although visible in principle – the vehicle type is hard to recognize in the range data only, due to the low spatial resolution. As can be seen from Figure 7, a combined image and range data analysis seems very promising.

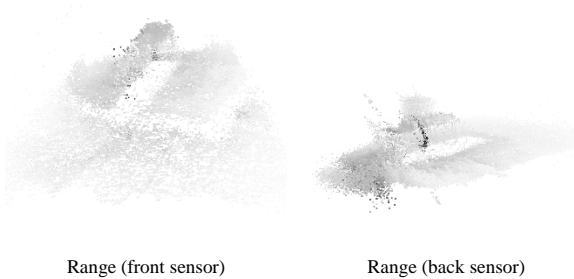


Figure 8. Single shot point cloud results of the convergent acquisition geometry.

6.2 Parallel acquisition geometry

Both sensors are nadir looking with a stereo base of a few decimetres. As expected the images look very similar. Active intensity, passive intensity, and phase are depicted in Figure 9.

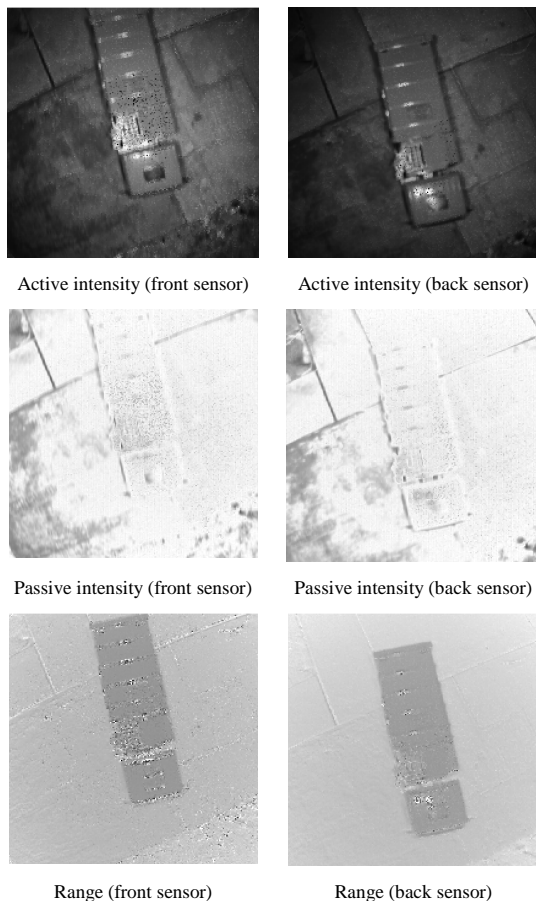


Figure 9. Image results of the parallel acquisition geometry.

Again the captured images are converted to a point cloud and visualized now in Figure 10. Similar to the convergent case, the

point cloud shows considerable noise and the vehicle type is hard to recognize due to the low image resolution. However, again, as can be seen from Figure 9, a combined image and range data analysis is very promising and, of course, easier as for the convergent case.

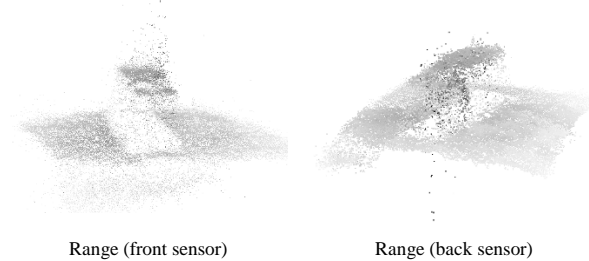


Figure 10. Single shot point cloud results of the parallel acquisition geometry.

7. CONCLUSION AND OUTLOOK

In this paper the first results for a multi-view range imaging device are presented. The captured data looks very promising. However the data has to be further investigated and it has to be shown that range imaging is superior to range scanning devices, especially for dynamic environments. Therefore a lot of tasks like, e.g., co-registration, have to be tackled in the future.

REFERENCES

- Jutzi B (2009) Investigations on ambiguity unwrapping of range images. In: Bretar F, Pierrot-Deseilligny M, Vosselman G (Eds.) *Laserscanning 2009*. International Archives of Photogrammetry and Remote Sensing 38 (Part 3/W8): 265-270.
- Jutzi B (2011) Extending the range measurement capabilities of modulated range imaging devices by time-frequency-multiplexing. *AVN - Allgemeine Vermessungs-Nachrichten*. In press.
- Shan J, Toth CK (2008) (Eds.) *Topographic Laser Ranging and Scanning: Principles and Processing*. Boca Raton, FL: Taylor & Francis.
- Toth CK, Grejner-Brzezinska D (2006) Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation. *ISPRS Journal of Photogrammetry & Remote Sensing*, 61(3-4): 137-148.
- Vosselman G, Maas HG (2009) (Eds.) *Airborne and Terrestrial Laser Scanning*. Whittles Publishing, Caithness, Scotland, UK.
- Weinmann Ma, Jutzi B (2011) Fully automatic image-based registration of unorganized TLS data. *International Archives of Photogrammetry and Remote Sensing*. Submitted.
- Weinmann Ma, Weinmann Mi, Hinz S, Jutzi B (2011) Fast and automatic image-based registration of TLS data. *ISPRS Journal of Photogrammetry & Remote Sensing*. Submitted.
- Yao W, Hinz S, Stilla U (2010) Airborne Analysis and Assessment of Urban Traffic Scenes from LiDAR Data - Theory and Experiments. In: *Proceedings of Workshop of IEEE Conference on Computer Vision and Pattern Recognition '10 (CVPR'10)*: 1-8.

RESEARCH ARTICLE

Weighted data fusion for UAV-borne 3D mapping with camera and line laser scanner

Boris Jutzi^{a*}, Martin Weinmann^a and Jochen Meidow^b

^aKarlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe, Germany; ^bFraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Ettlingen, Germany

(Received 3 November 2013; accepted 18 December 2013)

Unmanned aerial vehicles (UAVs) equipped with adequate sensors have nowadays become a powerful tool for capturing spatial information. In this article, we introduce a concept for weighted data fusion in order to enable an improved UAV-borne 3D mapping with a camera and a lightweight line laser scanner. For this purpose, we carry out geometric camera calibration as well as lever-arm and bore-sight calibration and subsequently present a new methodology for incorporating camera images and laser scanner data into an adjustment process. This adjustment is based on the concept of variance components in order to obtain a reasonable weight ratio for data fusion and accurately estimate the poses of the sensors. We demonstrate the feasibility of the proposed approach and show that the consideration of range measurements clearly improves the pose estimation.

Keywords: camera; laser scanner; UAV; sensor calibration; self-localisation; 3D mapping

1. Introduction

In recent years, unmanned aerial vehicles (UAVs) equipped with adequate sensors have become a powerful tool for capturing spatial information. They represent mobile and automatically operating low-cost solutions, are easy to handle and easy to transport to the surveying field. Typically, these devices are equipped with optical sensors in order to provide images for the operator, to support the localisation of the platform, or to facilitate 3D mapping of the environment. For mapping with optical sensors, a successful and precise localisation of the UAV and a simultaneous 3D mapping of the environment, referred to as simultaneous localisation and mapping (SLAM), can be gained by sensing distinctive elements of the environment. These elements are commonly known as 3D landmarks. Unfortunately, for 3D landmarks usually no prior knowledge about their location is given and therefore the 3D positions of the landmarks have to be estimated by utilising descriptive 2D image features from various observations as accurate as possible. Instead of estimating the 3D positions of the landmarks with passive sensors, an accurate measurement with active sensors could be practicable (Weinmann *et al.* 2011). For this purpose, laser range measurements are typically the first choice, but laser scanners are generally bulky and heavy.

Due to the rapid and substantial progress in miniaturising technology, the latest developments allow to mount suitable laser scanners on UAVs. For instance, Nagai

*Corresponding author. Email: boris.jutzi@kit.edu

et al. (2009) presented a UAV-borne 3D mapping system equipped with inertial measurement unit (IMU) and global positioning system (GPS) receiver, two digital cameras, two IR cameras and a laser scanner. Together, the components have a weight of more than 7 kg, and for this reason a helicopter-like UAV has been utilised. The system is able not only to simultaneously capture geometric 3D information and radiometric information, i.e. textures, but also to derive a vegetation index due to the use of IR cameras as well as to operate for a duration of about 1 h. However, the total weight of the platform is 330 kg. Hence, the system is neither low-cost nor easy to bring to the surveying field. A smaller UAV for rapid close-range monitoring which integrates different types of sensors and supporting modules has been proposed by Choi and Lee (2011). Among these components, there are GPS receiver and IMU as well as two digital cameras and a laser scanner. The whole system for data acquisition has a weight of above 10 kg. As a result, a high-quality digital elevation model and orthophotos can be obtained, but the use of a relatively large UAV is required due to the heavy payload. A further platform equipped with IMU, GPS receiver, camera and laser scanner has been presented by Wallace *et al.* (2012) and specially designed with respect to low costs and maximum flying time. However, the system only allows short flights of several minutes for a relatively heavy payload of up to 2.8 kg. Hence, the system is only suitable for acquiring point cloud data for a relatively small area. Recently, Conte *et al.* (2013) presented a system consisting of a high-precision GPS receiver, an IMU, a compass and a multi-echo lidar sensor for airborne terrain mapping. The components have a total weight of more than 4 kg, and they are carried by an industrial unmanned helicopter with a maximum take-off weight of 95 kg.

In contrast to all the aforementioned systems, lightweight systems are desirable for capturing larger 3D environments with low costs. Such systems can be established by reducing the number of components and, for being able to use mini UAVs, by selecting lightweight devices for the relevant components. In this context, the applied laser scanner has been a critical component for a long time. Since only a few months, the new lightweight single-line laser scanner Hokuyo UTM-30LX-EW is available (210 g without cable), which allows to capture multiple reflections and their corresponding intensity values for each transmitted laser pulse. Whereas line laser scanners are often used for navigation of mobile ground vehicles, particularly in indoor environments with flat ground, their use for UAVs has rarely been addressed in recent years. Devices providing 2D scans have, for instance, been used within indoor environments for obtaining 3D measurements in a plane around the UAV (Bachrach *et al.* 2009, Grzonka *et al.* 2009). For more complex environments, full 3D scans are required, or additional data has to be taken into account. In order to obtain 3D scans with a line laser scanner, a system exploiting an actuator for continuously rotating a Hokuyo UTM-30LX-EW laser range finder has recently been proposed for omnidirectional obstacle perception and flight trajectory estimation (Droeschel *et al.* 2013, Holz *et al.* 2013). Involving further devices, Kuhnert and Kuhnert (2013) focused on a system for mapping or monitoring purposes in outdoor environments, particularly for monitoring high-voltage power lines. Their system includes IMU, GPS receiver, digital camera and laser scanner. As the total weight of all components is relatively small, a mini UAV is used which is able to carry a payload of 1.5 kg. For laser scanning, the authors tested a Hokuyo UTM-30LX-EW with range measurements up to 30 m and a more expensive SICK LD-MRS-400001 with range measurements up to 200 m, both with an angular resolution of 0.25° .

In this article, an extension of the method presented by Jutzi *et al.* (2013) is proposed for fusing optical and laser scanner data. The main contribution is the introduction of a concept for weighted data fusion in order to enable an improved UAV-borne 3D mapping

with an appropriate camera and a lightweight line laser scanner. For this purpose, different kinds of data in the form of images and laser scanner measurements are fused by

- geometrically calibrating the camera for determining its intrinsic parameters,
- estimating the relative orientation between camera and laser scanner, i.e. lever-arm and bore-sight calibration,
- applying the Lucas-Kanade tracker for tracking distinctive image points across consecutive image frames, where a significant advantage is achieved by utilising laser range measurements for determining the scale of the observations, and
- bundle adjustment providing variance components for camera and laser scanner data in order to accurately estimate the pose of the UAV sensor platform.

The latter is essential for weighting measurements of different types appropriately. Already Helmert (1924) provided methods to check and improve the assumptions on the precision of the observations. Later various estimators for variance components based on statistical criteria have been devised, e.g. Koch (1986) or Rao (1973). The minimum norm quadratic unbiased estimation (MINQUE), for example, has successfully been applied for the analysis of GPS signals (Satirapod *et al.* 2002).

The article is organised as follows. In Section 2, we first describe the proposed methodology for weighted data fusion in order to enable UAV-borne 3D mapping with camera and line laser scanner. The weighting is achieved by introducing variance components and supports reliable pose estimation. Subsequently, in Section 3, we describe the automatic procedure for calibrating the sensor platform. This involves geometric camera calibration as well as lever-arm and bore-sight calibration. In Section 4, the indoor and outdoor experiments are described and the derived results, e.g. 3D laser points projected onto images, are presented. After discussing the results in Section 5, the concluding remarks and suggestions for future work are finally provided in Section 6.

2. Methodology

For weighted data fusion, we propose the incorporation of camera images and laser scanner data into an adjustment process which takes into account different types of observations by using the concept of variance components. Our methodology exploits

- range measurements,
- tracking of salient image points, and
- tracking of image points obtained by the projection of laser points onto the image plane.

In the following, we describe the workflow of our methodology in detail.

2.1. Adjustment model

For the proper fusion of measured image and range information, we utilise a standard bundle adjustment, augmented by the observational equations for the measured ranges. Since we have metric information at hand, we have to fix six degrees of freedom for the resulting scene model because of the gauge freedom. For the solution of the optimisation problem by exploiting the sparse structure of the involved matrices, we refer to Triggs *et al.* (2000). To account for the different kinds of measurements stemming from

heterogeneous sensors, we refine the stochastic model by introducing the so-called variance components which allow a proper weighting of the two groups of measurements (McGlone *et al.* 2004).

2.1.1. Functional model: observational equations

Image coordinates. For each extracted or tracked image point, the corresponding unknown 3D point \mathbf{X}_i in the object coordinate system S_o is mapped into the image plane by the collinearity equations, in homogeneous representation

$$\mathbf{x}_{it} = \mathbf{P}_t \mathbf{X}_i^{(o)} \quad (1)$$

for each pose t with the 3×4 projection matrix \mathbf{P}_t and the resulting image point $\mathbf{x}_{it} = [u_{it}, v_{it}, w_{it}]^T$ in homogeneous coordinates, too. The projection matrix $\mathbf{P}_t = \mathbf{K} \mathbf{R}_t [\mathbf{I}_{3 \times 3}, -\mathbf{C}_t^{(o)}]$ contains the exterior orientation of the camera, i.e. the pose represented by the position of the projection centre $\mathbf{C}_t^{(o)}$ in the world coordinate system S_o and the rotation matrix \mathbf{R}_t from the world to the camera coordinate system. The intrinsic camera parameters are compiled in the homogeneous calibration matrix \mathbf{K} , including a model for lens distortion.

Given a measured range $s_t^{(l)}$ and the corresponding direction, the 3D point \mathbf{X}_l is obtained by polar point determination in the coordinate system S_l of the laser scanner. The projection of such a 3D laser point into the image plane is then

$$\mathbf{x}_l = \mathbf{K} \mathbf{X}_l^{(c)} = \mathbf{K} \mathbf{R} [\mathbf{I}_{3 \times 3}, -\mathbf{T}] \mathbf{X}_l^{(l)} \quad (2)$$

with the six parameters of the laser-to-camera transformation, expressed by the rotation matrix \mathbf{R} and the lever-arm \mathbf{T} , with the laser point $\mathbf{X}_l^{(c)}$ in the camera coordinate system S_c as an intermediate result. In Section 3.2.2, we explicate the determination of the lever-arm and bore-sight parameters.

The Euclidean image points $\mathbf{x} = [x, y]^T$ are obtained by $x = u/w$ and $y = v/w$ for each 3D point, determined either by laser scanning or by spatial forward intersection.

Laser ranges. For convenience, we do not use the original measurements $s_t^{(l)}$ provided in the laser coordinate system but the slightly different ranges from the 3D laser points to the camera projection centres. These are identical in S_c and S_o and can easily be computed by

$$s_l^{(c)} = \|\mathbf{X}_l^{(c)}\| = s_l^{(o)} = \|\mathbf{X}_l^{(o)} - \mathbf{C}_t^{(o)}\| \quad (3)$$

with the laser-to-camera transformation $\mathbf{X}_l^{(c)} = \mathbf{R}(\mathbf{X}_l^{(l)} - \mathbf{T})$ used in Equation (2), too.

2.1.2. Stochastic model: variance components

The assembled sensor system provides heterogeneous observations: image coordinates and measured laser ranges (Section 2.1.1). For the variances of these observations, initially only a rough guess exists: the assumed uncertainty of the image coordinates is given by the precision of feature extraction and tracking, respectively; the uncertainty of the laser ranges is taken from the technical specifications. This implies that the ratio of the variances for both observation groups is unknown. To obtain reasonable variances and

weights, respectively, we apply the concept of variance components (Förstner 1987, McGlone *et al.* 2004).

The stochastic model is obtained by categorising the observations into image coordinates l_1 and laser ranges l_2 . The corresponding covariance matrix

$$l = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix}, \quad \Sigma_{ll} = \begin{bmatrix} \sigma_{01}^2 \Sigma_{11} & \mathbf{O} \\ \mathbf{O} & \sigma_{02}^2 \Sigma_{22} \end{bmatrix} \quad (4)$$

implies statistical independence between both groups, and the matrices Σ_{11} and Σ_{22} contain the initially assumed covariances for the observations. Without loss of generality, we assume $\Sigma_{11} = \sigma_1^2 \mathbf{I}$ and $\Sigma_{22} = \sigma_2^2 \mathbf{I}$ with an assumed identical standard deviation σ_1 for all image coordinates and σ_2 for all laser ranges, respectively.

During the iterative estimation process, these covariance matrices become corrected by the so-called estimated variance components which converge to 1. For the estimation of these factors, the covariance matrix

$$\Sigma_{\hat{v}\hat{v}} = \Sigma_{ll} - \mathbf{A} \Sigma_{\hat{x}\hat{x}} \mathbf{A}^T \quad (5)$$

of the residuals $\hat{v} = \Delta l - \mathbf{A} \Delta \hat{x}$ is needed, provided by the linear model $\Delta l + \hat{v} = -\mathbf{A} \Delta \hat{x}$ with the Jacobian \mathbf{A} and the update $\Delta \hat{x}$ for the unknown parameters \hat{x} . The estimates for the two variance factors are then

$$\hat{\sigma}_{0i}^2 = \frac{\hat{v}_i^T \Sigma_{ii}^{-1} \hat{v}_i}{\text{tr}(\Sigma_{\hat{v}_i \hat{v}_i} \Sigma_{ii}^{-1})} \quad (6)$$

within each iteration step with the residuals \hat{v}_i for both groups. The denominators in Equation (6) are the so-called redundancy numbers which sum up to the total redundancy of the problem, i.e. the model's degrees of freedom, see Förstner (1987) for details. With Equation (6) we get improved covariance matrices $\Sigma_{ii} \doteq \hat{\sigma}_{0i}^2 \Sigma_{ii}$ for the observation groups.

2.1.3. Parameterisation, approximate values and gauge freedom

The ratio of baseline length and depth of scene points is usually critical for mobile applications such as UAV scenarios with rather short tracks of image points. Therefore, we apply a bundle adjustment which allows for 3D points far away as proposed in Schneider *et al.* (2012).

Approximate values for the landmark positions can be obtained by spatial forward sectioning and for the 3D laser points by polar point determination. For the calculation of approximate poses, a simple motion model and the corresponding extrapolation are sufficient in most cases.

Using image observations only, the obtained photogrammetric model is determined up to a spatial similarity transformation with seven parameters. This gauge freedom can be fixed by applying the centroid constraints for the approximate parameters, i.e. scale, position and orientation of the approximate values are preserved (McGlone *et al.* 2004). But with the available range measurements, the scale of the reconstruction becomes fixed and we have only position and orientation at our disposal.

2.2. Feature extraction and tracking

The extraction of salient features in each image is the starting point for the determination of unknown 3D points, sometimes called landmarks, too. We utilise the Förstner operator (Förstner and Gülch 1987) to accomplish this task; however, derivatives and alternatives – especially scale invariant feature operators (Lowe 2004) – are conceivable, too. Since we are dealing with a video stream, the application of the pyramidal Lucas-Kanade tracker (Lucas and Kanade 1981) appears to be adequate to retrieve these points in the subsequent images.

The 3D laser points can be projected into each corresponding image with subpixel precision due to the known camera-to-laser scanner transformation (Equation (2)). Beside the extracted salient image points, this yields an additional set of image points, which have to be tracked. We cannot expect these points to appear in textured image regions, yielding points, which can easily be tracked. Therefore, an assessment of the ‘trackability’ of these image points is mandatory. This can be accomplished by considering the precision of the point coordinates. We truncate a track if an image point with a positional uncertainty above a threshold appears.

Of course, depending on the computational resources, the tracking can be made more robust and reliable by enforcing the epipolar constraints, potentially with a subsequent guided matching. This can be achieved by applying the random sample consensus (RANSAC) algorithm to account for outliers (Fischler and Bolles 1981). In general, the RANSAC algorithm iteratively estimates the parameters of a mathematical model from a given data set including outliers which do not fit to this model. Therefore, instead of using all data, a minimal subset of the data is randomly chosen to estimate the model parameters and the remaining data is checked for consistency. Data supporting the given model results in inliers, whereas data not supporting the model results in outliers which are discarded. Finally, the choice yielding the maximum number of inliers is selected (Hartley and Zisserman 2008). Furthermore, specific reliability measures have been proposed for handling multiple outliers (Knight *et al.* 2010).

2.3. Simultaneous localisation and mapping

The solution provided by the bundle adjustment yields the poses (exterior orientations) and the new 3D points (landmarks) in a batch process with a fixed set of observations and unknown parameters. But with an exploration task in mind, we have to cope with a continuous – virtual endless – data stream of images and laser ranges. Thus the task is to enable online processing for a dynamic system of observations and unknown parameters, a task which is commonly known as SLAM in robotics (Durrant-Whyte and Bailey 2006).

With the assumption of a static scene, we adopt a sliding window bundle adjustment (e.g. Beder and Steffen 2008) for the solution of the SLAM problem. An alternative is the incremental bundle adjustment for sparse nonlinear incremental optimisation proposed by Kaess *et al.* (2012). These approaches offer the possibility of re-linearisation within windows consisting of several consecutive frames with images and laser range measurements.

The ongoing incorporation of the range measurements into adjustment introduces scale information permanently and avoids a drift in the scale of the reconstruction. Furthermore, the result is expected to be more accurate due to the improved geometry and determinability of parameters. The measured laser ranges are considered simply by additional observational equations.

3. Sensor platform and calibration

In this section, the sensor platform (Section 3.1) with the utilised sensors and their calibration (Section 3.2) is described. For further processing, an adequate calibration is essential in order to fuse camera and laser scanner data.

3.1. Sensor platform

For the experiments, we select the involved sensors with respect to the application of UAV-borne 3D mapping where lightweight components are required. We propose to equip the UAV with a small sensor platform (Figure 1) carrying two sensors: a digital camera of type GoPro 1080p HD HERO2 and a laser scanner device of type Hokuyo UTM-30LX-EW.

3.1.1. Digital camera – GoPro 1080p HD HERO2

For obtaining optical data, a low-cost digital camera of type GoPro 1080p HD HERO2 is used. The device has a size of 42 mm × 60 mm × 30 mm, and a weight of 94 g including battery. Due to its low weight, it is often mounted on UAVs in order to capture a sequence of images during the flight. The GoPro 1080p HD HERO2 contains a 1/2.3" HD CMOS chip with rolling shutter exposure. Various photo capturing modes can be selected. For the experiments, we chose a moderate field-of-view with 127° and captured images with a size of 3200 × 2400 pixels.

3.1.2. Laser scanner device – Hokuyo UTM-30LX-EW

For obtaining accurate 3D measurements, a Hokuyo UTM-30LX-EW laser range finder is used. With a size of 62 mm × 62 mm × 87.5 mm and a weight of 210 g (without cable), the Hokuyo UTM-30LX-EW is well suited for UAV-borne scene monitoring and 3D mapping. According to the specifications, this device emits laser pulses with a wavelength of $\lambda = 905$ nm, and the laser safety is class 1. It provides 2D scans covering a scan angle of 270° with an angular resolution of 0.25°. The range measurement resolution is 1 mm,



Figure 1. Sensor platform with two sensors: laser scanner (left: Hokuyo UTM-30LX-EW) and digital camera (right: GoPro 1080p HD HERO2).

and the accuracy is specified with ± 30 mm within a range of 0.1–10 m and ± 50 mm within a range of 10–30 m (with an accuracy of $\sigma < 30$ mm for indoor environments with less than 1000 lx). The specified pulse repetition rate is 43 kHz. For each reflected laser pulse, the range and the corresponding intensity value are measured. Furthermore, the Hokuyo UTM-30LX-EW provides the capability to measure up to three echoes of a single-emitted laser pulse, where the number of echoes depends on the surface properties of the respective objects, i.e. shape and reflectivity. In particular, transparent materials, edges of objects or vegetation typically cause more than one echo. In these cases, the second echo can be assumed to result from a structure in the original pulse direction which is located behind a partially occluding object, i.e. the second echo can be treated as an additional range measurement (Droeschel *et al.* 2013).

3.2. Calibration

The calibration includes the estimation of the intrinsic behaviour, i.e. the interior orientation, of the camera as well as the estimation of the position and orientation offset between camera and laser scanner. In photogrammetry, this position and orientation offset is known as exterior orientation, whereas the synonyms lever-arm (position) and bore-sight (orientation) are commonly used for active vision systems such as laser scanners. Hence, we separate the calibration process into geometric camera calibration (Section 3.2.1) for estimating the intrinsic camera parameters, and lever-arm and bore-sight calibration (Section 3.2.2) for estimating the relative orientation between the two sensors.

3.2.1. Geometric camera calibration

For automatically estimating the intrinsic behaviour of the camera with respect to the geometric mapping of a scene onto the image plane, we assume a standard camera model considering both radial and tangential distortion (Brown 1971, Heikkilä and Silvén 1997). Accordingly, this geometric mapping is parameterised with the focal lengths in x - and y -direction, the image coordinates of the principal point, a skew coefficient, and the image distortion coefficients describing radial and tangential distortion.

3.2.2. Lever-arm and bore-sight calibration

As mentioned above, the laser of the laser scanner device operates in the near-infrared domain with a wavelength of $\lambda = 905$ nm, which is above the visibility of the utilised camera. Therefore, radiometric aspects might not be appropriate for calibration and consequently only geometric aspects are considered for lever-arm and bore-sight calibration.

The lever-arm and bore-sight calibration is based on augmenting an algorithm for automatic extraction of straight line segments in the line scans (Kassir and Peynot 2010) to a method for extrinsic calibration of a camera and laser scanner (Zhang and Pless 2004). Accordingly, the first step consists of automatically extracting straight lines from the line scans. For this purpose, a robust method is desirable which is able to reliably extract the longest lines. Hence, we follow the strategy proposed by Kassir and Peynot (2010) which is based on recursively testing every combination of two points in the scan with respect to a straightness criterion, and then recursively taking out the longest line. The selected straightness criterion involves all points between the two selected points. Since some of the extracted lines arise from walls, tables or other objects present in the structured environment, the next step focuses on the classification of the extracted lines

into (1) lines resulting from the calibration plane (i.e. from the checkerboard) and (2) lines resulting from background objects. Kassir and Peynot (2010) proposed to exploit an iterative scheme involving different metrics for successfully selecting the respective line from each scan which results from the calibration board. These metrics also deliver an estimate for the size of the checkerboard. From the board lines, an estimation of the exterior orientation is derived with the algorithm proposed by Zhang and Pless (2004) and iteratively refined until convergence.

Since the applied algorithm for estimating the relative orientation between the sensors is crucial for obtaining an adequate data fusion, we briefly summarise the basic principle according to Zhang and Pless (2004). Assuming a rigid transformation between points $\mathbf{X}_i^{(c)}$ in the camera coordinate frame and points $\mathbf{X}_i^{(l)}$ in the laser coordinate frame, the respective transformation can formally be described with

$$\mathbf{X}_i^{(l)} = \Phi \mathbf{X}_i^{(c)} + \mathbf{A} \quad (7)$$

where Φ represents a rotation matrix and \mathbf{A} is a translation vector. Since the measured points $\mathbf{X}_i^{(l)}$ are located on the calibration plane, we first transform them to camera coordinates with

$$\mathbf{X}_i^{(c)} = \Phi^{-1} (\mathbf{X}_i^{(l)} - \mathbf{A}) = \Phi^T (\mathbf{X}_i^{(l)} - \mathbf{A}) \quad (8)$$

and then substitute $\mathbf{X}_i^{(c)}$ in the equation

$$\mathbf{n}^T \mathbf{X}_i^{(c)} = N \quad (9)$$

describing the calibration plane in camera coordinates, where \mathbf{n} denotes the normal vector of the calibration plane and N describes the distance of the calibration plane from the origin of the respective camera coordinate frame:

$$\mathbf{n}^T \Phi^{-1} (\mathbf{X}_i^{(l)} - \mathbf{A}) = N \quad (10)$$

The normal vector \mathbf{n} and the distance N are already known as they result from geometric camera calibration. Further assuming that the points $\mathbf{X}_i^{(l)}$ are located on the plane $Y^{(l)} = 0$ in the laser coordinate system and representing laser points with $\hat{\mathbf{X}}_i^{(l)} = [X_i^{(l)}, Z_i^{(l)}, 1]^T$, we can transform this equation according to

$$\mathbf{n}^T \Phi^{-1} \left(\begin{bmatrix} X_i^{(l)} \\ 0 \\ Z_i^{(l)} \end{bmatrix} - \begin{bmatrix} \Delta_X \\ \Delta_Y \\ \Delta_Z \end{bmatrix} \right) = N \quad (11)$$

$$\underbrace{\mathbf{n}^T \Phi^{-1} \begin{bmatrix} 1 & 0 & -\Delta_X \\ 0 & 0 & -\Delta_Y \\ 0 & 1 & -\Delta_Z \end{bmatrix}}_H \begin{bmatrix} X_i^{(l)} \\ Z_i^{(l)} \\ 1 \end{bmatrix} = N \quad (12)$$

$$\mathbf{n}^T \mathbf{H} \widehat{\mathbf{X}}^{(l)} = N \quad (13)$$

where \mathbf{H} is a 3×3 matrix describing the transformation from the laser coordinate system to the camera coordinate system. Concatenating the respective equations for all the scans, a linear equation system can be derived which is solved for \mathbf{H} with a standard least squares estimation. From the resulting matrix \mathbf{H} , the rotation matrix Φ and the translation vector Δ are derived according to

$$\Phi = [\mathbf{h}_1, -\mathbf{h}_1 \times \mathbf{h}_2, \mathbf{h}_2]^T \quad (14)$$

$$\Delta = -[\mathbf{h}_1, -\mathbf{h}_1 \times \mathbf{h}_2, \mathbf{h}_2]^T \mathbf{h}_3 \quad (15)$$

where \mathbf{h}_i represents the i th column of \mathbf{H} . Since Φ may not meet the properties of a rotation matrix, an approximated rotation matrix $\hat{\Phi}$ can be derived by minimising the Frobenius norm of $\hat{\Phi} - \Phi$ subject to $\hat{\Phi} \hat{\Phi}^T = \mathbf{I}_{3 \times 3}$. Although other methods for extrinsic calibration have recently been proposed which provide more robustness in case of a lower number of scans, it has been demonstrated that the stability of the used method tends to become similar for $n > 8$ involved scans (Vasconcelos *et al.* 2012).

4. Experiments

By performing an indoor experiment (Section 4.1) and an outdoor experiment (Section 4.2), the proposed concept is proved.

4.1. Calibration

To calculate the image coordinates of the projected 3D laser points, a calibration has to be carried out. Therefore, the intrinsic parameters of the camera have to be estimated by geometric camera calibration (Section 4.1.1), and the position and orientation offset between the camera and the laser scanner is estimated by lever-arm and bore-sight calibration (Section 4.1.2). For this purpose, a synchronised pairwise capturing of a calibration plane (checkerboard) with the camera and the laser scanner has to be guaranteed (Figure 2). This can be achieved with a software or hardware trigger (Weinmann and Jutzi 2012).

4.1.1. Geometric camera calibration

For estimating the intrinsic camera parameters, we apply the camera calibration toolbox for Matlab (Bouguet 2010). The estimated parameters are used to derive undistorted images.

4.1.2. Lever-arm and bore-sight calibration

For calibrating the platform, 15 synchronised pairwise recordings of the calibration plane have been carried out with the camera and the laser scanner. It has to be mentioned that the sensors on the platform have been arranged with a minimum lateral shift and maximum alignment (Figure 1) to each other. Unfortunately, neither the laser scanner



Figure 2. Example of a distorted image with a calibration plane (checkerboard).

Table 1. Estimates of the lever-arm and bore-sight calibration.

	Estimated values	RMS error
Lever-arm (m)	-0.000625	± 0.009
	0.067100	± 0.047
	-0.006450	± 0.012
Bore-sight (°)	1.08	± 0.838
	1.69	± 0.373
	-89.3	± 0.140

origin nor the camera's image sensor location relative to the housing is provided by the manufacturer. Therefore, only a coarse evaluation of the estimated results can be provided. As expected, the results (Table 1) reflect the same trend.

4.2. Data fusion

For a proof of the concept and for convenience, we performed an experiment with our calibrated sensor platform. We used the sensor platform depicted in Figure 1 and captured a sequence of four images and corresponding laser scans in front of the Karlsruhe Palace located in the city of Karlsruhe, Germany. After each acquisition (pose), the system was moved 2 m along a straight line (Figure 3).

Figure 4 shows the first image of the sequence of images with the projected laser points. The red dots denote the extracted salient image points, the projected laser points are depicted in blue. Additionally, a scaled representation of range measurements is illustrated in green. Obviously, the range measurements fit well with the geometry of the building façade. The salient points and the laser points have been tracked through the sequence. Figures 5 and 6 show exemplary results.



Figure 3. Data acquisition.

The observations are summarised in [Table 2](#), with the number of observations, the redundancy parts summing up to the total redundancy of 2275, the assumed standard deviations, and the estimated standard deviations for observed image coordinates and laser ranges.

[Table 3](#) summarises the pairwise distances between the four estimated projection centres. The distances are slightly shorter than expected (preset 2 m shifts) pointing to the possible need of a scale factor larger than one for the range measurements. Apart from this, the distances indicate that the system had been moved along a straight line which is in accordance with the ground truth provided by the design of the experiment.

Histograms of the residuals for the image observations and for the laser ranges are provided in [Figure 7](#). The histogram of the laser ranges reveals at least two outliers, possibly stemming from reflections at window glass.



Figure 4. First image of the sequence captured in front of the Karlsruhe Palace (red: extracted salient image points; blue: laser points projected onto the image; green: scaled representation of range measurements).



Figure 5. The result for the tracking from the second to the third image using the features extracted at the first pose (red: tracked salient image points; blue: tracked laser points).

5. Discussion

The experiments revealed that additional aspects may be considered to further improve the proposed concept. In the following, we briefly discuss the main findings in this study.

Usually, a proper calibration of the sensor platform has to be done only once before using the platform for monitoring purposes. Due to external influences (e.g. mechanical

Table 2. Observations and their quality.

	Observations	
	Image coordinates	Laser ranges
Number of observations	3544	78
Redundancy parts	2199.1	75.9
A priori assumed standard deviations	0.50 pixel	0.10 m
Estimated standard deviations	0.44 pixel	0.12 m

Table 3. Estimated pairwise distances of the camera projection centres.

Pose	2	3	4
1	1.9561 m	3.9067 m	5.8622 m
2	0 m	1.9506 m	3.9061 m
3	–	0 m	1.9556 m

Note: The three pairwise distances 1-2, 2-3 and 3-4 sum up to 5.8623 m, which coincides with the estimated distance 5.8622 m for 1-4.

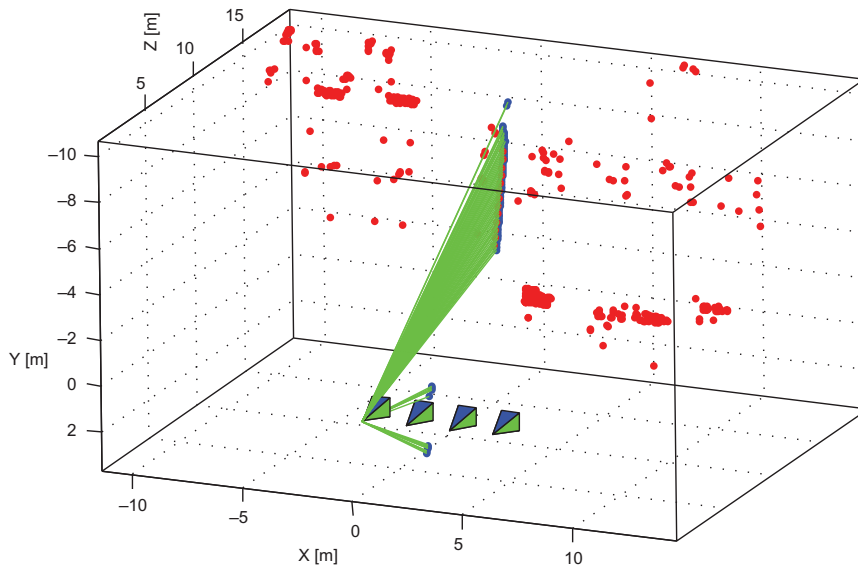


Figure 6. The results of the bundle adjustment for four poses, i.e. the estimated 3D points and the estimated camera poses. Furthermore, the rays of the laser scan are depicted (green lines).

stress or atmospheric variations), the relative orientation between the two sensors might significantly change during the recordings. Therefore, we recommend to apply multiple calibrations before and after the recordings, in particular for lever-arm and bore-sight estimation.

Furthermore, the results in Table 2 reveal that the assumed standard deviations of the image coordinates (a priori assumed: 0.5 pixels) had been slightly too pessimistic

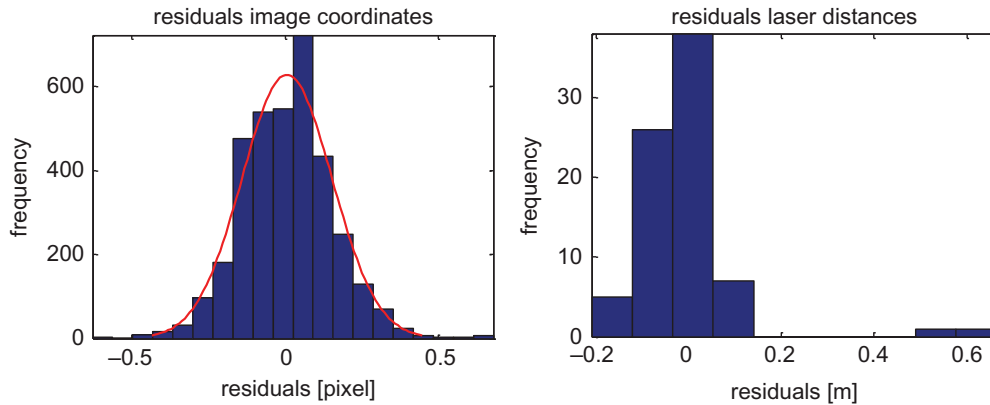


Figure 7. Histograms of the residuals for the image observations (left, with superimposed estimated normal distribution) and for the laser ranges (right).

(estimated: 0.44 pixels) while the assumed standard deviations of the laser range measurements (a priori assumed: 0.1 m) had been too optimistic (estimated: 0.12 m). It is noteworthy that these results only hold for valid models and error free observations. Actually, outliers and model violations affect the estimation of the variance components considerably so that a robust estimation is advisable.

It also has to be considered that the manufacturer of the lightweight laser scanner only provides information with respect to the range measurement, i.e. values for range accuracy, measurement resolution and repeated accuracy, but unfortunately not for the scan angles. Although the accuracies for the range measurements are specified in the sensor specifications, inconsistencies can be observed. In our experiments, obviously a systematic distance shift occurred (Table 3), which might be indirectly caused by range shifts. Therefore, further investigations on the range measurement accuracy of the utilised laser scanner might be worth to consider. Beside this, the scan angle accuracies should generally be deeply investigated, e.g. by also applying variance components.

6. Conclusions and outlook

Due to the availability of small and lightweight laser scanners for mobile applications, new innovative concepts are needed to fuse image and range measurements properly. Necessary prerequisites are an accurate camera calibration as well as lever-arm and bore-sight determination, respectively, for the sensor platform. In this contribution, we proposed a straightforward approach to integrate different kinds of measurements within a bundle adjustment. This framework is known to provide optimal results in terms of statistics since in principle different variances and correlations for the observations can be considered. To account for the different groups of measurements, we introduced a variance components estimation in order to obtain a reasonable weight ratio between both groups of observations.

The conducted experiments prove the feasibility and applicability of the proposed approach; the results are in accordance with the reference data provided by the design of the experiment with controlled conditions. The consideration of range measurements introduces and preserves scale information, eases initialisation procedures, and clearly

improves the adjustment result. Furthermore, the additional 3D laser points densify the scene information, and the points' distribution on the surfaces becomes more uniform.

Conceptually, the approach can be extended to cope with a continuous data stream. This leads to an incremental bundle adjustment within a sliding window of data frames. Furthermore, the approach can be made more robust in order to treat outliers properly.

With the proposed method, the main contribution for an improved UAV-borne mapping is given. Based on the accurate data fusion, this approach can be extended by considering dense matching techniques, e.g. given by the semi-global matching approach (Hirschmüller 2008) to gain a precise 3D model of the environment or by the use of different types of cameras such as the combination of RGB, multispectral and thermal infrared cameras (Lucieer *et al.* 2012) for mapping different environmental characteristics.

Acknowledgement

The authors would like to thank Matthias Kapica for assistance during the measurements.

References

- Bachrach, A., He, R., and Roy, N., 2009. Autonomous flight in unknown indoor environments. *International Journal of Micro Air Vehicles*, 1 (4), 217–228. doi:10.1260/175682909790291492.
- Beder, C. and Steffen, R., 2008. Incremental estimation without specifying a-priori covariance matrices for the novel parameters. *Proceedings of the IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW)*, 23–28 June, Anchorage, AK, 1–6.
- Bouguet, J.-Y., 2010. *Camera calibration toolbox for Matlab* [online]. Computer Vision Research Group, Department of Electrical Engineering, California Institute of Technology, Pasadena, USA. Available from: http://www.vision.caltech.edu/bouguetj/calib_doc/ [Accessed 31 October 2013].
- Brown, D.C., 1971. Close-range camera calibration. *Photogrammetric Engineering*, 37 (8), 855–866.
- Choi, K. and Lee, I., 2011. A UAV-based close-range rapid aerial monitoring system for emergency responses. In: H. Eisenbeiss, M. Kunz, and H. Ingensand, eds. *International conference on unmanned aerial vehicle in geomatics (UAV-g). The international archives of the photogrammetry, remote sensing and spatial information sciences*, Zürich, XXXVIII-1/C22, 247–252.
- Conte, G., *et al.*, 2013. Performance evaluation of a light-weight multi-echo lidar for unmanned rotorcraft applications. In: G. Grenzdörffer and R. Bill, eds. *UAV-g2013. The international archives of the photogrammetry, remote sensing and spatial information sciences*, Rostock, XL-1/W2, 87–92.
- Droeschel, D., Schreiber, M., and Behnke, S., 2013. Omnidirectional perception for lightweight UAVs using a continuously rotating 3D laser scanner. In: G. Grenzdörffer and R. Bill, eds. *UAV-g2013. The international archives of the photogrammetry, remote sensing and spatial information sciences*, Rostock, XL-1/W2, 107–112.
- Durrant-Whyte, H. and Bailey, T., 2006. Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine*, 13 (2), 99–110. doi:10.1109/MRA.2006.1638022.
- Fischler, M.A. and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24 (6), 381–395. doi:10.1145/358669.358692.
- Förstner, W., 1987. Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision. *Computer Vision, Graphics, and Image Processing*, 40 (3), 273–310. doi:10.1016/S0734-189X(87)80144-5.
- Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. *Proceedings of the ISPRS intercommission workshop*, June, Interlaken 281–305.

- Grzonka, S., Grisetti, G., and Burgard, W., 2009. Towards a navigation system for autonomous indoor flying. *Proceedings of the IEEE international conference on robotics and automation (ICRA'09)*, May, Kobe, 2878–2883.
- Hartley, R.I. and Zisserman, A., 2008. *Multiple view geometry in computer vision*. Cambridge University Press.
- Heikkilä, J. and Silvén, O., 1997. A four-step camera calibration procedure with implicit image correction. *Proceedings of the IEEE Computer society conference on computer vision and pattern recognition*, 17–19 June, San Juan, 1106–1112.
- Helmert, F.R., 1924. *Die ausgleichsrechnung nach der methode der kleinsten quadrate*. 3rd ed. Leipzig: Teubner.
- Hirschmüller, H., 2008. Stereo processing by semi-global matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (2), 328–341. doi:10.1109/TPAMI.2007.1166.
- Holz, D., et al., 2013. Towards multimodal omnidirectional obstacle detection for autonomous unmanned aerial vehicles. *UAV-g2013. The international archives of the photogrammetry, remote sensing and spatial information sciences*, XL-1/W2, 201–206.
- Jutzi, B., Weinmann, M., and Meidow, J., 2013. Improved UAV-borne 3D mapping by fusing optical and laserscanner data. In: G. Grenzdörffer and R. Bill, eds. *UAV-g2013. The international archives of the photogrammetry, remote sensing and spatial information sciences*, Rostock, XL-1/W2, 223–228.
- Kaess, M., et al., 2012. Isam2: incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research*, 31, 216–235. doi:10.1177/0278364911430419.
- Kassir, A. and Peynot, T., 2010. Reliable automatic camera-laser calibration. *Proceedings of the Australasian conference on robotics and automation (ACRA 2010)*, 1–3 December, Brisbane.
- Knight, N.L., Wang, J., and Rizos, C., 2010. Generalised measures of reliability for multiple outliers. *Journal of Geodesy*, 84 (10), 625–635. doi:10.1007/s00190-010-0392-4.
- Koch, K.R., 1986. Maximum likelihood estimate of variance components. *Bulletin Géodésique*, 60 (4), 329–338. doi:10.1007/BF02522340.
- Kuhnert, K.-D. and Kuhnert, L., 2013. Light-weight sensor package for precision 3D measurement with micro UAVs e.g. power-line monitoring. In: G. Grenzdörffer and R. Bill, eds. *UAV-g2013. The international archives of the photogrammetry, remote sensing and spatial information sciences*, Rostock, XL-1/W2, 235–240.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*, 60 (2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
- Lucas, B.T. and Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, 2, 674–679.
- Lucieer, A., et al., 2012. Using a micro-UAV for ultra-high resolution multisensor observations of antarctic moss beds. In: M. Shortis and N. El-Sheimy, eds. *XXII ISPRS congress – technical commission I. The international archives of the photogrammetry, remote sensing and spatial information sciences*, Melbourne, XXXIX-B1, 429–433.
- McGlone, J.C., Mikhail, E.M., and Bethel, J., 2004. *Manual of photogrammetry*. 5th ed. Bethesda, MD: American Society for Photogrammetry and Remote Sensing.
- Nagai, M., et al., 2009. UAV-borne 3-D mapping system by multisensor integration. *IEEE Transactions on Geoscience and Remote Sensing*, 47 (3), 701–708. doi:10.1109/TGRS.2008.2010314.
- Rao, C.R., 1973. *Linear statistical inference and its applications*. New York: Wiley.
- Satrapod, C., Wang, J., and Rizos, C., 2002. A simplified MINQUE procedure for the estimation of variance-covariance components of GPS observables. *Survey Review*, 36 (286), 582–590. doi:10.1179/sre.2002.36.286.582.
- Schneider, J., et al., 2012. Bundle adjustment for multi-camera systems with points at infinity. In: M. Shortis, N. Paparoditis, and C. Mallet, eds. *XXII ISPRS congress – technical commission III. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, Melbourne, I-3, 75–80.
- Triggs, B., et al., 2000. Bundle adjustment – a modern synthesis. In: B. Triggs, A. Zisserman, and R. Szeliski, eds. *Vision algorithms: theory and practice*. Lecture notes in computer science. Berlin: Springer, 1883, 298–372.

- Vasconcelos, F., Barreto, J.P., and Nunes, U., 2012. A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34 (11), 2097–2107.
- Wallace, L.O., Lucieer, A., and Watson, C.S., 2012. Assessing the feasibility of UAV-based lidar for high resolution forest change detection. In: M. Shortis, W. Wagner, and J. Hyypä, eds. *XXII ISPRS congress – technical commission VII. The international archives of the photogrammetry, remote sensing and spatial information sciences*, Melbourne, XXXIX-B7, 499–504.
- Weinmann, M. and Jutzi, B., 2012. A step towards dynamic scene analysis with active multi-view range imaging systems. In: M. Shortis, N. Paparoditis and C. Mallet, eds. *XXII ISPRS congress – technical commission III. The international archives of the photogrammetry, remote sensing and spatial information sciences*, 25 August–1 September, Melbourne, XXXIX-B3, 433–438.
- Weinmann, M., et al., 2011. Fast and automatic image-based registration of TLS data. *Advances in LIDAR data processing and applications. ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (6), S62–S70.
- Zhang, Q. and Pless, R., 2004. Extrinsic calibration of a camera and laser range finder (improves camera calibration). *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems*, 28 September–2 October, Sendai, 3, 2301–2306.

AUTOMATIC FEATURE-BASED POINT CLOUD REGISTRATION FOR A MOVING SENSOR PLATFORM

Martin Weinmann, André Dittrich, Stefan Hinz, and Boris Jutzi

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Englerstr. 7, 76131 Karlsruhe, Germany
{martin.weinmann, andre.dittrich, stefan.hinz, boris.jutzi}@kit.edu

Commission WG III/4

KEY WORDS: LIDAR, Point Cloud, Registration, Imagery, Dynamic, Close Range

ABSTRACT:

The automatic and accurate alignment of multiple point clouds is a basic requirement for an adequate digitization, reconstruction and interpretation of large 3D environments. Due to the recent technological advancements, modern devices are available which allow for simultaneously capturing intensity and range images with high update rates. Hence, such devices can even be used for dynamic scene analysis and for rapid mapping which is particularly required for environmental applications and disaster management, but unfortunately, they also reveal severe restrictions. Facing challenges with respect to noisy range measurements, a limited non-ambiguous range, a limited field of view and the occurrence of scene dynamics, the adequate alignment of captured point clouds has to satisfy additional constraints compared to the classical registration of terrestrial laser scanning (TLS) point clouds for describing static scenes. In this paper, we propose a new methodology for point cloud registration which considers such constraints while maintaining the fundamental properties of high accuracy and low computational effort without relying on a good initial alignment or human interaction. Exploiting 2D image features and 2D/2D correspondences, sparse point clouds of physically almost identical 3D points are derived. Subsequently, these point clouds are aligned with a fast procedure directly taking into account the reliability of the detected correspondences with respect to geometric and radiometric information. The proposed methodology is evaluated and its performance is demonstrated for data captured with a moving sensor platform which has been designed for monitoring from low altitudes. Due to the provided reliability and a fast processing scheme, the proposed methodology offers a high potential for dynamic scene capture and analysis.

1 INTRODUCTION

An adequate description of a 3D scene is typically derived in the form of point clouds consisting of a large number of measured 3D points and, optionally, different attributes for each point such as intensity or color. The sampling of observed object surfaces should be as dense and complete as possible. Due to occlusions resulting from objects in the scene or areas with low point density, typically multiple point clouds have to be captured from different locations in order to obtain complete objects and full scene coverage. However, as the spatial coordinates of each point cloud are only determined with respect to a local coordinate frame of the sensor, all captured point cloud data has to be transferred into a common coordinate frame which is commonly referred to as point cloud registration or 3D scan matching.

The approaches for point cloud registration can be categorized by considering the data they exploit. Standard approaches such as the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) or Least Squares 3D Surface Matching (LS3D) (Gruen and Akca, 2005) only exploit spatial 3D information and minimize either the difference between point clouds or the distance between matched surfaces. The use of point distributions or geometric primitives such as planes has also been proposed in literature (e.g. (Magnusson et al., 2007; Brenner et al., 2008)) and belongs to this category. Considering that the captured scans typically represent data measured on a regular scan grid, the spatial 3D information can also be represented as range image. Exploiting visual features in this range image significantly alleviates the registration process. As most of the modern active 3D sensors provide intensity or color information in addition to the spatial 3D information, respective intensity or color images may also be available. The intensity images are typically derived from reflectance information representing the respective energy of the backscattered

laser light, whereas color information is usually obtained from co-registered camera images. Such intensity or color images provide a higher level of distinctiveness and allow for detecting reliable correspondences between visual features.

Nowadays, many approaches for point cloud registration exploit visual features derived from intensity or color images in order to obtain sparse point clouds. Detected feature correspondences between the respective images indicate corresponding 3D points. Hence, the registration of such sparse point clouds may for instance be based on a standard rigid transformation (Eggert et al., 1997) which is typically combined with the RANSAC algorithm for increased robustness in case of existing outlier correspondences (Seo et al., 2005; Boehm and Becker, 2007; Barnea and Filin, 2007). As a powerful alternative, the transfer to solving the Perspective- n -Point (PnP) problem has been proposed (Weinmann et al., 2011; Weinmann and Jutzi, 2011).

Recently, an experimental setup for surveillance applications in indoor and outdoor environments has been presented which is suited for simulating airborne scene monitoring from low altitudes fairly realistically (Hinz et al., 2011). The first results presented in (Weinmann and Jutzi, 2012) indicate a high potential for dynamic scene capture and analysis, but they also reveal that additional effort is required to obtain satisfying results. For dynamic scene capture, highly accurate 3D measurements as provided by a terrestrial laser scanner cannot be assumed, but the modern scanning devices offer a simultaneous image-based acquisition of intensity and range information with high update rates. Hence, the registration process has to provide high-quality estimates of the transformation parameters, low computational effort and robustness with respect to noisy range measurements. Furthermore, the limited field of view and the limited non-ambiguous range have to be taken into account. An important and meanwhile commonly

used first step is the reduction of captured point cloud data to sparse point clouds by using visual features in 2D imagery. A reliable feature matching yields 2D/2D correspondences and the respective 3D/3D correspondences. In contrast to previous work (Weinmann and Jutzi, 2011; Weinmann and Jutzi, 2012), the proposed methodology involves an improved scheme for outlier rejection and estimation of inlier reliability. This scheme exploits information derived from the reliability of range measurements as well as the reliability of feature correspondences which is based on intensity measurements. With the further consideration of the plausibility of corresponding 3D points, a straightforward approach for aligning point clouds can be applied which would not be suitable without a reliable outlier removal. The contribution of this paper is a new methodology for fast and accurate point cloud registration which

- exploits sparse point clouds with additional point attributes in form of quality measures based on geometric and radiometric information,
- introduces an improved weighting scheme considering the derived quality measures, and
- involves a plausibility check taking into account the detected 3D/3D correspondences.

After presenting the methodology for successive pairwise registration in Section 2, the configuration of the sensor platform is described in Section 3. Subsequently, in Section 4, an evaluation is carried out which demonstrates the performance of the new approach for a realistic test scenario. In Section 5, the derived results are discussed with respect to basic requirements and other approaches. Finally, the content of the entire paper is concluded in Section 6 and suggestions for future work are outlined.

2 METHODOLOGY

The proposed methodology focuses on airborne scene monitoring with a moving sensor platform. After data acquisition (Section 2.1), a preprocessing is carried out in order to get normalized intensity images and the respective 3D point cloud (Section 2.2). As the captured point clouds are corrupted with noise, a quality measure is derived for each 3D point (Section 2.3). Subsequently, distinctive features are extracted from 2D intensity images (Section 2.4). A comparison of these features yields reliable 2D/2D correspondences between different frames as well as a quality measure taking into account the distinctiveness of matched features. Additionally, the projection of the respective 2D points into 3D space yields 3D/3D correspondences. As the influence of each 3D/3D correspondence on the registration process should rely on a respective quality measure, a weighting scheme considering geometric and radiometric information is introduced (Section 2.5). Finally, the point cloud registration is carried out by estimating the rigid transformation between two sparse point clouds with a weighted least squares alignment (Section 2.6).

2.1 Data Acquisition

The proposed concept focuses on the use of range imaging devices which are also referred to as range cameras, i.e. devices which provide 2D image representations of captured range and intensity/color. These devices should additionally provide a high update rate for capturing dynamic scenes or for rapid mapping.

2.2 Preprocessing

The first step consists of adapting the recorded data according to (Weinmann and Jutzi, 2012), where a histogram normalization is carried out which maps the captured intensity information to the

interval $[0, 255]$. For color images, a conversion to gray-valued images could be applied in order to obtain intensity images. Furthermore, the lens distortion has to be taken into account which involves an initial camera calibration and the respective correction of the captured 3D information.

2.3 Point Quality Assessment

As the range measurements might be corrupted with noise, it is suitable to add a quality measure as attribute for each measured 3D point. Considering the 2D representation of the measured range information, the variation of the range values within small and local image neighborhoods has a strong influence on the reliability of measured 3D points (Weinmann and Jutzi, 2011; Weinmann and Jutzi, 2012). Hence, for each point on the regular 2D grid, the reliability of the respective range information is described with the standard deviation $\sigma \in \mathbb{R}$ of all range values within a 3×3 neighborhood. Low values σ indicate a 3D point on a smooth surface and are assumed to be reliable, whereas high values indicate noisy and unreliable range measurements. Resulting from this, a confidence map M_C is available. In addition to this, the quality measure could further exploit the active intensity measurements representing the energy of the backscattered laser light if these are available, e.g. for range imaging devices such as PMD[vision] CamCube 2.0 or MESA Imaging SR4000.

2.4 2D Feature Extraction and Projection to 3D

Once the measured information has been assigned additional attributes, the registration process can rely on both range and intensity information, and a confidence map providing the respective quality measure. For detecting corresponding information, the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is applied on the intensity images. This yields distinctive keypoints at 2D image locations $\mathbf{x}_i \in \mathbb{R}^2$ as well as the respective local descriptors which are invariant to image scaling and image rotation, and robust with respect to image noise, changes in illumination and small changes in viewpoint. These properties of the descriptors allow a reliable feature matching relying on the ratio $\rho \in \mathbb{R}$ with

$$\rho = \frac{d(N_1)}{d(N_2)} \quad (1)$$

where $d(N_i)$ with $i = 1, 2$ denotes the Euclidean distance of a descriptor belonging to a keypoint in one image to the i -th nearest neighbor in the other image. A low value of ρ indicates a high similarity to only one of the derived descriptors belonging to the other image. Thus, the ratio $\rho \in [0, 1]$ describes the distinctiveness of the occurring features. Meaningful feature correspondences arise from a greater difference between $d(N_1)$ and $d(N_2)$ and hence, the ratio ρ has to satisfy the constraint $\rho \leq t_{des}$, where t_{des} is a certain threshold typically chosen within the interval $[0.6, 0.8]$. As the SIFT features are localized with subpixel accuracy, the assigned information has to be interpolated from the information available for the regular and discrete 2D grid, e.g. by applying a bilinear interpolation. Subsequently, the 2D/2D correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ between these visual features are used to reduce the captured point cloud data to sparse point clouds of physically almost identical 3D points $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ with $\mathbf{X}_i, \mathbf{X}'_i \in \mathbb{R}^3$. Including the assigned attributes, each correspondence can be described with two samples of corresponding information according to

$$s_i = (\mathbf{x}_i, \mathbf{X}_i, \sigma_i, \rho_i) \leftrightarrow s'_i = (\mathbf{x}'_i, \mathbf{X}'_i, \sigma'_i, \rho'_i) \quad (2)$$

where σ_i and σ'_i indicate the quality of the derived 3D points with respect to measured range information, and $\rho_i^* = \rho_i = \rho'_i$ are the assigned quality measures with respect to the distinctiveness of the used intensity information.

2.5 Weight Calculation

For weighting the influence of each 3D/3D correspondences on the estimated transformation, a weight parameter has to be derived for each 3D/3D correspondence. Given the calculated values $\sigma_i, \sigma'_i \in [0, \infty)$ and $\rho_i^* \in [0, t_{des}]$, which are considered as quality measures for the respective 3D points \mathbf{X}_i and \mathbf{X}'_i , the influence of the i -th 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ on the registration process can be weighted by applying a histogram-based approach (Weinmann and Jutzi, 2012). First, the interval $[0m, 1m]$ is divided into $n_b = 100$ bins of equal size, and the values σ_i and σ'_i are mapped to the respective bin. Those values above the upper boundary of 1m are mapped to the last bin. The occurrence of mappings to the different bins is stored in histograms $\mathbf{h} = [h_j]_{j=1, \dots, 100}$ and $\mathbf{h}' = [h'_j]_{j=1, \dots, 100}$. From these, the cumulative histograms \mathbf{h}_c and \mathbf{h}'_c are subsequently derived, where the entries reach from 0 to the number n of detected 3D/3D correspondences. For assigning those 3D points with a lower σ a higher weight, as these are more reliable according to the definition, the inverse cumulative histograms (ICHS)

$$\mathbf{h}_{c,inv} = \left[n - \sum_{j=1}^i h_j \right]_{i=1, \dots, 100} \quad (3)$$

and

$$\mathbf{h}'_{c,inv} = \left[n - \sum_{j=1}^i h'_j \right]_{i=1, \dots, 100} \quad (4)$$

are established. The quality measure ρ_i^* encapsulating the reliability of the matching process is also used for such a weighting scheme based on $n_b = 100$ bins of equal size covering the interval $[0, 1]$. Resulting from this, a histogram $\mathbf{h}^* = [h_j^*]_{j=1, \dots, 100}$ and a cumulative histogram \mathbf{h}_c^* are available as well as an inverse cumulative histogram $\mathbf{h}_{c,inv}^*$ with

$$\mathbf{h}_{c,inv}^* = \left[n - \sum_{j=1}^i h_j^* \right]_{i=1, \dots, 100} \quad (5)$$

from which additional weight parameters are derived. Thus, the new weighting scheme yields three weight parameters $\mathbf{h}_{c,inv}(\sigma_i)$, $\mathbf{h}'_{c,inv}(\sigma'_i)$ and $\mathbf{h}_{c,inv}^*(\rho_i^*)$ for each 3D/3D correspondence. From these, the respective weight w_i is finally determined with

$$w_i = \min \{ \mathbf{h}_{c,inv}(\sigma_i), \mathbf{h}'_{c,inv}(\sigma'_i), \mathbf{h}_{c,inv}^*(\rho_i^*) \} \quad (6)$$

as the minimum of these values.

2.6 Point Cloud Registration

Introducing a rotation matrix $\mathbf{R} = [r_{pq}] \in \mathbb{R}^{3 \times 3}$ and a translation vector $\mathbf{t} = [t_p] \in \mathbb{R}^3$, the spatial relation between two points $\mathbf{X}_i, \mathbf{X}'_i \in \mathbb{R}^3$ representing a 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ can formally be described as

$$\mathbf{X}'_i = \mathbf{R}\mathbf{X}_i + \mathbf{t} \quad (7)$$

or more detailed as

$$X'_i = r_{11}X_i + r_{12}Y_i + r_{13}Z_i + t_1 \quad (8)$$

$$Y'_i = r_{21}X_i + r_{22}Y_i + r_{23}Z_i + t_2 \quad (9)$$

$$Z'_i = r_{31}X_i + r_{32}Y_i + r_{33}Z_i + t_3 \quad (10)$$

where a perfect mapping is achieved in the ideal case. However, the 3D/3D correspondences typically do not fit perfectly and therefore, a fully automatic estimation of the transformation parameters can be derived by minimizing the error between the

point clouds in the least squares sense. For this purpose, the three equations resulting from each of the n 3D/3D correspondences are concatenated. Subsequently introducing vector-matrix notation and separating the vector $\mathbf{u} \in \mathbb{R}^{12}$ containing the unknown parameters according to

$$\mathbf{u} = [r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}, r_{33}, t_1, t_2, t_3]^T \quad (11)$$

yields a linear equation system of the form

$$\mathbf{l} = \mathbf{A}\mathbf{u} \quad (12)$$

with $\mathbf{l} \in \mathbb{R}^{3n}$ and $\mathbf{A} \in \mathbb{R}^{3n \times 12}$. For solving this linear equation system, the least squares estimate can be derived as

$$\hat{\mathbf{u}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{l} \quad (13)$$

where the matrix $\mathbf{P} \in \mathbb{R}^{3n \times 3n}$ is used for weighting the importance of the respective observations. As the observations are assumed to be independent, the weight matrix \mathbf{P} is considered as diagonal matrix. Exploiting the weights w_i defined in Section 2.5, which are stored in the vector $\mathbf{w} \in \mathbb{R}^n$, an initialization of the weight matrix \mathbf{P} according to

$$\mathbf{P} = \mathbf{P}_0 = \text{diag} \left\{ \left(\frac{w_i^*}{\|\mathbf{w}^*\|} \right)^2 \left(\frac{d_i}{\|\mathbf{d}\|} \right)^2 \right\}_{i=1, \dots, 3n} \quad (14)$$

is introduced, where $\mathbf{w}^* \in \mathbb{R}^{3n}$ contains the weights for each coordinate. The parameters d_i describe the difference between the respective change of a coordinate value and the mean change of the coordinate over all correspondences. The introduction of the parameters d_i is only possible for very small rotations of the sensor platform, but they ensure a weighting according to the plausibility of the respective 3D/3D correspondences which is derived from considering the major trend of all correspondences. Additionally, those correspondences whose 3D coordinates do not even fit to the confidence interval within at least two standard deviations of the mean coordinates, i.e. a confidence level of 95%, are removed. By considering the estimated improvement $\hat{\mathbf{v}}$ according to $\hat{\mathbf{v}} = \mathbf{A}\hat{\mathbf{u}} - \mathbf{l}$ with $\hat{\mathbf{v}} \in \mathbb{R}^{3n}$, an iterative update

$$\mathbf{P} = \mathbf{P}_k = \text{diag} \left\{ \frac{1}{\left(\frac{\hat{v}_i}{\|\hat{\mathbf{v}}\|} \right)^2} \right\}_{i=1, \dots, 3n} \quad (15)$$

is applied until the estimated transformation converges to changes below a certain threshold or until a maximum number of k_{max} iterations is reached.

3 ACTIVE MULTI-VIEW RANGE IMAGING SYSTEM

For demonstrating the performance of the proposed methodology, a sensor platform is used which allows for monitoring from low altitudes (Hinz et al., 2011). This platform is shown in Figure 1, and it is equipped with

- two range imaging devices (PMD[vision] CamCube 2.0) for data acquisition,
- a notebook with a solid state drive for efficient data storage, and
- a 12V battery with 6.5Ah for independent power supply.

As the relative orientation of the two range imaging devices can easily be changed, the system allows for different multi-view op-



Figure 1: Sensor platform.

tions with respect to parallel, convergent or divergent data acquisition geometries. Due to the large payload of several kilograms, mounting the components for data acquisition and data storage on an unmanned aerial vehicle (UAV) is still impracticable. Hence, for scene capture, the sensor platform is moved along a rope. Considering typical surveillance applications, the combination of the sensor platform with a scaled test scenario allows a fairly realistic simulation of a future operational system.

A closer consideration of the devices used for data acquisition reveals the potential as well as the challenges of the proposed system. A PMD[vision] CamCube 2.0 simultaneously captures geometric and radiometric information in form of images with a single shot, and hence, the captured information can be organized in frames. Each frame consists of a range image I_R , an active intensity image I_a and a passive intensity image I_p . The active intensity depends on the illumination emitted by the sensor, whereas the passive intensity depends on the background illumination arising from the sun or other external light sources. As the captured images have a size of 204×204 pixels which corresponds to a field of view of $40^\circ \times 40^\circ$, measurements with an angular resolution of approximately 0.2° are provided. The frame can be updated with high frame rates of more than 25 releases per second. Hence, the device is well-suited for capturing dynamic scenes. A great advantage of the chosen range imaging device is that it can also be used within outdoor environments, but due to the relatively large influence of noise effects arising from multipath scattering as well as the large amount of ambient radiation in comparison to the amount of emitted radiation, a limited absolute range accuracy of a few centimeters and thus noisy point clouds can be expected. A visualization of captured data is shown in Figure 2.

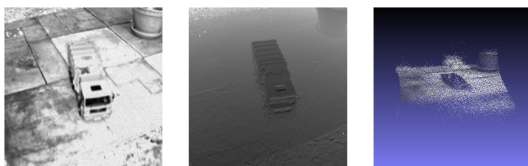


Figure 2: Visualization of captured data: normalized passive intensity image (left), range image (center) and textured 3D point cloud (right).

As the whole system involves multiple range imaging devices for extending the field of view, it has to be considered that these may influence each other and that interferences are likely to occur. This is overcome by choosing different modulation frequencies. Furthermore, a synchronization of both range imaging devices is required in order to obtain corresponding frames with respect to a temporal reference. For this purpose, a software-based trigger is introduced. Optionally, the range measurement restriction can

also be resolved with a hardware-based unwrapping procedure (Jutzi, 2012), which requires the use of different modulation frequencies for each of the two range imaging devices.

4 EXPERIMENTAL RESULTS

The experiments refer to the scene depicted in Figure 3. First, a local coordinate frame is defined in the center between both range imaging devices with fixed orientation with respect to the sensor platform. This coordinate frame is referred to as body frame (superscript b). The X^b -direction is oriented to the forward direction tangential to the rope, the Y^b -direction to the right and the Z^b -direction downwards. Subsequently, a global reference frame (superscript g) is defined which coincides with the initial position and orientation of the sensor platform. As the relative orientation between the devices and the platform is already known from a priori measurements, the projected 3D points \mathbf{X}_i^c which are related to the respective local coordinate frame of a range imaging device (superscript c) can directly be transformed into the body frame (superscript b) of the sensor platform according to

$$\mathbf{X}_i^b = \mathbf{R}_c^b \mathbf{X}_i^c + \mathbf{t}_c^b \quad (16)$$

where \mathbf{R}_c^b and \mathbf{t}_c^b describe the rotation and translation between the respective coordinate frames.



Figure 3: Illustration of the observed scene.

During the whole movement of the sensor platform, a total number of 116 frames is captured. Each frame contains simultaneous measurements of the two range imaging devices. Although there are no measured reference values for each single position in order to check the deviation of the position estimates from the real positions, a validation of the proposed methodology is still possible. As the sensor platform moves along a rope, the projection of the real trajectory onto the $X^g Y^g$ -plane represents a straight line. Furthermore, a loop closure constraint may be considered, where start position and end position of the sensor platform are identical. This is reached by repeating the first frame at the end of the movement. Additional criteria may involve a visual impression of several registered point clouds which yields insights about the scene and the quality of captured data. Hence, the evaluation of the proposed methodology involves

- the deviation σ_Y from the straight line in the $X^g Y^g$ -plane with $X^g = 0\text{m}$,
- the absolute error e_{loop} occurring when assuming an identical start and end position of the sensor platform, and
- a visual inspection of the registered point clouds.

For feature matching, a threshold of $t_{des} = 0.8$ is selected. The estimated trajectory obtained via successive pairwise registration is shown in Figure 4 in nadir view. The standard deviation σ_Y

of the position estimates projected into the $X^g Y^g$ -plane from the straight line with $X^g = 0m$ is $\sigma_Y = 0.0378m$, and the absolute error when considering a closed loop is $e_{loop} = 0.0967m$. A visualization of registered point clouds in a common coordinate frame is illustrated in Figure 5. The figure also shows that the raw point clouds contain many noisy 3D measurements which have to be removed for a subsequent scene analysis.

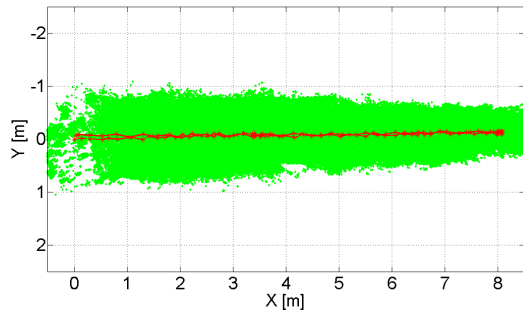


Figure 4: Projection of the estimated trajectory onto the $X^g Y^g$ -plane (red) and visualization of the distribution of measured 3D scene points (green).

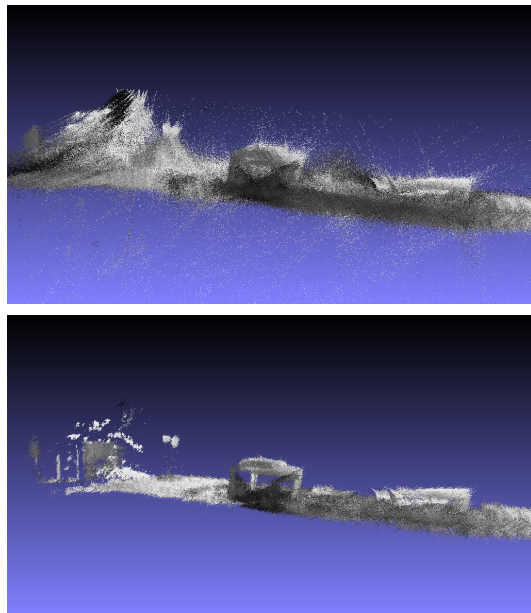


Figure 5: Visualization of registered point clouds: raw point cloud data (top) and those 3D points X_i satisfying $\sigma_i \leq 0.03m$ (bottom).

For data captured at two successive time steps during the movement of the sensor platform, the general appearance of the calculated histograms, cumulative histograms and inverse cumulative histograms is visualized in Figure 6, Figure 7 and Figure 8. The example is based on $n_1 = 480$ detected SIFT correspondences between the normalized active intensity images and $n_2 = 508$ detected SIFT correspondences between the normalized passive intensity images, i.e. a total number of $n = 928$ SIFT corre-

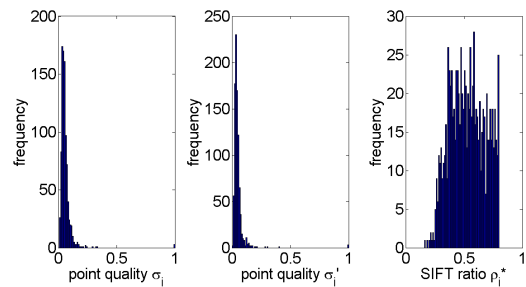


Figure 6: Histograms depicting the occurrence of range reliabilities σ_i and σ_i' (left and center), and the occurrence of SIFT correspondences with a SIFT ratio ρ_i^* (right).

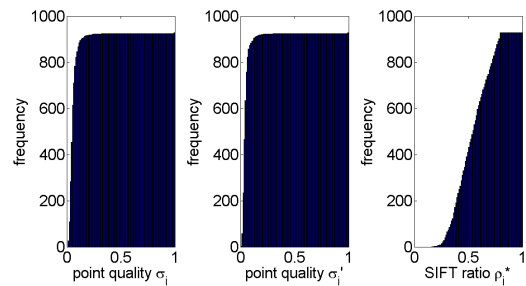


Figure 7: The respective cumulative histograms for the histograms depicted in Figure 6.

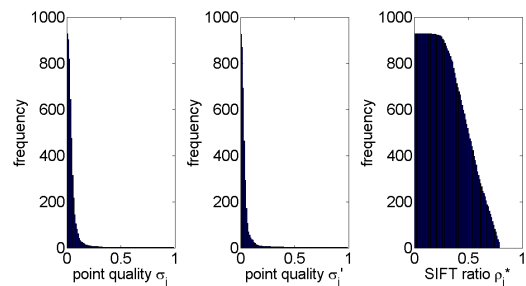


Figure 8: The respective inverse cumulative histograms (ICHS) for the histograms depicted in Figure 6.

spondences resulting in the same number of 3D/3D correspondences after the respective projection into 3D space. Finally, the required time effort for processing each frame and the relations between two frames has to be considered. The methodology has been implemented in Matlab and tested on a standard notebook (2.3GHz, 4GB RAM). For each frame, the average time required for preprocessing, point quality assessment, feature extraction and point projection is 1.6408s (Table 1). For obtaining the relations between two frames, the average time required for feature matching, calculation of weights and point cloud registration is 0.4821s (Table 2). The separate consideration of the average time required for the single tasks reveals that feature extraction and feature matching contribute to more than 90% of the whole time effort. This can significantly be reduced by applying a GPU-based implementation of SIFT or more efficient feature detectors and descriptors.

Task	Time
Preprocessing	0.0766s
Point Quality Assessment	0.0250s
Feature Extraction	1.5295s
Point Projection	0.0097s
Σ	1.6408s

Table 1: Average time effort for the processing of a single frame.

Task	Time
Feature Matching	0.4706s
Weighting Scheme	0.0022s
Registration Process	0.0092s
Σ	0.4821s

Table 2: Average time effort for aligning two point clouds.

5 DISCUSSION

In contrast to standard approaches such as the ICP algorithm, the presented approach does not require a good a priori alignment of the scans. Furthermore, it can be applied for very general scenes without assuming the presence of regular surfaces such as planes and can even cope with noisy measurements arising from monitoring within outdoor environments. The experimental results show that the proposed methodology is suited to recover the transformation parameters fast and accurately by exploiting synergies arising from the combined use of geometric and radiometric information. The proposed methodology directly takes into account the reliability of the 3D information captured with the two range cameras. Thus, it is even possible to stronger rely on the measurements of one range imaging device if the second range imaging device captures spatial information with more noise. Additionally considering how good the corresponding 3D/3D points fit together with respect to the gradient information of the local neighborhood in the 2D intensity representations strengthens the reliability with a further and complementary quality measure assigned to each 3D/3D correspondence. Subsequently, in the registration process, the plausibility of the respective 3D/3D correspondences and the iterative improvement of the geometric alignment of each coordinate are involved. If the estimated improvement \hat{v}_i is relatively large for a coordinate, the coordinate is considered as rather unreliable and the respective influence on the estimated transformation is assigned a lower weight. Due to all these considerations, a high reliability of the estimated flight trajectory can be expected. The only assumption of the registration process is the existence of structured 2D intensity representations in order to derive corresponding points via local image features. This, however, is a common assumption for all image-based approaches exploiting distinctive points or lines for point cloud registration.

6 CONCLUSIONS AND FUTURE WORK

In this paper, a new concept for data acquisition with a moving active multi-view range imaging system and a successive registration of the captured point clouds has been presented. Thus, the system is able to face the challenges arising from noisy range measurements, a limited non-ambiguous range and a limited field of view in dynamic environments. The proposed methodology focuses on the use of sparse point clouds and additional attributes from which a common quality measure considering geometric and radiometric information is derived via inverse cumulative histograms. This quality measure allows for weighting the influence of each 3D/3D correspondence on the estimated transformation

according to its reliability. Further applying a plausibility check for the detected 3D/3D correspondences, the registration process can be carried out with a weighted least squares adjustment. For future work, it would be desirable to detect relevant objects in the scene and describe their behavior by estimating the respective motion trajectories. This can also be achieved by exploiting the combined use of geometric and radiometric information as well as the respective consideration of 3D point cloud data and 2D image representations. Promising results can be expected.

REFERENCES

- Barnea, S. and Filin, S., 2007. Registration of terrestrial laser scans via image based features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), pp. 32–37.
- Besl, P. J. and McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), pp. 239–256.
- Boehm, J. and Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. *Optical 3-D Measurement Techniques VIII*, pp. 338–344.
- Brenner, C., Dold, C. and Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 63(1), pp. 4–18.
- Eggert, D. W., Lorusso, A. and Fisher, R. B., 1997. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications* 9(5-6), pp. 272–290.
- Gruen, A. and Akca, D., 2005. Least squares 3D surface and curve matching. *ISPRS Journal of Photogrammetry and Remote Sensing* 59(3), pp. 151–174.
- Hinz, S., Weinmann, Ma., Runge, P. and Jutzi, B., 2011. Potentials of image based active ranging to capture dynamic scenes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 4 / W19), pp. 143–147.
- Jutzi, B., 2012. Extending the range measurement capabilities of modulated range imaging devices by time-frequency multiplexing. *AVN - Allgemeine Vermessungs-Nachrichten* 02 / 2012, pp. 54–62.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.
- Magnusson, M., Lilienthal, A. and Duckett, T., 2007. Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics* 24(10), pp. 803–827.
- Seo, J. K., Sharp, G. C. and Lee, S. W., 2005. Range data registration using photometric features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, pp. 1140–1145.
- Weinmann, Ma. and Jutzi, B., 2011. Fully automatic image-based registration of unorganized TLS data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 5 / W12), pp. 55–60.
- Weinmann, Ma. and Jutzi, B., 2012. A step towards dynamic scene analysis with active multi-view range imaging systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39 (Part B3), pp. 433–438.
- Weinmann, Ma., Weinmann, Mi., Hinz, S. and Jutzi, B., 2011. Fast and automatic image-based registration of TLS data. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(6), pp. S62–S70.

Fast and Accurate Point Cloud Registration by Exploiting Inverse Cumulative Histograms (ICHs)

Martin Weinmann and Boris Jutzi
 Institute of Photogrammetry and Remote Sensing
 Karlsruhe Institute of Technology (KIT)
 Karlsruhe, Germany
 {martin.weinmann, boris.jutzi}@kit.edu

Abstract—The automatic and accurate alignment of captured point clouds is an important task for digitization, reconstruction and interpretation of 3D scenes. Standard approaches such as the ICP algorithm and Least Squares 3D Surface Matching require a good a priori alignment of the scans for obtaining satisfactory results. In this paper, we propose a new and fast methodology for automatic point cloud registration which does not require a good a priori alignment and is still able to recover the transformation parameters between two point clouds very accurately. The registration process is divided into coarse registration based on 3D/2D correspondences and fine registration exploiting 3D/3D correspondences. As the reliability of single 3D/2D correspondences is directly taken into account by applying Inverse Cumulative Histograms (ICHs), this approach is also capable to detect reliable tie points, even when using noisy raw point cloud data. The performance of the proposed methodology is demonstrated on a benchmark dataset and therefore allows for direct comparison with other already existing or future approaches.

I. INTRODUCTION

Terrestrial laser scanners (TLSs) are well-suited for dense and high-quality 3D reconstruction of static scenes and hence, for applications such as the digitization of urban environments or cultural heritage. A captured scan is typically represented as point cloud representing object surfaces in the local environment. Due to limitations with respect to range measurement and occlusions resulting from objects in the scene, however, a single scan is not sufficient for adequately describing complete objects. These circumstances demand for capturing multiple point clouds from different locations. As a consequence, these point clouds have to be aligned correctly in a common coordinate frame which is referred to as point cloud registration.

Standard approaches such as the Iterative Closest Point (ICP) algorithm [1] or Least Squares 3D Surface Matching (LS3D) [3] only exploit spatial 3D information. Whereas the ICP algorithm iteratively minimizes the difference between two point clouds, which can be very time-consuming for large point clouds, the LS3D approach minimizes the distance between matched surfaces. Both methods focus on a local optimization and require a good initial alignment in order to converge to the correct transformation. Other approaches are based on the distribution of 3D points (e.g. [5]) or on geometric primitives such as planes, spheres, cylinders or tori (e.g. [7]).

Whereas some techniques for point cloud registration only exploit 3D geometry, other techniques rely on either 3D geometry combined with its 2D representation as range image or 3D geometry combined with the respective 2D representation of intensity information. Involving additional information, which in fact is already available, is valuable and therefore significantly alleviates the registration process. The representations of the captured range and intensity information on the regular 2D scan grid, for instance, allow for applying image-based techniques. In general, features in the intensity image provide a higher level of distinctiveness than features in the respective range image [8] and, probably, information not represented in the range measurements. For both types of images, the image-based techniques are able to recover reliable correspondences between different images, and point correspondences are particularly useful for scenes without regular object surfaces. Projecting the information of distinctive 2D points to 3D space according to the respective range information yields sparse point clouds describing physically almost identical 3D points. Such reliable 3D/3D correspondences allow for applying the estimation of a rigid transformation which represents the least squares method in 3D and is typically combined with the RANSAC algorithm [2] for obtaining more robust results if there are still some outlier correspondences left [8]. Instead of using 3D/3D correspondences, the techniques presented in [9] and [10] introduce virtual planes and establish 3D/2D correspondences which allow for efficiently registering point clouds by applying a RANSAC-based scheme involving the EPnP algorithm [6].

In this paper, we propose a methodology which is based on a combined range and intensity representation. Similar to many techniques, the whole registration process is divided into coarse and fine registration. The coarse registration is comparable to the approach presented in [10] and uses 3D/2D correspondences, but it includes an additional measure indicating the quality of each point correspondence. This measure is used for weighting the influence of single point correspondences on the estimated transformation and already removing unreliable point correspondences. The resulting coarse estimate provides a high accuracy, so that a second outlier removal based on 3D distances between corresponding and coarsely aligned 3D points is sufficient to obtain the most promising 3D/3D correspondences. Hence, at this stage, a simple rigid

transformation is applicable for fine registration based on the remaining inlier correspondences and thus, the standard 3D Least Squares method which minimizes the difference between the respective point correspondences.

The contribution of this paper is a new methodology for point cloud registration which directly involves a quality measure for weighting the influence of corresponding information on the estimated transformation parameters in coarse registration, and replaces typical iterative schemes for fine registration by combining very reliable correspondences with a straight-forward approach.

II. METHODOLOGY

A. Feature Extraction and Matching

In a first step, the measured intensity information of each scan is adapted to the interval $[0, 255]$ by applying a histogram normalization. Image representations of this intensity information allow for efficiently comparing the scans via visual 2D features and, consequently, for preparing an unorganized set of TLS point clouds for a successive pairwise registration scheme. In order to ensure a reliable detection of corresponding information, the Scale Invariant Feature Transform (SIFT) [4] is used. This algorithm yields distinctive keypoints at blob-like structures of an image as well as the respective local feature descriptors which are invariant to image scaling and image rotation, and robust with respect to image noise, changes in illumination and small changes in viewpoint. The vector representation of the descriptors allows a consideration of the Euclidean distance between two descriptors for deriving their similarity. In order to obtain only the distinctive descriptors, the ratio of the Euclidean distances of a descriptor belonging to a keypoint in one image to the nearest neighbor and the second nearest neighbor in the other image is considered. This ratio can vary between 0 and 1. If it is below a certain threshold $t_{des} = 0.6 \dots 0.8$, there is only one similar descriptor in the other image and hence, the respective feature correspondence is considered to be distinctive.

Comparing any pair of intensity image representations belonging to different scans i and j , the confusion matrix $\mathbf{C} = [c_{ij}]$ can be derived where the entry c_{ij} represents the total number of detected feature correspondences. Once the confusion matrix is available, the graph-based algorithm presented in [10] is used to derive the minimum spanning tree and sort the scans according to their similarity for the successive pairwise registration process.

B. Point Quality Assessment and Interpolation

The raw data of a typical scan is partially corrupted with noise, i.e. only points which arise from objects in the scene will probably provide a smooth surface whereas points corresponding to the sky or points along edges of the objects might be very noisy. In order to detect such noisy measurements, the standard deviation σ of all range values within a 3×3 neighborhood is calculated for each point on the regular 2D grid and used as a measure describing the reliability of the

respective range information [10]. This yields a confidence map M_C .

In feature extraction, the image locations of SIFT features are determined with subpixel accuracy. The spatial 3D information, the intensity information and the measure for the point quality however are only available for discrete points on a regular scan grid. For this reason, the respective 3D points belonging to the SIFT feature correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ and the assigned quality measures σ_i and σ'_i have to be interpolated from the existing data, e.g. by applying a bilinear transformation. Thus, by exploiting meaningful 2D points \mathbf{x}_i and \mathbf{x}'_i , the large and dense point cloud data is reduced to sparse point clouds where 3D/3D point correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ indicate physically almost identical 3D points \mathbf{X}_i and \mathbf{X}'_i . Besides, the reliability of each of the remaining 3D points \mathbf{X}_i and \mathbf{X}'_i is specified with the quality measure σ_i and σ'_i .

C. Weight Calculation with Inverse Cumulative Histograms

Given a set of derived 2D points \mathbf{x}_i , the respective 3D points \mathbf{X}_i and a measure σ_i describing the respective reliability, the importance of each of these points can be weighted. A measure suited for such a weighting can be derived by applying Inverse Cumulative Histograms (ICHs) [11]. The creation of an ICH is initialized by dividing the interval $[0m, 1m]$ into $n_b = 100$ bins of equal size. For all points, the respective quality measure σ_i is mapped to the respective bin b_j . Due to the upper limitation, points with $\sigma_i > 1m$ are mapped to the last bin. The occurrence of mappings to the different bins is stored in a histogram $\mathbf{h} = [h_j]_{j=1, \dots, 100}$. The cumulative histograms

$$\mathbf{h}_c = \left[\sum_{j=1}^i h_j \right]_{i=1, \dots, 100} \quad (1)$$

then reach from 0 to the number n of detected correspondences. However, points with a low standard deviation σ_i are more reliable and should therefore be assigned a higher weight than those points with a higher standard deviation. For this reason, the ICH

$$\mathbf{h}_{c,inv} = \left[n - \sum_{j=1}^i h_j \right]_{i=1, \dots, 100} \quad (2)$$

is derived which satisfies this constraint. As point correspondences have already been derived at this stage, this scheme is applied for samples $s_i = (\mathbf{X}_i, \sigma_i)$ and the corresponding samples $s'_i = (\mathbf{X}'_i, \sigma'_i)$. For finally weighting the influence of each correspondence on the estimated transformation, the minimum weight of both values for a correspondence is used, and in order to keep the difference between maximum and minimum possible weight reasonable, this measure is modified to

$$w_i = \sqrt{\min\{\mathbf{h}_{c,inv}(\sigma_i), \mathbf{h}'_{c,inv}(\sigma'_i)\}}. \quad (3)$$

D. Point Cloud Registration

The spatial transformation between two point clouds with 3D/3D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ can formally be described as $\mathbf{X}'_i = \mathbf{R}\mathbf{X}_i + \mathbf{t}$, where \mathbf{R} and \mathbf{t} denote the difference in rotation and translation between two point clouds. Additional information, however, can be exploited to reduce the effect of outlier correspondences which might still be present at this stage. Hence, similar to [9] and [10], the coarse registration is based on transferring the point cloud registration to solving the PnP problem. This can be achieved by introducing virtual image planes \mathcal{P}_C assigned to each scan according to

$$\mathbf{x}_i = \mathbf{K}_C [\mathbf{R}_C | \mathbf{t}_C] \mathbf{X}_i, \quad (4)$$

where \mathbf{K}_C can be considered as calibration matrix of a virtual camera C and the parameters \mathbf{R}_C and \mathbf{t}_C describe its fixed relative orientation with respect to the local coordinate frame of the scanning device. As a virtual camera is assumed, any values for focal length or principal point are possible and the image plane does not necessarily have to be limited on a finite area. Without loss of generality, one position of the scanner can be assumed to be known, and hence, the projections of the 2D locations of SIFT features to 3D space are also known. Combining the known 3D points of this scan with the 2D observations on the virtual plane assigned to a new scan yields 3D/2D correspondences. These 3D/2D correspondences are required for solving the PnP problem, and the Efficient-Perspective- n -Point (EP nP) algorithm [6] provides an accurate solution with only linear complexity. For increased robustness, the EP nP algorithm is typically combined with a RANSAC-based scheme. The resulting estimate usually provides a high reliability.

However, as there might still be some outlier correspondences due to periodic shapes of façades, a subsequent outlier removal is carried out which focuses on the 3D distances $d(\mathbf{X}_i, \mathbf{X}'_i)$ after aligning the sparse point clouds according to the coarse estimate. As small 3D distances indicate the most promising point correspondences, only the reliable point correspondences with $d(\mathbf{X}_i, \mathbf{X}'_i) < \bar{d}$, where \bar{d} denotes the median value of all 3D distances, are used for the subsequent fine registration. Here, the high reliability of the remaining 3D/3D correspondences is sufficient to apply a standard rigid transformation minimizing the difference between the sparse point clouds in the standard Least Squares sense. For calculating a rigid transformation, at least three non-collinear point correspondences are required. The presence of more 3D/3D correspondences increases the accuracy of the transformation, and the presence of outlier correspondences causing erroneous estimates is very unlikely at this stage.

III. EVALUATION AND DISCUSSION

A. Dataset

The performance of the proposed methodology is demonstrated on a benchmark TLS dataset acquired in the German city of Hanover in an area called Holzmarkt [12]. It consists of several scans and the respective reference values for the

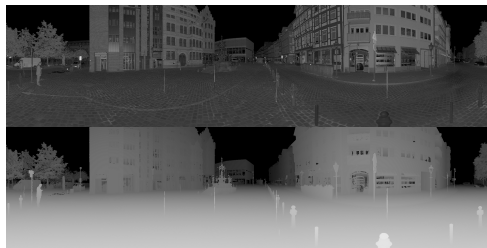


Fig. 1. Image representations for intensity and range information for the scan with the scan ID 01.

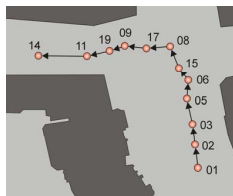


Fig. 2. Scene with the different scan positions (left) and the derived minimum spanning tree for successive pairwise registration (right).

relative orientation between these scans. The scans have been recorded with a Riegl LMS-Z360i and contain information about the spatial 3D coordinates of object points as well as the corresponding intensity information (Fig. 1). Each scan covers 360° in the horizontal direction and 90° in the vertical direction with a single shot measurement accuracy of 12mm and an angular resolution of 0.12° up to a range of approximately 200m. Thus, a scan is represented by 2.25 million 3D points at a regular scan grid of 3000×750 points which allows a 2D representation for intensity and range data. For the tests, the subset of the 12 upright scans is used.

B. Experimental Results

First, the scans are ordered according to their similarity for the successive pairwise registration process. The scene, the scan positions named according to the respective scan ID, and the scheme derived via the minimum spanning tree are illustrated in Fig. 2. The arrows in the right part of Fig. 2 are labeled with the number of SIFT correspondences between the respective intensity images. For the first pair of scans, 217 SIFT correspondences have been detected with $t_{des} = 0.66$. As only the filtered point clouds are available, only 182 of them can be assigned the corresponding range information. The point quality of the 182 respective 3D points belonging to scan 01 is depicted in Fig. 3 as well as the derived cumulative histogram and the ICH. As the entry of the last bin of the ICH equals 0 (Fig. 3, right), very unreliable point correspondences have no influence on the registration process. The results after coarse and fine registration are shown in Fig. 4 and Fig. 5 (blue and red curve). It becomes visible that the coarse registration already provides a high accuracy, except for the last scan pair where the distance between the respective scan positions is almost 12m. After fine registration, the absolute position error can further be reduced to $0.5 \dots 3.2$ cm.

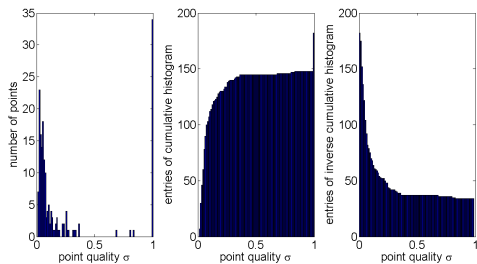


Fig. 3. Histogram based on the range reliability of derived 3D points for the scan with the scan ID 01 (left), the respective cumulative histogram (center) and the respective ICH (right).

C. Discussion

Whereas the ICP algorithm is not applicable due to the required, but not available good a priori alignment, the standard rigid transformation is strongly influenced by outlier correspondences. These might be caused by repetitive patterns on building façades, e.g. around windows, or by unreliable range information assigned to a SIFT feature. The resulting estimates hence are not reliable (Fig. 4, black curve). Combining the proposed weighting scheme with a rigid transformation improves the results (Fig. 4, green curve) as the outlier correspondences providing a low quality of range information get reduced influence. In contrast to this, the presented methodology recovers the transformation parameters much more accurately for both coarse registration (Fig. 4 and Fig. 5, blue curve) and fine registration (Fig. 4 and Fig. 5, red curve).

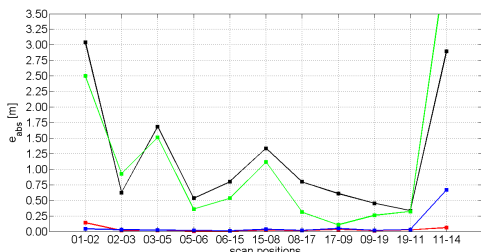


Fig. 4. Absolute position error when applying a rigid transformation (black curve), a rigid transformation combined with the new weighting scheme (green curve), and the new methodology (blue curve: result after coarse registration; red curve: result after fine registration).

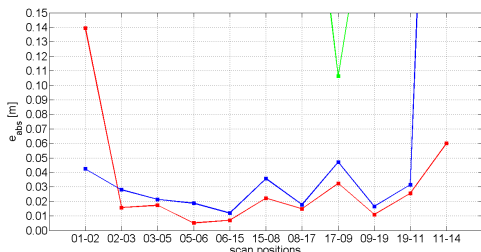


Fig. 5. Closer consideration of the results after coarse registration (blue curve) and fine registration (red curve).

Additionally, the registration process for two point clouds with millions of points takes 7.97s for feature matching, which can significantly be reduced by using a GPU-implementation, and only 7.14s for weight calculation, coarse registration, outlier removal and fine registration on a standard notebook (2.3GHz, 4GB RAM, Matlab implementation).

IV. CONCLUSION

In this paper, we propose a new and fast methodology for point cloud registration which provides reliable and accurate results without the need of a priori information about the order of scans, the presence of regular surfaces or human interaction. The registration process exploits a quality measure derived from the 2D range representation. By applying a weighting scheme based on Inverse Cumulative Histograms (ICHs), the proposed approach is even capable to detect reliable tie points when only noisy raw point cloud data is available. This new weighting scheme simultaneously allows for weighting the importance of single point correspondences and already removing very unreliable ones. Without the need for adaptations, this approach can directly be applied when using different scanning devices such as Microsoft Kinect or Time-of-Flight cameras. For future work, it would be desirable to compare different approaches for registering point clouds on a benchmark dataset.

REFERENCES

- [1] P. J. Besl, N. D. McKay, "A Method for Registration of 3-D Shapes", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, 1992, pp. 239-256
- [2] M. A. Fischler, R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, Vol. 24, No. 6, 1981, pp. 381-395
- [3] A. Gruen, D. Akca, "Least Squares 3D Surface and Curve Matching", ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 59, No. 3, 2005, pp. 151-174
- [4] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol. 60, No. 2, 2004, pp. 91-110
- [5] M. Magnusson, A. Lilienthal, T. Duckett, "Scan Registration for Autonomous Mining Vehicles Using 3D-NDT", Journal of Field Robotics, Vol. 24, No. 10, 2007, pp. 803-827
- [6] F. Moreno-Noguer, V. Lepetit, P. Fua, "Accurate Non-Iterative O(n) Solution to the PnP Problem", In: Proceedings of the IEEE 11th International Conference on Computer Vision, 2007, pp. 1-8
- [7] T. Rabbani, S. Dijkman, F. van den Heuvel, G. Vosselman, "An Integrated Approach for Modelling and Global Registration of Point Clouds", ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 61, No. 6, 2007, pp. 355-370
- [8] J. K. Seo, G. C. Sharp, S. W. Lee, "Range Data Registration Using Photometric Features", In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 1140-1145
- [9] Ma. Weinmann, Mi. Weinmann, S. Hinz, B. Jutzi, "Fast and Automatic Image-Based Registration of TLS Data", ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 66, No. 6, 2011, pp. S62-S70
- [10] M. Weinmann, B. Jutzi, "Fully Automatic Image-Based Registration of Unorganized TLS Data", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, No. 5, 2011
- [11] M. Weinmann, B. Jutzi, "A Step Towards Dynamic Scene Analysis With Active Multi-View Range Imaging Systems", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 39, No. B3, 2012, pp. 433-438
- [12] <http://www.ikg.uni-hannover.de/index.php?id=413&L=de>

FULLY AUTOMATIC IMAGE-BASED REGISTRATION OF UNORGANIZED TLS DATA

Martin Weinmann, Boris Jutzi

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Kaiserstr. 12, 76128 Karlsruhe, Germany
{martin.weinmann, boris.jutzi}@kit.edu

KEY WORDS: Laser Scanning, TLS, point cloud, registration, image-based, graph-based, automation.

ABSTRACT:

The estimation of the transformation parameters between different point clouds is still a crucial task as it is usually followed by scene reconstruction, object detection or object recognition. Therefore, the estimates should be as accurate as possible. Recent developments show that it is feasible to utilize both the measured range information and the reflectance information sampled as image, as 2D imagery provides additional information. In this paper, an image-based registration approach for TLS data is presented which consists of two major steps. In the first step, the order of the scans is calculated by checking the similarity of the respective reflectance images via the total number of SIFT correspondences between them. Subsequently, in the second step, for each SIFT correspondence the respective SIFT features are filtered with respect to their reliability concerning the range information and projected to 3D space. Combining the 3D points with 2D observations on a virtual plane yields 3D-to-2D correspondences from which the coarse transformation parameters can be estimated via a RANSAC-based registration scheme including the EPnP algorithm. After this coarse registration, the 3D points are again checked for consistency by using constraints based on the 3D distance, and, finally, the remaining 3D points are used for an ICP-based fine registration. Thus, the proposed methodology provides a fast, reliable, accurate and fully automatic image-based approach for the registration of unorganized point clouds without the need of a priori information about the order of the scans, the presence of regular surfaces or human interaction.

1. INTRODUCTION

The automatic registration of point clouds acquired with a terrestrial laser scanner (TLS) is still of great interest. Each point cloud represents dense and accurate 3D information about surfaces of objects in the local area around the scanner with respect to a local coordinate frame. However, usually multiple scans from different locations have to be recorded to obtain a full scene coverage. Hence, a registration process has to be carried out which transforms all point clouds into a common coordinate frame.

Standard approaches for calculating the transformation parameters between two partially overlapping point clouds are based on the Iterative Closest Point (ICP) algorithm (Besl & McKay, 1992) and different variants of it (Rusinkiewicz & Levoy, 2001). The ICP algorithm minimizes the difference between two point clouds. For large numbers of points, however, the ICP algorithm shows a high computational effort which is due to the iterative processing scheme. Hence, it seems quite feasible to extract relevant information from the point clouds which can be used for registration. Such relevant information may for example be derived via the distribution of the points within each point cloud by using the normal distributions transform (NDT) either on 2D scan slices (Brenner et al., 2008) or in 3D (Magnusson et al., 2007).

Urban environments or scenes containing industrial installations usually contain regular surfaces of which various types of geometric features might arise. Simple features which are likely to occur and useful for registration are lines (Stamos & Leordeanu, 2003) derived from the range information sampled as range images. Other commonly used features which are extracted directly from the point clouds are planes (Dold & Brenner, 2004; Brenner et al., 2008; Pathak et al. 2010a; Pathak et al. 2010b) or more complex geometric features like spheres, cylinders or tori (Rabbani et al., 2007). However, all these feature types representing geometric primitives are not suited in scenes without regular surfaces. In addition to regular surfaces, scans in urban scenes might also contain a typical skyline. This border between the sky and a set of buildings shows special

features when using a cylindrical projection model for sampling the range information to panoramic range images covering 360° in the horizontal direction, e.g. extrema or flat regions which are suited for a coarse alignment of two scans (Nüchter et al., 2011). As such features strongly depend on the scene content, this approach is not suited if the skyline is less distinctive and thus not sufficient for registration purposes.

In the presence of cluttered scenes, descriptors representing local surface patches are more appropriate. Such descriptors may be derived from geometric curvature or normal vectors of the local surface (Bae & Lichti, 2004; Bae & Lichti 2008). Further approaches which are suitable for more complex scenes are based on extracting special feature points in the range images in order to support the registration process (Barnea & Filin, 2008; Steder et al., 2010).

Currently, most of the terrestrial laser scanners can not only measure the distance to 3D scene points but also capture either co-registered camera images or panoramic reflectance images representing the respective energy of the backscattered laser light. Therefore, several approaches are based on the use of 2D imagery as the images provide additional information about the local area around the scanner which might not always be represented in the range measurements. Hence, the registration of two point clouds can be supported by using reliable feature correspondences between the respective camera or reflectance images. For this purpose, different kinds of features can be used, but most of the current image-based approaches are based on the use of feature points or keypoints. Many of the image-based approaches use SIFT features to detect distinctive 2D feature points by which point correspondences between two images can be detected. These features can be extracted from the co-registered camera images (Bendels et al., 2004; Al-Manasir & Fraser, 2006; Barnea & Filin, 2007) or from the reflectance images (Böhm & Becker, 2007; Wang & Brenner 2008; Kang et al., 2009). For all point correspondences, the respective feature points are projected into 3D space and thus lead to a much smaller set of 3D points used for registration.

In this paper, a method for a fully automatic registration of a large number of unorganized scans is proposed. Reaching a

high level of automation also including the sorting of the scans is essential as most of the current approaches are based on pairwise registration for which an already known order of the scans is assumed. For pairwise registration, a modification of a fast and automatic image-based registration approach which has been published recently (Weinmann et al., 2011) and which is also suited for scenes without regular surfaces is presented. In contrast to this approach, only those correspondences with reliable range information are used and, instead of a refinement step, an ICP-based fine registration is introduced. Thus, besides being very fast, the proposed algorithm does neither depend on a priori information about the order of the scans nor on the presence of regular surfaces.

The paper is organized as follows. In Section 2, the processing chain of the proposed algorithm is outlined. As it will be shown, the algorithm can be divided into the two major steps of organizing the TLS data and carrying out a successive pairwise registration which are presented in Section 3 and Section 4 in detail. In Section 5, the performance of the proposed algorithm is proved by processing 11 point clouds of a benchmark TLS data set. The capability of the proposed method is discussed in Section 6 with respect to accuracy, reliability and performance. Finally, conclusions and suggestions for future work are outlined in Section 7.

2. METHODOLOGY

The registration approach proposed in this paper aims at reaching a high level of automation and simultaneously getting fast, reliable and accurate results. As illustrated in Figure 1, the approach can be divided into two major parts. The first part deals with the organization of unorganized point clouds which is later required for a successive pairwise registration. Hence after the acquisition of TLS data, special features have to be extracted which are suitable for organizing the scans. The second part focuses on the registration of the point clouds. Using reliable 3D points derived via the previously extracted features and introducing their projections onto a virtual plane yields 3D-to-2D correspondences. These are used for receiving a coarse registration of the point clouds which is followed by a fine registration in order to improve the accuracy of the results. The scheme of this briefly summarized methodology is presented in detail in Section 3 and Section 4.

3. ORGANIZATION OF TLS DATA

The current registration approaches address a fully automatic registration of different scans of a scene. To further increase the level of automation, the proposed algorithm first organizes the scans automatically, which yields a structure for successive pairwise registration. This is done by considering the recorded scans (Section 3.1), extracting distinctive features and feature correspondences between the scans (Section 3.2) and using a graph-based approach (Section 3.3).

3.1 Data Set and Reference Values

In the following, the quality of the proposed registration approach is demonstrated with a benchmark TLS data set provided by the University of Hanover. This set consists of 12 upright scans and 8 tilted scans which were acquired in the German city of Hanover in an area called Holzmarkt, and the respective reference values for the relative orientation between the scans. The scans were recorded with a Riegl LMS-Z360i scanner and contain information about the 3D coordinates of object points as well as the corresponding reflectance

information. Covering 360° in the horizontal direction and 90° in the vertical direction with a single shot measurement accuracy of 12mm and an angular resolution of 0.12° up to a range of approximately 200m, each scan returns 2.25 million 3D points from a regular scan area of 3000×750 points being represented as panoramic reflectance image (Wang & Brenner, 2008). The reflectance and range information derived from the scan at scan position 01 are visualized in Figure 2.

In order to check the quality of the automatically calculated registration results, accurate reference values are needed. The provided reference values are based on the use of artificial targets and a manual alignment which yields an expected accuracy of the scan positions in the low millimeter range. For testing the proposed algorithm, a subset consisting of 11 upright scans with a spacing of approximately 5m is used.

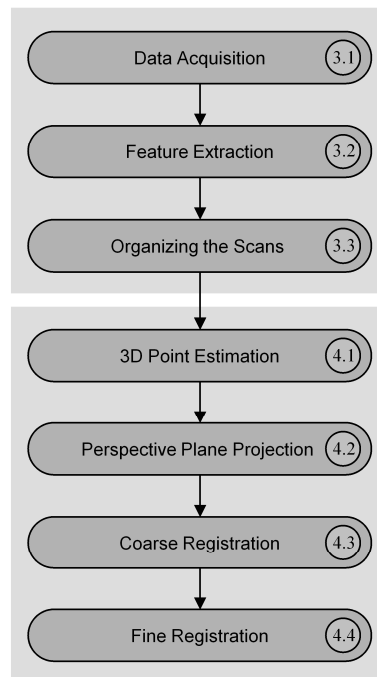


Figure 1. Processing chain of the proposed approach.

3.2 Feature Extraction

Once several scans have been acquired, the next step consists of extracting distinctive features. Here, the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is utilized for detecting such distinctive keypoints in an image derived from the TLS data and extracting local feature descriptors which are invariant to image scaling and image rotation, and robust with respect to image noise, changes in illumination and small changes in viewpoint. These descriptors allow for locating correspondences between different images and, finally, to derive common image objects. As the descriptors are represented as vectors, they can be compared by considering Euclidean distances. An effective measure describing the distinctiveness of a keypoint can be derived from the ratio of the Euclidean distances of a descriptor belonging to a keypoint in one image to the nearest neighbor and the second nearest neighbor in the other image. This ratio has to be below a given threshold t_{des} , which can vary between 0 and 1. For practical purposes and different applications, distinctive features arise when using a threshold between $t_{des} = 0.6$ and $t_{des} = 0.8$. As the feature correspondences used for

registration should be reliable, a threshold of $t_{des} = 0.66$ is used. This means that the distance of a descriptor belonging to a SIFT feature in image i to the nearest neighbor in image j is only about 2/3 of the distance to the second nearest neighbor.

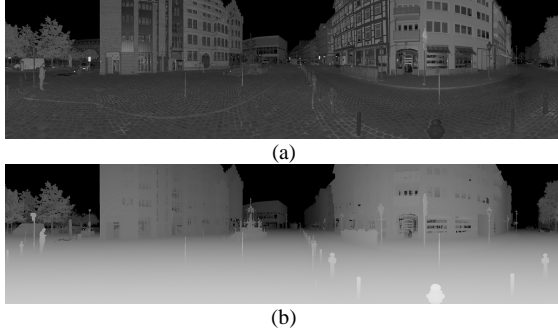


Figure 2. Visualization of the captured TLS data: (a) reflectance and (b) range information.

In order to check the similarity of the scans, the number of SIFT features between image pairs from all available positions is calculated and stored in the confusion matrix \mathbf{C} . For this purpose, the same scan IDs are used as in the provided data set. The diagonal elements $C(i,i)$ represent the total number of SIFT features extracted in the respective reflectance image i . As can be seen in Table 1, the confusion matrix is not necessarily symmetric which depends on the calculated ratio of the Euclidean distances of a feature descriptor to the nearest and second nearest neighbor. If, for a feature descriptor derived from image i , the nearest neighbor and the second nearest neighbor in image j are a little more distinctive as required, this ratio is below the threshold t_{des} and thus meets the constraint. In the reverse case, when comparing a feature descriptor derived from image j to feature descriptors derived from image i , it might occur that the nearest neighbor and the second nearest neighbor are more similar which causes a ratio above the threshold t_{des} .

3.3 Organizing Large Numbers of Scans by Similarity

When dealing with a large number of scans, it might be desirable to reach a high level of automation. This will also include automatically sorting the scans for pairwise registration so that the error between estimated and real position is minimal. Therefore, a graph-based algorithm is proposed here.

Any set of unorganized point clouds can directly be represented as a graph, where the nodes represent the scans and the edges

are weighted with the total number of SIFT correspondences between the respective scans. In the most general case, every node is connected with every other node which results in a complete graph. As mentioned before, the confusion matrix \mathbf{C} is not necessarily symmetric and therefore, a directed graph is used instead of an undirected graph. Hence, the entry $C(i,j)$ of the confusion matrix represents the weight of an unidirectional edge from node i to node j .

The first step towards organizing the point clouds consists of an initialization which can be done via selecting a defined initial scan. Alternatively, it would be possible to use other criteria if only the relations between the scans are of importance, e.g. the node from which the edge with the maximum weight within the graph starts. This initial set containing exactly one node is then iteratively expanded until it contains all nodes of the graph. Each iteration starts with searching unidirectional edges from the actual set of nodes to the remaining nodes and the edge with the maximum weight leads to the node by which the actual set is expanded. Resulting from the selected connections, a structure can be generated which represents the order of the scans for the automatic pairwise registration. For the given confusion matrix \mathbf{C} (Table 1), the resulting structure for a successive pairwise registration process is shown in Figure 3.

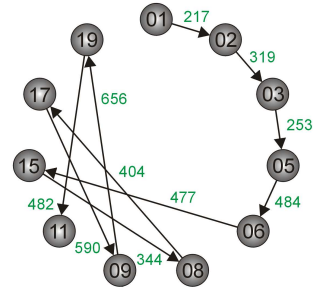


Figure 3. Resulting scheme for successive pairwise registration: The scans are labeled with their ID and the connections used for further calculations are labeled with the number of detected SIFT correspondences between the respective reflectance images.

4. REGISTRATION OF TLS DATA

After calculating the order of the scans by checking the similarity of the respective reflectance images, a pairwise registration of successive scans can be carried out. For this purpose, the already calculated 2D SIFT features leading to

Scan ID	01	02	03	05	06	08	09	11	15	17	19
01	4986	217	63	45	33	58	41	28	44	62	39
02	229	5663	319	100	59	80	43	48	35	46	38
03	88	308	5967	253	120	56	47	68	57	38	56
05	70	114	277	6200	484	78	58	68	131	68	84
06	31	70	124	466	6682	169	68	56	477	134	71
08	86	96	53	78	163	6867	205	64	328	404	99
09	39	34	37	56	68	158	5571	330	78	577	656
11	17	24	37	40	44	41	277	4061	24	134	408
15	61	40	59	129	503	344	82	30	7154	211	53
17	53	56	34	60	121	379	590	169	240	6159	361
19	21	25	43	51	54	84	629	482	42	344	4852

Table 1. Number of SIFT correspondences between the reflectance images of different scans within the chosen subset. The values can be summarized in the confusion matrix \mathbf{C} and the entry $C(i,j)$ of this matrix denotes the number of point correspondences found when all descriptors derived from image i are compared to the nearest neighbor and the second nearest neighbor derived from image j .

correspondences have to be projected to 3D space, and the reliability of the calculated 3D points has to be checked with respect to range information (Section 4.1) before registration. The following registration is based on 2D projections of the reliable 3D points onto a virtual plane (Section 4.2), and divided into a coarse registration (Section 4.3) and a fine registration (Section 4.4).

4.1 3D Point Estimation

As SIFT features are determined with subpixel accuracy, the respective 3D information has to be interpolated as the measured values are only available on the regular scan raster. A reliable 3D point corresponding to a SIFT feature can however only be generated, if all of the four nearest points on the scan raster contain valid range information. The measured points which arise from objects in the scene will probably provide a smooth surface whereas points corresponding to the sky or points along edges of the objects might be very noisy. Therefore, points have to be discarded if they do not lie on the surface of any object in the scene.

The provided scans are already filtered with respect to minimum values of the backscattered energy (Figure 4a). Additionally, the proposed algorithm considers the standard deviation σ of the values within a 3×3 neighborhood of each pixel in the range image in order to avoid unreliable range information at edges of scene objects. If the standard deviation σ of the respective range values is larger than a predefined threshold t_{std} which is selected to $t_{std} = 0.1m$, the range information of the center pixel is not reliable, otherwise the range information of the center pixel is assumed to be reliable (Figure 4b). Combining these constraints yields a 2D confidence map M_C which is illustrated in Figure 4c.

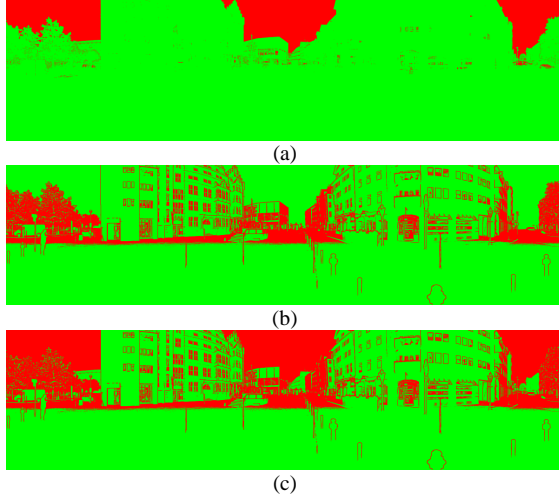


Figure 4. Confidence map for the scan at scan position 01: (a) information filtered with respect to reflectance, (b) information filtered with respect to the standard deviation σ using a threshold value of $t_{std} = 0.1m$ and (c) the resulting confidence map M_C . The reliable points are shown in green, the unreliable ones in red.

4.2 Perspective Plane Projection

The reflectance images have been sampled using a spherical projection. For the registration, however, it is useful to get the coordinates of the extracted and reliable 3D points X_i projected

onto a 2D image plane of a virtual camera in order to use powerful algorithms of computer vision applications. The respective transformation can be described via

$$x_i = K[R|t]X_i \quad (1)$$

where the matrix K is the calibration matrix of a virtual camera, and the matrix R and the vector t describe the rotation and the translation of this virtual camera with respect to the local coordinate frame of the laser scanner. In the registration process, R refers to the local coordinate frame so that the virtual camera has the same orientation as the laser scanner and looks into the horizontal direction (Weinmann et al., 2011). Besides, the position of the virtual camera is assumed to equal the location of the laser scanner and therefore, the translation vector is set to $t = 0$. As a consequence of introducing a virtual camera plane, any parameters can be used for the focal lengths of the camera in x - and y - direction as well as for the coordinates of the principal point. Furthermore, the image plane has not necessarily to be limited on a finite area and all points behind the camera may also be included by mapping them onto the virtual plane via symmetric constraints as they represent the 2D projections onto the virtual plane of a second camera looking in the opposite direction. Thus, instead of creating synthetic camera images and using these for registration (Forkuo & King, 2004), only a few points are projected with subpixel accuracy.

4.3 Coarse Registration using EPnP and RANSAC

Once 3D-to-2D correspondences are known, the problem of pose estimation is the same as when using a camera instead of a laser scanner. Recently, the Efficient Perspective-n-Point (EPnP) algorithm has been proposed as a non-iterative method to estimate the exterior orientation or pose of a camera from a set of n correspondences between 3D points X_i of a scene and their 2D projections x_i onto the image plane (Moreno-Noguer et al., 2007; Lepetit et al., 2009). The EPnP algorithm is based on the idea of expressing the n known 3D scene points X_i as a weighted sum of four virtual and non-coplanar control points C_j for general configurations. The weights α_{ij} remain unchanged when transferring this relation to camera coordinates and therefore, the points X_i^c can be expressed via the control points C_j^c which leads to

$$w_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix} = K X_i^c = K \sum_{j=1}^4 \alpha_{ij} C_j^c \quad (2)$$

for $i = 1, \dots, n$, where K describes the camera matrix. The scalar projective parameters w_i can be substituted by

$$w_i = \sum_{j=1}^4 \alpha_{ij} Z_j^c \quad (3)$$

via the Z^c coordinates of the control points. Concatenating the resulting equations for all n 3D-to-2D correspondences yields a linear system $Mx = 0$ with $x = [C_1^{cT}, C_2^{cT}, C_3^{cT}, C_4^{cT}]^T$ and a $2n \times 12$ matrix M . The solution x then leads to the camera coordinates X_i^c . Once the world coordinates and the camera coordinates of the 3D points are known, the rotation and translation parameters aligning both coordinate systems can be retrieved via standard methods (Horn et al., 1988). As the EPnP algorithm considers all 3D-to-2D correspondences without checking their reliability, the quality of the registration results

can be increased by introducing further constraints. The RANSAC algorithm (Fischler & Bolles, 1981) provides a good possibility for eliminating outliers and thus reaching a more robust pose estimation. This combination of EPnP and RANSAC is based on randomly selecting small, but not minimal subsets of seven correspondences for estimating the model parameters, and checking the whole set of correspondences for consistent sample points (Moreno-Noguer et al., 2007; Lepetit et al., 2009).

4.4 Fine Registration using Outlier Removal and ICP

The results from the previous coarse registration provide a good a priori alignment which is required for using the ICP algorithm in order to get a fine registration. However, the RANSAC algorithm only considers the quality of the 3D-to-2D correspondences. Hence, the quality of the 3D points has to be considered separately. This is done by eliminating those 3D-to-3D correspondences for which at least one 3D point arises from periodic shapes of façades and thus from ambiguities in the scene via geometric constraints (Weinmann et al., 2011). The resulting points used for fine registration are very reliable as they have been checked with respect to the reliability of their range information, the quality of the 3D-to-2D correspondences and the quality of the 3D-to-3D correspondences. Therefore, the ICP algorithm is expected to yield very accurate results.

5. EVALUATION

The first part of the presented approach can easily be verified if the nodes belonging to the scans are projected onto the reference positions of the respective scans in the scene, which is done in Figure 5.

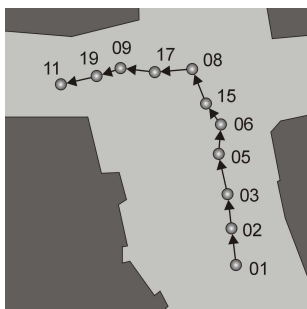


Figure 5. Scans sorted with respect to their reference position in the nadir view of the scene: The streets are colored in bright gray, buildings in dark gray.

Once the scans are sorted, a successive pairwise registration can be carried out. Between the reflectance images of the scans 01 and 02, a total number of 217 SIFT correspondences has been detected (Table 1) of which 89 are reliable with respect to the range information of the corresponding SIFT features. The respective 3D points of those reliable SIFT correspondences are projected to 3D space using bilinear interpolation. If for one scan, the absolute transformation parameters with respect to the world coordinate frame are known which is assumed for scan position 01, the 3D world coordinates of the calculated 3D points can easily be determined. For a new scan, the corresponding 2D features with reliable range information are also projected to 3D space and backprojected onto a virtual plane assigned to the local coordinate frame which yields 2D observations. Establishing 3D-to-2D correspondences from the 3D information derived from the first scan and the 2D observations derived from the new scan allows for using the

EPnP algorithm which has been extended by the RANSAC algorithm for an increased robustness. Subsequently, a consistency check with respect to 3D distances between the 3D points transformed into a common coordinate frame via the coarse estimate of the transformation parameters is carried out. After this geometric outlier removal, the remaining 3D-to-3D correspondences (29 between the scans 01 and 02) are used for an ICP-based fine registration. As shown in Figure 6, the absolute position errors after coarse registration are in the range between 12mm and 49mm, and the fine registration yields accurate results with absolute position errors between 9mm and 32mm.

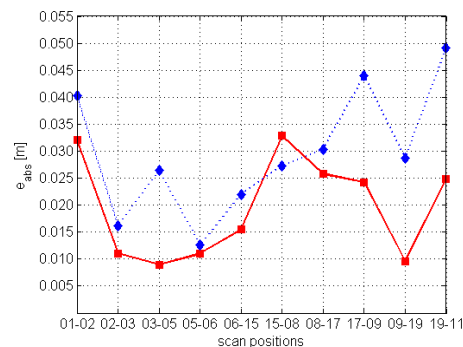


Figure 6. Absolute error between reference and estimated positions for coarse (dotted line with diamonds) and fine registration (solid line with squares).

6. DISCUSSION

The presented registration approach was tested in Matlab on a standard PC with 2.83GHz. Although the code is not fully optimized with respect to a possible parallelization on multiple cores and thus only one core is used, the average time required for pairwise registration is about 13s. Of this time, about 5s are needed for calculating the SIFT correspondences, about 7s for coarse estimation using a RANSAC-based scheme including EPnP and only 1s for consistency checks and ICP on two relatively small subsets each consisting of approximately 100 points. If the ICP algorithm is used for larger subsets, the computational effort increases highly.

Concerning accuracy and performance, the proposed approach is comparable to other image-based approaches (Wang & Brenner, 2008; Weinmann et al., 2011). As the approach focuses on using only reliable information concerning range, 3D-to-2D correspondences and 3D-to-3D correspondences, the estimated transformation parameters are very reliable which can be seen when comparing them to the reference values (Figure 6). The approach is suited for both urban environments and scenes containing vegetation and does neither depend on regular surfaces nor human interaction. However, one constraint concerning the scene arises as point-like features have to be extracted. Hence, the scene has to be well-structured which is assumed in all image-based approaches using SIFT features.

As the total number of SIFT correspondences decreases with an increasing distance between the respective scan positions which can be seen when considering the entries in the confusion matrix (Table 1) and the reference positions (Figure 5), the presented approach as well as other image-based approaches will not lead to optimal results for larger distances between the scans. For this purpose, approaches based on geometric primitives (Brenner et al., 2008; Rabbani et al., 2007) might be

more robust in direct comparison, but they assume that regular surfaces can be found in the scene and thus less general scenes.

7. CONCLUSION AND OUTLOOK

In this paper, a fully automatic registration approach is presented which is based on both the range information and the reflectance information of terrestrial laser scans. Automatically sorting any number of unorganized scans by means of their similarity and then carrying out a successive fast and accurate pairwise registration, the approach provides a powerful framework suited for typical environments. The approach has been successfully applied to a benchmark TLS data set containing millions of points and been discussed concerning accuracy, reliability and performance. For future work, the approach could be extended by introducing a final global registration over all scans or at least considering those parts of the confusion matrix arising from the similarity of a new scan to all of the already registered scans. This might improve the quality of the estimated transformation parameters and yield an even further increased robustness.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Claus Brenner and the Institute of Cartography and Geoinformatics at the University of Hanover for providing the TLS data. The data is available at <http://www.ikg.uni-hannover.de/index.php?id=413&L=de> and has been accessed in March 2011.

REFERENCES

- Al-Manasir, K., Fraser, C. S., 2006. Registration of terrestrial laser scanner data using imagery. *The Photogrammetric Record*, 21 (115), pp. 255-268.
- Bae, K.-H., Lichti, D. D., 2004. Automated registration of unorganised point clouds from terrestrial laser scanners. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B5), pp. 222-227.
- Bae, K.-H., Lichti, D. D., 2008. A method for automated registration of unorganised point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), pp. 36-54.
- Barnea, S., Filin, S., 2007. Registration of terrestrial laser scans via image based features. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), pp. 32-37.
- Barnea, S., Filin, S., 2008. Keypoint based autonomous registration of terrestrial laser point-clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), 19-35.
- Bendels, G. H., Degener, P., Körtgen, M., Klein, R., 2004. Image-based registration of 3D-range data using feature surface elements. In: Chrysanthou, Y., Cain, K., Silberman, N., Niccolucci, F. (Eds.), *The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, pp. 115-124.
- Besl, P. J., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), pp. 239-256.
- Böhm, J., Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. In: Gruen, A., Kahmen, H. (Eds), *Optical 3-D Measurement Techniques VIII*, pp. 338-344.
- Brenner, C., Dold, C., Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), pp. 4-18.
- Dold, C., Brenner, C., 2004. Automatic matching of terrestrial scan data as a basis for the generation of detailed 3D city models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B3), pp. 1091-1096.
- Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24 (6), pp. 381-395.
- Forkuo, E. K., King, B., 2004. Automatic fusion of photogrammetric imagery and laser scanner point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B4), pp. 921-926.
- Horn, B. K. P., Hilden, H. M., Negahdaripour, S., 1988. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A* (5), pp. 1127-1135.
- Kang, Z., Li, J., Zhang, L., Zhao, Q., Zlatanova, S., 2009. Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. *Sensors* 9 (4), pp. 2621-2646.
- Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. EPnP: An accurate O(n) solution to the PnP problem. *International Journal of Computer Vision* 81 (2), pp. 155-166.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60 (2), pp. 91-110.
- Magnusson, M., Lilienthal, A., Duckett, T., 2007. Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics* 24 (10), pp. 803-827.
- Moreno-Noguer, F., Lepetit, V., Fua, P., 2007. Accurate non-iterative O(n) solution to the PnP problem. *IEEE 11th International Conference on Computer Vision*, pp. 1-8.
- Nüchter, A., Güttev, S., Borrmann, D., Elseberg, J., 2011. Skyline-based registration of 3D laser scans. *Journal of Geo-spatial Information Science*. In Press.
- Pathak, K., Birk, A., Vaskevicius, N., Poppinga, J., 2010a. Fast registration based on noisy planes with unknown correspondences for 3-D mapping. *IEEE Transactions on Robotics* 26 (3), pp. 424-441.
- Pathak, K., Borrmann, D., Elseberg, J., Vaskevicius, N., Birk, A., Nüchter, A., 2010b. Evaluation of the robustness of planar-patches based 3D-registration using marker-based ground-truth in an outdoor urban scenario. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5725-5730.
- Rabbani, T., Dijkman, S., van den Heuvel, F., Vosselman, G., 2007. An integrated approach for modelling and global registration of point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 61 (6), pp. 355-370.
- Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the ICP algorithm. *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling*, pp. 145-152.
- Stamos, I., Leordeanu, M., 2003. Automated feature-based range registration of urban scenes of large scale. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. II, pp. 555-561.
- Steder, B., Grisetti, G., Burgard, W., 2010. Robust place recognition for 3D range data based on point features. *IEEE International Conference on Robotics and Automation*, pp. 1400-1405.
- Wang, Z., Brenner, C., 2008. Point based registration of terrestrial laser data using intensity and geometry features. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (Part B5), pp. 583-589.
- Weinmann, Ma., Weinmann, Mi., Hinz, S., Jutzi, B., 2011. Fast and automatic image-based registration of TLS data. *ISPRS Journal of Photogrammetry and Remote Sensing*.

THERMAL 3D MAPPING FOR OBJECT DETECTION IN DYNAMIC SCENES

Martin Weinmann ^a, Jens Leitloff ^a, Ludwig Hoegner ^b, Boris Jutzi ^a, Uwe Stilla ^b and Stefan Hinz ^a

^a Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
{martin.weinmann, jens.leitloff, boris.jutzi, stefan.hinz}@kit.edu

^b Photogrammetry and Remote Sensing, Technische Universität München (TUM)
ludwig.hoegner@bv.tum.de, stilla@tum.de

Commission I, WG I/2

KEY WORDS: Multisensor, point cloud, thermal imaging, 3D mapping, dynamic, object detection

ABSTRACT:

The automatic analysis of 3D point clouds has become a crucial task in photogrammetry, remote sensing and computer vision. Whereas modern range cameras simultaneously provide both range and intensity images with high frame rates, other devices can be used to obtain further information which could be quite valuable for tasks such as object detection or scene interpretation. In particular thermal information offers many advantages, since people can easily be detected as heat sources in typical indoor or outdoor environments and, furthermore, a variety of concealed objects such as heating pipes as well as structural properties such as defects in isolation may be observed. In this paper, we focus on thermal 3D mapping which allows to observe the evolution of a dynamic 3D scene over time. We present a fully automatic methodology consisting of four successive steps: (i) a radiometric correction, (ii) a geometric calibration, (iii) a robust approach for detecting reliable feature correspondences and (iv) a co-registration of 3D point cloud data and thermal information via a RANSAC-based EPnP scheme. For an indoor scene, we demonstrate that our methodology outperforms other recent approaches in terms of both accuracy and applicability. We additionally show that efficient straightforward techniques allow a categorization according to background, people, passive scene manipulation and active scene manipulation.

1. INTRODUCTION

The automated description and analysis of static and dynamic 3D scenes represents a topic of major interest in photogrammetry, remote sensing and computer vision. Due to the recent technological advancements, a variety of devices is currently available which can be used for acquiring different types of information such as color, temperature or spatial 3D geometry. Thus, the use of different devices on a common sensor platform allows to collect multidimensional spatial data.

In particular those devices delivering complementary types of information offer a high potential for numerous applications. Modern range cameras such as Microsoft Kinect, PMD[vision] CamCube 2.0 or MESA Imaging SR4000 simultaneously provide geometric information as well as radiometric information in form of range and intensity images, and they are also applicable for adequately capturing dynamic scenes. Whereas the radiometric information is typically represented as color or gray-valued images, other devices such as thermal cameras offer to capture complementary information which can be helpful for describing and analyzing the observed scene and its evolution over time.

A fusion of intensity and thermal information for instance seems desirable, since the respective images reveal a very different behavior:

- Standard intensity images (i.e. color or gray-valued images) typically represent information of the visual domain and thus radiometric surface properties of observed objects. This information may strongly depend on the material and reflectance behavior of respective objects as well as the relative geometric orientation of the surface to the camera.
- Thermal infrared images represent thermal radiation in the infrared spectrum. This radiation is emitted by objects in the

scene and not visible in the visual domain. Consequently, the thermal infrared images allow a different look on objects and the extraction of additional information like temperature and different materials of observed objects.

Accordingly, objects visible in the visual domain may be invisible in the infrared domain if they have the same temperature and emissivity coefficient as the respective background. In contrast, in infrared images, even further objects below the surface of an object may be visible which certainly remain invisible in the visual domain. Note that two different materials with the same temperature can appear with different intensity if they have a significantly different emissivity coefficient. Interestingly, two objects with different temperature and emissivity coefficient can even coincidentally appear with very similar intensity in thermal images. As a consequence, a fusion of intensity images and thermal infrared images can reveal information which may not be present in either intensity images or thermal infrared images (Chen and Leung, 2009; Bai et al., 2011). This is for instance of special interest for enhancing contrast in environments of poor visibility or inadequate illumination (Liu and Laganière, 2007), for target detection (Yao and Sekmen, 2008) or for concealed weapon detection (Xue et al., 2002). More importantly, however, infrared thermography allows for building diagnostics (Balaras and Argiriou, 2002) which, due to current attempts for saving energy, has become a research topic itself.

Whereas the fusion of intensity images and thermal infrared images provides different types of information, it does not account for the respective spatial dimensions. Recent technological advancements however allow a thermal 3D mapping in terms of projecting the captured image data onto 3D point clouds acquired with a terrestrial laser scanner, mobile laser scanner or range camera. The metric information, in turn, allows the quantification of thermal studies if the point clouds are texturized with thermal information (Lagüela et al., 2011b). Especially in building

observation (Hoegner et al., 2007a; Iwaszczuk et al., 2011), the joint representation of building façades and thermal information allows a reconstruction of the surface temperature and thus a look into the interior behavior of a wall. Valuable insights about different materials, heating pipes and leakages can easily be obtained and added as semantic or geometric information to the respective building model.

In order to capture co-registered intensity information, thermal information and spatial 3D geometry, the combination of data captured with a terrestrial laser scanner and images acquired by a bi-camera system, i.e. a system consisting of an optical camera and a thermal camera, has recently been proposed (Alba et al., 2011). Furthermore, a robot equipped with a 3D laser scanner, a thermal camera and a digital camera has been presented which is capable to generate a precise 3D model showing the heat distribution in the scene (Borrmann et al., 2012a; Borrmann et al., 2012b). The additional color information can for instance be used for identifying heat sources or for obtaining photo-realistic 3D models. Focusing on the acquisition of dynamic scenes, respective systems addressing thermal 3D mapping involve range cameras such as RGB-D cameras or Time-of-Flight (ToF) cameras. For instance, the use of a low-cost RGB-D camera in combination with a thermal camera has recently been proposed for acquiring dense 3D models of environments with both appearance and temperature information (Vidas et al., 2013). Involving a Time-of-Flight camera, an approach for gaining infrared-textured 3D models has been proposed (Weinmann et al., 2012) which additionally focuses on sharpening the blurry thermal information by exploiting the high quality of intensity information acquired with the range camera.

In this paper, we focus on thermal 3D mapping which allows to observe the evolution of a dynamic 3D scene over time. The main contribution presented in our paper is a fully automatic methodology for thermal 3D mapping which involves

- a geometric calibration of both the range camera and the thermal camera based on a common strategy,
- a robust approach for detecting 2D/2D feature correspondences via shape-based matching,
- a removal of unreliable 2D/2D feature correspondences by considering the reliability of the respective range information, and
- a novel approach exploiting 3D/2D feature correspondences for the co-registration of 3D point cloud data and thermal information.

After reflecting the related work in Section 2, we provide a detailed explanation of the different steps of our new methodology in Section 3. For an experimental setup described in Section 4, we present the derived results in Section 5 and discuss these with respect to different aspects in Section 6. Finally, Section 7 contains conclusions as well as ideas for future work.

2. RELATED WORK

For thermal 3D mapping involving a range camera, respective 3D coordinates are available for each pixel of the intensity image. Accordingly, feature correspondences have to be derived either between standard intensity images (i.e. color or gray-valued images) and thermal infrared images (Section 2.1), or between 3D point clouds and thermal infrared images (Section 2.2). For the sake of completeness, we also briefly reflect approaches generating 3D point clouds from thermal infrared images (Section 2.3).

2.1 Co-Registration of Intensity and Thermal IR Images

One category of approaches directly leads to basic image processing in terms of image registration (Zitová and Flusser, 2003). Once the respective transformation model has been estimated, the thermal infrared image can be warped onto the intensity image of the range camera, and thus also be projected to 3D space by forward projection according to the respective 3D information.

Among the approaches for image registration, the feature-based approaches are most widely spread. In general, we may consider any feature present in an image as visual feature. More specifically, according to a recent survey on visual features (Weinmann, 2013), different types of visual features can be categorized, where texture, shape and local features are the most prominent types due to their applicability for numerous applications such as image registration, data retrieval, scene analysis and the autonomous navigation of aerial and ground vehicles. Typically, approaches for feature extraction and matching are tailored for images acquired by either identical or similar sensors from various positions. However, a matching between different image domains – e.g. a co-registration of satellite imagery and LiDAR intensity images (Toth et al., 2011) – can be quite challenging as the respective images may have very different characteristics due to which many standard approaches for deriving feature correspondences tend to fail.

In particular, an automatic matching between the visual domain and the thermal domain still remains challenging. Intensity images in the visual domain often provide abrupt changes of properties like intensity or texture at the edges of objects in the observed scene. In thermal infrared images, however, we might face challenges arising from (i) the low geometric resolution compared to classical optical camera systems and (ii) the fact that features such as lines or contours do not show strong edges, but rather appear blurred. As a consequence, even powerful standard methods for deriving feature correspondences, e.g. the Scale Invariant Feature Transform (SIFT) (Lowe, 1999; Lowe, 2004), are not applicable for automatically detecting feature correspondences between the visual domain and the thermal domain.

For registration of image data representing information in these different spectral bands, different approaches have been proposed such as a segment-based approach (Coiras et al., 2000) or an approach involving normalized mutual information (Park et al., 2008). Furthermore, approaches for generating textured 3D point clouds often rely on a matching between images followed by forward projection to 3D space. For flat building façades and almost planar scenes, the transformation model of a homography (Hartley and Zisserman, 2008) can be applied. Such a transformation may for instance be exploited for mapping thermal IR information on existing building models (Hoegner et al., 2007b) or thermal 3D mapping involving a range camera (Weinmann et al., 2012). The latter approach relies on deriving feature correspondences between the respective images via shape-based matching, a respective image registration and a subsequent forward projection to 3D space. Considering mutual information between images has also been proposed for mapping multispectral texture information onto 3D models (Pelagotti et al., 2009) and for co-registration of intensity images and 3D LiDAR data (Parmehr et al., 2013).

2.2 Direct Co-Registration of Point Clouds and Thermal Information

One of the simplest approaches for directly acquiring colored 3D objects consists of using co-registered devices with known fixed

relative orientation (Pulli et al., 1998). If the relative orientation is unknown, the standard approach consists of the manual selection of tie points and a subsequent bundle adjustment based on the collinearity equations, which has recently been used for co-registering 3D point clouds and thermal infrared images (Hoegner et al., 2014).

For automatically estimating the relative orientation between the used devices, silhouette-based approaches may be applied, which focus on minimizing the error between the contour of an object in the image and the contour of the respective projected 3D model (Lowe, 1991). In particular, linear features are often used as they typically occur in man-made environments such as urban areas. For such scenes, lines can be extracted at edges in 2D imagery, and clusters of vertical and horizontal lines can be detected in the respective 3D point cloud. This allows a registration based on the matching of corresponding 2D and 3D linear features (Liu and Stamos, 2005; Liu and Stamos, 2012).

As interesting alternative, which could potentially be applicable also for co-registering thermal information and 3D point cloud data, a recent approach focused on the mapping of iPhone images onto LiDAR point clouds (Sirmacek et al., 2013). In this work, the extraction of shape features from iPhone images is based on conducting mean shift segmentation, nonlinear smoothing and applying steerable filters in different orientations. Further exploiting the metafile, the GPS position and the looking angle of the iPhone camera are available. These are used for defining the orientation of a virtual plane. The extraction of shape features from point cloud data is based on the perpendicular projection of LiDAR points onto the derived virtual plane, which yields a representation for those façade and roof points of a building which are the closest to the virtual plane. Extracting the alpha shape around the points, sharp corners of the alpha shape may be used as discriminative features. Finally, a co-registration of image and point cloud data is conducted by estimating an affine transformation based on the extracted features.

2.3 Direct Generation of Point Clouds from Thermal IR Images

In contrast to approaches involving range cameras, the direct generation of 3D models from thermal images via Structure-from-Motion (SfM) techniques has been proposed (Markov and Birk, 2007). Such image-based techniques may be applied for simultaneously recovering both 3D structure of the scene and the respective pose (i.e. position and orientation) of the camera (Szeliski, 2010). Furthermore, a thermal stereo system for recovering a 3D surface temperature map of the scene has been presented (Prakash et al., 2006) which is even able to estimate object depth within a dark environment. This system is based on exploiting isotherms (i.e. lines connecting points of equal temperature) and epipolar geometry (Hartley and Zisserman, 2008), whereby the epipolar constraints reduce the correspondence search space to the intersecting points between epipolar lines and isotherms.

3. METHODOLOGY

For thermal 3D mapping, we propose a methodology which automatically recovers the relative orientation between a range camera and a thermal camera (Figure 1). After conducting a radiometric correction (Section 3.1), a geometric calibration (Section 3.2) is applied in order to obtain undistorted 2D imagery. Subsequently, a shape-based technique for detecting feature correspondences between intensity images and thermal infrared images is

exploited (Section 3.3) and, due to the acquired range information, each 2D/2D correspondence can directly be assigned a respective 3D point which yields 3D/2D correspondences. Those correspondences containing unreliable range information are discarded via an image-based rejection strategy. The remaining correspondences can efficiently be used for co-registering 3D point cloud data and thermal information (Section 3.4), i.e. mapping the thermal information onto the 3D point cloud.

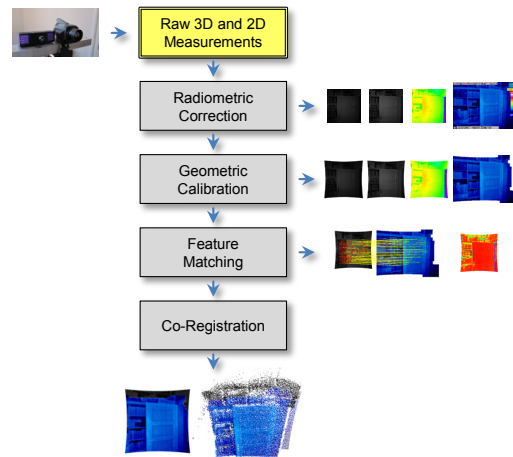


Figure 1: The components of our methodology.

3.1 Radiometric Correction

For range cameras, the captured intensity information corresponds to the energy of the laser light reaching the device. Due to internal processes such as conversion to a digital signal and signal amplification (which are not identical for different range cameras), this intensity information has to be adapted for applying standard image processing techniques. This is done by applying a histogram normalization of the form

$$I_n = \frac{I - I_{min}}{I_{max} - I_{min}} \cdot 255 \quad (1)$$

which adapts the intensity information I of each pixel to the interval $[0, 255]$ and thus yields gray-valued images. Here, the minimal and maximal intensity within all intensity measurements on the regular 2D grid are denoted as I_{min} and I_{max} , respectively. The histogram normalization can be conducted for both active and passive intensity images. For the example of an indoor scene, the recorded information is visualized in Figure 2.

For thermal cameras, there is no need to apply a histogram normalization as the captured thermal information of each pixel is already assigned a color value according to a certain colorbar. A respective visualization of thermal information for the provided example is depicted in Figure 3.

3.2 Geometric Calibration

For calibration, we follow a recent strategy focusing on the generation of infrared-textured 3D models (Weinmann and Jutzi, 2012). This strategy involves a geometric calibration for both range camera and thermal camera.

When using range cameras, it has to be taken into account that the captured range and intensity images are distorted (Figure 2). Consequently, a camera calibration has to be carried out for such



Figure 2: Visualization of the data captured with a PMD[vision] CamCube 2.0: Normalized active intensity image, normalized passive intensity image and range image (from left to right). The range is encoded in terms of a color scale reaching from red (near) via yellow, green, cyan and blue to violet (far).



Figure 3: Visualization of thermal information captured with an InfraTec VarioCAM hr.

devices. For this purpose, we assume that the geometric mapping of a scene onto the image plane can be described with a standard camera model representing the intrinsic behavior of a digital camera. This standard model considers both radial and tangential distortion (Brown, 1971; Heikkilä and Silvén, 1997). Accordingly, the geometric mapping can be parameterized with the focal lengths in x - and y -direction, the image coordinates (x_0, y_0) of the principal point, a skew coefficient s , and the image distortion coefficients describing radial and tangential distortion. The well-known standard methodology then consists of using a rectangular checkerboard pattern with known size, capturing respective intensity images with the device, extracting the grid corners of the checkerboard pattern in the intensity images, and finally applying the calibration procedure (Bouguet, 2010). For the example depicted in Figure 2, the derived undistorted images are provided in Figure 4. Subsequently, the respective 3D information in the local coordinate frame can be derived (Weinmann and Jutzi, 2012).



Figure 4: Visualization of the undistorted images for normalized active intensity, normalized passive intensity and range information (from left to right).

Since the checkerboard pattern is not visible in the thermal infrared domain, the calibration of thermal cameras is based on a planar testfield with lamps (Luhmann et al., 2010; Lagüela et al., 2011a; Borrmann et al., 2012b; Hoegner et al., 2014). The lamps are clearly visible in the thermal infrared images (Figure 5) and can thus easily be detected. Using a regular grid of lamps also allows for using the aforementioned standard calibration procedure for digital cameras (Figure 6).

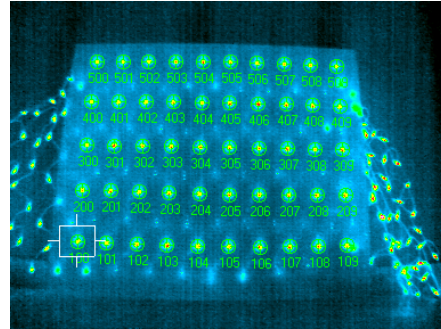


Figure 5: Captured thermal information for the planar testfield with lamps. The distortion of the thermal infrared image is clearly visible.

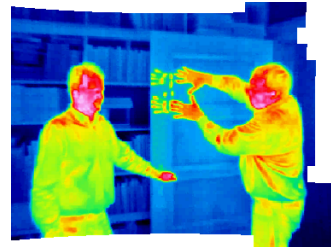


Figure 6: Visualization of the undistorted thermal infrared image.

3.3 Detection of 2D/2D and 3D/2D Correspondences

For deriving feature correspondences between the image representations of intensity and thermal infrared information, we use local features in terms of keypoints with characteristic descriptors. Due to the different characteristics of the compared images, commonly used keypoint descriptors obtained from algorithms such as SIFT (Lowe, 1999; Lowe, 2004) or SURF (Bay et al., 2011) fail in the automatic detection of point correspondences. Hence, we apply the shape-based matching algorithm proposed in (Steger, 2001) and (Ulrich, 2003) which matches image patches of a user-defined size by exploiting wavelength independent properties in terms of shape information. Thereby, the surrounding of a keypoint is described by a generated model. Thus, the algorithm is able to derive feature correspondences of which a certain percentage provides a high reliability.

In our case, a model image is generated by selecting quadratic areas of 100×100 pixels around points on a discrete grid with a spacing of 10 pixels in the intensity image provided by the range camera. The values for the size of the model image and the grid spacing were selected empirically (Weinmann et al., 2012). Subsequently, a Sobel filter is applied to the model image (Figure 7, left), and the associated gradient directions are determined for pixels with high gradient magnitude (Figure 7, center). Finally, the model image is matched to the gradients of the search image – which is represented by the thermal infrared image – by comparing the respective gradient directions (Figure 7, right). More specifically, a similarity measure is calculated which represents the normalized dot product of vector representations for the gradient directions of the transformed model and the search image (Ulrich, 2003), according to which a score can be obtained for each pixel in the search image. This similarity measure is robust in case of noise, illumination changes and partial occlusions, but not in case of changes in rotation and scale. Hence, the search

space is extended to a predefined range of rotations and scales. If the derived similarity measure is above a certain threshold, a point correspondence is detected (Figure 8). For each correspondence, the coordinates of the center, the rotation angle and the similarity measure itself are assigned. We use the HALCON 11 implementation (MVTec Software) of the shape-based matching algorithm for our experiments.

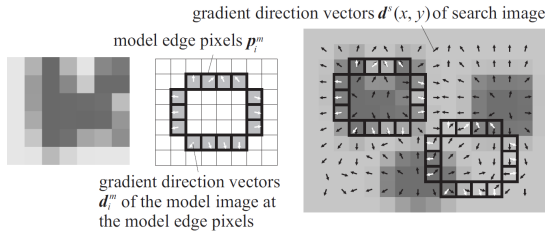


Figure 7: Principle of shape-based matching (Ulrich, 2003): model image (left), model edges (center) and search image (right).

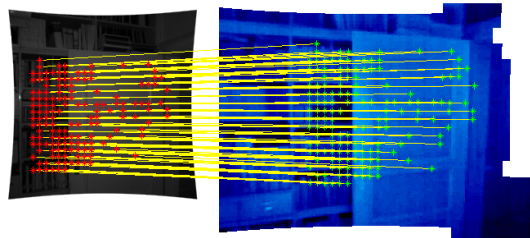


Figure 8: Automatically detected feature correspondences between an intensity image and a thermal infrared image for a static scene.

Since the range camera also provides range information in terms of a range image, we can easily assign each of the derived 2D/2D correspondences a respective 3D scene point whose coordinates are known in the local coordinate frame of the range camera. For estimating the relative orientation between the two utilized devices, we do not want to make strong assumptions on the scene structure, but rather exploit a general solution which relies on 3D/2D correspondences, where the 3D information is derived via the range camera and the 2D information is derived via the thermal camera. Depending on the surface properties of the respective objects, i.e. shape and reflectivity, the range camera provides more or less noisy range information. Note that many points which arise from objects in the scene will probably provide a smooth surface. However, range information of points along edges of the respective objects might be very noisy. In order to detect such noisy measurements, the standard deviation σ of all range values within a 3×3 neighborhood is calculated for each point on the regular 2D grid and used as a measure describing the reliability of the respective range information (Weinmann and Jutzi, 2011). Combining these constraints for all points on the 2D grid yields a 2D confidence map which is illustrated in Figure 9 for two exemplary scenes.

In order to remove unreliable feature correspondences, a simple thresholding is conducted. If the parameter σ of the respective range values is larger than a predefined threshold t_σ , the range information of the center pixel is assumed to be unreliable, otherwise the range information of the center pixel is assumed to be reliable. Following (Weinmann and Jutzi, 2012), we select this threshold to $t_\sigma = 0.05\text{m}$.

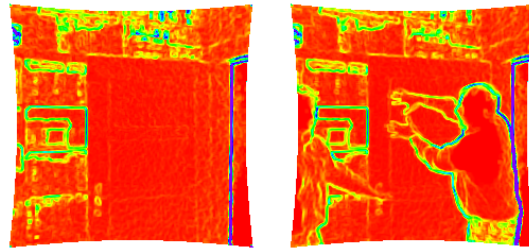


Figure 9: Confidence maps indicating the reliability of the respective range information. The reliability is encoded in terms of a color scale reaching from red (reliable) via yellow, green, cyan and blue to violet (unreliable).

3.4 Co-Registration of 3D and 2D Information

Once 3D/2D correspondences have been detected, the task of co-registering 3D and 2D information may be related to the well-known Perspective- n -Point (PnP) problem where the aim is to estimate the exterior orientation or pose of a camera from a set of n correspondences between 3D points \mathbf{X}_i of a scene and their 2D projections \mathbf{x}_i onto the image plane, where $i = 1, \dots, n$. In recent years, the Efficient Perspective- n -Point (EPnP) algorithm (Moreno-Noguer et al., 2007) has been proposed as a non-iterative method which provides an accurate solution to the PnP problem with only linear complexity. Compared to other approaches for solving the PnP problem, this algorithm is not only fast and accurate, but also designed to work with a large number of correspondences and it does not require an initial estimate.

The EPnP algorithm is based on the idea of expressing the n known 3D scene points \mathbf{X}_i as a weighted sum of four virtual and non-coplanar control points \mathbf{C}_j for general configurations. Denoting the weights as α_{ij} and introducing a superscript c which indicates coordinates in the camera coordinate frame, each 3D/2D correspondence provides a relation of the form

$$w_i \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{X}_i^c = \mathbf{K} \sum_{j=1}^4 \alpha_{ij} \mathbf{C}_j^c \quad (2)$$

where \mathbf{K} describes the calibration matrix. Considering the respective three equations, the scalar projective parameters w_i can be determined according to the third equation and substituted into the other two equations. Concatenating the two modified equations for $i = 1, \dots, n$ yields a linear equation system $\mathbf{M} \mathbf{x} = \mathbf{0}$, where \mathbf{x} contains the 3D coordinates of the four control points \mathbf{C}_j . For more details on the efficient solution of this equation system, we refer to the respective paper (Moreno-Noguer et al., 2007). Once both world coordinates and camera coordinates of the 3D points are known, the transformation parameters aligning both coordinate systems can be retrieved via standard methods involving a closed-form solution in the least-squares sense (Horn et al., 1988; Arun et al., 1987; Umeyama, 1991).

For a robust estimation in case of existing outlier correspondences, the RANSAC algorithm (Fischler and Bolles, 1981) represents the method of choice as it completely eliminates the influence of outlier correspondences which are not in accordance with the largest consensus set supporting the given transformation model (i.e. a 3D rigid-body transformation). Following the original implementation (Moreno-Noguer et al., 2007), the RANSAC-based EPnP scheme relies on selecting small, but not minimal subsets of seven correspondences for estimating the model parameters and checking the whole set of correspondences for consistent

samples. In comparison to minimal subsets, this further reduces the sensitivity to noise. In order to avoid testing all possible subsets, which would be very time-consuming, we exploit an efficient variant, where the number of iterations – which equals the number of randomly chosen subsets – is selected high enough, so that a subset including only inlier correspondences is selected with a certain probability p (Hartley and Zisserman, 2008).

In our case, we may assume that the local coordinate frame of the range camera represents the reference coordinate frame. Consequently, the derived 3D coordinates are known with respect to the reference frame. Together with the respective observed 2D image locations in the thermal infrared image, they form the required 3D/2D correspondences. Note that, for this reason, the matrix \mathbf{K} in Equation (2) refers to the thermal camera.

4. EXPERIMENTAL SETUP

For thermal 3D mapping, we use a sensor platform equipped with a range camera and a thermal camera (Figure 10), where a fixed relative orientation between the two devices is preserved.



Figure 10: Sensor platform equipped with a range camera (left) and a thermal camera (right).

4.1 Range Camera: PMD[vision] CamCube 2.0

Modern range cameras capture geometric information as well as radiometric information and thus complementary types of data in form of images, where the information is acquired simultaneously for all pixels on the regular 2D grid. Due to the high frame rates of more than 25 releases per second, range cameras also allow to capture dynamic scenes. For our experiments, the choice of the utilized device is motivated by a possible use in outdoor environments, where range cameras based on the use of structured light are not applicable. We hence use a Time-of-Flight camera of type PMD[vision] CamCube 2.0, which measures three features for each pixel: range, active intensity and passive intensity. Note that the active intensity depends on the illumination emitted by the sensor, whereas the passive intensity depends on the background illumination arising from the sun or other external light sources. The resulting images have a size of 204×204 pixels which corresponds to a field-of-view of $40^\circ \times 40^\circ$. This in turn represents an angular resolution of approximately 0.2° .

Due to the measurement principle of such Time-of-Flight cameras, the non-ambiguous range depends on the modulation frequency. A modulation frequency of 20MHz for instance corresponds to a non-ambiguous range of 7.5m. In order to overcome this range measurement restriction, image- or hardware-based unwrapping procedures have recently been introduced (Jutzi, 2009; Jutzi, 2012).

4.2 Thermal Camera: InfraTec VarioCAM hr

As thermal camera, we use a bolometer-based InfraTec VarioCAM hr which records in the wavelength interval from $7.5\text{-}14\mu\text{m}$

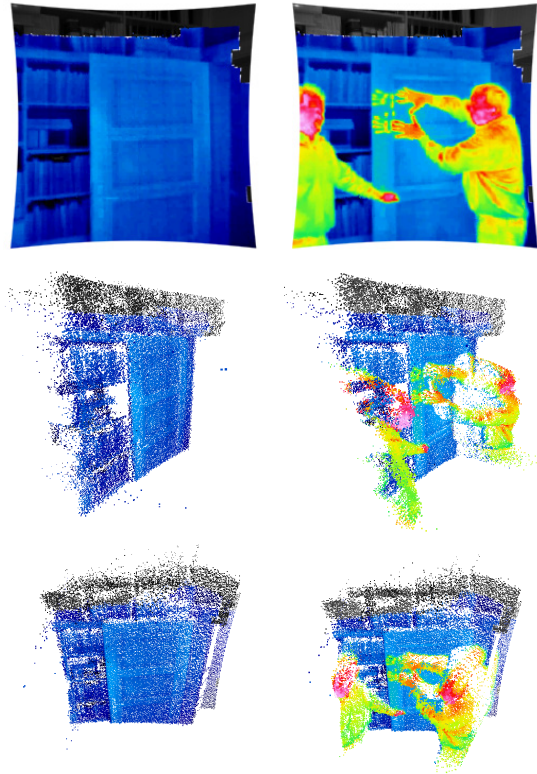


Figure 11: Visualization of thermal information mapped onto the respective 3D point cloud for two different scenes: 2D image projections (top) and colored 3D point clouds (center and bottom).

with a radiometric resolution of 0.05K. The captured thermal information is represented as images with a size of 384×288 pixels. The angular resolution is approximately 0.16° , and thus a captured thermal infrared image corresponds to a field-of-view of $61^\circ \times 46^\circ$. As the frame rate is 25fps, this device can also be applied for observing dynamic scenes.

5. EXPERIMENTAL RESULTS

For testing the proposed methodology, we only have to estimate the relative orientation between the utilized devices once, which is done for the example of a static scene (Figure 1). Afterwards, co-registered data can be recorded continuously when using a (hardware or software) trigger for synchronization, since the relative orientation between the utilized devices is known.

When using the sensor platform shown in Figure 10 for observing the static indoor scene, we obtain a total number of 151 point correspondences via shape-based matching as shown in Figure 8. Of these, 14 point correspondences are discarded since the respective range information is considered to be unreliable according to the applied rejection strategy (Figure 9, left). The remaining point correspondences are exploited by the RANSAC-based EPnP scheme, and the results in terms of thermal 3D mapping are shown in Figure 11 (left). Once the relative orientation is known, we can also conduct thermal 3D mapping by exploiting the known transformation parameters for other, possibly also dynamic scenes, e.g. as shown in Figure 11 (right).

6. DISCUSSION

The derived results (Figure 11) clearly reveal that the presented methodology is well-suited for thermal 3D mapping. In particular, the quality is rather high due to a robust mapping between image data with different characteristics, where standard approaches typically tend to fail in finding reliable feature correspondences. Only a few pixels at the edges between the right person and the background are observable as being erroneous.

Without making strong assumptions on the 3D scene structure and without human interaction for selecting corresponding features, our proposed methodology is capable to fully automatically perform thermal 3D mapping in dynamic scenes. Thus, the proposed methodology outperforms other recent approaches which, in turn, reveal limitations as they partially rely on human interaction (Hoegner et al., 2014) or as they are only tailored for planar scene structures (Weinmann et al., 2012). Note that – without loss of generality – the methodology could also be applied for co-registering 3D point cloud data and color information, where the latter is acquired with a digital camera. This may even be significantly easier since the contours in both intensity images of the range camera and color images of the digital camera are not as blurry as in thermal infrared images.

For the sake of clarity, we also want to demonstrate that the proposed methodology for thermal 3D mapping for instance allows to observe the evolution of a 3D scene over time. Considering the two scenes depicted in Figure 11 (center and bottom), the respective projections onto the image plane of the range camera (Figure 11, top) and further involving the undistorted intensity and range images acquired with the range camera (e.g. Figure 4) allows a categorization of (i) background where no significant changes with respect to intensity, range or thermal information are observed, (ii) people in the scene which can be observed from a change in intensity, range and thermal information, (iii) passive scene manipulation which is indicated only by a change of intensity information, and (iv) active scene manipulation caused by interactions between people and scene objects which is indicated only by a change of thermal information. Note that already the exploitation of thresholded difference images and a connection of logical operations allows a respective statement on change detection (Figure 12). From the respective rule-based classification results according to the four considered classes of changes (Figure 13), a small error in the co-registration becomes visible which can also be observed in Figure 11. Accordingly, a respective segmentation of people in the 3D point cloud becomes trivial.



Figure 12: Thresholded difference images for intensity, range and thermal information (from left to right). Changes are indicated in white. Note the noisy behavior of range measurements.

7. CONCLUSIONS

In this paper, we focused on thermal 3D mapping as an important prerequisite for object detection in dynamic scenes. For this purpose, we presented a fully automatic methodology which involves (i) a radiometric correction, (ii) a common geometric calibration procedure for both range camera and thermal camera, (iii)

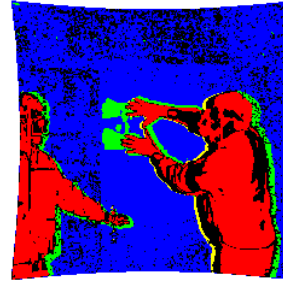


Figure 13: Results of rule-based classification based on logical operators: a clear distinction between background (blue), people (red), passive scene manipulation (yellow) and active scene manipulation (green) is possible. Note that passive and active scene manipulation are also detected at edges which indicates a small error in the co-registration.

a robust approach for detecting reliable feature correspondences between different image domains by exploiting wavelength independent properties as well as an image-based rejection strategy, and (iv) a co-registration of 3D point cloud data and thermal information based on an efficient and robust technique designed to work with a large number of 3D/2D correspondences. We have demonstrated that our methodology outperforms other recent approaches in terms of both applicability and accuracy, due to avoiding both human interaction and strong assumptions on the 3D scene structure. We have also provided an example for a possible use in order to observe the evolution of a 3D scene over time which can easily be conducted by involving straightforward techniques in terms of thresholding followed by rule-based classification.

For future research, it would be desirable to not only observe a static or dynamic scene with devices mounted on a static sensor platform, but also move the sensor platform in order to capture 3D environments of larger scale and extract complete 3D models. This could for instance be valuable for building diagnostics in both indoor and outdoor environments.

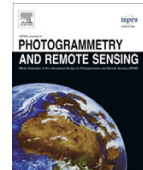
REFERENCES

- Alba, M. I., Barazzetti, L., Scaioni, M., Rosina, E. and Previtali, M., 2011. Mapping infrared data on terrestrial laser scanning 3D models of buildings. *Remote Sensing* 3(9), pp. 1847–1870.
- Arun, K. S., Huang, T. S. and Blostein, S. D., 1987. Least-squares fitting of two 3-D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9(5), pp. 698–700.
- Bai, X., Zhou, F. and Xue, B., 2011. Fusion of infrared and visual images through region extraction by using multi scale center-surround top-hat transform. *Optics Express* 19(9), pp. 8444–8457.
- Balaras, C. A. and Argiriou, A. A., 2002. Infrared thermography for building diagnostics. *Energy and Buildings* 34(2), pp. 171–183.
- Bay, H., Tuytelaars, T. and Van Gool, L., 2011. SURF: speeded up robust features. In: Leonardis, A., Bischof, H. and Pinz, A. (Eds.), *ECCV 2006, Part I. Lecture Notes in Computer Science*, Vol. 3951, Springer, Heidelberg, Germany, pp. 404–417.
- Borrmann, D., Afzal, H., Elseberg, J. and Nüchter, A., 2012a. Mutual calibration for 3D thermal mapping. *Proceedings of the 10th International IFAC Symposium on Robot Control*.
- Borrmann, D., Elseberg, J. and Nüchter, A., 2012b. Thermal 3D mapping of building façades. *Proceedings of the 12th International Conference on Intelligent Autonomous Systems*, pp. 173–182.
- Bouguet, J.-Y., 2010. Camera calibration toolbox for Matlab [online]. Computer Vision Research Group, Department of Electrical Engineering, California Institute of Technology, Pasadena, USA. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html (Accessed: 30 May 2014).

- Brown, D. C., 1971. Close-range camera calibration. *Photogrammetric Engineering* 37(8), pp. 855–866.
- Chen, S. and Leung, H., 2009. An EM-CI based approach to fusion of IR and visual images. *Proceedings of the 12th International Conference on Information Fusion*, pp. 1325–1330.
- Coiras, E., Santamaría, J. and Miravet, C., 2000. A segment-based registration technique for visual-IR images. *Optical Engineering* 39, pp. 282–289.
- Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), pp. 381–395.
- Hartley, R. I. and Zisserman, A., 2008. *Multiple view geometry in computer vision*. University Press, Cambridge, UK.
- Heikkilä, J. and Silvén, O., 1997. A four-step camera calibration procedure with implicit image correction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1106–1112.
- Hoegner, L., Kumke, H., Meng, L. and Stilla, U., 2007a. Automatic extraction of textures from infrared image sequences and database integration for 3D building models. *PFG Photogrammetrie, Fernerkundung, Geoinformation* 6 / 2007, pp. 459–468.
- Hoegner, L., Kumke, H., Schwarz, A., Meng, L. and Stilla, U., 2007b. Strategies for texturing building models with low resolution infrared image sequences. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI-5/C55.
- Hoegner, L., Roth, L., Weinmann, M., Jutzi, B., Hinz, S. and Stilla, U., 2014. Fusion von Time-of-Flight-Entfernungsdaten und thermalen IR-Bildern. *AVN - Allgemeine Vermessungs-Nachrichten* 5 / 2014, pp. 192–197.
- Horn, B. K. P., Hilden, H. M. and Negahdaripour, S., 1988. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A* 5(7), pp. 1127–1135.
- Iwasczuk, D., Hoegner, L. and Stilla, U., 2011. Detection of windows in IR building textures using masked correlation. In: Stilla, U., Rottensteiner, F., Mayer, H., Jutzi, B. and Butenuth, M. (Eds.), *Photogrammetric Image Analysis*, ISPRS Conference - Proceedings. Lecture Notes in Computer Science, Vol. 6952, Springer, Heidelberg, Germany, pp. 133–146.
- Jutzi, B., 2009. Investigations on ambiguity unwrapping of range images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-3/W8, pp. 265–270.
- Jutzi, B., 2012. Extending the range measurement capabilities of modulated range imaging devices by time-frequency multiplexing. *AVN - Allgemeine Vermessungs-Nachrichten* 2 / 2012, pp. 54–62.
- Lagüela, S., González-Jorge, H., Armesto, J. and Arias, P., 2011a. Calibration and verification of thermographic cameras for geometric measurements. *Infrared Physics & Technology* 54, pp. 92–99.
- Lagüela, S., Martínez, J., Armesto, J. and Arias, P., 2011b. Energy efficiency studies through 3D laser scanning and thermographic technologies. *Energy and Buildings* 43, pp. 1216–1221.
- Liu, L. and Stamos, I., 2005. Automatic 3D to 2D registration for the photorealistic rendering of urban scenes. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 137–143.
- Liu, L. and Stamos, I., 2012. A systematic approach for 2D-image to 3D-range registration in urban environments. *Computer Vision and Image Understanding* 116(1), pp. 25–37.
- Liu, Z. and Laganière, R., 2007. Context enhancement through infrared vision: a modified fusion scheme. *Signal, Image and Video Processing* 1(4), pp. 293–301.
- Lowe, D. G., 1991. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(5), pp. 441–450.
- Lowe, D. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision*, pp. 1150–1157.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.
- Luhmann, T., Ohm, J., Piechel, J. and Roelfs, T., 2010. Geometric calibration of thermographic cameras. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-5, pp. 411–416.
- Markov, S. and Birk, A., 2007. Detecting humans in 2D thermal images by generating 3D models. In: Hertzberg, J., Bectz, M. and Englert, R. (Eds.), *KI 2007: Advances in Artificial Intelligence*. Lecture Notes in Artificial Intelligence, Vol. 4667, Springer, Heidelberg, Germany, pp. 293–307.
- Moreno-Noguer, F., Lepetit, V. and Fua, P., 2007. Accurate non-iterative O(n) solution to the PnP problem. *Proceedings of the International Conference on Computer Vision*, pp. 1–8.
- Park, C., Bae, K.-H., Choi, S. and Jung, J.-H., 2008. Image fusion in infrared image and visual image using normalized mutual information. *Proceedings of SPIE*, Vol. 6968, pp. 69681Q–1–9.
- Parmehr, E. G., Fraser, C. S., Zhang, C. and Leach, J., 2013. Automatic registration of optical imagery with 3D lidar data using local combined mutual information. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-5/W2, pp. 229–234.
- Pelagotti, A., Mastio, A. D., Uccheddu, F. and Remondino, F., 2009. Automated multispectral texture mapping of 3D models. *Proceedings of the 17th European Signal Processing Conference*, pp. 1215–1219.
- Prakash, S., Pei, Y. L. and Caelli, T., 2006. 3D mapping of surface temperature using thermal stereo. *Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision*, pp. 1–4.
- Pulli, K., Abi-Rached, H., Duchamp, T., Shapiro, L. G. and Stuetzle, W., 1998. Acquisition and visualization of colored 3D objects. *Proceedings of the 14th International Conference on Pattern Recognition*, pp. 11–15.
- Sirmacek, B., Lindenbergh, R. C. and Menenti, M., 2013. Automatic registration of iPhone images to laser point clouds of urban structures using shape features. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-5/W2, pp. 265–270.
- Steger, C., 2001. Similarity measures for occlusion, clutter, and illumination invariant object recognition. In: Radig, B. and Florczyk, S. (Eds.), *Pattern Recognition, DAGM 2001*. Lecture Notes in Computer Science, Vol. 2191, Springer, Heidelberg, Germany, pp. 148–154.
- Szeliski, R., 2010. *Computer vision: algorithms and applications*. Springer, New York, USA.
- Toth, C., Ju, H. and Grejner-Brzezinska, D., 2011. Matching between different image domains. In: Stilla, U., Rottensteiner, F., Mayer, H., Jutzi, B. and Butenuth, M. (Eds.), *Photogrammetric Image Analysis*, ISPRS Conference - Proceedings. Lecture Notes in Computer Science, Vol. 6952, Springer, Heidelberg, Germany, pp. 37–47.
- Ulrich, M., 2003. Hierarchical real-time recognition of compound objects in images. Dissertation, German Geodetic Commission (DGK), Vol. C.
- Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(4), pp. 376–380.
- Vidas, S., Moghadam, P. and Bosse, M., 2013. 3D thermal mapping of building interiors using an RGB-D and thermal camera. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2311–2318.
- Weinmann, M., 2013. Visual features – From early concepts to modern computer vision. In: Farinella, G. M., Battiato, S. and Cipolla, R. (Eds.), *Advanced Topics in Computer Vision*. Advances in Computer Vision and Pattern Recognition, Springer, London, UK, pp. 1–34.
- Weinmann, M. and Jutzi, B., 2011. Fully automatic image-based registration of unorganized TLS data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-5/W12, pp. 55–60.
- Weinmann, M. and Jutzi, B., 2012. A step towards dynamic scene analysis with active multi-view range imaging systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX-B3, pp. 433–438.
- Weinmann, M., Hoegner, L., Leitloff, J., Stilla, U., Hinz, S. and Jutzi, B., 2012. Fusing passive and active sensed images to gain infrared-textured 3D models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX-B1, pp. 71–76.
- Xue, Z., Blum, R. S. and Li, Y., 2002. Fusion of visual and IR images for concealed weapon detection. *Proceedings of the International Conference on Image Fusion*, pp. 1198–1205.
- Yao, F. and Sekmen, A., 2008. Multi-source airborne IR and optical image fusion and its application to target detection. *Proceedings of the 4th International Symposium on Advances in Visual Computing*, pp. 651–660.
- Zitová, B. and Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing* 21(11), pp. 977–1000.

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Fast and automatic image-based registration of TLS data

Ma. Weinmann*, Mi. Weinmann, S. Hinz, B. Jutzi

Karlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing, Kaiserstr. 12, 76128 Karlsruhe, Germany

ARTICLE INFO

Article history:

Available online 13 October 2011

Keywords:

Laser scanning
TLS
Point cloud
Image
Registration
Automation

ABSTRACT

The fast and automatic registration of laser scanner data is of great interest in photogrammetric research. Recent developments show that for registration purposes characteristic 3D points can be extracted from the measured laser data, where the data are also represented as image. In this paper, radiometric and geometric information derived from TLS data are utilized for estimating the transformation parameters between two unregistered point clouds. After the extraction of characteristic 2D points based on *SIFT* features, these points are projected into 3D space by using interpolated range information. From these 3D conjugate points and their corresponding 2D projections onto a virtual plane 3D-to-2D correspondences are established. The fast, accurate and robust *RANSAC*-based registration scheme including the *EPnP* algorithm provides a framework to estimate the coarse transformation parameters from these 3D-to-2D correspondences. The coarse estimates are further refined by a single step outlier removal to gain a higher accuracy by introducing additional geometric constraints. These new constraints are based on 3D-to-3D correspondences which are much stronger than the 3D-to-2D correspondences alone. It will be shown that the presented approach is successfully applied to a benchmarked data set with millions of points resulting in a fast and accurate estimation of the transformation parameters with a processing speed of several seconds on a standard PC and an accuracy in the low centimeter range.

© 2011 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.

1. Introduction

The automatic generation of accurate 3D models from laser scanning data for a detailed description of objects is of great interest in photogrammetric research (Wehr and Lohr, 1999; Lichti et al., 2008). The description of objects requires a sampling of the surface as a point cloud as dense and complete as possible. Terrestrial laser scanners (TLSs) provide a dense and accurate three-dimensional representation of object surfaces in the local environment. In most cases, one scan is not sufficient to obtain a full scene coverage, and therefore, partially overlapping point clouds with high spatial information are typically collected from different locations via multiple scans. As each of the captured data sets refers to its own local coordinate frame, usually a registration step transforming all data into a common reference frame has to be carried out.

This registration can either be done fully automatically or by placing artificial markers into the scene, which provide clearly demarcated corresponding points in the different scans, and a successive manual alignment. For a large number of scans, such a manual alignment is very time-consuming and therefore, it is desirable to use a fully automatic registration. The automatic registration is usually divided into a coarse registration, which

leads to a pre-alignment of the scans, and a fine registration, which increases the accuracy of the estimated parameters. For fine registration, the *Iterative Closest Point (ICP)* algorithm and its variants, being described in Besl and McKay (1992) and Rusinkiewicz and Levoy (2001), are the commonly used standard approaches. The *ICP* algorithm iteratively estimates the transformation parameters between two partially overlapping point clouds and converges to a local minimum, but a good a priori alignment of the point clouds is required in order to reach an appropriate solution. Due to the iterative processing of the usually large number of points, the algorithm shows a high computational effort.

In general, the registration approaches use distinctive features to recover the translation and rotation parameters for the alignment of two point clouds, as such features may provide additional and useful information. These features have to be extracted from the TLS data and a whole variety of feature types may be considered. In the following, the prior works will be distinguished by range-based approaches (Section 1.1) and image-based approaches (Section 1.2).

1.1. Related work focusing on range-based registration

Characteristic point features are extracted in Barnea and Filin (2008) by using the so-called min-max operator adapted to panoramic range images, and thus to 3D point clouds being

* Corresponding author.

E-mail address: martin.weinmann@kit.edu (Ma. Weinmann).URL: <http://www.ipf.kit.edu> (Ma. Weinmann).

projected onto a 2D plane showing the distance of each point, which yields features being located at corners. The resulting subset of points with distinctive characteristics in their surrounding areas is then used for a RANSAC-based matching of the point clouds which, however, only leads to a coarse estimation of the transformation parameters. Hence, an additional ICP-based refinement starting with this estimate is utilized.

Instead of point features, line features can also be used for the registration process like in Stamos and Leordeanu (2003). Therefore, planar surfaces are extracted from the point clouds and the intersection of neighboring planar regions yields line features. Moreover, line features representing the boundaries of the recovered planes are also taken into consideration and, finally, the transformation parameters between two point clouds are calculated from at least two corresponding line pairs.

According to Dold and Brenner (2004), the transformation parameters between two point clouds can also be calculated based on three plane matches. A comparison of this method to a method which is based on a modification of the *Normal Distributions Transform (NDT)* is carried out in Brenner et al. (2008). Whereas the latter one is conceptually simple and therefore fast, the approach based on planar patches tends to be more accurate and works for larger distances between the scans. In Dold (2005), the normal vectors of local or segmented planes of the TLS data are projected onto a unit sphere and clustered through their tessellation which yields extended Gaussian images. These are used to recover the rotation parameters between different scans. Further improvements to reduce the search area by excluding implausible correspondences based on geometric constraints are described in Dold and Brenner (2006) as well as the use of a calibrated hybrid sensor system in order to get textured patches. The additional texture information is used for determining the translation parameters between two scans via correlation.

With some additional assumptions made in von Hansen (2006), single plane correspondences are sufficient to recover the transformation parameters between two scans. So single surface elements are extracted from the TLS data and grouped to planes. Then, a coarse registration is carried out returning a set of corresponding planes and coarse parameter estimates, which are finally refined by using a least-squares adjustment over all scan positions.

In some applications, it is feasible to extract several different geometric primitives from the TLS data. Scenes containing industrial installations offer a variety of features like planes, spheres, cylinders and tori which are used in Rabbani et al. (2007). Once such primitives have been extracted from the point clouds, the registration is carried out via corresponding object models in different scans. This indirect method, however, leads to a coarse estimate of the transformation parameters and therefore, the successive stage consists of a direct method extending the idea of bundle adjustment for global registration. As the refinement stage estimates the transformation parameters of all scans simultaneously as well as the shape parameters of all the objects used for registration, an accumulation of errors resulting from pairwise registration is avoided.

Extracting line features, plane patches or geometric objects from TLS data, however, can be a quite difficult task in the presence of cluttered scenes and variable point density. Therefore, other approaches use local descriptors representing surface patches. In Bae and Lichti (2004) for example, the change of geometric curvature and approximate normal vector of the local surface formed by a point and its neighborhood are used for the registration of two partially overlapping point clouds yielding only small registration errors without the need of a good a priori alignment. Modifications based on these features and neighborhood search are presented in Bae and Lichti (2008) and include the radial and angular uncertainty of laser scanner measurements. However, the pro-

posed *Geometric Primitive ICP (GP-ICP)* and its extension using the RANSAC algorithm require an initial alignment.

1.2. Related work focusing on image-based registration

Recent developments show that local features can also be extracted from the point clouds by using images which are usually acquired simultaneously with the range data by recording either co-registered camera images or reflectance images representing the energy of the backscattered laser light of an active sensor. These images provide additional information about the local area around the scanner which might not always be represented in the range measurements. The fusion of range data and 2D imagery is investigated in Forkuo and King (2004). There, three different approaches are presented. The first one is based on using the 3D range data and synthetic camera images (SCIs) which are generated from the respective reflectance values via backprojection into a regular grid. By contrast, the second approach is based on comparing synthetic camera images to real camera images (RCIs) acquired with a digital camera from a different viewpoint. This involves feature detection as well as feature matching. Finally, a third approach is presented for relating each pixel in the 2D camera image to the corresponding 3D point in space which can be used to produce a photo-realistic model. These approaches show that the registration of two point clouds can be supported by using reliable features of the images but, in general, it has to be taken into account that the images might be based on polar geometry instead of central projection. As the features have to be robust in case of changes in viewpoint resulting in changes of scale, rotation and various image distortions as well as changes in illumination, features resulting from the *SIFT* algorithm described in Lowe (2004) are considered to be well-suited and therefore widely used. These features are based on a close approximation to the scale-normalized Laplacian-of-Gaussian function (Lindeberg, 1994) and, according to Mikolajczyk (2002), the extrema of this function produce the most stable feature points in 2D images.

SIFT features have been introduced for the registration of TLS data in Bendels et al. (2004) where both range information and information from co-registered camera images are used. Combining the information of *SIFT* features extracted from the color images and the corresponding surface elements extracted from the range data leads to special tie points representing the centers of gravity of the respective surface elements. The calculated points are used for registration which is divided into a coarse registration using a least-squares alignment of the point clouds and a fine registration by an ICP-based alignment of the feature surface elements. Finally, a graph relaxation algorithm is presented for multiview registration.

In Al-Manasir and Fraser (2006) as well, a digital camera mounted on the TLS is used for image-based registration. The approach is based on the extraction of *SIFT* features as reliable feature points, the model of coplanarity for calculating the relative orientation between two images and the Helmert transformation for determining the relationship between two scans. Assuming coplanarity, however, is not possible in the case of panoramic images.

Such panoramic images are used for supporting the registration process in Barnea and Filin (2007). There, *SIFT* features being extracted with respect to image information are projected into the local 3D space by using range information. Subsequently, the rotation and translation parameters between the scans are estimated using the RANSAC algorithm.

However, traditional *SIFT* features can also be modified via image rectification which is presented in Seo et al. (2005) and reduces effects of various distortions caused by changes in viewpoint and illumination.

The transfer of the *SIFT* algorithm to reflectance images derived directly from the reflectance values of the TLS data is proposed in Böhm and Becker (2007). There, the features are filtered with respect to both their distinctiveness and their geometric relationship, and the remaining feature points are used as tie points for registration.

Reflectance images are also used in Wang and Brenner (2008), where the *SIFT* feature descriptor is modified with additional entries. Each entry added represents the mean geometric curvature of a surface formed by the projection of a 2D *SIFT* feature into local 3D space and its 3D neighborhood in the point cloud. The transformation parameters are then estimated by using the *RANSAC* algorithm and can be refined by using the *ICP* algorithm.

An approach including both pairwise and global registration is presented in Kang et al. (2009). Based on corresponding *SIFT* features, the respective 3D points and the removal of outliers by a geometric invariance check, a least-squares adjustment of the point clouds is carried out. The following step for global registration utilizes bundle adjustment strategies in form of self-closure constraints.

1.3. Contribution of this paper

In this paper, we propose a method for fast and automatic image-based registration of terrestrial laser scanning data. For estimating the transformation parameters between two point clouds, radiometric and geometric information are utilized. After the extraction of characteristic feature points from the reflectance images of two scans, these features are projected into 3D space using range information interpolated from surrounding points. The projection of these 3D points onto virtual planes allows for a fast and fully automatic registration which utilizes both 2D and 3D observations of the respective point clouds. In order to detect outliers, a refinement based on the idea of using the Euclidean distances between the matched point pairs in 3D space is applied. The presented approach cuts down the processing speed for the registration of large point clouds, consisting of millions of points, to a few seconds on a standard computer. Furthermore, the fully automatic approach handles a marker-less registration without any prior knowledge of the scene.

The paper is organized as follows. In Section 2, the scheme of the above briefly mentioned methodology is presented in detail. The capability of the proposed approach is proved by processing 11 point clouds of an urban environment in Section 3. In Section 4, the received transformation parameters for the registration are discussed concerning accuracy, reliability and performance. Conclusions and suggestions for future work are outlined in Section 5.

2. Methodology

The general overview of the registration approach proposed in this paper is illustrated in Fig. 1. First, from the two unregistered point clouds PC_n and PC_{n+1} the corresponding reflectance images I_n and I_{n+1} and range images R_n and R_{n+1} have to be generated. Then, *SIFT* features are extracted from both reflectance images and only the matching features, i.e. the homologous points, are considered for the further processing. As these features are located with sub-pixel accuracy and only each pixel represents a measured 3D point around the scan position, a bilinear transformation based on the information of a local neighborhood is carried out in order to assign each 2D location of a *SIFT* feature the corresponding 3D location. This is done for I_n and R_n as well as for I_{n+1} and R_{n+1} . The resulting 3D points are the conjugate points $CP_{i,n}$ and $CP_{i,n+1}$.

Furthermore, a step called *Projective Plane Projection (PPP)* which yields 2D backprojections $\mathbf{x}_{i,n}$ and $\mathbf{x}_{i,n+1}$ of the conjugate points $CP_{i,n}$ and $CP_{i,n+1}$ onto virtual planes P_n and P_{n+1} is introduced.

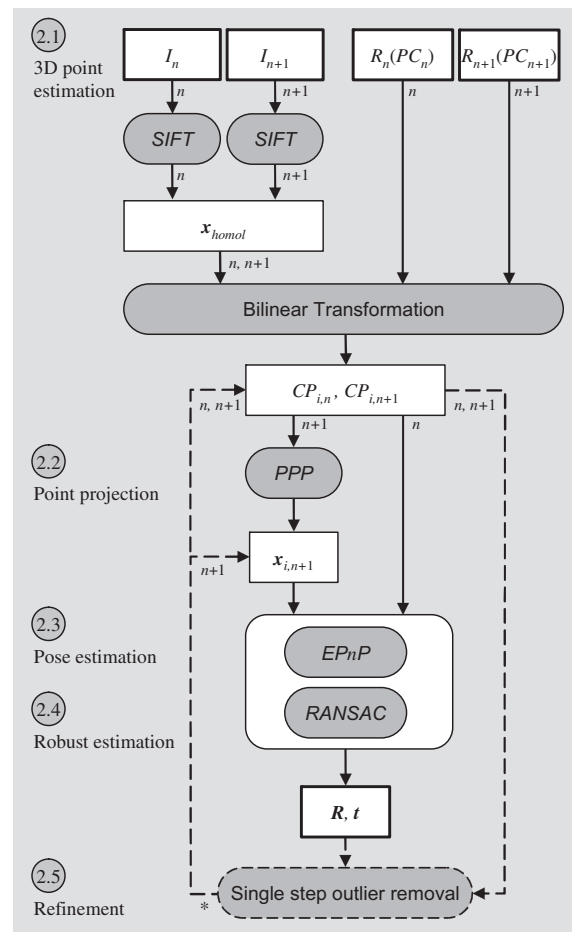


Fig. 1. Processing chain.

Together with the conjugate points $CP_{i,n}$, the backprojections $\mathbf{x}_{i,n+1}$ are then used for a pre-alignment of both scans involving the *EPnP* algorithm combined with the *RANSAC* algorithm.

The coarse estimate allows for a propagation of the conjugate points $CP_{i,n}$ and therefore, both sets of conjugate points can be transformed into one common coordinate frame. In order to refine the previous estimate, the geometric distance between matching 3D points is considered and outlying correspondences are removed from both sets of conjugate points resulting in their modifications $CP_{i,n}^*$ and $CP_{i,n+1}^*$ and the respective 2D backprojections $\mathbf{x}_{i,n}^*$ and $\mathbf{x}_{i,n+1}^*$. As before, 3D-to-2D correspondences are built from the conjugate points $CP_{i,n}^*$ and the corresponding backprojections $\mathbf{x}_{i,n+1}^*$. These 3D-to-2D correspondences are of a significant better quality than those utilized in the coarse estimation. Hence, using the *EPnP* algorithm combined with the *RANSAC* algorithm yields a more accurate result for the estimated transformation parameters.

The basic ideas of the more complex algorithms used here are described in the following paragraphs.

2.1. 3D point estimation using SIFT

The *Scale Invariant Feature Transform (SIFT)* introduced in Lowe (1999) and later improved in Lowe (2004) can be utilized for detecting distinctive keypoints in an image and extracting local feature descriptors which are invariant to image scaling and image

rotation, and robust with respect to image noise, changes in illumination and small changes in viewpoint. With these descriptors it is possible to locate correspondences between different images and, finally, to derive common image objects. Therefore, the *SIFT* algorithm can be applied to images derived from TLS data. In general, the computation of *SIFT* features is carried out in four major steps:

- **Scale-space extrema detection**

At first, the image I is convolved with Gaussian kernels of variable scale, reduced in size and again convolved with Gaussian kernels of variable scale in order to build the Gaussian scale-space. The blurred images are grouped according to their size and sorted according to their scale. Subtracting neighboring images yields the scale-space of Gaussian differences and stable keypoint locations can then be determined by comparing each sample point to its eight neighbors in the same scale and its nine neighbors in both neighboring scales. Resulting from this, local extrema are selected as keypoint candidates.

- **Keypoint localization**

Once keypoint candidates have been detected, the next step is to improve their location to subpixel accuracy and to remove keypoint candidates with a low stability.

- **Orientation assignment**

Based on the local image gradient directions of the closest smoothed image, each remaining keypoint is assigned a characteristic orientation.

- **Generation of keypoint descriptors**

The descriptor of each remaining keypoint is also derived from the local image gradients, but in order to achieve rotational invariance, this gradient information is aligned with the assigned keypoint orientation.

For two images of a scene, corresponding *SIFT* features can then be detected by comparing the Euclidean distances of a keypoint descriptor in the first image to the nearest neighbor and to the second nearest neighbor in the second image. The ratio of these distances has to be below a given threshold t , which can vary between 0 and 1. In order to get reliable correspondences, distinctive features have to be detected and therefore, a threshold of $t = 0.66$ is used. This means that the distance of a descriptor belonging to a *SIFT* feature in the first image to the nearest neighbor in the second image is only about $2/3$ of the distance to the second nearest neighbor. Using a higher threshold means that the feature descriptors of the nearest and second nearest neighbor in the second image may be more similar and thus, the considered feature in the first image is less distinctive. So a higher threshold usually leads to additional, but less reliable correspondences.

As these homologous points are located with subpixel accuracy, the corresponding 3D location has to be determined via nearby range measurements. Therefore, a bilinear interpolation is carried out in order to derive 3D points which are not necessarily equal to the measured points within a given scan grid. If the range information is not available for all four neighboring points in the regular scan grid, the corresponding point pair is discarded. The resulting 3D points for two successive scans are introduced as conjugate points $CP_{i,n}$ and $CP_{i,n+1}$.

2.2. Projective Plane Projection

Unlike the transformation used by the laser scanner, which maps each measured 3D point with its reflectance information

onto a cylinder or a sphere, the *Projective Plane Projection (PPP)* is a camera-like projection, mapping a set of 3D points \mathbf{X}_i , which are actually the conjugate points in our case, onto points \mathbf{x}_i of a virtual plane. Hence, this mapping can formally be described via a projective projection model of the form

$$\mathbf{x}_i = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X}_i, \quad (1)$$

where the matrix \mathbf{K} is the calibration matrix of a virtual camera. The matrix \mathbf{R} and the vector \mathbf{t} describe the rotation and the translation of the virtual camera with respect to the local coordinate system of the laser scanner. In the registration process, the matrix \mathbf{R} is chosen so that the virtual camera has the same orientation as the laser scanner with respect to the local scan and that the camera looks into the horizontal direction. As it can be assumed that there is no difference between the position of the virtual camera and the laser scanner, the translation vector is set to $\mathbf{t} = \mathbf{0}$. Using a virtual plane means that any parameters can be used for the focal lengths in x - and y -direction as well as for the coordinates of the principal point. Therefore, the image plane has not necessarily to be limited on a finite area. Instead of creating a synthetic camera image, as it is done in Forkuo and King (2004), and then extracting features from this synthetic camera image, only subpixel accurate 2D projections of specific 3D points are considered. Thus, the virtual plane only serves for getting 2D observations of 3D points. Expanding the virtual plane to infinity leads to visible 3D points within the whole area in front of the virtual camera. All points behind the camera may also be included in this approach as they represent the 2D projections onto the virtual plane of a second camera looking in the opposite direction. These 2D projections are mapped onto the first virtual plane via symmetric constraints.

2.3. Laser scanner pose estimation using EPnP

In Moreno-Noguer et al. (2007) and Lepetit et al. (2009), the *Efficient Perspective-n-Point (EPnP)* algorithm has been proposed as a non-iterative solution to the *Perspective-n-Point (PnP)* problem where the aim is to estimate the exterior orientation or pose of a camera from a set of n correspondences between 3D points \mathbf{X}_i of a scene and their 2D projections \mathbf{x}_i onto the image plane. Showing a computational complexity of $O(n)$, using a system with both linear and quadratic equations and being applicable for both planar and non-planar configurations, the *EPnP* algorithm is much faster and even more accurate than other non-iterative methods and much faster than iterative methods with only little loss in accuracy. Furthermore, this algorithm is less sensitive to noise and it does not require an initial estimate.

The *EPnP* algorithm is based on the idea of expressing the n 3D scene points \mathbf{X}_i whose coordinates are known in the world coordinate system as a weighted sum of four virtual and non-coplanar control points \mathbf{C}_j for general configurations, which means that

$$\mathbf{X}_i = \sum_{j=1}^4 \alpha_{ij} \mathbf{C}_j \quad (2)$$

with

$$\sum_{j=1}^4 \alpha_{ij} = 1 \quad (3)$$

holds true, where the α_{ij} are homogeneous barycentric coordinates. As, in the world coordinate system, the points \mathbf{X}_i^w are known as well as the control points \mathbf{C}_j^w , the parameters α_{ij} are known for the same equation in camera coordinates. Hence, the points \mathbf{X}_i^c can be expressed via the control points \mathbf{C}_j^c which leads to

$$\mathbf{w}_i \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{X}_i^c = \mathbf{K} \sum_{j=1}^4 \alpha_{ij} \mathbf{C}_j^c \quad (4)$$

for $i = 1, \dots, n$, where w_i are scalar projective parameters and \mathbf{K} describes the camera matrix. This relationship between 3D points and 2D points is illustrated in Fig. 2. Expanding the previous expression yields

$$w_i \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^4 \alpha_{ij} \begin{bmatrix} X_j^c \\ Y_j^c \\ Z_j^c \end{bmatrix} \quad (5)$$

with the focal length coefficients f_x and f_y , the coordinates x_0 and y_0 of the principal point and the coordinates $[X_j^c, Y_j^c, Z_j^c]^T$ of the control points. This linear system has 12 unknown parameters from the control points and, additionally, n unknown parameters w_i . As can be seen from the third row,

$$w_i = \sum_{j=1}^4 \alpha_{ij} Z_j^c \quad (6)$$

holds true which can be substituted into the other two rows. Concatenating the resulting equations for all n 3D-to-2D correspondences yields a linear system

$$\mathbf{M}\mathbf{x} = \mathbf{0} \quad (7)$$

with $\mathbf{x} = [\mathbf{C}_1^t, \mathbf{C}_2^t, \mathbf{C}_3^t, \mathbf{C}_4^t]^T$ and a $2n \times 12$ matrix \mathbf{M} . The solution \mathbf{x} then leads to the camera coordinates \mathbf{X}_i^c of the 3D points. Once the world coordinates and the camera coordinates of the 3D points are known, the rotation and translation parameters aligning both coordinate systems can be retrieved via standard methods (Horn et al., 1988).

Thus, only the coordinates of the control points in the camera coordinate system have to be estimated as they are directly linked with the coordinates of all 3D points which is the main reason for the efficiency of the *EPnP* algorithm. A further optimization step using the Gauss–Newton algorithm is described in Lepetit et al. (2009).

2.4. Robust estimation using RANSAC

The data used for testing is derived from measurements and so the quality of the used 3D-to-2D correspondences depends on the scanner accuracy with respect to both distance and angular resolution. Hence, the performance increases if only the reliable 3D-to-2D correspondences are used instead of all correspondences. For this purpose, the RANSAC algorithm presented in Fischler and Bolles (1981) is adapted to the *EPnP* algorithm. In this step, it is important to notice that the distinctive 3D conjugate points $CP_{i,n}$ are detected via *SIFT* features in the reflectance image I_n , whereas

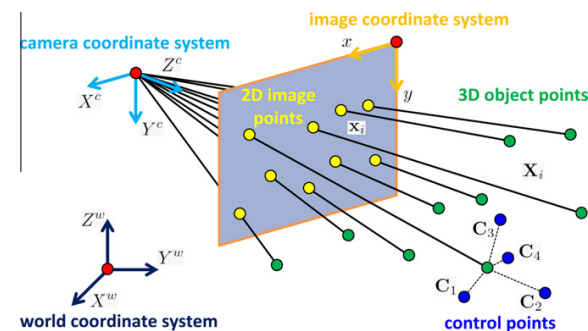


Fig. 2. Relationship between 3D points and the respective 2D projections onto the image plane.

the observed 2D backprojections of these distinct 3D points are derived from projecting the matching 3D conjugate points $CP_{i,n+1}$ onto the virtual plane P_{n+1} . The geometrical quality of corresponding 3D points in both point clouds has to be considered separately.

Generally, the RANSAC algorithm iteratively estimates the parameters of a mathematical model from a data set including outliers which do not fit to this model. Instead of using all the data available to get an initial solution and then attempting to eliminate the outliers, the RANSAC algorithm starts with a minimal subset of the data. This subset is sufficient to estimate the model parameters and randomly chosen, and then it is enlarged with consistent sample points when possible. According to Moreno-Noguer et al. (2007), a small, but not minimal subset of seven correspondences is randomly selected to reduce the sensitivity to noise. In order to avoid testing all possible subsets, which would be time-consuming, an efficient variant described in Hartley and Zisserman (2008) is used. There, the number of iterations which equals the number of randomly chosen subsets is selected high enough, so that a subset including only inliers is selected with a certain probability p .

2.5. Single step outlier removal

When dealing with measurements and applying the *SIFT* algorithm on the reflectance images, it seems quite feasible to check if all detected correspondences are really suitable with respect to the corresponding 3D points or if there are uncertainties concerning some of the correspondences. In the registration process, such uncertainties might arise from *SIFT* features being detected in the reflectance images at periodic shapes of façades which are likely to occur in urban environments. As a threshold value of $t = 0.66$ is used for a reliable matching of *SIFT* features, a feature detected in one image is very similar to only one feature detected in the other image. Despite this visual similarity, however, periodically appearing features could lead to obviously wrong correspondences in case of the local 3D geometry around the scan positions and so this kind of wrong correspondences has to be excluded from the set of all correspondences. This can be done by introducing geometric constraints based on the 3D distances between the propagated conjugate points $CP_{i,n}$ and $CP_{i,n+1}$.

The propagation transforms both sets of conjugate points into the world coordinate system. As the transformation matrix containing \mathbf{R} and \mathbf{t} with respect to the world coordinate system is assumed to be already known for the point cloud PC_n , the world coordinates of the conjugate points $CP_{i,n}$ are also known. For calculating the world coordinates of the points $CP_{i,n+1}$, the transformation matrix used is derived directly from the rotation and translation parameters which are estimated in the coarse registration step with respect to the world coordinate system. Thus, a comparison of the 3D points with respect to their 3D distances is possible. Especially large distances caused by erroneous correspondences have to be discarded as they influence the robustness of the *EPnP* algorithm and the time needed to find a useful solution.

Hence, the border between smaller and larger distances has to be determined and all matching 3D-to-2D correspondences with a large distance between their 3D points have to be rejected. To get the most promising correspondences for further calculations, a criterion based on the distribution of the distance between matching 3D points is introduced. Sorting the matching points with respect to their 3D distances, building a set consisting of the third of the smallest distances and iteratively adding the correspondence with the next greater distance allows for checking whether the new sample takes a strong influence on the mean value of the set which is done by considering the first derivative of

$$q(i) = \frac{\text{mean}_{\text{actual set}(i)}}{\text{mean}_{\text{complete set}}} \quad (8)$$

3. Experiments

In the following, the TLS data (Section 3.1) used for testing the presented approach to pairwise registration is described as well as the registration results (Section 3.2).

3.1. Data set

For processing and evaluating the proposed approach, the benchmark TLS data set from the University of Hanover is used which has been acquired at a district in Hanover called *Holzmarkt*. The complete data set was obtained by using a Riegl LMS-Z360i scanner and consists of several scans being captured from various positions. It includes the 3D coordinates of object points as well as the corresponding reflectance information. According to Wang and Brenner (2008), the laser scanner covers a field of view of 360 degrees in the horizontal direction and only 90 degrees in the vertical direction up to a range of approximately 200 m for each scan. The angular resolution of the TLS is 0.12 degrees and a measurement accuracy of 12 mm can be expected. Hence, a single scan returns 2.25 million 3D points from a regular scan area of 3000×750 points being represented as panoramic reflectance image. The provided reference values for the rotation and translation parameters of the TLS are based on a manual alignment using artificial targets which were placed in the scene. Thus, an accuracy in the low millimeter range can be expected for the reference positions.

3.2. Pairwise registration

As illustrated in Fig. 3, the scans recorded at 11 different locations were used and the distance between successive scanner positions is approximately 5 m. The figure shows the single scans after registration into a common coordinate system using the reference data. However, only a few points which are determined via *SIFT* features in the reflectance images are used for registration. To illustrate the density of these points, Fig. 4 shows the projections of the 2D *SIFT* features of each scan to 3D space separating neighboring scans by color. As a result from this nadir view onto the scene, the rough structure of the scene can be interpreted showing façades of houses and the coarse direction of streets. In the area between these plain structures, features from static objects or moving people are detected, but these features do not show a larger regular surface patch. For comparison purposes, all absolute distances between the scanner positions are shown in Table 1.

For the pairwise registration of two successive scans, the proposed algorithm first extracts *SIFT* features in the reflectance

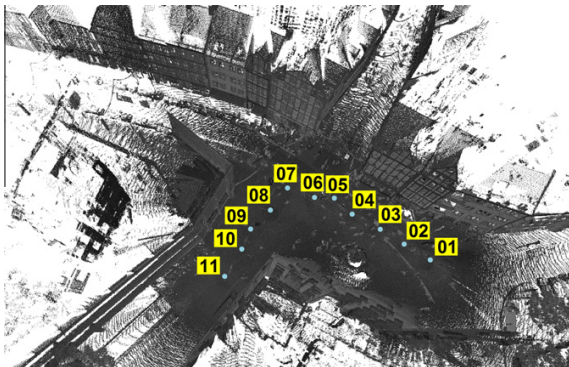


Fig. 3. View on the scene using the scans after registration into a common coordinate system via the reference data.

images and searches for reliable point correspondences by using a threshold value of $t = 0.66$. Using the scans from scan position 01 and scan position 02, a total number of 217 *SIFT* correspondences is found which is illustrated in Fig. 5. The number of matched *SIFT* features between image pairs from all available positions is shown in Table 2, where the diagonal elements represent the total number of extracted *SIFT* features in the respective image. As can be seen in Table 2, the confusion matrix is not necessarily symmetric. This depends on the ratio of the distances of a feature descriptor in one image to the nearest neighbor and the second nearest neighbor in the other image. If a feature descriptor derived from the first image is compared to the feature descriptors derived from the second image, it might occur that the nearest neighbor and the second nearest neighbor in the second image are a little more distinctive as required and therefore, the ratio is below the threshold t and meets the constraint. In the reverse case, in which a feature descriptor derived from the second image is compared to the feature descriptors derived from the first image, the nearest neighbor and the second nearest neighbor in the first image might be more similar which causes a ratio above the threshold t .

Subsequently, the respective *SIFT* features are projected to 3D space yielding a point cloud much smaller than those point clouds of the scans. This however can only be done if the corresponding range information is available. As the *SIFT* features are located with subpixel accuracy and a bilinear transformation is carried out, the range information has to be available for all four surrounding points in the regular grid. This constraint is met by 175 of the 217 conjugate point pairs. Assuming that position and orientation are known for the first scan at scan position 01 results in known coordinates of the 3D conjugate points $CP_{i,n}$. The 2D backprojections of the conjugate points $CP_{i,n+1}$ onto the camera-like virtual plane assigned to scan position 02 are observed. In this step, it is assumed that the virtual plane is expanded to infinity and that 3D points in front of the virtual camera and behind the virtual camera are included. Hence, all of the previously extracted 3D points are observed on the virtual plane. The 3D coordinates of these points are known from the first scan and their 2D backprojections from the observation in the second scan. Thus, the $EPnP$

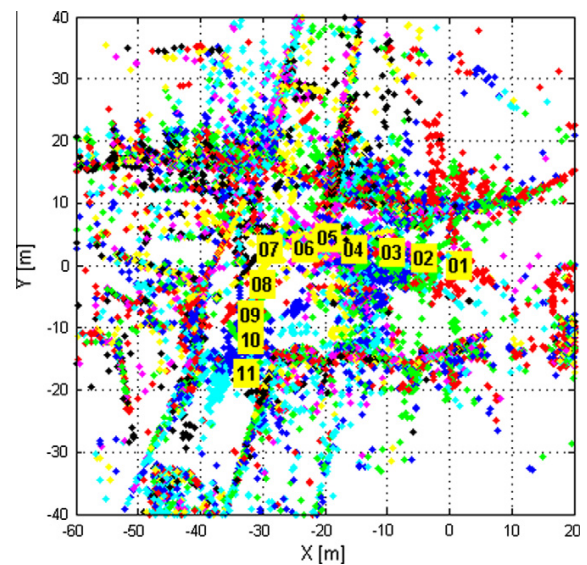


Fig. 4. Nadir view depicting the spatial distribution of the significant points in the investigated 3D environment and the numbered scan locations (01–11) of the TLS after registration into a common coordinate system via the reference data.

Table 1
Absolute 3D distance between the single scan positions in meters.

Position	01	02	03	04	05	06	07	08	09	10	11
01	0	5.58	10.85	16.95	21.47	24.86	30.40	31.79	34.45	35.55	38.32
02	5.58	0	5.27	11.37	15.89	19.30	24.85	26.46	29.41	30.82	34.02
03	10.85	5.27	0	6.11	10.62	14.05	19.62	21.54	24.86	26.66	30.34
04	16.95	11.37	6.11	0	4.60	7.95	13.53	15.93	19.78	22.12	26.41
05	21.47	15.89	10.62	4.60	0	3.96	9.42	12.93	17.48	20.42	25.24
06	24.86	19.30	14.05	7.95	3.96	0	5.59	9.11	13.90	17.11	22.18
07	30.40	24.85	19.62	13.53	9.42	5.59	0	5.81	10.97	14.77	20.16
08	31.79	26.46	21.54	15.93	12.93	9.11	5.81	0	5.18	8.95	14.35
09	34.45	29.41	24.86	19.78	17.48	13.90	10.97	5.18	0	3.92	9.31
10	35.55	30.82	26.66	22.12	20.42	17.11	14.77	8.95	3.92	0	5.40
11	38.32	34.02	30.34	26.41	25.24	22.18	20.16	14.35	9.31	5.40	0

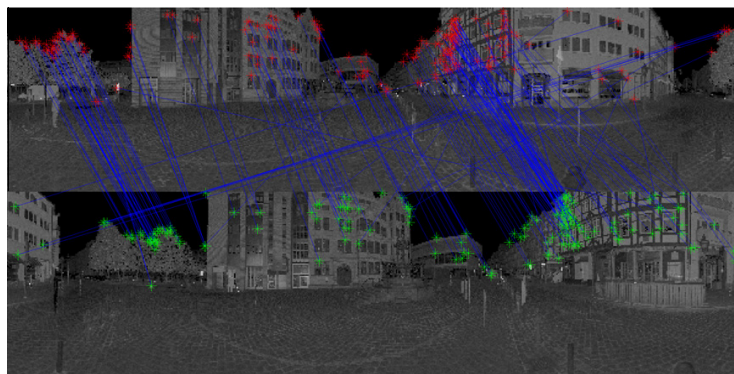


Fig. 5. Visualization of the connected interest points between two consecutive scans.

Table 2
Number of SIFT correspondences between the reflectance images of different scans.

Position	01	02	03	04	05	06	07	08	09	10	11
01	4986	217	63	45	33	44	58	62	41	39	28
02	229	5663	319	100	59	35	80	46	43	38	48
03	88	308	5967	253	120	57	56	38	47	56	68
04	70	114	277	6200	484	131	78	68	58	84	68
05	31	70	124	466	6682	477	169	134	68	71	56
06	61	40	59	129	503	7154	344	211	82	53	30
07	86	96	53	78	163	328	6867	404	205	99	64
08	53	56	34	60	121	240	379	6159	590	361	169
09	39	34	37	56	68	78	158	577	5571	656	330
10	21	25	43	51	54	42	84	344	629	4852	482
11	17	24	37	40	44	24	41	134	277	408	4061

algorithm can be used for estimating the actual transformation parameters. The results from this coarse estimation are shown in Table 3.

This coarse estimate is used to compare the 3D points in a common coordinate frame. According to the criterion introduced for removing outliers based on the geometric 3D distance, only the most promising 3D points remain, in this case 74 between scan position 01 and scan position 02. Based on the 2D and 3D observations of these points, the *EPnP* algorithm is again used to get a refinement of the previous estimate which is shown in Table 4. The result after the registration of the scans from scan position 01 and scan position 02 is illustrated in Fig. 6. For a better visualization, the point clouds were thinned and points below the scan positions were removed. The improvements of the refinement step are finally shown in Fig. 7.

Table 3
Error between reference and estimated transformation parameters for coarse registration.

Position	$\Delta\omega$ [°]	$\Delta\phi$ [°]	$\Delta\kappa$ [°]	ΔX [m]	ΔY [m]	ΔZ [m]	e_{abs} [m]
01–02	-0.001	-0.041	0.057	-0.023	0.030	-0.002	0.038
02–03	0.009	0.007	0.053	-0.011	-0.014	-0.003	0.018
03–04	-0.014	-0.072	0.029	0.015	0.024	-0.022	0.036
04–05	0.002	-0.007	0.002	0.016	-0.025	-0.003	0.030
05–06	0.038	-0.009	0.003	-0.004	-0.008	-0.013	0.016
06–07	0.046	0.035	0.018	0.012	-0.023	0.005	0.026
07–08	0.037	-0.000	-0.017	0.027	-0.013	0.010	0.032
08–09	-0.013	-0.002	0.031	-0.017	-0.010	0.007	0.021
09–10	-0.067	-0.009	0.043	-0.035	0.008	-0.022	0.042
10–11	0.025	-0.064	-0.011	0.037	-0.001	0.007	0.038

Table 4
Error between reference and estimated transformation parameters for refined registration.

Position	$\Delta\omega$ [°]	$\Delta\phi$ [°]	$\Delta\kappa$ [°]	ΔX [m]	ΔY [m]	ΔZ [m]	e_{abs} [m]
01–02	-0.047	-0.011	0.004	0.019	-0.001	-0.005	0.019
02–03	0.044	-0.013	0.035	-0.004	-0.009	-0.003	0.010
03–04	0.034	-0.004	0.079	-0.015	-0.006	0.010	0.019
04–05	-0.021	0.022	0.008	-0.002	-0.012	0.005	0.013
05–06	0.031	-0.002	0.009	-0.001	-0.005	-0.008	0.010
06–07	0.017	0.032	0.053	0.004	-0.011	0.010	0.015
07–08	0.054	-0.044	0.012	0.002	0.006	-0.005	0.008
08–09	0.015	-0.027	0.031	0.009	0.005	-0.008	0.013
09–10	-0.043	-0.004	0.029	-0.008	0.003	-0.012	0.015
10–11	0.016	-0.065	-0.014	0.025	-0.012	0.009	0.029

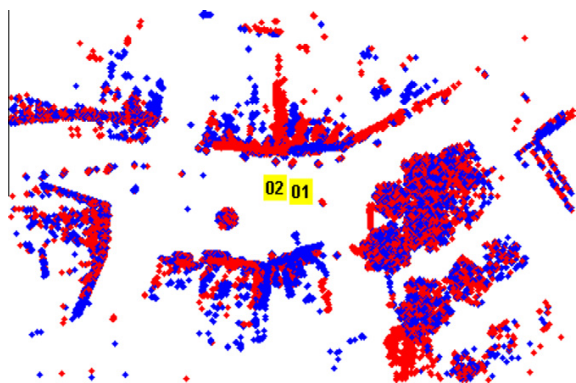


Fig. 6. Nadir view depicting two point clouds after registration. The 3D points of the first scan from scan position 01 are illustrated as red points and the 3D points of the second scan from scan position 02 as blue points.

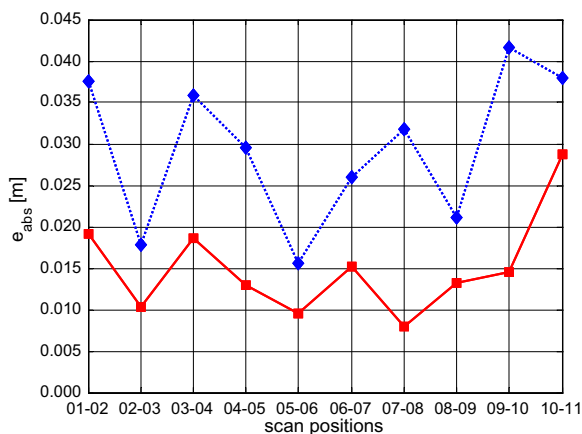


Fig. 7. Absolute error between reference and estimated position for coarse (dotted line with diamonds) and refined (solid line with squares) registration.

4. Discussion

Although the proposed algorithm only uses a refinement and no fine registration method like *ICP*, the quality of the resulting estimate for pairwise registration of two consecutive scans is in the

same range as in Wang and Brenner (2008). For larger distances between the scan positions (Table 1), however, the total number of corresponding *SIFT* features decreases which is summarized in the confusion matrix (Table 2). It has also to be mentioned that the number of detected and corresponding *SIFT* features strongly depends on the scene content. Therefore, as all other methods based on *SIFT* features, the proposed algorithm is suited for well-textured scenes, whereas scenes with less correspondences will probably not lead to optimal results.

The first approach presented in Brenner et al. (2008) can cope with larger distances, but it is based on extracting planes and therefore, this approach is restricted on environments with regular surfaces like urban environments. Such regular surfaces are also used in von Hansen (2006) and Rabbani et al. (2007), a variety of geometric primitives is used for registration. By contrast, the proposed approach is not restricted on environments with regular surfaces and can also be used in areas with vegetation.

Furthermore, it is an advantage to use distinctive feature points with subpixel accuracy derived from the regular measured grid to avoid scanning limitations. For an accurate subpixel localization, the interpolated range estimation could be further improved by substituting the bilinear interpolation with a higher-order interpolation, e.g. using splines or geometric curvature as suggested in Bae and Lichti (2004).

Generally, well-distributed points in 3D space are essential to obtain a high-quality estimate of the transformation parameters. With an already given distribution of the processed data, the coarse registration with an absolute position error in the range between 16 and 42 mm is achieved (Table 3, last column). This can be refined to an absolute position error in the range between 8 and 29 mm (Table 4, last column) which is close to the TLS measurement accuracy of 12 mm. For all results shown, an improvement in the absolute position estimate is obtained by the refinement step (Fig. 7). However, it has to be stated that the solutions depend on the *RANSAC* algorithm estimates. If the coarse estimate already shows a high accuracy, the refinement may lead to an improvement, but not necessarily. In some cases, despite an improvement of the position estimate single rotation parameters are affected in a negative way. But still most of the deviations in orientation are below 0.065 degrees except one value of 0.079 degrees.

The proposed algorithm was tested on a standard PC with 2.83 GHz and, for the implementation is not fully optimized with respect to a possible parallelization on multiple cores, only one core was used. Thus, the matching of corresponding *SIFT* features takes an average time of about 5 s and the total time required for pairwise registration of two scans is approximately 18 s. As Matlab was used, the speed can further be optimized. A comparison with a fully Matlab-based implementation of the standard *ICP* algorithm with 10 iterations was carried out on the same PC using a single core again. Assuming a good a priori alignment of the point clouds, the *ICP* algorithm needed a time of 56.1 s for two subsets consisting of 22,500 points, i.e. only 1% of the respective point clouds, and approximately 2.8 h for two subsets consisting of 225,000 points, i.e. 10% of the respective point clouds.

Concerning *EPnP* and dealing with noisy points, in Moreno-Noguer et al. (2007) it is suggested to use the *RANSAC* algorithm on small subsets consisting of seven correspondences which is used in the proposed algorithm. Larger subsets are motivated as they introduce redundancy and therefore reduce the sensitivity to noise. In this context, the *RANSAC* scheme is utilized for selecting the most reliable 3D-to-2D correspondences.

The advantage of the proposed refinement is that additional geometric constraints are introduced with respect to 3D distances in order to check the *SIFT* correspondences for consistency, for instance in case of periodic shapes of façades in urban environments. Thus, reliable 3D-to-3D correspondences are remaining. These are

much stronger than reliable 3D-to-2D correspondences which can be observed by comparing the improvements between coarse registration and its refinement.

5. Conclusions

In this paper, a fast and fully automatic method for pairwise registration of TLS data is presented. The method is based on both reflectance images and range information which are commonly used to derive simple, but characteristic points in 3D space. For a new scan, the observation of these points on a virtual plane aligned with the scanner allows for a fast and accurate RANSAC-based registration scheme including a single step outlier removal for checking consistency.

Instead of a registration for consecutive scans where the registration of each scan is based on comparing it to the previous scan, the idea of the proposed algorithm might be extended to multiview registration. This means that for each new scan all previous scans could be considered which will probably improve the registration results. Thus, a growing subset of scans is considered as the number of available scans increases. The 3D coordinates of specific points, detected by comparing the actual reflectance image with all previous reflectance images and generating the respective conjugate points via bilinear transformation, are known from the previous scans and their 2D projections onto the actual virtual plane are observed. This concept including parts of the confusion matrix instead of single entries increases the total number of conjugate points and, probably, also improves their distribution in local space, which should result in a higher accuracy.

Acknowledgement

The authors would like to thank Dr. Claus Brenner from the Institute of Cartography and Geoinformatics at the University of Hanover for providing the TLS data. The data is available at <http://www.ikg.uni-hannover.de/index.php?id=413&L=de> (Accessed October, 2010).

References

- Al-Manasir, K., Fraser, C.S., 2006. Registration of terrestrial laser scanner data using imagery. *The Photogrammetric Record* 21 (115), 255–268.
- Bae, K.-H., Lichti, D.D., 2004. Automated registration of unorganised point clouds from terrestrial laser scanners. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B5), 222–227.
- Bae, K.-H., Lichti, D.D., 2008. A method for automated registration of unorganised point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), 36–54.
- Barnea, S., Filin, S., 2007. Registration of terrestrial laser scans via image based features. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), 32–37.
- Barnea, S., Filin, S., 2008. Keypoint based autonomous registration of terrestrial laser point-clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), 19–35.
- Bendels, G.H., Degener, P., Körtgen, M., Klein, R., 2004. Image-based registration of 3D-range data using feature surface elements. In: Chrysanthou, Y., Cain, K., Silberman, N., Niccolucci, F. (Eds.), *The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, pp. 115–124.
- Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), 239–256.
- Böhm, J., Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. In: Gruen, A., Kahmen, H. (Eds.), *Optical 3-D Measurement Techniques VIII*, pp. 338–344.
- Brenner, C., Dold, C., Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), 4–18.
- Dold, C., 2005. Extended Gaussian images for the registration of terrestrial scan data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3/W19), 180–185.
- Dold, C., Brenner, C., 2004. Automatic matching of terrestrial scan data as a basis for the generation of detailed 3D city models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B3), 1091–1096.
- Dold, C., Brenner, C., 2006. Registration of terrestrial laser scanning data using planar patches and image data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 5), 78–83.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24 (6), 381–395.
- Forkuo, E.K., King, B., 2004. Automatic fusion of photogrammetric imagery and laser scanner point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B4), 921–926.
- Hartley, R.I., Zisserman, A., 2008. *Multiple view geometry in computer vision*. University Press, Cambridge.
- Horn, B.K.P., Hilden, H.M., Negahdaripour, S., 1988. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A* 5, 1127–1135.
- Kang, Z., Li, J., Zhang, L., Zhao, Q., Zlatanova, S., 2009. Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. *Sensors* 9 (4), 2621–2646.
- Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. EPnP: an accurate O(n) solution to the PnP problem. *International Journal of Computer Vision* 81 (2), 155–166.
- Lichti, D.D., Pfeifer, N., Maas, H.-G. (Eds.), 2008. *Terrestrial laser scanning*. ISPRS Journal of Photogrammetry and Remote Sensing 63 (1).
- Lindeberg, T., 1994. Scale-space theory: a basic tool for analysing structures at different scales. *Journal of Applied Statistics* 21 (2), 224–270.
- Lowe, D.G., 1999. Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision*, 1150–1157. doi:10.1109/ICCV.1999.790410.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60 (2), 91–110.
- Mikolajczyk, K., 2002. *Detection of local features invariant to affine transformations*. Institut National Polytechnique de Grenoble, France, Dissertation.
- Moreno-Noguer, F., Lepetit, V., Fua, P., 2007. Accurate non-iterative O(n) solution to the PnP problem. In: *IEEE 11th International Conference on Computer Vision*, pp. 1–8.
- Rabbani, T., Dijkman, S., van den Heuvel, F., Vosselman, G., 2007. An integrated approach for modelling and global registration of point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 61 (6), 355–370.
- Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the ICP algorithm. In: *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling (3DIM 2001)*, 28 May–1 June 2001, Quebec City, Canada. IEEE Computer Society, Los Alamitos, CA, pp. 145–152.
- Seo, J.K., Sharp, G.C., Lee, S.W., 2005. Range data registration using photometric features. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, pp. 1140–1145.
- Stamos, I., Leordeanu, M., 2003. Automated feature-based range registration of urban scenes of large scale. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. II, IEEE CS Press 2003, pp. 555–561.
- von Hansen, W., 2006. Robust automatic marker-free registration of terrestrial scan data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), 105–110.
- Wang, Z., Brenner, C., 2008. Point based registration of terrestrial laser data using intensity and geometry features. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (Part B5), 583–589.
- Wehr, A., Lohr, U. (Eds.), 1999. *Airborne laser scanning*. ISPRS Journal of Photogrammetry and Remote Sensing 54 (2–3).

C. Publikationen zu 2.3 Szenenanalyse

- [C1] Gross H, **Jutzi B**, Thoennessen U (2009) Classification of Elevation Data based on analytical versus trained Feature Values to determine Object Boundaries. In: Seyfert E (Ed) Geoinformatik und Erdbeobachtung: 29. Wissenschaftlich-Technische Jahrestagung der DGPF, 2009 (18): 315-326
- [C2] Gross H, **Jutzi B**, Thoennessen U (2007) Segmentation of tree regions using data of a full-waveform laser. In: Stilla U, Mayer H, Rottensteiner F, Heipke C, Hinz S (Eds) Photogrammetric Image Analysis PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W49A): 57-62
- [C3] **Jutzi B**, Gross H (2009) Nearest neighbour classification on Laser point clouds to gain object structures from buildings. In: Heipke C, Jacobsen K, Müller S, Sörgel U (Eds) High-resolution Earth Imaging for Geospatial Information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 1-4-7/W5) (on CD)
- [C4] Weinmann M, **Jutzi B**, Hinz S, Mallet C (2015) Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry & Remote Sensing 105: 286-304 doi:10.1016/j.isprsjprs.2015.01.016
- [C5] Weinmann M, Urban S, Hinz S, **Jutzi B**, Mallet C (2015) Distinctive 2D and 3D Features for Automated Large-Scale Scene Analysis in Urban Areas. Special Section on Processing Large Geospatial Data. Computers & Graphics 49: 47-57 doi:10.1016/j.cag.2015.01.006
- [C6] Weinmann M, **Jutzi B**, Mallet C (2013) Feature relevance assessment for the semantic interpretation of 3D point cloud data. In: Scaioni M, Lindenberg RC, Oude Elberink S, Schneider D, Pirotti F (Eds) ISPRS Workshop Laserscanning 2013. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-5/W2, 2013: 313-318
- [C7] Weinmann M, **Jutzi B**, Mallet C (2014) Semantic 3D Scene Interpretation: A Framework Combining Optimal Neighborhood Size Selection with Relevant Features. In: Schindler K, Papanastasiou N (Eds) ISPRS Technical Commission III Symposium. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3, 2014: 181-188 PCV 2014 Best Paper - Honourable Mention

- [C8] Weinmann M, Mallet C, Hinz S, **Jutzi B** (2015) Efficient interpretation of 3D point clouds by assessing feature relevance. AVN - Allgemeine Vermessungsnachrichten
- [C9] Weinmann M, Schmidt A, Mallet C, Hinz S, Rottensteiner F, **Jutzi B** (2015) Contextual classification of point cloud data by exploiting individual 3D neighborhoods. In: Stilla U, Heipke C (Eds) Photogrammetric Image Analysis PIA15 + High-Resolution Earth Imaging for Geospatial Information HRIGI15 - Joint ISPRS conference. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W4, 2015: 271-278

NEAREST NEIGHBOUR CLASSIFICATION ON LASER POINT CLOUDS TO GAIN OBJECT STRUCTURES FROM BUILDINGS

B. Jutzi^a, H. Gross^b

^a Institute of Photogrammetry and Remote Sensing, Universität Karlsruhe, Englerstr. 7, 76128 Karlsruhe, Germany
boris.jutzi@ipf.uni-karlsruhe.de

^b FGAN-FOM, Research Institute for Optronics and Pattern Recognition, Gutleuthausstraße 1, 76275 Ettlingen, Germany
gross@fom.fgan.de

KEY WORDS: Laser data, point cloud, classification, nearest neighbour, covariance, eigenvalues.

ABSTRACT:

The application of three dimensional building models has become more and more important for urban planning, enhanced navigation and visualization of touristic or historic objects. 3D models can be used to describe complex urban scenes. The automatic generation of 3D models using elevation data is a challenge for actual research. Especially extracting planes edges and corners of man made objects is of great interest. This paper deals with the automatic classification of points by utilizing the eigenvalues of the covariance within the close neighbourhood. The method is based on the analysis of 3D point clouds derived from Laser scanner data. For each 3D point additional structural features by considering the neighbourhood are calculated. Invariance with respect to position, scale and rotation is achieved by normalization of the features. For classification the derived features are compared with analytical calculated as well as trained feature values for typical object structures. For the generation of a training data set several point sets with different density and varying noise are generated and exploited. The result of the investigations is that the quality of the classification using the analytical eigenvalues as reference is not harmful in comparison to the trained data set for a small noise. Therefore for all structures presented here it is not necessary to use training data sets instead of an unsupervised classification based on the analytical eigenvalues. Weighting the calculated distances in the eigenvalue space dependent on the structure type improves the classification result. Due to this classification all points which may belong to a building edge are selected. Assembling these points to lines the 3D borders of the objects were achieved. The algorithm is tested for several urban scenes and the results are discussed.

1. INTRODUCTION

Three-dimensional building models have become important during the past for various applications like urban planning, enhanced navigation or visualization of touristic or historic objects. They can increase the understanding and explanation of complex scenes and support the decision process of operation planning. The benefit for several applications by utilizing LIDAR data was demonstrated for instance by Brenner et al. (2001). For decision support and operation planning the real urban environment should be available. In most cases the object models of interest are not obtainable and especially in time critical situations the 3D models must be generated as fast and accurate as possible.

Different approaches to generate the 3D models of urban scenes are discussed in the literature (Shan & Toth, 2008). Building models are typically acquired by (semi-) automatic processing of Laser scanner elevation data or aerial imagery (Baillard et al., 1999; Geibel & Stilla, 2000). LIDAR data can be utilized for large urban scenes (Gross & Thoennessen, 2005). The processing of raw full-waveform data to gain object structures of buildings was investigated by Jutzi et al. (2005) and the iterative processing to increase the set of 3D points of buildings by Kirchhof et al. (2008). Pollefeys (1999) uses projective geometry for a 3D reconstruction from image sequences. Fraser et al. (2002) use stereo approaches for 3D building reconstruction. Vosselman et al. (2004) describes a scan line segmentation method grouping points in a 3D proximity. Airborne systems are widely used but also terrestrial Laser scanners are increasingly available. The latter ones provide a much higher geometrical resolution and accuracy (mm vs. dm) and they are able to acquire fine building facade details which are an essential requirement for a realistic virtual visualization.

In Section 2 the calculation of additional point features is described. The features are normalized with respect to translation, scale and rotation. In Section 3 typical constellations of points are discussed and discriminating features are presented. Examples for the combination of eigenvalues and structure tensor are shown. For typical situations analytical feature values are derived. For the classification procedure the results of the trained feature values are discussed in Section 4 and the trained values are compared with the analytical values. The generation of lines is described in Section 5. Points with the same eigenvectors are assembled and approximated by lines. The resulting 3D structures (boundaries) of objects are shown for the selected laser point cloud. In Section 6 the possibilities using additional features are summarized. Outstanding topics and aspects of the realized method are discussed.

2. EIGENVALUE ESTIMATION TO GAIN OBJECT STRUCTURES

A Laserscanning device delivers 3D point measurements in an Euclidian coordinate system. For airborne systems mostly the height information is stored in a raster grid with a predefined resolution. Image cells without a measurement are interpolated by considering their neighbourhood.

An example data set gathered by an airborne Laser scanner system (TopoSys®) as 3D points is shown in Figure 1a. The color corresponds to the height. A transformation to a raster image, selecting the highest value for each pixel and after filling missing pixels with a Median operation, yields to Figure 1b. Due to the filtering the image does not represent the original 3D information anymore. The horizontal position is slightly different and some of the height values are interpolated to fill the gaps even if there was no measured value available.

Additionally, sometimes more than one measurement for a resolution cell exists considering first and last echo or combining data of several measurement campaigns.

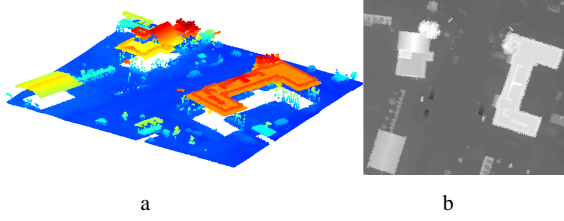


Figure 1. Point clouds measured with TopoSys® Laser scanner
a) colored by height, b) raster image based on point clouds with interpolated values.

An example of data received by a terrestrial Laser scanner (Z+F sensor) for a dense point cloud colored by intensity is shown in Figure 2.

In contrary to the airborne data the projection of terrestrial Laser data along any direction is not very reasonable. Especially the combination of airborne (Figure 1) and terrestrial (Figure 2) Laserscanning data requires directly the analysis in the 3D data.



Figure 2. Point clouds of a Z+F sensor colored by intensity.

2.1 Calculation of the covariance matrix utilizing a 3D spherical volume cell

A 3D spherical volume cell with radius R is assigned to each point of the cloud. All points in a spherical cell will be analyzed. The 3D covariance matrix as described by Maas & Vosselman (1999) are discussed and further improved as described in Gross & Thoennessen (2006).

In a continuous domain, moments are defined by:

$$m_{ijk} = \int_V x^i y^j z^k f(x, y, z) dv, \quad (1)$$

where $i, j, k \in \mathbb{N}$, and $i + j + k$ is the order of the moments integrated over a predefined volume weighted by $f(x, y, z)$. As weighting function the mass density can be used. It reduces to a constant value if homogeneous material is assumed. Another possibility is to use the measured intensity as weighting function as discussed in earlier works. To normalize the terms they have to divide by the volume $m_{000} = \int_V f(x, y, z) dv$.

Considering only surfaces of objects all moments have to be calculated with a constant but small thickness for the volume vanishing by the normalization. After discretization of the integrand and setting $f(x, y, z) = 1 \quad \forall$ points the integral is approximated by a sum. The mean values $\bar{x}, \bar{y}, \bar{z}$ and the moments of the second order $i + j + k = 2$ have been calculated.

The normalized and dimensionless moments of second order for discrete points are given by

$$\tilde{m}_{ijk} = \frac{\sum_{l=1}^N (x_l - \bar{x})^i (y_l - \bar{y})^j (z_l - \bar{z})^k}{R^{i+j+k} N}. \quad (2)$$

Neither the number of points nor the chosen physical unit for the coordinates, the radius and the weighting factor influences the values of the covariance matrix.

For each point of the whole data set a symmetrical covariance matrix is calculated by

$$M = \begin{pmatrix} \tilde{m}_{200} & \tilde{m}_{110} & \tilde{m}_{101} \\ \tilde{m}_{110} & \tilde{m}_{020} & \tilde{m}_{011} \\ \tilde{m}_{101} & \tilde{m}_{011} & \tilde{m}_{002} \end{pmatrix}. \quad (3)$$

The calculation of the eigenvalues λ_i and eigenvectors \vec{e}_i with $i=1,2,3$ delivers additional features for each point. The eigenvalues are invariant concerning translation, rotation, and scaling.

2.2 Point distribution in 3D space

In this section the influence of the measurement and the related point distribution on the investigated structures is described.

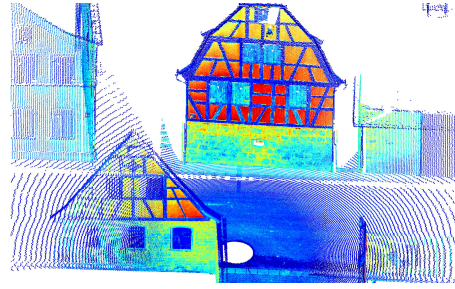


Figure 3. Illustration of a point cloud captured by a terrestrial Laser scanner with typical scan pattern (color indicates the reflected intensity).

Figure 3 shows as an example for the point distribution derived by a terrestrial Laser scanner (Zoller+Fröhlich). The point density depends on the distance of the object to the sensor and also on the incidence angle between laser beam and normal vector onto the object surface. For the airborne Laser scanner (TopoSys®) mounted on an aircraft the point density can be much higher in flight direction than perpendicular to the flight direction. In both cases there is no uniform distribution of the measured points.

The investigations show that an inhomogeneous distribution does not influence the eigenvalues essentially as long as the radius of the neighbourhood is large enough. This means points inside a plane are characterized as plane points if the neighbourhood encloses at least five points in all directions and the rate of the point distances for any two different directions is smaller than 5:1.

2.3 Analytical eigenvalues for object structures

For specific object structures analytical eigenvalues can be determined. Table 1 show some typical object structures with their corresponding eigenvalues, where all values are determined by utilizing all required integrations of formula (1).










Eigenvalues		λ_1	λ_2	λ_3
	Isolated point	0	0	0
	End of a line	$\frac{1}{12}$	0	0
	Line	$\frac{1}{3}$	0	0
	Half plane	$\frac{1}{4}$	$\frac{1}{4}\left(1-\frac{64}{9\pi^2}\right)=0.0$	0
	Plane	$\frac{1}{4}$	$\frac{1}{4}$	0
	Quarter plane	$\frac{1}{4}\left(1-\frac{2}{\pi}\right)=0.09$	$\frac{1}{4}+\frac{1}{2\pi}-\frac{32}{9\pi^2}=0$	0
	Two planes	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}-\frac{8}{9\pi^2}=0.03$
	Three planes	$\frac{1}{6}\left(1-\frac{1}{\pi}\right)=0.11$	$\frac{1}{6}\left(1-\frac{1}{\pi}\right)=0.11$	$\frac{1}{6}\left(1+\frac{2}{\pi}\right)-\frac{2^2}{3^2\pi^2}=0.03$
	Two planes 30°	0.25	0.1875	0.01747

Table 1. Eigenvalues for some selected object structures.

For all possible values of the roof slope the eigenvalues are drawn in Figure 4. The greatest eigenvalue is 0.25 and constant. The second eigenvalue starts from 0.125 and increases with increasing slope until 0.25. The smallest eigenvalue decreases from 0.03 to zero. For a slope of 30° the eigenvalues reaches the mean values for a flat roof and a plane. Therefore an own class for this structure is defined.

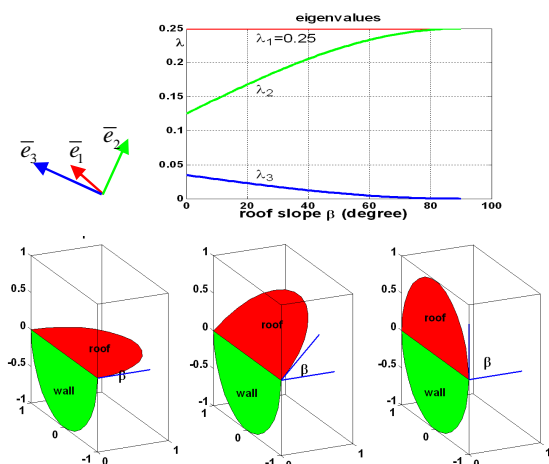


Figure 4. Eigenvalues of the eave points for different roof slopes (0°, 30°, and 90°); the colored arrows visualize the direction of the eigenvectors.

3. MONTE CARLO SIMULATION

The analytical calculated values in Table 1 do not correspond to the statistical averages, which can be expected for the relevant structures of real data. Usually, for an example, the smallest eigenvalue of points belonging to a plane do not converge to $\lambda_3=0$. Already very small deviations of points from a flat surface yield to $\lambda_3>0$. Therefore for all the structures in Table 1 inside a spherical neighbourhood with radius R points

with the different distances, normalized by the radius of the sphere $dx/R \in \{0.03, 0.1, 0.2, 0.3, 0.4\}$ are generated. Each coordinate of the position of the points is modified by a Gaussian distributed noise with the normalized standard deviation $\sigma/R \in \{0.0, 0.01, 0.02, 0.03, 0.04\}$.

For each parameter combination and structure 1000 point clouds have been generated by random 3D points. The mean value and the standard deviation of every 3 eigenvalues were determined. The histograms of one test set for each structure are drawn in Figure 5. The distribution of the eigenvalues seems to be Gaussian with center near by the analytical values.

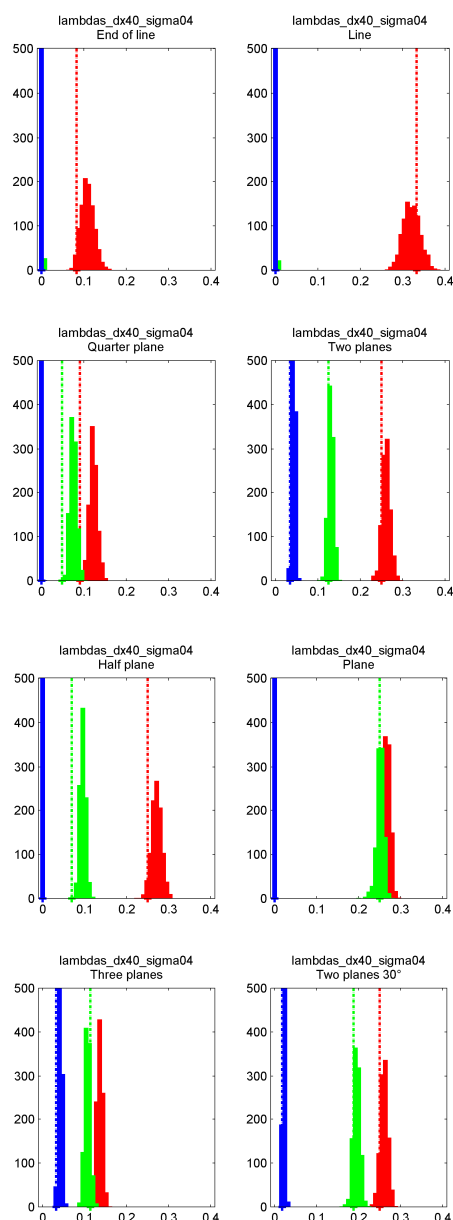


Figure 5. Histograms of the eigenvalues and comparison with the analytical values (dashed lines) for $dx=0.4R$ and $\sigma=0.04R$ for all structures (red: first, green: second, blue: third eigenvalue).

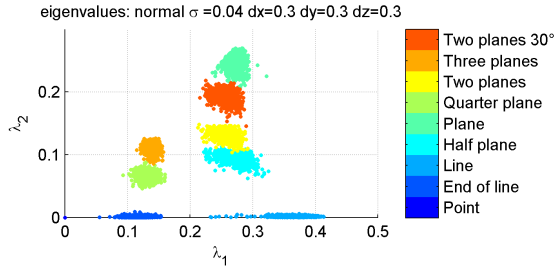


Figure 6. Eigenvalue point cloud projection along the axis of the smallest eigenvalue.

In the next steps the 3 eigenvalues are considered as a point of a 3D space. For a small standard deviation σ the point cloud of eigenvalues results in a small accumulation of points. If σ is increasing the clusters are extending and nearby clusters may overlap. Figure 6 shows the 2D-projection along the axis of the eigenvalue λ_3 . Projections along the two other eigenvalues demonstrate the separability of the cluster for each structure.

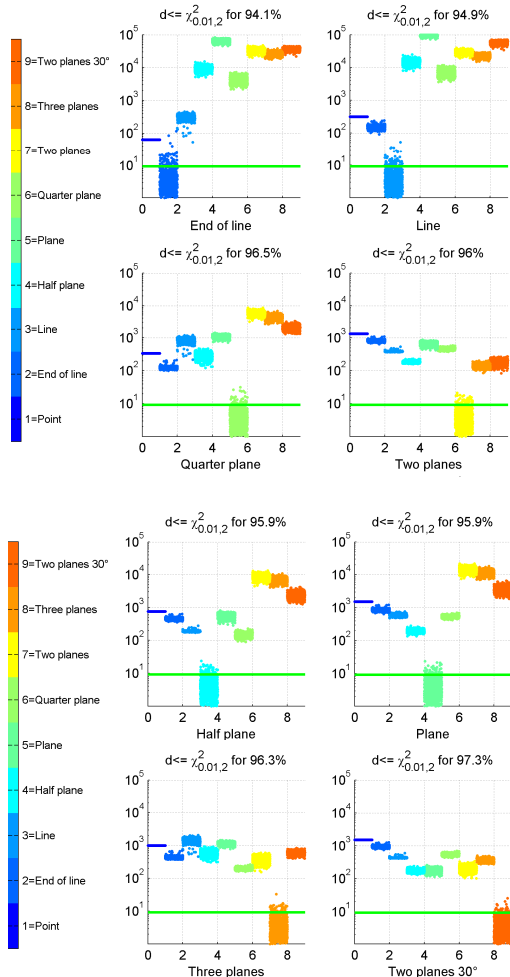


Figure 7. Distances of the eigenvalue points to all classes.

The eigenvalues of the points for each structure define a training record from which the three mean values $\bar{\lambda}_S = 1/N \sum_{p \in S} \lambda_p$ as well as the associated eigenvalue-covariance-

matrix $C_S = \sum_{p \in S} (\lambda_p - \bar{\lambda})(\lambda_p - \bar{\lambda})^T$ can be calculated, where N is the number of eigenvalue-points of the structure. The distance of any test point λ of the eigenvalue space is determined by using the Mahalanobis-distance $d(\lambda, S) = (\lambda - \bar{\lambda}_S)^T C_S^{-1} (\lambda - \bar{\lambda}_S)$. This measure gives a distance for any test eigenvalue-point to the different structures. These eight distances of every point against their own and all other structures (except for isolated point) are listed in the Figure 7. The points of a structure are plotted and colored in accordance to their membership S and drawn in the interval $[S-1, S]$ (horizontal axis). The vertical axis represents the logarithm of the distance of each eigenvalue point to each structure. In the 1st picture the distances between the eigenvectors of all test records of all structures against the structure "End of line" are drawn. The remaining pictures show the respective distance of all test points to the other structures. The green line mark the value of the Chi-square tests $\chi^2_{0.01,2}$. The percentage number of points of each structure with a smaller distance has been indicated. With increasing noise the distance of a point of a structure to a different structure decreases. Therefore false classification increases.

Figure 8 shows the mean value and the standard deviation of the eigenvalues of the training set for a plane dependent on the point density and the noise. The mean values approximate the analytical eigenvalues with a very small standard deviation.

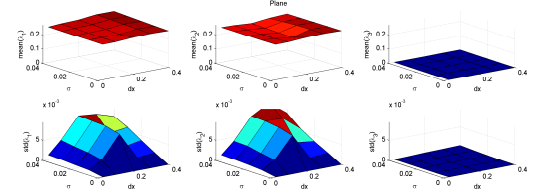


Figure 8. Mean value and standard deviation of the three eigenvalues of the training set for a plane.

A comparison between the mean value of the eigenvalues of the training set for a plane and the analytical values is shown in Figure 9. The differences depend on the point density and the noise. A high point density delivers nearly the analytical eigenvalues. The non monotonic behaviour of the curve for λ_2 may be caused by the approximation of a plane by nearly equidistant points (discretization effects). The mean value of the third eigenvalues is positive but very small.

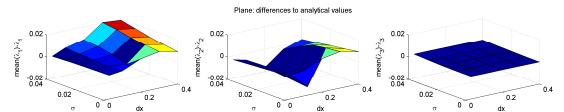


Figure 9. Differences between the mean value of the eigenvalues of the training set for a plane and the analytical values.

For the same points the Euclidean distances in the eigenvalue space against the analytical eigenvalues were calculated. Within the tested mean point intervals and the investigated noise all the points were assigned to the correct structure. Based on this investigation the classification of elevation points can be realized by nearest neighbour classification in the eigenvalue space of the structures of Table 1. This is possible as far the noise is lower than 4% of the radius of the neighbourhood environment.

4. NEAREST NEIGHBOUR CLASSIFICATION OF 3D POINTS

After calculating the covariance matrix for each point in the data set by considering the local environment defined by a sphere additional features for each point are derived. These features are the centre of gravity, the geometrical distance between centre of gravity to the point, the eigenvectors, the eigenvalues and the number of points inside the sphere. The same features can be used to determinate the object characteristics.

Table 1 shows the eigenvalues of the covariance matrix of some special point configurations. The first six rows present 2D and the last three rows 3D object structures. The eigenvalues for the typical object structures are calculated analytically. For an ideal line two eigenvalues are zero and one of it is greater than zero. If test points inside a plane are of interest their eigenvalues have to be compared with the analytical eigenvalues $\lambda_1 = \lambda_2 = 0.25 \wedge \lambda_3 = 0$ for a correct plane.

The eigenvalues in Table 1 are considered as reference points in the 3D eigenvalue space for each structure. The classification of any test point by the nearest neighbour method was performed, were all distances were measured in the eigenvalue space.

For the following steps we define the dimensionality $\dim(S)$ for each structure, which means the dimension of all points belonging to the same structure of a contiguous object. The dimensionalities for each structure are given in Table 2. Corner like points have the dimensionality 0, edge like points 1 and plane like points 2.

Structure	Dimensionality
Isolated point	0
End of a line	0
Line	1
Half plane	1
Plane	2
Quarter plane	0
Two planes	1
Three planes	0
Two planes 30°	1

Table 2. Dimensionality for each structure.

By utilizing the empirically derived weighting factors $w(S) = 1/(1 + \dim(S))$ for the distance $d(S)$ between the test point and the analytically calculated eigenvalues of structure S the classification result was refined. This weighting of the distances between test and reference points introduces non-planar separation surfaces defined by $d(S_i)w(S_i) = d(S_j)w(S_j)$ between two structures. Ignoring the influence of all other structures, the separation surface between the structures i and j is given by the constant ratio of both distances $d(S_i)/d(S_j) = w(S_j)/w(S_i) = w_{ji}$. For $w_{ji} = 1$ we get the intermediate plane between both structures as separation surface. For $w_{ji} \neq 1$ the separation is described by a sphere. Radius and centre point depend only on w_{ji} and the distance between the two structures in the 3D eigenvalue space.

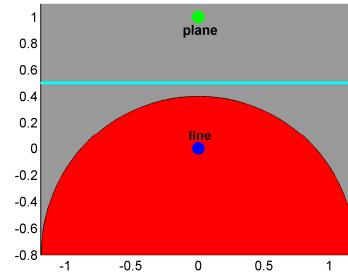


Figure 10. Equipotential surface between line and plane in the eigenvalue space.

As an example Figure 10 illustrates the situation between the structures *line* and *plane* with weighted distances. All test points with eigenvalues inside the red region are classified as line points meanwhile all points in the grey region are classified as points belonging to a plane. Without weighting the cyan marked horizontal line (hyper plane) separates the two classes.

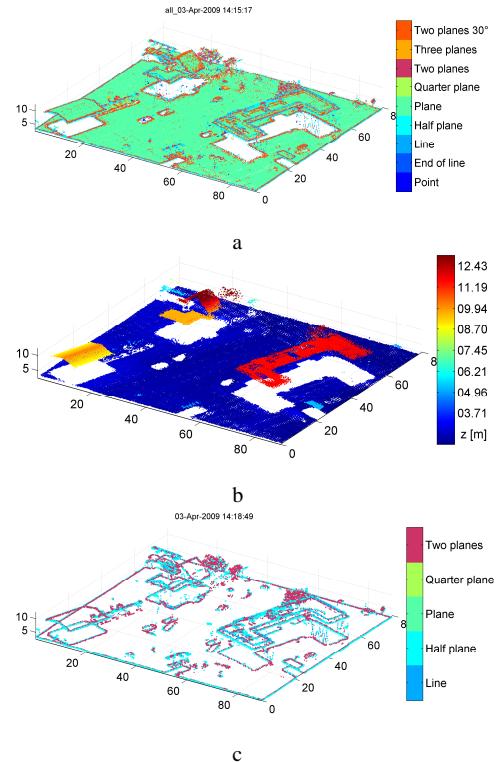


Figure 11. Classified object points. a) All points colored by their classification, b) Points identified as plane points (colored by their height), c) Points with one high and two small eigenvalues representing edges of objects.

By utilizing the weighted distance calculation during the classification procedure for all points the derived results are shown in Figure 11a. Figure 11b shows all points with eigenvalues fulfilling the criteria for planes. The color indicates the object height. In Figure 11c only the edge points are depicted corresponding to Table 1 rows 3, 4, and 7.

For the introduced classification further results are shown for comparison purposes of a more complex building. The results are depicted in Figure 12 with an oblique view to demonstrate the geometrical relation of the 3D points.

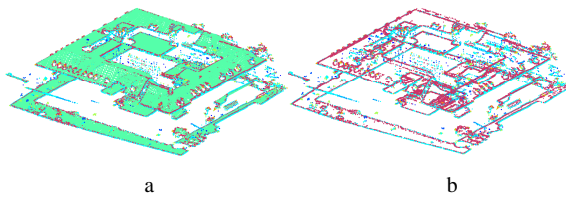


Figure 12. Classification result of a laser point cloud for a complex urban building. a) with all points, b) without points inside a plane.

5. LINE GENERATION

All points marked as edge point may belong to a line. These points are assembled to lines by a grouping process (Gross & Thoennessen, 2006). Therefore the greatest eigenvalue and its eigenvector are considered. Consecutive points with a similar eigenvector, lying inside a small cylinder are grouped together and approximated by a line.

The procedure starts with any arbitrary point of the point cloud classified as edge-like point (*line, halfplane, two_planes*). This trigger point is compared with all points which have nearly the same or opposite eigenvector of the largest eigenvalue. Furthermore only points with very small distance to the straight line defined by the trigger point and its first eigenvector are included in the next consideration. Finally it is focused on the first two gaps starting from the trigger point going along the first eigenvector and also its opposite direction. Only points inside these gaps and fulfilling all those conditions are selected and used to determine a regression line and its endpoints.

The same procedure is repeated for all points not assigned to a line until each point belongs to a line or can not generate an acceptable line.

Figure 13 shows the results of the line generation for the data set shown in figure 1. The color indicates the length of the lines. The eaves as well as the ground plan of the buildings are approximated by lines. For the detection of the ridge of the saddle roof a readjustment of the thresholds for the eigenvalues might be recommended to improve the results especially for roofs with small inclination.

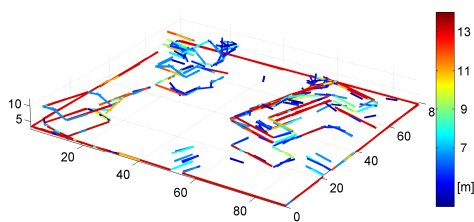


Figure 13. Lines generated by using the classified laser elevation points.

6. CONCLUSION AND OUTLOOK

For exploiting Laser scanning data the processing of the original 3D point clouds is proposed. Additional features for each point of the cloud can be calculated from the covariance matrix including all neighbour points. The neighbourhood can be investigated by considering a sphere. The quality of the resulting eigenvalues and the eigenvectors of the matrix strongly depend on the spatial resolution and the number of points inside the sphere. The new features are invariant with respect to position, rotation and scale.

The additional features are appropriate for classification of the points as edge, corner, plane or tree points. For some typical situations analytically determined eigenvalues are opposed to calculated eigenvalues of real data for comparison. The greatest eigenvalue can be used for filtering edge like points.

The described method for generation of lines combines consecutive points with the same eigenvector inside a small cylinder without any gap. The presented results are promising.

Further investigations are planned concerning the fusion of the data on basis of the point clouds and/or on a higher level of lines. Especially the construction of planes assembling plane like points should be investigated in future.

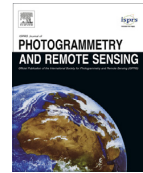
REFERENCES

- Baillard, C., Schmid, C., Zisserman, A., Fitzgibbon, A., 1999. Automatic line matching and 3D reconstruction from multiple views. In: ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery 32, pp. 69-80.
- Brenner, C., Haala, N., Fritsch, D., 2001. Towards fully automated 3D city model generation. In: Baltsavias, E., Gruen, A., van Gool, L., (Eds), Proc. 3rd Int. Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images, pp. 47-57.
- Fraser, C.S., Baltsavias, E., Gruen, A., 2002. Processing of IKONOS Imagery for Submetre 3D Positioning and Building Extraction. ISPRS Journal of Photogrammetry and Remote Sensing 56 (3), pp. 177-194.
- Geibel, R., Stilla, U., 2000. Segmentation of Laseraltimeter data for building reconstruction: Comparison of different procedures. International Archives of Photogrammetry and Remote Sensing 33 (Part B3), pp. 326-334.
- Gross, H., Thoennessen, U., 2005. 3D Modeling of Urban Structures. Joint Workshop of ISPRS/DAGM Object Extraction for 3D City Models, Road Databases, and Traffic Monitoring CMRT05, International Archives of Photogrammetry and Remote Sensing 36 (Part 3/W24), pp. 137-142.
- Gross, H., Thoennessen, U., 2006. Extraction of Lines from Laser Point Clouds. In: Förstner, W., Steffen, R., (Eds) Symposium of ISPRS Commission III: Photogrammetric Computer Vision PCV06. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3), pp. 86-91.
- Jutzi, B., Neulist, J., Stilla, U., (2005) High-Resolution waveform acquisition and analysis for pulsed laser. In: Heipke, C., Jacobsen, K., Gerke, M. (Eds.) High-resolution earth imaging for geospatial information. International Archives of Photogrammetry and Remote Sensing 36 (Part 1 W3) (on CD).
- Kirchhof, M., Jutzi, B., Stilla, U., (2008) Iterative processing of laser scanning data by full waveform analysis in close neighborhood. In: Lichti, D., Pfeifer, N., Maas, H.-G., (Eds.) ISPRS Journal of Photogrammetry & Remote Sensing 63 (1): pp. 99-114. [doi:10.1016/j.isprsjprs.2007.08.006]
- Maas, H.-G., Vosselman, G., 1999. Two algorithms for extracting building models from raw Laser altimetry data. ISPRS Journal of Photogrammetry & Remote Sensing 54 (2-3), pp. 153-163.
- Pollefeys, M., 1999. Self-Calibration and Metric 3D-Reconstruction from Uncalibrated Image Sequences, PhD-Thesis, K. U. Leuven.
- Shan, J., Toth, C.K., (Eds.) 2008. Topographic Laser Ranging and Scanning: Principles and Processing. Boca Raton, FL: Taylor & Francis.
- Vosselman, G., Gorte, B., Sithole, G., Rabbani, T., 2004. Recognizing structure in Laser scanner point clouds. Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 46 (Part 8/W2), pp. 33-38.



Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers



Martin Weinmann^{a,*}, Boris Jutzi^a, Stefan Hinz^a, Clément Mallet^b

^a Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Englerstraße 7, 76131 Karlsruhe, Germany

^b Université Paris-Est, IGN, SRIG, MATIS, 73 avenue de Paris, 94160 Saint-Mandé, France

ARTICLE INFO

Article history:

Received 31 October 2014
Received in revised form 28 January 2015
Accepted 30 January 2015
Available online 27 February 2015

Keywords:

Point cloud
Neighborhood selection
Feature extraction
Feature selection
Classification
3D scene analysis

ABSTRACT

3D scene analysis in terms of automatically assigning 3D points a respective semantic label has become a topic of great importance in photogrammetry, remote sensing, computer vision and robotics. In this paper, we address the issue of how to increase the distinctiveness of geometric features and select the most relevant ones among these for 3D scene analysis. We present a new, fully automated and versatile framework composed of four components: (i) neighborhood selection, (ii) feature extraction, (iii) feature selection and (iv) classification. For each component, we consider a variety of approaches which allow applicability in terms of simplicity, efficiency and reproducibility, so that end-users can easily apply the different components and do not require expert knowledge in the respective domains. In a detailed evaluation involving 7 neighborhood definitions, 21 geometric features, 7 approaches for feature selection, 10 classifiers and 2 benchmark datasets, we demonstrate that the selection of optimal neighborhoods for individual 3D points significantly improves the results of 3D scene analysis. Additionally, we show that the selection of adequate feature subsets may even further increase the quality of the derived results while significantly reducing both processing time and memory consumption.

© 2015 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

Due to the increasing availability of 3D point cloud data and respective acquisition systems, the automated analysis of 3D point clouds has become a topic of great importance in photogrammetry, remote sensing, computer vision and robotics. Exploiting such data, recent investigations address a variety of different tasks such as the extraction of building structures (Vanegas et al., 2012), the recognition of power-line objects (Kim and Sohn, 2011), the extraction of roads and curbstones or road markings (Boyko and Funkhouser, 2011; Zhou and Vosselman, 2012; Guan et al., 2014), the mapping of vegetation (Wurm et al., 2014), the detection of numerous different objects (Kim and Medioni, 2011; Pu et al., 2011; Velizhev et al., 2012; Bremer et al., 2013; Serna and Marcotegui, 2014), the accessibility analysis in urban environments (Serna and Marcotegui, 2013), the creation of large-scale city models (Poullis and You, 2009; Lafarge and Mallet, 2012; Zhou and Neumann, 2013), or the semantic perception for ground robotics (Hebert et al., 2012). However, many of these tasks are

based on the results of a 3D scene analysis in terms of uniquely assigning a semantic label (e.g. *ground*, *building* or *vegetation*) to each 3D point of a given point cloud.

When addressing the task of 3D scene analysis, we have to account for the general ideas shared by many respective approaches. Typically, 3D scene analysis involves (i) the recovery of a local neighborhood for each 3D point, (ii) the extraction of geometric features based on all 3D points within the local neighborhood, and (iii) the classification of all 3D points based on the respective features. Since often as many features as possible are exploited due to a lack of knowledge, recent investigations also addressed the selection of meaningful features as additional step between feature extraction and classification (Chehata et al., 2009; Mallet et al., 2011; Khoshelham and Oude Elberink, 2012; Weinmann et al., 2013; Weinmann et al., 2014). For all steps, however, a variety of challenges results from the complexity of 3D scenes caused by irregular point sampling, varying point density and very different types of objects. Furthermore, the computational burden arising from large 3D point clouds and many available features has to be taken into account.

In this paper, we focus on individual point classification, i.e. we only exploit feature vectors for assigning class labels, since respec-

* Corresponding author.

E-mail addresses: martin.weinmann@kit.edu (M. Weinmann), boris.jutzi@kit.edu (B. Jutzi), stefan.hinz@kit.edu (S. Hinz), clement.mallet@ign.fr (C. Mallet).

<http://dx.doi.org/10.1016/j.isprsjprs.2015.01.016>

0924-2716/© 2015 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

tive improvements also represent an important issue for methods involving contextual information. Besides revisiting foundations and trends in 3D scene analysis, we also provide new insights addressing all major steps in the respective data processing. These insights are based on our previous work involving feature relevance assessment (Weinmann et al., 2013), recovery of optimal 3D neighborhoods (Weinmann et al., 2014) and large-scale capability (Weinmann et al., 2015). Resulting from these investigations, we may easily derive a fully automatic, efficient and general framework for 3D scene analysis which involves

- neighborhoods of optimal size,
- low-level geometric 3D and 2D features,
- different strategies for feature selection, and
- efficient methods for supervised classification

while preserving both reproducibility and applicability of the involved methods. Hereby, we want to emphasize that neighborhood size selection and feature extraction are strongly interleaved issues, since the distinctiveness of geometric features strongly depends on the respective neighborhood encapsulating those 3D points which are taken into consideration for feature extraction. We further extend the framework by adding several approaches to different components, so that a variety of approaches is available for each component (Fig. 1). By providing a detailed evaluation involving two standard benchmark datasets, we are able to derive general statements on the suitability of the different approaches. Since only the spatial 3D geometry in terms of an appropriate representation of object surfaces as measured counterpart of the real world serves as input, our framework is generally applicable for interpreting 3D point cloud data obtained via different acquisition techniques such as terrestrial laser scanning (TLS), mobile laser scanning (MLS) or airborne laser scanning (ALS). Furthermore, the framework may be applied for point clouds captured with 3D cameras or point clouds obtained via 3D reconstruction from images.

In the following, we first reflect related work in Section 2. Subsequently, in Section 3, we explain the single components of our framework and respective methods in detail. For demonstrating the performance of our framework, we describe the involved publicly available datasets, the conducted experiments and the respective results in Section 4. Additionally, we discuss the derived results in Section 5. Finally, in Section 6, we provide concluding remarks and suggestions for future work.

2. Related work

In this section, we reflect the related work on 3D scene analysis and group respective approaches according to the single steps of 3D scene analysis.

2.1. Neighborhood selection

For being able to describe the local 3D structure around a given point \mathbf{X} via geometric features, a respective neighborhood

definition encapsulating all considered 3D points is required. Generally, different strategies may be applied for defining the local neighborhood \mathcal{N} around a given 3D point \mathbf{X} . Among these, the most commonly applied neighborhood definitions are represented by

- a spherical neighborhood definition \mathcal{N}_s , where the neighborhood is formed by all 3D points in a sphere of fixed radius $r_s \in \mathbb{R}$ around the point \mathbf{X} (Lee and Schenk, 2002),
- a cylindrical neighborhood definition \mathcal{N}_c , where the neighborhood is formed by all those 3D points whose 2D projections onto a plane (e.g. the ground plane) are within a circle of fixed radius $r_c \in \mathbb{R}$ around the projection of \mathbf{X} (Filin and Pfeifer, 2005), and
- a neighborhood definition \mathcal{N}_k based on a fixed number of the $k \in \mathbb{N}$ closest neighbors of \mathbf{X} in 3D (Linsen and Prutzsch, 2001) or in 2D (Niemeyer et al., 2014).

Hereby, the third definition also results in a spherical neighborhood if 3D distances are evaluated for finding the closest neighbors, but – in contrast to the first definition – a variable absolute size is taken into account. Whereas these definitions with a constant scale parameter (i.e. either a fixed radius or a constant value k) across all 3D points provide a straightforward solution to neighborhood selection, it has to be taken into account that the scale parameter is typically selected with respect to heuristic or empiric knowledge on the scene and thus specific for each dataset. Furthermore, the scale parameter may not be identical across all considered 3D points, since it intuitively rather depends on the local 3D structure and point density. This holds particularly for MLS data, where due to the process of data acquisition dense and accurate 3D point clouds with significant variations in point density may be expected.

In order to avoid strong assumptions on local 3D neighborhoods, more recent investigations focused on introducing an optimal neighborhood size for each individual 3D point and thus increasing the distinctiveness of derived features. Most of the presented approaches exploit the idea of a neighborhood based on the k closest 3D points and optimize k for each individual 3D point. This optimization may for instance be based on iterative schemes relating neighborhood size to curvature, point density and noise of normal estimation (Mitra and Nguyen, 2003; Lalonde et al., 2005) which is particularly relevant for rather densely sampled and thus almost continuous surfaces. Other alternatives also account for more cluttered surface representations and are based on surface variation (Pauly et al., 2003; Belton and Lichti, 2006), dimensionality-based scale selection (Demantké et al., 2011) or eigenentropy-based scale selection (Weinmann et al., 2014). Even though deriving individual neighborhoods causes additional effort, the need for such concepts clearly becomes visible when considering the suitability of respective geometric features for neighborhoods of different size (Blomley et al., 2014) or the significant improvement in comparison to neighborhoods with a constant scale parameter across all 3D points (Weinmann et al., 2014).

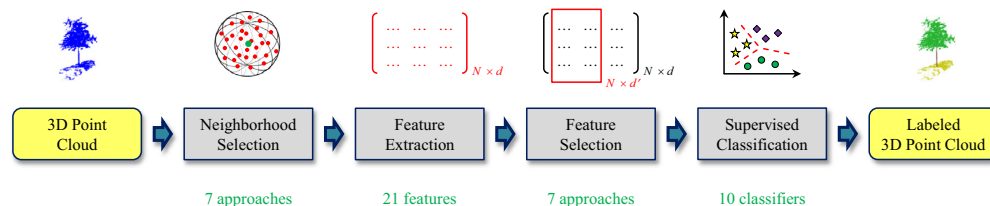


Fig. 1. The proposed framework and the quantity of attributes/approaches taken into account for evaluation.

Instead of focusing on the concept of an optimal 3D neighborhood for each individual 3D point, geometric features may also be derived for different scales, since – by varying the size of the local neighborhood and calculating all features for each scale – it is possible to involve information on how the local 3D geometry behaves across scales (Brodu and Lague, 2012) which, in turn, may support the discrimination between different classes and thus improve the classification results (Niemeyer et al., 2014). In this context, neighborhoods of small size are suitable to analyze fine details of the local 3D structure, whereas increasing the size of the local neighborhood is similar to applying a smoothing filter (Pauly et al., 2003), since with increasing neighborhood size each individual 3D point will contribute less to the surface variation estimate. By directly providing all features over a range of scales to the classifier, a training procedure can later be used to define which combination of scales allows the best separation of different classes (Brodu and Lague, 2012). However, the number of scales and their distance are typically selected based on heuristic or empirical knowledge on the scene and thus specific for each dataset (Brodu and Lague, 2012; Niemeyer et al., 2014; Schmidt et al., 2014). Consequently, it would be desirable to combine generic approaches for selecting a single optimal scale with a generic multi-scale representation. For the latter, a transfer of concepts from 2D scale space theory could be applied in analogy to scale space representations for detecting 3D interest points (Yu et al., 2013) or 3D interest regions (Unnikrishnan and Hebert, 2008). More generally, the extraction of geometric features may not only rely on different scale parameters, but also on different entities such as points and regions (Xiong et al., 2011; Xu et al., 2014). Note that this is in analogy to the labeling paradigm, where the labels of single 3D points and their closest neighbors are exploited for a structured prediction.

In contrast to considerations on point-level, 3D scene analysis may also be based on voxels or 3D segments. For instance, it has recently been proposed to exploit voxels which carry geometric, textural and color information collected from airborne imagery and point clouds derived via dense image matching (Gerke and Xiao, 2013). Furthermore, a multi-stage inference procedure exploiting a hierarchical segmentation based on voxels, blocks and pillars has recently been presented (Hu et al., 2013). Alternatively, a 3D segmentation of the given point cloud data could be introduced in order to support 3D scene analysis (Vosselman, 2013), where attributes of single 3D points can be exploited for a better separation of neighboring objects (Oude Elberink and Kemboi, 2014). Even the combination of generic scale selection and voxelization has been proposed with an approach based on supervoxels resulting from an oversegmentation of a 3D point cloud (Lim and Suter, 2009), where the size of supervoxels is determined based on an iterative scheme involving curvature, point density, noise and point intensities. However, results based on a simple voxelization of 3D space strongly depend on the empirically or heuristically selected voxel size and, typically, a generic segmentation significantly increases the computational burden.

2.2. Feature extraction

For extracting discriminative features from point clouds with possibly varying point density, a variety of approaches has been proposed. Focusing on the characterization of a local 3D neighborhood around a given point \mathbf{X} , one strategy consists of deriving suitable 3D descriptors such as spin image descriptors (Johnson and Hebert, 1999) which are based on spinning a planar image patch around the surface normal at a given 3D point \mathbf{X} and counting the number of points per pixel. A powerful alternative is represented by shape distributions (Osada et al., 2002) which are based on randomly sampling geometric relations such as distances and

angles and may thus be applied for characterizing the local neighborhood of a 3D point \mathbf{X} (Blomley et al., 2014). A similar strategy has been presented with Point Feature Histograms (PFHs) (Rusu et al., 2008; Rusu et al., 2009) which are based on sampling geometric relations such as point distances and angular variations between the closest neighbors relative to the local normal vector at \mathbf{X} into histograms. In contrast, the Signature of Histograms of Orientations (SHOT) descriptor (Tombari et al., 2010) is based on exploiting a spherical grid centered on the point \mathbf{X} , where each 3D bin of the grid is represented as weighted histogram of normals. For all these approaches, however, single entries of the resulting feature vector are hardly interpretable.

Consequently, a variety of approaches for 3D scene analysis relies on the 3D structure tensor which is represented by the 3D covariance matrix derived from the 3D coordinates of all points within the local neighborhood of a given 3D point \mathbf{X} . Based on this 3D structure tensor, the analytical consideration of its eigenvalues has been proposed for characterizing specific shape primitives (Jutzi and Gross, 2009). Furthermore, the eigenvalues may be exploited to derive a set of local 3D shape features (West et al., 2004; Pauly et al., 2003) which contains more intuitive descriptions (e.g. about linear, planar or volumetric structures) and is therefore nowadays commonly applied in lidar data processing. Such features are often complemented with other geometric features based on angular statistics (Munoz et al., 2009), height and local plane characteristics (Mallet et al., 2011), basic properties of the neighborhood and characteristics of a 2D projection (Weinmann et al., 2013), descriptors involving surface properties, slope, height characteristics, vertical profiles and 2D projections (Guo et al., 2014) or point distances and height differences (Waldhauser et al., 2014). Moreover, eigenvalue-based 3D features are often combined with additional full-waveform and echo-based features (Chehata et al., 2009; Mallet et al., 2011; Niemeyer et al., 2012; Schmidt et al., 2014; Waldhauser et al., 2014).

2.3. Feature selection

Whereas often as many features as possible are exploited in order to compensate a lack of knowledge, it has to be considered that some of these features may be more and others less suitable. Consequently, the interest in feature selection techniques emerged for finding compact and robust subsets of relevant and informative features in order to gain predictive accuracy, improve computational efficiency with respect to both time and memory consumption, and retain meaningful features (Guyon and Elisseeff, 2003; Saeys et al., 2007; Liu et al., 2010). Respective strategies for feature selection can be categorized into filter-based methods, wrapper-based methods and embedded methods. The main characteristics, advantages and disadvantages of these strategies are provided in Table 1. For more details and further reading, we recommend a comprehensive review of feature selection techniques (Saeys et al., 2007).

In 3D point cloud processing, however, feature selection has only rarely been applied and different aspects have to be taken into account. Since both wrapper-based methods and embedded methods involve a classifier, they tend to yield a better performance than filter-based methods. Wrapper-based methods typically involve a classifier either via Sequential Forward Selection (SFS) which is based on finding the feature yielding the highest predictive accuracy and successively adding the feature that improves performance the most, or via Sequential Backward Elimination (SBE) which is based on starting with the whole feature set and repeatedly deleting the feature that reduces performance the least (Mallet et al., 2011; Khoshelham and Oude Elberink, 2012). Whereas SFS tends to result in smaller feature subsets, SBE tends to yield larger feature subsets. Due to the involved classifier, however, such

Table 1
Feature selection (FS) techniques and their main characteristics.

Strategy	Advantages	Disadvantages	Examples
Filter-based FS			
–Univariate	Simple Fast Classifier-independent	No feature dependencies No interaction with the classifier	Fisher score Information gain Symmetrical uncertainty
–Multivariate	Classifier-independent Models feature dependencies Faster than wrapper-based FS	Slower than univariate techniques No interaction with the classifier	CFS FCBF mRMR
Wrapper-based FS	Interaction with the classifier Models feature dependencies	Classifier-dependent Computationally intensive	SFS SBE Simulated annealing
Embedded FS	Interaction with the classifier Models feature dependencies Faster than wrapper-based FS	Classifier-dependent	Random Forests AdaBoost

methods reveal a rather high computational effort. In contrast, embedded methods such as Random Forests or AdaBoost provide the capability of dealing with exhaustive feature sets as input and letting the classifier internally select a suitable feature subset during the training phase (Chehata et al., 2009; Tokarczyk et al., 2013). However, both wrapper-based methods and embedded methods provide feature subsets which are only optimized with respect to the applied classifier. In contrast, filter-based methods are classifier-independent (thus they tend to provide a slightly weaker performance) and only exploit a score function directly based on the training data which, in turn, results in simplicity and efficiency (Weinmann et al., 2013; Weinmann et al., 2014). Depending on the use of (i) only feature-class relations for determining relevant features or (ii) both feature-class relations and feature-feature relations for also removing redundancy, such techniques may further be categorized into univariate and multivariate techniques.

2.4. Classification

Concerning 3D scene analysis, the most effort has recently been spent in the design and improvement of the classification procedure. In order to obtain interpretable results, supervised classification is commonly applied which involves a set of training examples. Generally, approaches for supervised classification can be categorized into (i) individual point classification where each 3D point is classified based on the respective feature vector or (ii) contextual classification where each 3D point is classified based on the respective feature vector as well as the labels of neighboring 3D points.

For 3D scene analysis based on individual point classification, a variety of standard classification approaches has been applied such as classical Maximum Likelihood classifiers based on Gaussian Mixture Models (Lalonde et al., 2005), Support Vector Machines (Secord and Zakhor, 2007), AdaBoost (Lodha et al., 2007), a cascade of binary classifiers (Carlberg et al., 2009), Random Forests (Chehata et al., 2009) and Bayesian Discriminant Classifiers (Khoshelham and Oude Elberink, 2012). Whereas many respective approaches are publicly available in various software tools and furthermore easily applicable, the derived results appear to be noisy since it is not taken into account that semantic labels of neighboring 3D points tend to be correlated.

In order to account for correlated labels of neighboring 3D points, contextual classification approaches also involve relationships among 3D points in a local neighborhood which have to be inferred from the training data. Hereby, it is important to notice that the local neighborhood for inferring relationships among 3D points is typically different from the one used for feature extraction. Addressing different strategies for modeling interactions

between neighboring 3D points, for 3D scene analysis, respective methods have been proposed with Associative and non-Associative Markov Networks (Munoz et al., 2009; Shapovalov et al., 2010; Najafi et al., 2014), Conditional Random Fields (Niemeyer et al., 2012; Niemeyer et al., 2014), multi-stage inference procedures focusing on point cloud statistics and relational information over different scales (Xiong et al., 2011), and spatial inference machines modeling mid- and long-range dependencies inherent in the data (Shapovalov et al., 2013). Due to focusing on a smooth labeling, contextual classification approaches tend to yield higher classification accuracies than approaches for individual point classification.

Even though a spatially smooth labeling is desirable, the use of contextual classification approaches typically causes a high computational effort for modeling interactions between neighboring 3D points and thus often tends to be impracticable. More specifically, an exact inference is computationally intractable, since the training data is limited and modeling relationships across the whole training data results in too many degrees of freedom. For this reason, approximate inference techniques have been proposed which infer the spatial relationships among 3D points within a local 3D neighborhood. Whereas modeling long-range dependencies is feasible in order to improve the predictive power of the classifier and rather time-consuming, many approaches focus on modeling only short-range interactions in order to improve computational efficiency. Generally, however, there is no indication towards an optimal inference strategy, and an efficient alternative solution may consist of a decoupling in terms of first labeling without smoothness constraint via individual point classification and subsequently smoothing the labels via probabilistic relaxation or smooth labeling techniques (Schindler, 2012). This further motivates to investigate sources for potential improvements in individual point classification.

3. Methodology

In this section, we present our framework shown in Fig. 1 and explain its components as well as respective methods in detail. These components address neighborhood selection (Section 3.1), feature extraction (Section 3.2), feature selection (Section 3.3) and classification (Section 3.4).

3.1. Neighborhood selection

Generally, neighborhood selection may focus on the selection of a single scale or the selection of multiple scales. While the latter strategy accounts for the behavior of features across different scales by delaying the decision about the suitability of a neighborhood to the classifier, there are many further aspects which have to be taken into consideration. For instance, it is important to know

how the scale space is designed, how many scales are involved and how the distance between scales is determined. Such decisions are still typically based on a heuristic or empiric selection (Brodu and Lague, 2012; Niemeyer et al., 2014; Schmidt et al., 2014) and thus specific for each dataset. In contrast, the selection of a single scale also offers a generic selection which even allows a generalization between datasets since variations in point density may be handled. For these reasons, we focus on a single-scale approach.

Accordingly, for being able to describe the local 3D structure around a given 3D point $\mathbf{X} = (X, Y, Z)^T \in \mathbb{R}^3$, the fundamental task consists of recovering the neighboring 3D points. This, in turn, involves (i) a suitable neighborhood definition, (ii) an efficient recovery of neighboring 3D points and (iii) an adequate parameterization in terms of neighborhood size. In the following, we consider these three issues in detail.

3.1.1. Neighborhood definition

The adequate choice of a neighborhood definition and the respective scale parameter certainly depends on the characteristics of the respective point cloud data. Whereas, according to Section 2, the spherical and cylindrical neighborhood definitions \mathcal{N}_s and \mathcal{N}_c directly involve empiric or heuristic knowledge on the scene in order to obtain a suitable radius, the neighborhood definition \mathcal{N}_k accounts for more flexibility in case of varying point density. Since we intend to provide a versatile framework without being restricted to a specific dataset, we employ the neighborhood definition \mathcal{N}_k based on the k closest neighbors of a given 3D point \mathbf{X} . Thereby, we select the closest neighbors based on 3D distances and thus a spherical neighborhood with flexible radius. Consequently, we first need to involve a strategy to recover the closest neighbors of \mathbf{X} and then select a suitable number k of considered neighbors.

3.1.2. Neighborhood recovery

In order to find the closest neighbors for a given 3D point \mathbf{X} , the most commonly used approach is based on a Kd-tree (Friedman et al., 1977). Generally, a Kd-tree represents a compact hierarchical data structure for point sets sampled from a K -dimensional manifold and thus allows an efficient recovery of neighboring points. If required, the computational efficiency can further be increased by replacing the exact nearest neighbor search by an approximate nearest neighbor search (Arya et al., 1998). This, in turn, introduces little loss in accuracy since non-optimal neighbors may be returned. A good trade-off between computational efficiency and accuracy has been presented with the Fast Library for Approximate Nearest Neighbors (FLANN) (Muja and Lowe, 2009) which is publicly available (e.g. in OpenCV) and based on either (i) searching hierarchical K -means trees with a priority search order or (ii) using multiple randomized Kd-trees.

3.1.3. Neighborhood parameterization

When selecting a suitable number k of considered neighbors, the simplest and straightforward approach would be to select a fixed value k for all points of the point cloud. This way, the choice of k still relies on empiric or heuristic knowledge on the scene. Intuitively, however, we may prefer a choice where the parameter k (which is also commonly referred to as scale) is more flexible and thus allowed to vary within a dataset. This idea is further strengthened by the fact that k certainly depends on the respective 3D structures and the local point density within a dataset. Consequently, a generic method for deriving locally optimal neighborhoods would be desirable. A respective method would even completely avoid the use of a priori knowledge on the scene.

In order to obtain suitable and individual neighborhoods, seminal work (Pauly et al., 2003; Demantké et al., 2011) is based on the well-known 3D structure tensor $\mathbf{S} \in \mathbb{R}^{3 \times 3}$ with

$$\mathbf{S} = \frac{1}{k+1} \sum_{i=0}^k (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \quad (1)$$

which represents a 3D covariance matrix constructed for a given 3D point $\mathbf{X} = \mathbf{X}_0$ by involving its k closest neighbors \mathbf{X}_i with $i = 1, \dots, k$. The geometric center $\bar{\mathbf{X}}$ is thereby defined as

$$\bar{\mathbf{X}} = \frac{1}{k+1} \sum_{i=0}^k \mathbf{X}_i \quad (2)$$

and may thus slightly vary from the considered 3D point \mathbf{X}_0 . Since the 3D structure tensor represents a symmetric positive-definite matrix, its three eigenvalues exist, are non-negative and correspond to an orthogonal system of eigenvectors. Further assuming that there may not necessarily be a preferred variation with respect to the eigenvectors, we consider the general case of a structure tensor with rank 3. Consequently, the three eigenvalues λ_1, λ_2 and λ_3 with $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ represent the extent of a 3D ellipsoid along its principal axes. Thus, the eigenvalues may be exploited in order to characterize the local 3D shape. In the context of neighborhood size selection, three approaches based on the eigenvalues of the structure tensor have been proposed.

Firstly, the eigenvalues can be exploited in order to estimate the local surface variation with

$$C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \quad (3)$$

which is also referred to as change of curvature. By starting with small values and successively increasing the neighborhood size, i.e. the scale parameter $k \in \mathbb{N}$, a critical neighborhood size and thus a respective value for k corresponds to a significant increase of C_λ (Pauly et al., 2003; Belton and Lichti, 2006), since occurring jumps indicate strong deviations in the normal direction.

Secondly, the eigenvalues can be exploited in order to derive the dimensionality features represented by linearity L_λ , planarity P_λ and scattering S_λ with

$$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1} \quad (4)$$

$$P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1} \quad (5)$$

$$S_\lambda = \frac{\lambda_3}{\lambda_1} \quad (6)$$

which represent 1D, 2D and 3D features. More specifically, the dimensionality features $L_\lambda, P_\lambda, S_\lambda \in \mathbb{R}$ with $L_\lambda, P_\lambda, S_\lambda \in [0, 1]$ sum up to 1 and thus satisfy two of three probability axioms according to (Kolmogorov, 1933). Further taking into account that quasiprobability distributions generally relax the third axiom addressing the junction of mutually disjoint random events, the dimensionality features may be considered as the “probabilities” of a 3D point to be labeled as 1D, 2D or 3D structure (Demantké et al., 2011). Accordingly, the task of finding a suitable neighborhood size may be transferred to favoring one dimensionality the most which, in turn, corresponds to minimizing a measure of unpredictability given by the Shannon entropy (Shannon, 1948) as

$$E_{\text{dim}} = -L_\lambda \ln(L_\lambda) - P_\lambda \ln(P_\lambda) - S_\lambda \ln(S_\lambda) \quad (7)$$

across different scales $k \in \mathbb{N}$. For this purpose, in the original implementation (Demantké et al., 2011), the neighborhood radius r has been taken into account and the interval $r \in [r_{\min}, r_{\max}]$ has been sampled in 16 scales. Thereby, the radii are not linearly increased since the radius of interest is usually closer to r_{\min} . Thus, the optimal neighborhood size corresponds to the radius which yields the minimum Shannon entropy. Since the values r_{\min} and r_{\max} depend on various characteristics of the given data, they are specific for each dataset. In order to avoid a heuristic parameter selection, directly

varying the scale parameter k (e.g. between $k_{\min} = 10$ and $k_{\max} = 100$ with $\Delta k = 1$) has recently been proposed (Weinmann et al., 2014) which results in an increase of the computational effort.

Thirdly, the eigenvalues can directly be exploited in order to estimate the order/disorder of 3D points within the local 3D neighborhood (Weinmann et al., 2014). For this purpose, the three eigenvalues λ_1, λ_2 and λ_3 are normalized by their sum Σ_i . The normalized eigenvalues $e_i = \lambda_i / \Sigma_i$ with $i \in \{1, 2, 3\}$ and $e_i \in [0, 1]$ thus sum up to 1 and, consequently, the measure of eigenentropy is defined via the Shannon entropy as

$$E_i = -e_1 \ln(e_1) - e_2 \ln(e_2) - e_3 \ln(e_3) \quad (8)$$

where the occurrence of eigenvalues identical to zero has to be avoided by adding an infinitesimal small value ε . The eigenentropy represents a measure describing the order/disorder of 3D points within the local 3D neighborhood. Favoring a minimum disorder of 3D points corresponds to minimizing E_i across different scales $k \in \mathbb{N}$. Accordingly, the optimal neighborhood size can be determined by varying the scale parameter k and selecting the value which yields the minimum Shannon entropy. In accordance with other investigations (Demantké et al., 2011; Weinmann et al., 2014), we consider relevant statistics to start with $k_{\min} = 10$ neighbors and vary k with $\Delta k = 1$ up to a relatively high number of $k_{\max} = 100$ neighbors. This approach for neighborhood size selection is generally applicable, since it does neither involve parameters which are specific for each dataset nor rely on the assumption of particular shapes being present in the observed scene.

Even though such approaches for optimal neighborhood size selection cause additional computational effort, involving optimal neighborhoods has a significantly positive impact on 3D scene analysis (Weinmann et al., 2014) and should therefore be taken into account. More specifically, by exploiting optimal neighborhoods – which may be different for each individual 3D point – we may expect that, in the subsequent step of feature extraction, the distinctiveness of geometric features calculated from the neighboring points is increased.

3.2. Feature extraction

Most of the publicly available 3D point cloud datasets only contain geometric information in terms of spatial 3D coordinates. For this reason, we only involve geometric features in our investigations. Consequently, all types of point clouds with an adequate point density may serve as input for our framework. In particular, point clouds acquired via mobile laser scanning or dense matching provide an appropriate representation of object surfaces as measured counterpart of the real world. Further information such as intensity/color or full-waveform features can easily be added, but this does not influence the presented methodology and is therefore not in the scope of this work.

Considering a given 3D point \mathbf{X} , respective geometric features typically rely on a local neighborhood. Thus, neighborhood selection and feature extraction are interleaved issues, since the distinctiveness of geometric features strongly depends on the respective neighborhood encapsulating those 3D points which are taken into consideration for feature extraction. Whereas some features will be more distinctive for larger neighborhoods (e.g. features addressing planarity or curvature), other features will be more distinctive for smaller neighborhoods (e.g. features addressing fine details of the local 3D structure). In order to avoid optimizing the neighborhood size for each feature, we focus on the concept of decoupling neighborhood selection and feature extraction. Consequently, we treat feature extraction independent from neighborhood selection since an interaction of these components in terms of optimization is not taken into account, and we furthermore assume that the

consideration of locally adaptive neighborhoods is sufficient for retrieving distinctive features. Thus, in analogy to our previous work (Weinmann et al., 2013; Weinmann et al., 2014; Weinmann et al., 2015), we exploit eigenentropy-based scale selection and derive fundamental geometric 3D properties as well as local 3D shape features from this local 3D neighborhood (Section 3.2.1). Furthermore, we also consider 2D neighborhoods resulting from a 2D projection and derive geometric 2D features (Section 3.2.2). In total, this yields a set of 21 low-level geometric 3D and 2D features which are briefly described in the following subsections. A respective toolbox¹ (Matlab, C++ and binaries) for calculating these geometric features has recently been released (Weinmann et al., 2015). Since the geometric features address different quantities with possibly different units, a normalization across all feature vectors is involved which maps the values of each dimension to the interval $[0, 1]$.

3.2.1. 3D features

A variety of 3D features can be derived by considering basic properties of the local 3D neighborhood and local 3D shape features arising from the spatial arrangement of 3D points within the neighborhood.

3.2.1.1. Geometric 3D properties. For 3D scene analysis, valuable information about a given 3D point $\mathbf{X} = (X, Y, Z)^T$ for which the XY -plane represents a horizontally oriented plane might arise from its absolute height Z . Additionally, fundamental geometric 3D properties of the local 3D neighborhood are represented by the radius r_{k-NN} of the spherical neighborhood encapsulating the k closest neighbors as well as the maximum difference ΔZ_{k-NN} and standard deviation $\sigma_{Z,k-NN}$ of height values within the neighborhood. Further basic properties of the local neighborhood arise from the local point density D (Weinmann et al., 2013) given by

$$D = \frac{k+1}{\frac{4}{3} \pi r_{k-NN}^3} \quad (9)$$

and the verticality $V = 1 - n_z$ (Demantké et al., 2012) which is derived from the vertical component n_z of the normal vector $\mathbf{n} \in \mathbb{R}^3$.

3.2.1.2. Local 3D shape features. For a given 3D point \mathbf{X} and its k closest neighbors, the respective derived normalized eigenvalues e_i with $i \in \{1, 2, 3\}$ may be exploited in order to obtain a set of 8 local 3D shape features (West et al., 2004; Pauly et al., 2003). This feature set encapsulates linearity L_i , planarity P_i , scattering S_i , omnivariance O_i , anisotropy A_i , eigenentropy E_i , sum Σ_i of eigenvalues and change of curvature C_i according to

$$L_i = \frac{e_1 - e_2}{e_1} \quad (10)$$

$$P_i = \frac{e_2 - e_3}{e_1} \quad (11)$$

$$S_i = \frac{e_3}{e_1} \quad (12)$$

$$O_i = \sqrt[3]{e_1 e_2 e_3} \quad (13)$$

$$A_i = \frac{e_1 - e_3}{e_1} \quad (14)$$

$$E_i = -\sum_{i=1}^3 e_i \ln(e_i) \quad (15)$$

$$\Sigma_i = e_1 + e_2 + e_3 \quad (16)$$

$$C_i = \frac{e_3}{e_1 + e_2 + e_3} \quad (17)$$

¹ This toolbox is available at <http://www.ipf.kit.edu/code.php>.

which are meanwhile commonly applied in 3D lidar data processing.

3.2.2. 2D features

Generally, point clouds representing an observed 3D scene do not provide a completely random point distribution since the 3D points are the measured or derived counterpart of real object surfaces. For instance, urban environments are composed of a variety of man-made objects, where geometric constraints in terms of symmetry and orthogonality are likely to be satisfied. Since such man-made objects often tend to provide almost perfectly vertical structures (e.g. building façades, poles, traffic signs or curbstone edges), we may also involve geometric features resulting from a 2D projection of the 3D point cloud onto a horizontally oriented plane, i.e. the XY -plane. Such 2D features might reveal complementary information compared to the aforementioned 3D features and can also be categorized into different groups.

3.2.2.1. Geometric 2D properties. In analogy to the 3D case, basic geometric properties are given by the radius $r_{k-NN,2D}$ of the circular neighborhood defined by a 2D point and its k closest neighbors or the local point density D_{2D} (Lari and Habib, 2012). In order to assess 2D properties corresponding to the optimized neighborhood, we again use the closest neighbors based on 3D distances.

3.2.2.2. Local 2D shape features. Exploiting the XY -coordinates of a point \mathbf{X} and its k closest neighbors, the 2D structure tensor \mathbf{S}_{2D} can be derived in analogy to the 3D structure tensor. From its eigenvalues, the sum $\Sigma_{\lambda,2D}$ and the ratio $R_{\lambda,2D}$ of eigenvalues may be calculated and exploited as 2D features.

3.2.2.3. Features based on an accumulation map. Since the aforementioned 2D features are based on the spherical neighborhood encapsulating a point \mathbf{X} and its k closest neighbors, we may also involve neighborhoods resulting from a spatial binning. For this purpose, it has been proposed to introduce a second neighborhood definition by discretizing the 2D projection plane and deriving a 2D accumulation map with quadratic bins (Monnier et al., 2012; Weinmann et al., 2013), e.g. with a side length of 0.20... 0.25 m. Within each bin, respective features arise from the number M of points falling into the bin as well as the maximum height difference ΔZ and the standard deviation σ_z of height values within the bin.

3.3. Feature selection

Particularly when dealing with many features or many training examples, simpler and more efficient methods are favorable and, consequently, filter-based methods are often applied for selecting a subset of relevant features. These filter-based methods address simple and more intuitive relations between features and classes and possibly also among features (i.e. relations based on well-known concepts of distance, information, dependency or consistency), whereas the relations exploited by embedded methods are more sophisticated and thus hardly interpretable. By representing these relations in the form of score functions, one could argue that – in the sense of statistical learning or machine learning – embedded methods are most appropriate, since the respective score function focuses on minimizing the classification error. However, embedded methods would directly introduce a dependency between selected features and the settings of a classifier, e.g. the number of involved weak learners, their type and the (ideally high) number of considered choices per variable. In order to avoid an exhaustive classifier tuning and thus preserve applicability for non-expert users, we focus on filter-based methods and accept if these tend to provide a (slightly) weaker performance. Since we

apply both univariate and multivariate filter-based methods, we briefly explain the basic ideas in the following subsections. Most of these techniques, however, should be conducted for training data with an equal number of training examples per class in order to avoid a bias in feature selection.

3.3.1. Univariate filter-based feature selection

A univariate filter-based feature selection method relies on a score function which evaluates feature-class relations in order to discriminate between relevant and irrelevant features. More specifically, the score function evaluates the relation between the vector containing values of a single feature across all observations and the respective label vector. Thereby, the score function may address different intrinsic properties of the given training data such as distance, information, dependency or consistency. Among a variety of possible score functions addressing a specific intrinsic property (Guyon and Elisseeff, 2003; Zhao et al., 2010), the most popular ones are represented by simple metrics such as Pearson correlation coefficient (Pearson, 1896), Gini index (Gini, 1912), Fisher score (Fisher, 1936), information gain (Quinlan, 1986) or symmetrical uncertainty (Press et al., 1988). Since some of these score functions are only defined for discrete features, a respective discretization of continuous-valued features is introduced if required (Fayyad and Irani, 1993). Thus, a variety of score functions allows to rank the extracted features according to their relevance. By exploiting a separate ranking with respect to different score functions and subsequently assessing the mean rank, we recently proposed a general relevance metric addressing different intrinsic properties of the given training data (Weinmann et al., 2013).

3.3.2. Multivariate filter-based feature selection

A multivariate filter-based feature selection method relies on a score function which evaluates both feature-class and feature-feature relations in order to discriminate between relevant, irrelevant and redundant features. A respective score function has been proposed with Relief (Kononenko, 1994). However, the score function may also be based on standard score functions applied for univariate filter-based feature selection. For instance, an approach taking the symmetrical uncertainty as correlation metric is represented by Correlation-based Feature Selection (CFS) (Hall, 1999). Considering random variables X_i for the features and C for the class labels and further defining $\bar{\rho}_{XC}$ as average correlation between features and classes as well as $\bar{\rho}_{XX}$ as average correlation between different features, the relevance R of a feature subset comprising n features results in

$$R(X_{1..n}, C) = \frac{n\bar{\rho}_{XC}}{\sqrt{n + n(n-1)\bar{\rho}_{XX}}} \quad (18)$$

which can be maximized by searching the feature subset space (Hall, 1999), i.e. by iteratively adding a feature to the feature subset (forward selection) or removing a feature from the feature subset (backward elimination) until R converges to a stable value.

For comparison only, we also consider feature selection exploiting a Fast Correlation-Based Filter (FCBF) (Yu and Liu, 2003) which involves heuristics and thus does not meet our intention of a fully generic methodology. For deciding whether features are relevant to the class or not, a typical feature ranking based on symmetrical uncertainty is conducted in order to determine the feature-class correlation. If the symmetrical uncertainty is above a certain threshold, the respective feature is considered to be relevant. For deciding whether a relevant feature is redundant or not, the symmetrical uncertainty among features is compared to the symmetrical uncertainty between features and classes in order to remove redundant features and only keep predominant features.

Finally, we also apply an approach addressing the aims of both univariate and multivariate filter-based feature selection methods. Whereas univariate filter-based methods focus on selecting the best-ranked features with the highest relevance (i.e. maximal relevance selection), multivariate filter-based methods focus on selecting features with the minimal redundancy (i.e. minimal redundancy selection). Accordingly, an approach combining two constraints for (i) minimal redundancy selection and (ii) maximal relevance selection has been presented with the minimal-redundancy-maximal-relevance (mRMR) criterion (Peng et al., 2005). However, a remaining issue is how to determine the optimal number of features which can be done either heuristically or by applying wrapper-based schemes. Since we focus on efficiency and therefore do not want to apply a classifier-dependent feature selection, the method of choice involves heuristics and we select a set comprising 10 features. Thus, the approach does also not meet our intention of a fully generic, classifier-independent method and, consequently, it only serves for comparison.

3.4. Classification

In the last step, either all or only the selected features are provided to a classifier which returns a respective assignment to one of the specified (semantic) classes. For a wide range of applications, most approaches focus on a supervised classification scheme, where the fundamental idea consists of exploiting given training data in order to train a classifier which afterwards is able to generalize to new data. Thereby, the training data is represented by a set \mathcal{X} of training examples which, in turn, consist of an assignment between a feature vector in a d -dimensional feature space and a respective class label. In contrast, the test set \mathcal{Y} containing new data to be classified may only consist of feature vectors in the d -dimensional feature space.

For the sake of applicability in terms of simplicity, efficiency and reproducibility, we focus on individual point classification for which many respective approaches are available in a variety of software tools. Since different learning principles may be involved for inferring a function between feature vectors and class labels in the training phase, we use a variety of classifiers and briefly present the main ideas in the following subsections. Hereby, we take into account that an unbalanced distribution of training examples per class in the training set may often have a detrimental effect on the training process (Criminisi and Shotton, 2013). In order to avoid this, we introduce a class re-balancing by randomly sampling the same number of training examples per class which yields a reduced training set. Thus, end-users will not only get an impression of the performance of single approaches, but also a comprehensive comparison.

3.4.1. Instance-based learning

Instance-based learning does neither require parameter estimation nor the assumption of a certain model, since unseen feature vectors are directly compared to the feature vectors in the training set. Accordingly, a similarity metric has to be defined which may be based on the Euclidean distance, a general Minkowski metric or other distance metrics. A very simple method and straightforward example for instance-based learning is represented by a Nearest Neighbor (NN) classifier which assigns each feature vector the class label of the most similar training example. The more general definition represented by a k Nearest Neighbor (k -NN) classifier (Cover and Hart, 1967) selects the k nearest samples in the training data and classifies according to the majority vote of their class labels.

3.4.2. Rule-based learning

Rule-based learning focuses on the representation of acquired knowledge in terms of (mostly binary) decisions. As most prominent example, decision trees (DTs) conduct a series of simple tests which are organized hierarchically in a tree structure (Quinlan, 1986). The construction of a decision tree is typically based on a top-down strategy, where at each step the variable is chosen which best splits the given training examples. This recursive partitioning strongly depends on the definition of a split function and a respective stopping criterion. For both criteria, we use standard settings.

3.4.3. Probabilistic learning

Probabilistic learning focuses on deriving an explicit underlying probabilistic model and inferring the most probable class label for each observed feature vector. The Naive Bayesian (NB) classifier (John and Langley, 1995), for instance, is a probabilistic classifier which is based on Bayes' theorem and the naive assumption of all features being conditionally independent. Consequently, a set of class probabilities and conditional probabilities for the occurrence of a class given a specific class label have to be determined based on the training set \mathcal{X} . Thus, in the classification process, a new feature vector of a test set \mathcal{Y} is assigned the most likely class label. However, since conditional independence is assumed, correlated features cannot be modeled appropriately. Alternatively, a classical maximum likelihood (ML) classifier can be derived by considering distribution-based Bayesian Discriminant Analysis. In the training phase, a multivariate Gaussian distribution is fitted to the given training data, i.e. the parameters of a Gaussian distribution are estimated for each class by parameter fitting. For a Linear Discriminant Analysis (LDA) classifier, the same covariance matrix is assumed for each class and therefore only the means vary. For a Quadratic Discriminant Analysis (QDA) classifier, the covariance matrix of each class may also vary. For classifying a new feature vector, the probability of belonging to the different classes is evaluated, and the class with maximum probability is assigned.

3.4.4. Max-margin learning

Max-margin learning focuses on maximizing the distance between samples of different classes in the feature space. A respective approach has been presented with Support Vector Machines (SVMs) (Cortes and Vapnik, 1995). In general, a SVM is a binary classifier trained to linearly separate two classes by constructing a hyperplane or a set of hyperplanes in a high-dimensional feature space. However, often a linear separation in the feature space is not possible and hence a kernel function is introduced which implicitly maps the training data into a new feature space of higher dimensionality where the data is linearly separable. For solving the problem of multi-class classification, we apply a SVM classifier composed of several binary SVMs and provided in the LIBSVM package (Chang and Lin, 2011). This classifier is based on a one-against-one approach and a (Gaussian) radial basis function (RBF) as kernel. Thus, for each pair of classes, a SVM is trained to distinguish samples of one class from samples of the other class. Since with TLS and MLS, many objects with similar shapes or at least similar geometrical behavior are typically acquired (e.g. poles, wires, trunks or traffic lights), such a strategy may allow a better training and subsequent discrimination of classes closely located in the feature space. However, the classification results strongly depend on (i) the parameter γ representing the width of the RBF kernel and (ii) the parameter C penalizing classification errors. In order to optimally select these parameters, we conduct a grid search in a suitable subspace (γ, C) .

3.4.5. Ensemble learning

Ensemble learning is based on the idea of strategically generating a set of weak learners and combining them in order to create a single strong learner. An intuitive way for this combination is represented by bagging (Breiman, 1996). Using bootstrapped replica of the training data, i.e. subsets of the complete training data which are randomly drawn with replacement (Efron, 1979), diversity is obtained by training a weak learner of the same type for each subset of the training data. Consequently, the weak learners are all randomly different from one another which results in a de-correlation between individual hypotheses and thus improved generalization and robustness when taking the respective majority vote over all hypotheses (Criminisi and Shotton, 2013). The most popular example for bagging is represented by a Random Forest (RF) classifier (Breiman, 2001) which relies on decision trees as weak learners. A respective modification in terms of a non-hierarchical structure consisting of a set of ferns as weak learners whose hypotheses are combined in a Naive Bayesian way has been presented with a Random Fern (RFe) classifier (Özuysal et al., 2007), where a fern can be considered as simplified decision tree. For both classifiers, the settings have been determined via respective experiments. Accordingly, we use 100 decision trees for the Random Forest classifier and 100 ferns for the Random Fern classifier.

In contrast to bagging, boosting (Schapire, 1990) is based on incrementally generating a set of weak learners over consecutive iterations and different distributions of the training data. In each iteration, a subset of the complete training data is selected to train a weak learner and get a weak hypothesis with low error with respect to the true labels. After a certain number of iterations, all hypotheses are combined by taking the majority vote over all hypotheses. Since boosting was originally proposed for binary classification problems, an extension to multiclass classification has been proposed with Adaptive Boosting (Freund and Schapire, 1997) which is commonly referred to as AdaBoost (AB). As a result of respective tests, we use AdaBoost based on 100 decision trees as weak learners.

3.4.6. Deep learning

Deep learning has been inspired by biological neural networks which are capable to model high-level abstractions in given data. As most prominent example, the Multi-Layer Perceptron (MLP) consists of multiple layers of neurons: an input layer, one or two hidden layers and an output layer. Each layer is fully connected to the next one, and each connection is characterized by a weight factor. Thus, a number of weighted inputs is provided to each neuron which, in turn, maps these inputs to its output via an activation function. Whereas the number of neurons for input and output layer is given with the respective training examples, a suitable number of neurons in the hidden layer has to be determined heuristically. In the training phase, the weights are learned via backpropagation (Rumelhart et al., 1986) which represents a gradient descent technique for minimizing an error function in a high-dimensional space. Based on various tests, we select a Multi-Layer Perceptron with 11 neurons in the hidden layer, linear activation functions for input and output layer, logistic sigmoid functions for the hidden layer and the Resilient Backpropagation algorithm (Riedmiller and Braun, 1993) for learning the parameters in the training phase.

4. Experimental results

In the following, we focus on the performance of our framework. For this purpose, we first describe the two involved and publicly available benchmark datasets in Section 4.1. Subsequently, we outline the conducted experiments in Section 4.2. Accordingly, in

Section 4.3, we provide the derived results for optimal neighborhood size selection in comparison to standard neighborhood definitions and thereby focus on a comparison of single approaches for both individual point classification and feature selection.

4.1. Datasets

Since our main goal consists of applicability of involved methods and reproducibility of derived results, we want to facilitate an objective comparison to other methodologies. Hence, we test our framework on two publicly available and labeled 3D point cloud datasets which are described in the following subsections.

4.1.1. Oakland 3D Point Cloud Dataset

One of the most widely used MLS datasets has been presented with the Oakland 3D Point Cloud Dataset² (Munoz et al., 2009). This dataset represents an urban environment and it has been acquired with a mobile platform equipped with side looking SICK LMS laser scanners used in push-broom mode. A separation of the dataset into training set \mathcal{X} , validation set \mathcal{V} and test set \mathcal{Y} is provided, and each 3D point is assigned one of the five semantic labels *wire*, *pole/trunk*, *façade*, *ground* and *vegetation*. In order to get an impression of the dataset, the respective number of samples per class is provided in Table 2. After class re-balancing, the reduced training set encapsulates 1000 training examples per class.

4.1.2. Paris-rue-Madame database

As second dataset, we consider the Paris-rue-Madame database³ (Serna et al., 2014) which has been acquired in the city of Paris, France. This dataset consists of 20 million 3D points and corresponds to a street section with a length of approximately 160 m. For data acquisition, the Mobile Laser Scanning system L3D2 (Goulette et al., 2006) equipped with a Velodyne HDL32 was used, and annotation has been conducted in a manually assisted way. Since the annotation includes both point labels and segmented objects, the database contains 642 objects which, in turn, are categorized in 26 classes. For our experiments, we only exploit those 3D points belonging to one of the six dominant semantic classes *façade*, *ground*, *cars*, *motorcycles*, *traffic signs* and *pedestrians*, since the remaining classes are smaller than 0.05% of the complete dataset (Table 3). Again, we conduct a class re-balancing and randomly select a training set \mathcal{X} with 1000 training examples per class, while the remaining data is used as test set \mathcal{Y} .

4.2. Experiments

In the experiments, we first consider the general behavior of the proposed method for optimal neighborhood size selection in Section 4.3.1. Subsequently, in Section 4.3.2, we focus on the impact of 7 different neighborhood definitions on the classification results of 10 standard classifiers of different categories:

- the neighborhood \mathcal{N}_{10} formed by the 10 closest neighbors,
- the neighborhood \mathcal{N}_{25} formed by the 25 closest neighbors,
- the neighborhood \mathcal{N}_{50} formed by the 50 closest neighbors,
- the neighborhood \mathcal{N}_{75} formed by the 75 closest neighbors,
- the neighborhood \mathcal{N}_{100} formed by the 100 closest neighbors,

² The Oakland 3D Point Cloud Dataset is publicly available at http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/ (last access: 30 October 2014).

³ Paris-rue-Madame database: MINES ParisTech 3D mobile laser scanner dataset from Madame street in Paris. © 2014 MINES ParisTech. MINES ParisTech created this special set of 3D MLS data for the purpose of detection-segmentation-classification research activities, but does not endorse the way they are used in this project or the conclusions put forward. The database is publicly available at <http://cmm.ensmp.fr/~serna/rueMadameDataset.html> (last access: 30 October 2014).

Table 2
Number of samples per class for the Oakland 3D Point Cloud Dataset.

Class	Training set	Test set
Wire	2571	3794
Pole/trunk	1086	7933
Façade	4713	111,112
Ground	14,121	934,146
Vegetation	14,441	267,325
Σ	36,932	1,324,310

Table 3
Number of samples per class for the Paris-rue-Madame database: both training set and test set have been derived by splitting the whole dataset.

Class	Training set	Test set
Façade	1000	9,977,435
Ground	1000	8,023,295
Cars	1000	1,834,383
Motorcycles	1000	97,867
Traffic signs	1000	14,480
Pedestrians	1000	9048
Σ	6000	19,956,508

- the optimal neighborhood $\mathcal{N}_{\text{opt,dim}}$ for each individual 3D point when considering dimensionality-based scale selection, and
- the optimal neighborhood $\mathcal{N}_{\text{opt},\lambda}$ for each individual 3D point when considering our proposed approach of eigenentropy-based scale selection.⁴

The latter two definitions involving optimal neighborhoods are based on varying the scale parameter k between $k_{\min} = 10$ and $k_{\max} = 100$ with a step size of $\Delta k = 1$, and selecting the value with the minimum Shannon entropy of the respective criterion. Since training a classifier strongly depends on the given training data, we further consider the influence of a varying amount of training data on the classification results in Section 4.3.3. Subsequently, in Section 4.3.4, we focus on feature selection and test 7 different feature sets for each neighborhood definition:

- the whole feature set \mathcal{S}_{all} with all 21 features,
- the feature subset \mathcal{S}_{dim} covering the 3 dimensionality features L_λ, P_λ and S_λ ,
- the feature subset $\mathcal{S}_{\lambda,3D}$ covering the 8 eigenvalue-based 3D features,
- the feature subset \mathcal{S}_5 consisting of the 5 best-ranked features according to a general relevance metric (Weinmann et al., 2013),
- the feature subset \mathcal{S}_{CFS} derived via Correlation-based Feature Selection,
- the feature subset $\mathcal{S}_{\text{FCBF}}$ derived via the Fast Correlation-Based Filter, and
- the feature subset $\mathcal{S}_{\text{mRMR}}$ derived via the minimal-redundancy-maximal-relevance (mRMR) criterion.

The latter four feature subsets are based on either explicitly or implicitly assessing feature relevance. Note that the full feature set only has to be calculated and stored for the training data, whereas a smaller feature subset automatically selected during the training phase has to be calculated for the test data. Finally, in Section 4.3.5, we focus on the transfer of the derived feature selection results to other datasets.

⁴ The code is available at <http://www.ipf.kit.edu/code.php>.

For evaluation, we consider five commonly used measures: (i) precision which represents a measure of exactness or quality, (ii) recall which represents a measure of completeness or quantity, (iii) F_1 -score which combines precision and recall with equal weights, (iv) overall accuracy which reflects the overall performance of the respective classifier on the test set, and (v) mean class recall which reflects the capability of the respective classifier to detect instances of different classes. In order to facilitate an objective comparison, all results are averaged over 20 runs since the results for classification may slightly vary for different runs. Additionally, we consider that, when involving filter-based feature selection, the derived feature subsets may slightly vary due to the random selection of training data in each run, and hence determine them as the most often occurring feature subsets over 20 runs.

All implementation and processing was done in Matlab. In the following, the main focus is put on the impact of both optimal neighborhood size selection and feature selection on the classification results. We may expect that (i) optimal neighborhoods for individual 3D points significantly improve the classification results and (ii) feature subsets selected via feature relevance assessment provide an increase in classification accuracy.

4.3. Results

In the following, we present the results derived by applying our framework.

4.3.1. Insights in the process of neighborhood selection

First, we want to provide more insights in the process of optimal neighborhood size selection. For this purpose, we utilize the Oakland 3D Point Cloud Dataset. Since our approach for selecting an optimal scale parameter k involves an upper boundary of $k_{\max} = 100$ in order to limit the computational costs, we might expect that it is likely to have a certain percentage of points which favor a higher value of k . Consequently, we consider the distribution of k across the full dataset over the interval between $k_{\min} = 10$ and $k_{\max} = 100$ with $\Delta k = 1$. The respective distribution for $\mathcal{N}_{\text{opt},\lambda}$ is shown in Fig. 2 and quite similar for $\mathcal{N}_{\text{opt,dim}}$. The figure reveals a clear trend towards smaller neighborhoods, and the percentage of 3D points which are assigned neighborhoods with $k < 100$ is 98.08% and 98.13% for $\mathcal{N}_{\text{opt},\lambda}$ and $\mathcal{N}_{\text{opt,dim}}$. For the last bin corresponding to $k_{\max} = 100$, a slight increase can be observed. The distributions per class are provided in Fig. 3 for $\mathcal{N}_{\text{opt},\lambda}$ and follow the major trend with only a slight difference between different classes.

However, it has to be taken into account that – when considering optimal neighborhoods $\mathcal{N}_{\text{opt},\lambda}$ and $\mathcal{N}_{\text{opt,dim}}$ – an additional processing time of approximately 21 s and 758 s is required on a high-performance computer (Intel Core i7-3820, 3.6 GHz, 64 GB RAM)

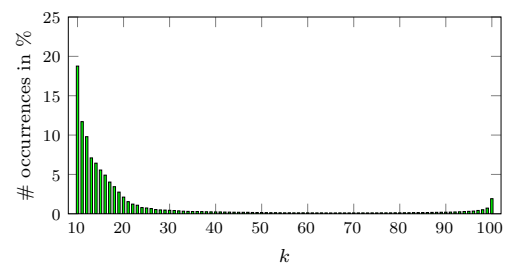


Fig. 2. Distribution of the assigned optimal neighborhood size k for all 3D points in the Oakland 3D Point Cloud Dataset.

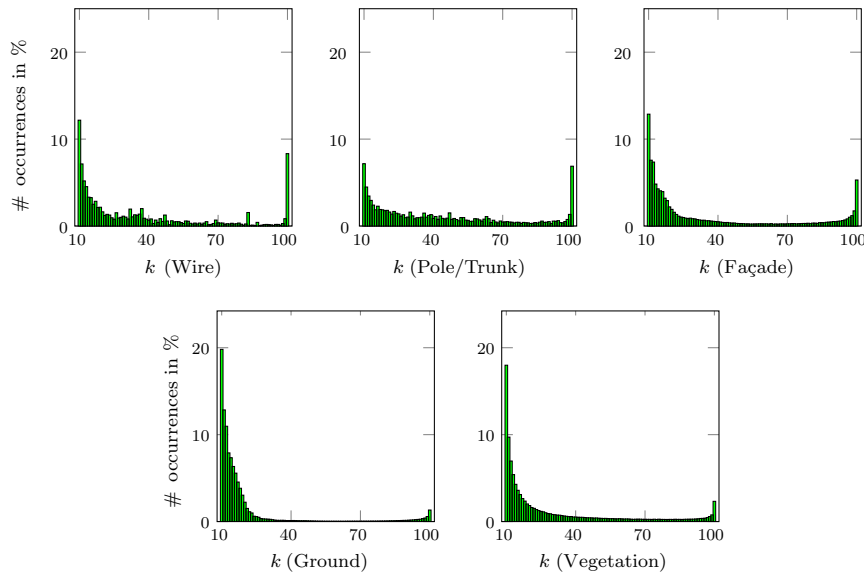


Fig. 3. Distribution of the assigned optimal neighborhood size k for the different classes of the Oakland 3D Point Cloud Dataset.

for the full training set and the test set, respectively. The additional effort is significant in comparison to feature extraction, where approximately 4 s and 2793 s are required for calculating all features for the training set and the test set. This raises the question if individual neighborhoods of optimal size are really necessary. We focus on this issue in the next subsection.

4.3.2. Impact of neighborhood selection

In order to reason about the impact of neighborhood selection on individual point classification, we consider all 21 geometric features for each of the 7 different neighborhood definitions and provide them to 10 classifiers of different categories. The obtained overall accuracy and mean class recall values are provided in Tables 4 and 5. For an in-depth analysis concerning the impact of neighborhood selection on the classification results, the recall

and precision values for the different neighborhood definitions and different classifiers are provided in Tables 6 and 7. The corresponding F_1 -scores are visualized in Fig. 4. In order to argue about the efficiency of the involved classifiers, we also provide the absolute and relative processing times for training phase and testing phase in Table 8.

Note that the combination of optimal neighborhoods $\mathcal{N}_{opt,\lambda}$ and a Random Forest provides a good solution with respect to both accuracy and computational efficiency. A visualization of respective classification results is provided in Fig. 5 for different parts of the scene.

4.3.3. Impact of training data

Since we introduce a class re-balancing by randomly sampling the same number of 1000 training examples per class, we also have

Table 4

Overall accuracy (in %) for different neighborhood definitions and different classifiers. Bold values indicate the highest overall accuracy obtained with the respective classifier.

Oakland	NN	DT	NB	LDA	QDA	SVM	RF	RFe	AB	MLP
\mathcal{N}_{10}	73.86	65.64	78.88	87.38	78.93	82.93	87.53	81.94	86.78	80.54
\mathcal{N}_{25}	86.25	69.30	83.64	90.08	83.62	88.88	90.50	88.77	89.99	78.59
\mathcal{N}_{50}	88.89	75.47	85.03	92.83	84.95	92.00	91.54	90.42	91.80	85.68
\mathcal{N}_{75}	89.97	76.87	85.00	93.05	84.99	91.99	91.06	91.16	90.56	87.07
\mathcal{N}_{100}	89.90	84.45	84.33	92.60	84.43	91.76	90.16	90.59	87.01	84.39
$\mathcal{N}_{opt,dim}$	79.34	70.71	83.75	91.01	83.80	90.15	91.89	90.12	91.62	85.69
$\mathcal{N}_{opt,\lambda}$	79.87	75.76	85.63	90.39	85.69	89.10	92.25	90.45	92.28	87.29

Table 5

Mean class recall values (in %) for different neighborhood definitions and different classifiers. Bold values indicate the highest mean class recall value obtained with the respective classifier.

Oakland	NN	DT	NB	LDA	QDA	SVM	RF	RFe	AB	MLP
\mathcal{N}_{10}	63.40	54.19	62.29	70.68	62.33	58.86	70.78	63.75	67.52	64.20
\mathcal{N}_{25}	70.01	57.41	68.46	75.54	68.47	68.50	74.65	71.48	68.64	68.03
\mathcal{N}_{50}	69.47	59.99	67.12	72.76	66.98	68.47	72.72	69.22	71.46	69.13
\mathcal{N}_{75}	68.29	57.82	65.49	73.05	65.44	68.00	69.99	68.88	68.19	70.47
\mathcal{N}_{100}	66.66	57.96	63.44	72.35	63.46	64.76	68.51	67.16	59.58	68.98
$\mathcal{N}_{opt,dim}$	74.17	62.15	74.49	81.36	74.35	79.58	81.70	78.35	77.63	78.61
$\mathcal{N}_{opt,\lambda}$	73.98	66.99	76.19	82.05	76.15	79.97	82.59	78.70	79.49	79.92

Table 6

Recall values (in %) for different neighborhood definitions and different classifiers. Bold values indicate the highest recall value obtained with the respective classifier for the respective class.

Oakland	\mathcal{N}	NN	DT	NB	LDA	QDA	SVM	RF	RFe	AB	MLP
Wire	\mathcal{N}_{10}	65.51	55.20	65.92	77.06	65.69	62.21	70.46	72.24	58.66	65.61
	\mathcal{N}_{25}	64.63	50.50	70.78	73.30	70.97	60.11	69.48	68.14	52.24	63.70
	\mathcal{N}_{50}	50.40	42.09	66.88	48.57	66.69	41.13	56.86	55.82	53.55	51.81
	\mathcal{N}_{75}	49.68	35.22	63.10	48.48	63.29	43.06	49.71	51.72	51.94	48.84
	\mathcal{N}_{100}	50.09	26.62	60.54	46.74	60.35	34.91	49.67	52.51	54.69	47.41
	$\mathcal{N}_{opt,dim}$	77.91	69.30	75.72	82.51	75.24	80.19	85.16	76.58	78.58	78.54
	$\mathcal{N}_{opt,\lambda}$	80.13	73.48	79.15	87.00	79.13	82.99	86.05	82.01	80.40	81.11
Pole/trunk	\mathcal{N}_{10}	61.74	60.45	52.05	61.55	52.29	35.88	68.49	47.56	66.92	58.24
	\mathcal{N}_{25}	66.72	63.62	57.98	69.92	58.21	55.67	69.59	63.79	58.97	71.77
	\mathcal{N}_{50}	61.11	60.23	50.70	60.56	50.62	55.32	62.64	56.39	53.76	67.73
	\mathcal{N}_{75}	55.96	52.98	45.81	59.80	45.95	50.83	58.63	50.82	46.89	69.23
	\mathcal{N}_{100}	47.35	49.51	41.76	60.62	41.60	41.49	58.27	45.24	38.91	69.60
	$\mathcal{N}_{opt,dim}$	73.91	70.57	75.22	78.81	75.11	78.42	78.90	75.41	65.90	81.61
	$\mathcal{N}_{opt,\lambda}$	74.49	71.36	76.33	79.85	76.28	78.10	79.99	73.25	70.17	82.07
Façade	\mathcal{N}_{10}	46.16	29.42	43.35	48.21	43.45	43.45	50.29	41.94	48.62	42.42
	\mathcal{N}_{25}	52.10	39.51	46.70	59.97	46.49	54.15	60.98	52.35	55.89	49.72
	\mathcal{N}_{50}	65.27	51.29	47.32	73.81	47.29	65.00	68.13	55.96	74.19	58.94
	\mathcal{N}_{75}	65.19	51.16	48.47	75.85	48.36	65.22	67.51	63.75	70.07	66.78
	\mathcal{N}_{100}	65.93	52.90	47.06	73.97	47.28	66.14	62.69	61.09	34.41	65.56
	$\mathcal{N}_{opt,dim}$	61.65	33.57	54.37	68.82	54.34	64.48	65.90	65.77	63.91	65.75
	$\mathcal{N}_{opt,\lambda}$	55.82	46.63	51.88	67.12	52.15	64.39	67.01	58.87	65.73	66.43
Ground	\mathcal{N}_{10}	80.32	73.96	88.26	97.55	88.30	95.44	98.23	91.84	97.61	89.79
	\mathcal{N}_{25}	94.66	76.50	90.34	97.69	90.42	97.31	98.91	97.02	97.81	84.45
	\mathcal{N}_{50}	96.25	81.71	91.58	98.55	91.56	98.47	98.84	97.93	98.52	92.23
	\mathcal{N}_{75}	98.48	83.27	91.66	98.55	91.83	98.47	98.81	98.12	98.35	93.55
	\mathcal{N}_{100}	98.58	93.72	91.51	98.25	91.59	97.86	98.71	97.97	98.19	90.49
	$\mathcal{N}_{opt,dim}$	82.88	78.55	88.95	97.04	89.07	96.83	98.52	96.93	98.00	90.80
	$\mathcal{N}_{opt,\lambda}$	84.06	83.90	90.47	96.24	90.70	94.92	98.48	96.58	98.41	92.70
Vegetation	\mathcal{N}_{10}	63.27	51.90	61.86	69.03	61.91	57.31	66.45	65.15	65.79	64.94
	\mathcal{N}_{25}	71.95	56.95	76.49	76.83	76.24	75.27	74.29	76.11	78.28	70.53
	\mathcal{N}_{50}	74.32	64.62	79.09	82.33	78.76	82.44	77.10	80.00	77.29	74.92
	\mathcal{N}_{75}	72.13	66.48	78.41	82.59	77.76	82.43	75.31	79.96	73.70	73.92
	\mathcal{N}_{100}	71.36	67.06	76.32	82.19	76.48	83.38	73.20	78.96	71.69	71.87
	$\mathcal{N}_{opt,dim}$	74.51	58.76	78.17	79.62	78.00	77.99	79.99	77.05	81.78	76.34
	$\mathcal{N}_{opt,\lambda}$	75.38	59.56	83.09	80.02	82.49	79.46	81.41	82.79	82.74	77.26

to consider what happens when involving more or less training examples. This is of particular interest when dealing with datasets where small classes are only represented by a few hundreds of 3D points.

From the previous experiment, it becomes visible that a Random Forest provides a good solution when considering a trade-off between accuracy and efficiency. Hence, we take such a classifier and exploit various numbers of training examples per class. The respective recall and precision values are given in Tables 9 and 10. Note that, for less training examples per class, the recall values tend to decrease for the smaller classes of *wire* and *pole/trunk* while the respective precision values tend to increase. The same characteristic holds when involving all training examples per class, but with a more significant impact.

4.3.4. Impact of feature selection

Focusing on the capability towards large-scale 3D scene analysis with significantly larger datasets, we also want to provide a solution for selecting relevant features and discarding irrelevant ones in order to reduce the computational burden with respect to processing time and memory consumption. Accordingly, we again use a Random Forest classifier and provide the results obtained when using the 7 different neighborhood definitions and 7 different feature sets. This yields a total number of 49 possible combinations. For each combination, the respective overall accuracy and mean class recall values are provided in Tables 11 and 12. Note that for this experiment

- \mathcal{S}_{all} contains all 21 features,
- \mathcal{S}_{dim} contains 3 features (which represents about 14% of the available features),
- $\mathcal{S}_{z,3D}$ contains 8 features (about 38%),
- \mathcal{S}_5 contains 5 features (about 24%),
- \mathcal{S}_{CF5} contains between 12 and 16 features (about 57–76%),
- \mathcal{S}_{FCBF} contains between 6 and 9 features (about 29–43%), and
- \mathcal{S}_{mRMR} contains 10 features (about 48%).

The latter four subsets contain features which are distributed across both 3D and 2D features.

4.3.5. Transfer between datasets

Finally, we focus on the transfer of the feature selection results for classifying a different point cloud represented by the Paris-rue-Madame database. For this purpose, we again select a Random Forest representing an efficient classifier with respect to both accuracy and time consumption. Since in all previous experiments, our approach for selecting individual neighborhoods of optimal size ($\mathcal{N}_{opt,\lambda}$) has proven to have a beneficial impact on the respective results, we select this neighborhood definition for extracting the geometric 3D and 2D features. For the same training data and test data, we compare the results obtained for the full feature set \mathcal{S}_{all} and the most powerful feature selection approaches yielding the feature sets \mathcal{S}_{CF5} and \mathcal{S}_{FCBF} in Table 13. Corresponding to the provided recall and precision values, we obtain

Table 7

Precision values (in %) for different neighborhood definitions and different classifiers. Bold values indicate the highest precision value obtained with the respective classifier for the respective class.

Oakland	\mathcal{N}	NN	DT	NB	LDA	QDA	SVM	RF	RFe	AB	MLP
Wire	\mathcal{N}_{10}	1.61	1.00	3.21	6.14	3.24	5.15	5.51	4.15	4.89	2.59
	\mathcal{N}_{25}	4.28	1.05	3.83	5.93	3.87	4.54	7.12	5.65	6.62	4.89
	\mathcal{N}_{50}	3.63	1.34	3.50	5.17	3.50	5.06	4.81	4.91	6.16	4.79
	\mathcal{N}_{75}	5.19	1.47	3.34	5.18	3.35	5.78	4.00	4.88	4.25	5.29
	\mathcal{N}_{100}	5.29	2.72	3.26	4.95	3.24	5.79	3.98	5.15	4.13	5.49
	$\mathcal{N}_{\text{opt.dim}}$	1.85	1.70	5.02	6.27	4.93	5.63	7.98	5.90	8.09	4.09
	$\mathcal{N}_{\text{opt.}\lambda}$	2.97	1.72	6.49	5.76	6.46	5.60	9.03	7.86	9.34	5.01
Pole/trunk	\mathcal{N}_{10}	4.48	5.18	3.75	8.10	3.67	2.43	7.99	3.00	6.60	5.51
	\mathcal{N}_{25}	6.28	6.38	5.78	9.32	5.68	5.99	9.46	6.24	6.36	4.77
	\mathcal{N}_{50}	8.66	8.46	7.97	16.61	7.90	9.02	19.47	7.53	10.02	6.71
	\mathcal{N}_{75}	8.17	11.09	7.91	19.26	7.89	8.85	18.25	8.84	11.10	8.14
	\mathcal{N}_{100}	7.19	9.04	8.27	18.55	7.97	8.17	13.55	7.42	5.06	7.50
	$\mathcal{N}_{\text{opt.dim}}$	9.97	6.85	18.90	34.65	18.88	12.91	22.09	11.58	14.86	13.55
	$\mathcal{N}_{\text{opt.}\lambda}$	9.13	11.62	18.22	34.52	18.38	11.46	24.13	11.10	18.71	14.54
Façade	\mathcal{N}_{10}	63.54	47.59	60.86	66.03	61.41	52.59	77.62	62.50	72.56	60.01
	\mathcal{N}_{25}	73.20	49.82	79.54	78.49	79.22	79.85	83.88	79.40	80.79	54.33
	\mathcal{N}_{50}	78.61	64.97	82.78	83.42	82.83	83.16	83.43	81.41	89.05	74.60
	\mathcal{N}_{75}	77.38	65.10	78.84	87.48	77.71	81.60	80.28	80.22	85.71	76.78
	\mathcal{N}_{100}	76.56	68.74	65.27	84.63	67.69	72.93	76.19	74.78	71.15	68.03
	$\mathcal{N}_{\text{opt.dim}}$	73.45	45.95	79.62	79.73	79.63	77.27	83.71	77.04	82.92	74.88
	$\mathcal{N}_{\text{opt.}\lambda}$	71.56	51.84	82.71	80.75	81.80	76.14	84.69	76.69	84.21	76.42
Ground	\mathcal{N}_{10}	98.76	98.40	96.68	96.78	96.73	98.68	96.82	99.64	97.57	96.47
	\mathcal{N}_{25}	99.01	98.24	99.67	99.58	99.67	99.68	98.58	99.75	99.16	98.57
	\mathcal{N}_{50}	98.84	98.23	98.27	99.51	97.98	99.75	97.77	99.74	98.66	98.83
	\mathcal{N}_{75}	98.66	97.70	98.63	99.42	98.27	99.73	97.86	99.66	97.67	99.28
	\mathcal{N}_{100}	98.56	96.67	97.84	99.64	97.93	99.82	97.92	99.74	96.45	99.19
	$\mathcal{N}_{\text{opt.dim}}$	99.16	98.78	96.58	96.62	96.60	99.39	97.67	99.76	98.15	97.97
	$\mathcal{N}_{\text{opt.}\lambda}$	99.04	98.45	96.72	96.25	96.60	99.22	97.18	99.60	97.57	97.57
Vegetation	\mathcal{N}_{10}	77.23	76.16	81.06	95.34	81.59	92.46	94.79	91.73	93.32	93.84
	\mathcal{N}_{25}	91.55	77.87	78.06	94.89	78.20	94.47	94.87	94.79	94.29	90.67
	\mathcal{N}_{50}	93.06	82.52	79.23	93.16	79.71	93.47	94.40	94.18	93.36	90.37
	\mathcal{N}_{75}	92.93	80.16	77.69	91.32	78.79	92.08	93.84	93.30	93.34	86.55
	\mathcal{N}_{100}	92.58	79.16	78.91	89.95	78.69	91.14	93.55	92.88	92.82	81.99
	$\mathcal{N}_{\text{opt.dim}}$	87.28	78.70	69.87	95.87	70.11	95.12	94.97	93.75	94.65	93.06
	$\mathcal{N}_{\text{opt.}\lambda}$	75.77	84.53	73.77	96.27	74.32	93.96	95.87	92.83	95.19	93.96

- an overall accuracy of 88.76% and a mean class recall of 83.56% for the full feature set \mathcal{S}_{all} ,
- an overall accuracy of 88.98% and a mean class recall of 84.66% for the feature set \mathcal{S}_{CFE} , and
- an overall accuracy of 89.16% and a mean class recall of 83.83% for the feature set $\mathcal{S}_{\text{FCBF}}$.

For obtaining an impression on the quality of the derived results, the results when involving the full feature set \mathcal{S}_{all} are visualized in Fig. 6.

5. Discussion

Certainly, a huge advantage of our framework consists of its composition of four successive components, where each component may be treated independently from the others which, in turn, allows to exhaustively test all conceivable configurations. For each of these components, we briefly discuss the main conclusions derived from our experiments.

For the first component of neighborhood selection, the use of individual neighborhoods of optimal size provides a general approach, since it completely avoids the use of empiric or heuristic a priori knowledge on the scene which would be necessary when specifying neighborhoods according to standard approaches. The use of individual 3D neighborhoods is also in accordance with the idea that the optimal neighborhood size may not be the same for different classes and that it may furthermore depend on the

respective point density. Note that the class-specific classification results clearly reveal that, for neighborhood definitions with fixed scale parameter, the suitability may vary from one class to the other (Tables 6 and 7). In contrast, the proposed method of eigenentropy-based scale selection directly adapts to the given 3D point cloud data. Consequently, this method significantly improves the classification results, in particular when considering the mean class recall values. This improvement becomes visible for a variety of different classifiers (Tables 4 and 5). Note that for all classifiers, the significantly beneficial impact of individual neighborhoods of optimal size on the mean class recall values results from a considerable impact for the smaller classes *wire* and *pole/trunk* (Tables 6 and 7). Consequently, we may state that the respective classifiers are less prone to overfitting when introducing individual neighborhoods of optimal size. This is in accordance with the fact that, for the Oakland 3D Point Cloud Dataset, we have an unbalanced test set and an overall accuracy of 70.5% could be obtained if only the instances of the class *ground* are correctly classified. This clear trend to overfitting becomes visible when considering the respective mean class recall of only 20.0%. Thus, in our experiments, the strongest indicator for the quality of the derived results is represented by the mean class recall, as only a high overall accuracy may not be sufficient.

For the second component of feature extraction, we focus on both 3D and 2D features. Whereas the 3D features provide information about the spatial arrangement of neighboring 3D points in terms of linear, planar or volumetric behavior, the projection

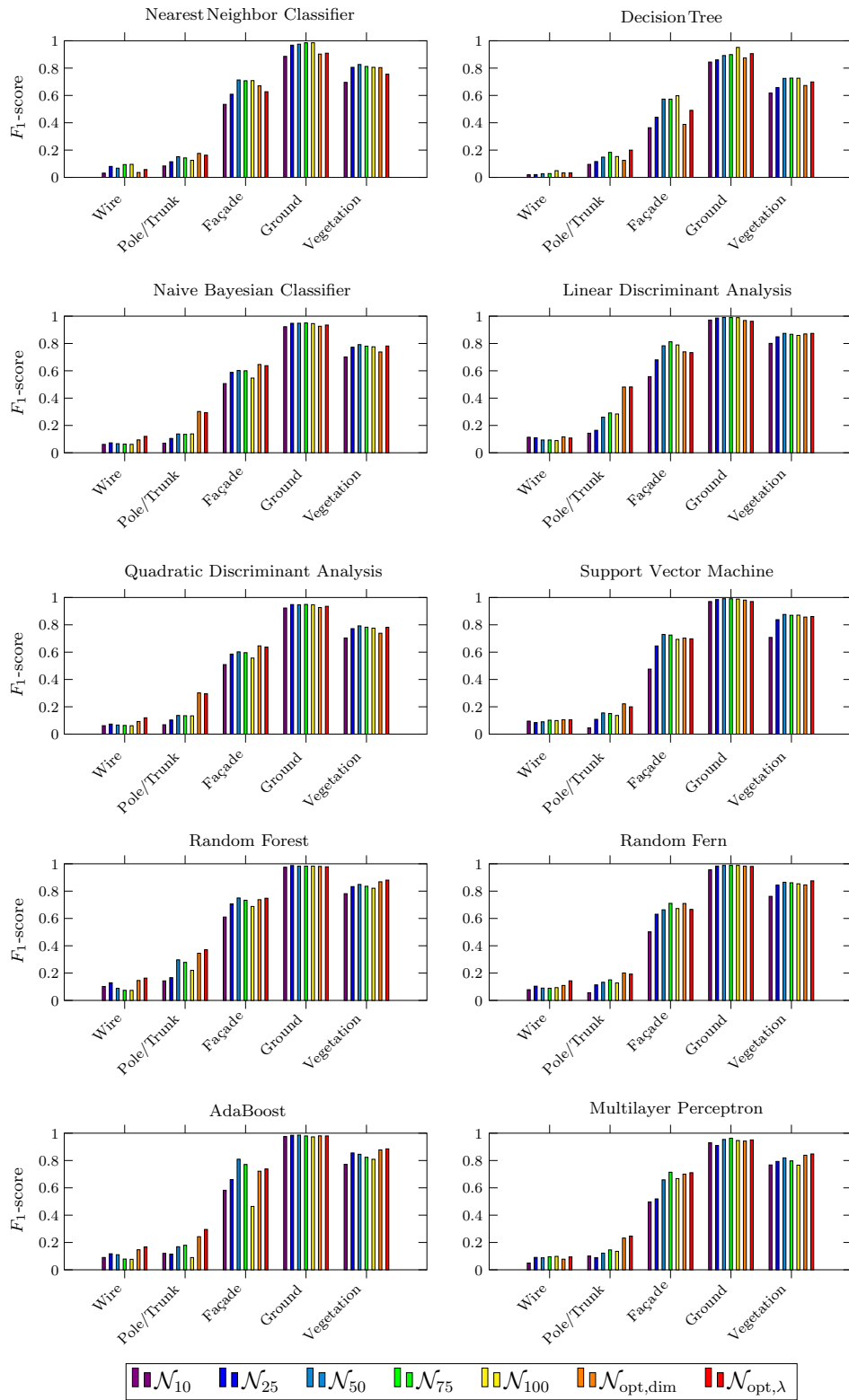


Fig. 4. F_1 -scores for different classifiers applied on the Oakland 3D Point Cloud Dataset.

Table 8

Absolute and relative processing times for training and testing when using different classifiers on a standard notebook (Intel Core i5-2410M, 2.3 GHz, 4 GB RAM). The reference for relative values is represented by the Random Forest classifier. Note that, for the training, additional time is required for tuning the settings of some classifiers (SVM, RF, RFe, AB and MLP).

Oakland	NN	DT	NB	LDA	QDA	SVM	RF	RFe	AB	MLP
t_{train} (s)	00.00	0.11	0.01	0.05	0.07	1.39	0.44	0.03	6.20	2.28
t_{train} (%)	00.00	24.54	3.13	10.74	15.18	317.19	100.00	6.96	1410.66	518.13
t_{test} (s)	167.52	0.65	3.71	4.45	3.92	319.48	6.33	8.12	76.31	1.80
t_{test} (%)	2645.77	10.22	58.64	70.34	61.96	5045.68	100.00	128.22	1205.24	28.45

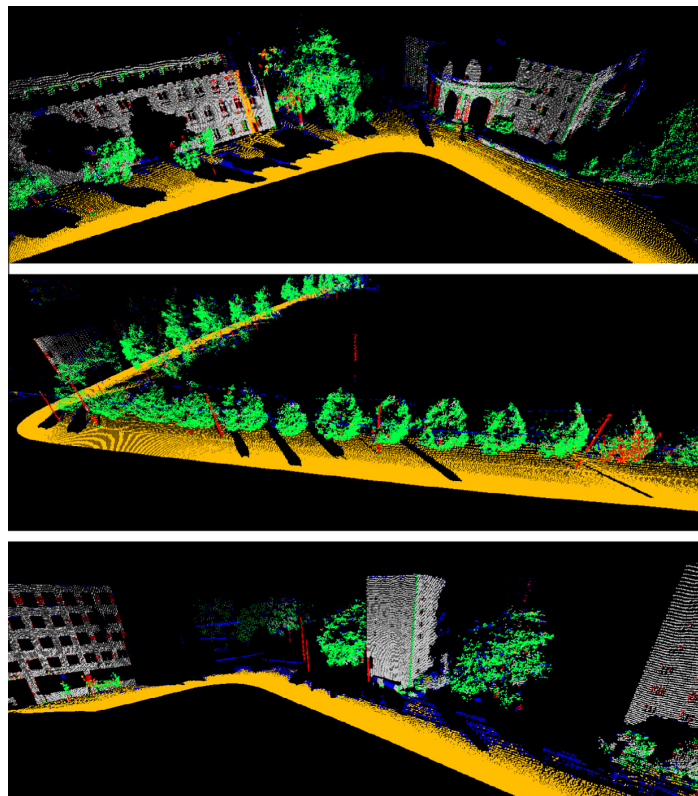


Fig. 5. Exemplary classification results when using optimal neighborhoods $\mathcal{N}_{\text{opt},z}$ and a Random Forest (wire: blue, pole/trunk: red, façade: gray, ground: brown, vegetation: green).

Table 9

Recall values (in %) for eigenentropy-based scale selection combined with a Random Forest classifier.

Oakland	250 samples	500 samples	750 samples	1000 samples	All samples
Wire	84.22	85.07	85.91	86.05	82.31
Pole/trunk	76.47	77.98	78.77	79.99	70.52
Façade	66.50	67.14	67.47	67.01	65.80
Ground	98.50	98.55	98.49	98.48	97.89
Vegetation	81.82	81.34	81.78	81.40	93.35

Table 10

Precision values (in %) for eigenentropy-based scale selection combined with a Random Forest classifier.

Oakland	250 samples	500 samples	750 samples	1000 samples	All samples
Wire	9.06	9.21	9.05	9.03	10.19
Pole/trunk	32.38	28.53	27.83	24.14	45.75
Façade	82.22	82.52	84.32	84.69	92.73
Ground	96.77	96.94	97.01	97.18	98.98
Vegetation	95.48	95.72	95.73	95.87	89.82

Table 11

Overall accuracy (in %) for a Random Forest, different neighborhood definitions and different feature sets. Bold values indicate the highest overall accuracy obtained with the respective feature set.

Oakland	S_{all}	S_{dim}	$S_{\lambda,3D}$	S_5	S_{CFS}	S_{FCBF}	S_{mRMR}
N_{10}	87.50	58.36	74.33	85.66	87.43	87.29	82.11
N_{25}	90.78	68.80	82.48	89.60	90.59	91.78	84.70
N_{50}	91.64	73.19	81.38	91.01	91.71	92.69	85.64
N_{75}	91.00	73.63	80.12	90.24	91.17	91.47	85.99
N_{100}	90.11	72.99	81.96	89.84	90.31	90.94	85.76
$N_{\text{opt,dim}}$	91.92	69.59	77.69	91.41	91.83	91.55	86.82
$N_{\text{opt},\lambda}$	92.28	63.61	84.88	91.44	92.27	92.78	84.28

Table 12

Mean class recall values (in %) for a Random Forest, different neighborhood definitions and different feature sets. Bold values indicate the highest mean class recall value obtained with the respective feature set.

Oakland	S_{all}	S_{dim}	$S_{\lambda,3D}$	S_5	S_{CFS}	S_{FCBF}	S_{mRMR}
N_{10}	70.83	48.41	59.95	58.24	69.28	70.08	62.46
N_{25}	75.48	55.68	65.22	73.30	74.24	76.58	63.61
N_{50}	72.71	54.41	64.43	65.91	72.64	74.14	60.90
N_{75}	69.75	52.12	61.37	59.86	70.19	68.66	58.35
N_{100}	68.49	50.33	61.37	60.22	69.02	66.34	56.05
$N_{\text{opt,dim}}$	81.79	61.53	67.65	75.57	81.33	80.83	74.72
$N_{\text{opt},\lambda}$	82.60	59.48	69.17	78.50	82.39	82.93	69.37

Table 13

Recall R and precision P (in %) for eigenentropy-based scale selection, different feature sets and a Random Forest classifier.

Paris	R_{all}	P_{all}	R_{CFS}	P_{CFS}	R_{FCBF}	P_{FCBF}
Façade	95.29	96.45	95.11	96.61	95.23	96.89
Ground	86.69	97.92	86.38	98.70	87.44	98.66
Cars	63.13	78.96	67.71	77.42	64.42	79.82
Motorcycles	71.52	9.25	73.68	9.51	73.30	8.45
Traffic signs	95.82	4.76	95.88	4.90	96.26	5.34
Pedestrians	88.89	1.67	89.21	1.74	86.32	1.75

onto a horizontally oriented plane clearly provides evidence about the presence of building façades, which appear as a line in the 2D projection. Furthermore, the sampling via the discrete accumulation map with quadratic bins introduces a second neighborhood definition with infinite extent in the vertical direction. In this neighborhood definition, the maximal difference and standard deviation of height values provide further insights about the local 3D structure, which are not represented by the other features.

For the third component of feature selection, it becomes visible that the use of individual neighborhoods tends to provide the best

classification results for all feature sets. Furthermore, the results clearly reveal that, in comparison to the whole feature set S_{all} , the feature set S_{dim} consisting of the three dimensionality features L_λ , P_λ and S_λ is not sufficient for obtaining adequate classification results (Tables 11 and 12). This might be due to ambiguities, since the classes *wire* and *pole/trunk* provide a linear behavior, whereas the classes *façade* and *ground* provide a planar behavior. For adequately handling this issue, additional features have to be taken into account. The feature set $S_{\lambda,3D}$ of the eigenvalue-based 3D features performs significantly better with respect to both overall accuracy and mean class recall, but the results are still not comparable to those obtained for S_{all} . In contrast, the feature sets derived via the four filter-based methods for feature selection provide classification results of better quality. Whereas the feature set S_{mRMR} performs worst of the filter-based feature selection methods, the feature set S_5 performs considerably well when taking into account that only 5 features of all 21 features are used (which reduces the required memory for data storage to only about 24%). The feature sets S_{CFS} and S_{FCBF} provide a performance close to the full feature set S_{all} or even better while simultaneously reducing the required memory for data storage to about 29–76% which, in turn, is an important aspect for large-scale considerations. These results are in accordance with the general aim of feature selection to improve the classification results while reducing the computational effort. Since the selection of S_{FCBF} is based on heuristics, the feature set S_{CFS} derived via Correlation-based Feature Selection provides the method of choice. Note that we only account for filter-based feature selection, since the use of classifier-independent filter-based methods results in simplicity and efficiency compared to other methods interacting with a classifier.

For the fourth component of supervised classification, it becomes visible that the classifiers based on rule-based learning cannot compete with the other classifiers (Tables 4–7). Instance-based learning significantly improves the classification results, but the computational effort for the testing phase is relatively high due to the delayed induction process instead of a training phase (Table 8). The more sophisticated classifiers based on probabilistic learning, max-margin learning, ensemble learning and deep learning yield classification results of better quality. However, it has to be taken into account that max-margin learning, ensemble learning and deep learning require additional time for tuning the settings of a respective classifier. Thus, the use of Support Vector Machines – for which the computational effort is already high without a parameter tuning – does not really satisfy the constraint with respect to efficiency. Since deep learning relies on heuristicaly determining the number of nodes in the hidden layer, probabilistic learning and ensemble learning via bagging seem to

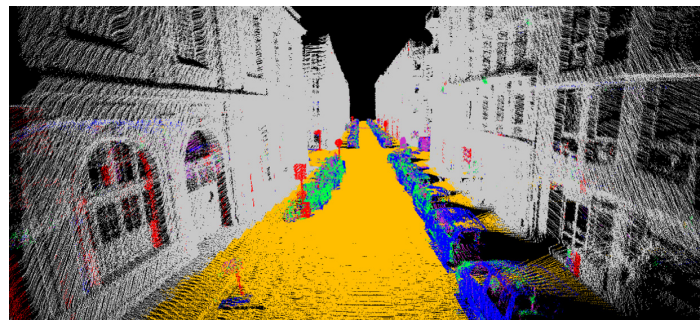


Fig. 6. Exemplary classification results when using optimal neighborhoods $N_{\text{opt},\lambda}$, the full feature set S_{all} and a Random Forest (façade: gray, ground: brown, cars: blue, motorcycles: green, traffic signs: red, pedestrians: pink).

be favorable. Considering the derived results, the Random Forest classifier provides a good trade-off between accuracy and efficiency. Note that the selection of a Random Forest as classifier can further be motivated by its simplicity, since it is relatively easy to understand and use for non-expert users.

In total, our framework reveals that, based on fully generic solutions, the consideration of optimal neighborhoods improves the classification results in terms of accuracy and less overfitting, whereas the selection of relevant features reduces the computational burden with respect to both processing time and memory consumption without reducing the quality of the classification results. By providing our implementations for neighborhood recovery and feature extraction (in Matlab, C++ and as binaries), we allow end-users to apply the code on their platform and experienced users to involve the code in their investigations. These components may not only be used for point cloud classification, but also for a variety of other applications such as object segmentation or urban accessibility analysis.

6. Conclusions

In this paper, we addressed the issue of 3D scene analysis in terms of 3D point cloud classification. We presented a new, fully automated and versatile framework involving four successive components and a variety of approaches per component which, in turn, satisfy the constraints of simplicity, efficiency and reproducibility. As main focus of our work, we considered the interleaved issue of (i) using individual neighborhoods of optimal size for extracting features with increased distinctiveness and (ii) selecting a subset consisting of the most relevant features. In a detailed evaluation involving 7 neighborhood definitions, 21 geometric features, 7 approaches for feature selection and 10 classifiers, we demonstrated the significantly beneficial impact of using individual neighborhoods of optimal size as well as the advantages of feature selection in terms of increasing the classification accuracy while simultaneously reducing the computational burden. In particular, the neighborhood selection based on minimizing the measure of eigentropy over varying scales provided a positive impact, independent of the respective classifier. Furthermore, those approaches for feature selection which are based on the measure of symmetrical uncertainty for evaluating both feature-class and feature-feature relations proved to provide the most suitable feature subsets, since they do not only discard irrelevant features but also reduce redundancy among features.

For future work, we plan to further improve the results of 3D scene analysis by introducing a spatially smooth labeling. This may be achieved by involving either smoothing techniques or techniques exploiting contextual information. Since both options are based on the results of individual point classification, however, the presented methodology provides an important prerequisite for these.

Acknowledgements

The project was partially supported by KIT-GRACE, the Graduate School for Climate and Environment at the Karlsruhe Institute of Technology (KIT), and by the FP7 Project IQmulus (FP7-ICT-2011-318787). Furthermore, a funding for a stay abroad of the first author was provided by the Karlsruhe House of Young Scientists (KHYS) at KIT in order to support the collaboration.

References

Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y., 1998. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *J. ACM* 45 (6), 891–923.

- Belton, D., Lichti, D.D., 2006. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI-5*, pp. 44–49.
- Blomley, R., Weinmann, M., Leitloff, J., Jutzi, B., 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 II-3*, pp. 9–16.
- Boyko, A., Funkhouser, T., 2011. Extracting roads from dense point clouds in large scale urban environment. *ISPRS J. Photogr. Remote Sens.* 66 (6), S02–S12.
- Breiman, L., 1996. Bagging predictors. *Machine Learn.* 24 (2), 123–140.
- Breiman, L., 2001. Random forests. *Machine Learn.* 45 (1), 5–32.
- Bremer, M., Wichmann, V., Rutzinger, M., 2013. Eigenvalue and graph-based object extraction from mobile laser scanning point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-5/W2*, pp. 55–60.
- Brodu, N., Lague, D., 2012. 3d terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology. *ISPRS J. Photogr. Remote Sens.* 68, 121–134.
- Carlberg, M., Gao, P., Chen, G., Zakhor, A., 2009. Classifying urban landscape in aerial lidar using 3d shape analysis. In: *Proceedings of the IEEE International Conference on Image Processing, IEEE, Cairo, Egypt, 7–10 November*, pp. 1701–1704.
- Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (3), 27:1–27:27.
- Chehata, N., Guo, L., Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-3/W8*, pp. 207–212.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learn.* 20 (3), 273–297.
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* 13 (1), 21–27.
- Criminisi, A., Shotton, J., 2013. Decision forests for computer vision and medical image analysis. In: *Advances in Computer Vision and Pattern Recognition*. Springer, London, UK.
- Demantké, J., Mallet, C., David, N., Vallet, B., 2011. Dimensionality based scale selection in 3d lidar point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-5/W12*, pp. 97–102.
- Demantké, J., Vallet, B., Paparoditis, N., 2012. Streamed vertical rectangle detection in terrestrial laser scans for facade database production. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3*, pp. 99–104.
- Efron, B., 1979. Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7 (1), 1–26.
- Fayyad, U.M., Irani, K.B., 1993. Multi-interval discretization of continuous-valued attributes for classification learning. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. Morgan Kaufman, Chambéry, France, 28 August – 3 September, pp. 1022–1027.
- Filin, S., Pfeifer, N., 2005. Neighborhood systems for airborne laser data. *Photogr. Eng. Remote Sens.* 71 (6), 743–755.
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7 (2), 179–188.
- Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55 (1), 119–139.
- Friedman, J.H., Bentley, J.L., Finkel, R.A., 1977. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw.* 3 (3), 209–226.
- Gerke, M., Xiao, J., 2013. Supervised and unsupervised MRF based 3d scene classification in multiple view airborne oblique images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W3*, pp. 25–30.
- Gini, C., 1912. Variabilit  e mutabilit  a. *Memorie di metodologia statistica*.
- Goulette F., Nashashibi, F., Abuhadrous, I., Ammoun, S., Laugeau, C., 2006. An integrated on-board laser range sensing system for on-the-way city and road modelling. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI-1*.
- Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M., Yang, B., 2014. Using mobile laser scanning data for automated extraction of road markings. *ISPRS J. Photogr. Remote Sens.* 87, 93–107.
- Guo, B., Huang, X., Zhang, F., Sohn, G., 2014. Classification of airborne laser scanning data using JointBoost. *ISPRS J. Photogr. Remote Sens.* 92, 124–136.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *J. Machine Learn. Res.* 3, 1157–1182.
- Hall, M.A., 1999. Correlation-based feature subset selection for machine learning. Ph.D. thesis, Department of Computer Science, University of Waikato, New Zealand.
- Hebert, M., Bagnell, J.A., Bajracharya, M., Daniilidis, K., Matthies, L.H., Mianzo, L., Navarro-Serment, L., Shi, J., Wellfare, M., 2012. Semantic perception for ground robotics. In: *Proceedings of SPIE 8387, Unmanned Systems Technology XIV*, SPIE, Baltimore, USA, 23 April, pp. 83870Y:1–12.
- Hu, H., Munoz, D., Bagnell, J.A., Hebert, M., 2013. Efficient 3-d scene analysis from streaming data. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, Karlsruhe, Germany, 6–10 May, pp. 2297–2304.
- John, G.H., Langley, P., 1995. Estimating continuous distributions in Bayesian classifiers. In: *Proceedings of the Eleventh Conference on Uncertainty in*

- Artificial Intelligence. Morgan Kaufman, Montreal, Canada, 18–20 August, pp. 338–345.
- Johnson, A.E., Hebert, M., 1999. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Machine Intell.* 21 (5), 433–449.
- Jutzi, B., Gross, H., 2009. Nearest neighbour classification on laser point clouds to gain object structures from buildings. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-1-4-7/W5*.
- Khoshelham, K., Oude Elberink, S.J., 2012. Role of dimensionality reduction in segment-based classification of damaged building roofs in airborne laser scanning data. In: *Proceedings of the International Conference on Geographic Object Based Image Analysis*, Rio de Janeiro, Brazil, 7–9 May, pp. 372–377.
- Kim, E., Medioni, G., 2011. Urban scene understanding from aerial and ground lidar data. *Machine Vis. Appl.* 22 (4), 691–703.
- Kim, H.B., Sohn, G., 2011. Random forests based multiple classifier system for power-line scene classification. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-5/W12*, pp. 253–258.
- Kolmogorov, A.N., 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, Germany.
- Kononenko, I., 1994. Estimating attributes: analysis and extensions of RELIEF. In: *Proceedings of the European Conference on Machine Learning*. Springer, Catania, Italy, 6–8 April, pp. 171–182.
- Lafarge, F., Mallet, C., 2012. Creating large-scale city models from 3d-point clouds: a robust approach with hybrid representation. *Int. J. Comput. Vis.* 99 (1), 69–85.
- Lalonde, J.F., Unnikrishnan, R., Vandapel, N., Hebert, M., 2005. Scale selection for classification of point-sampled 3d surfaces. In: *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*. IEEE, Ottawa, Canada, 13–16 June, pp. 285–292.
- Lari, Z., Habib, A., 2012. Alternative methodologies for estimation of local point density index: moving towards adaptive lidar data processing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B3*, pp. 127–132.
- Lee, I., Schenk, T., 2002. Perceptual organization of 3d surface points. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIV-3A*, pp. 193–198.
- Lim, E.H., Suter, D., 2009. 3d terrestrial lidar classifications with super-voxels and multi-scale conditional random fields. *Comput.-Aid. Des.* 41 (10), 701–710.
- Linsen, L., Prautzsch, H., 2001. Local versus global triangulations. In: *Proceedings of Eurographics*, Manchester, UK, 5–7 September, pp. 257–263.
- Liu, H., Motoda, H., Setiono, R., Zhao, Z., 2010. Feature selection: an ever evolving frontier in data mining. In: *Proceedings of the Fourth International Workshop on Feature Selection in Data Mining*, Hyderabad, India, 21 June, pp. 4–13.
- Lodha, S.K., Fitzpatrick, D.M., Helmbold, D.P., 2007. Aerial lidar data classification using AdaBoost. In: *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*. IEEE, Montreal, Canada, 21–23 August, pp. 435–442.
- Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS J. Photogr. Remote Sens.* 66 (6), S71–S84.
- Mitra, N.J., Nguyen, A., 2003. Estimating surface normals in noisy point cloud data. In: *Proceedings of the Annual Symposium on Computational Geometry*. ACM, San Diego, USA, 8–10 June, pp. 322–328.
- Monnier, F., Vallet, B., Soheilian, B., 2012. Trees detection from laser point clouds acquired in dense urban areas by a mobile mapping system. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3*, pp. 245–250.
- Muja, M., Lowe, D.G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: *Proceedings of the International Conference on Computer Vision Theory and Applications*, Lisbon, Portugal, 5–8 February, pp. 331–340.
- Munoz, D., Bagnell, J.A., Vandapel, N., Hebert, M., 2009. Contextual classification with functional max-margin Markov networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Miami, USA, 20–25 June, pp. 975–982.
- Najafi, M., Taghavi Namin, S., Salzmann, M., Petersson, L., 2014. Non-associative higher-order Markov networks for point cloud classification. In: *Proceedings of the European Conference on Computer Vision*. Springer, Zurich, Switzerland, 6–12 September, pp. 500–515.
- Niemeyer, J., Rottensteiner, F., Soergel, U., 2012. Conditional random fields for lidar point cloud classification in complex urban areas. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3*, pp. 263–268.
- Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS J. Photogr. Remote Sens.* 87, 152–165.
- Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D., 2002. Shape distributions. *ACM Trans. Graph.* 21 (4), 807–832.
- Oude Elberink, S., Kemboi, B., 2014. User-assisted object detection by segment based similarity measures in mobile laser scanner data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-3*, pp. 239–246.
- Özuysal, M., Fua, P., Lepetit, V., 2007. Fast keypoint recognition in ten lines of code. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Minneapolis, USA, 17–22 June, pp. 1–8.
- Pauly, M., Keiser, R., Gross, M., 2003. Multi-scale feature extraction on point-sampled surfaces. *Comput. Graph. Forum* 22 (3), 281–289.
- Pearson, K., 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. Roy. Soc. Lond. A* 187, 253–318.
- Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Machine Intell.* 27 (8), 1226–1238.
- Poullis, C., You, S., 2009. Automatic reconstruction of cities from remote sensor data. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Miami, USA, 20–25 June, pp. 2775–2782.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1988. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK.
- Pu, S., Rutzinger, M., Vosselman, G., Oude Elberink, S., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS J. Photogr. Remote Sens.* 66 (6), S28–S39.
- Quinlan, J.R., 1986. Induction of decision trees. *Machine Learn.* 1 (1), 81–106.
- Riedmiller, M., Braun, H., 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, USA, 28 March–1 April, pp. 586–591.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Rusu, R.B., Marton, Z.C., Blodow, N., Beetz, M., 2008. Persistent point feature histograms for 3d point clouds. In: *Proceedings of the International Conference on Intelligent Autonomous Systems*, Baden-Baden, Germany, 23–25 July, pp. 119–128.
- Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3d registration. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, Kobe, Japan, 12–17 May, pp. 3212–3217.
- Saeyns, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics* 23 (19), 2507–2517.
- Schapire, R.E., 1990. The strength of weak learnability. *Machine Learn.* 5 (2), 197–227.
- Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Trans. Geosci. Remote Sens.* 50 (11), 4534–4545.
- Schmidt, A., Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of full waveform lidar data in the Wadden Sea. *IEEE Geosci. Remote Sens. Lett.* 11 (9), 1614–1618.
- Secord, J., Zakhor, A., 2007. Tree detection in urban regions using aerial lidar and image data. *IEEE Geosci. Remote Sens. Lett.* 4 (2), 196–200.
- Serna, A., Marcotegui, B., 2013. Urban accessibility diagnosis from mobile laser scanning data. *ISPRS J. Photogr. Remote Sens.* 84, 23–32.
- Serna, A., Marcotegui, B., 2014. Detection, segmentation and classification of 3d urban objects using mathematical morphology and supervised learning. *ISPRS J. Photogr. Remote Sens.* 93, 243–255.
- Serna, A., Marcotegui, B., Goulette, F., Deschard, J.E., 2014. Paris-rue-Madame database: a 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. In: *Proceedings of the International Conference on Pattern Recognition Applications and Methods*. ACM, Angers, France, 6–8 March, pp. 819–824.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (3), 379–423.
- Shapovalov, R., Velizhev, A., Barinova, O., 2010. Non-associative Markov networks for 3d point cloud classification. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-3A*, pp. 103–108.
- Shapovalov, R., Vetrov, D., Kohli, P., 2013. Spatial inference machines. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Portland, USA, 23–28 June, pp. 2985–2992.
- Tokarczyk, P., Wegner, J.D., Walk, S., Schindler, K., 2013. Beyond hand-crafted features in remote sensing. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W1*, pp. 35–40.
- Tombari, F., Salti, S., Di Stefano, L., 2010. Unique signatures of histograms for local surface description. In: *Proceedings of the European Conference on Computer Vision*. Springer, Heraklion, Greece, 5–11 September, pp. 356–369.
- Unnikrishnan, R., Hebert, M., 2008. Multi-scale interest regions from unorganized point clouds. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, Anchorage, USA, 23–28 June, pp. 1–8.
- Vanegas, C.A., Aliaga, D.G., Benes, B., 2012. Automatic extraction of manhattan-world building masses from 3d laser range scans. *IEEE Trans. Visual. Comput. Graph.* 18 (10), 1627–1637.
- Velizhev, A., Shapovalov, R., Schindler, K., 2012. Implicit shape models for object detection in 3d point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3*, pp. 179–184.
- Vosselman, G., 2013. Point cloud segmentation for urban scene classification. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-7/W2*, pp. 257–262.
- Waldhauser, C., Hochreiter, R., Otepka, J., Pfeifer, N., Ghuffar, S., Korzeniowska, K., Wagner, G., 2014. Automated classification of airborne laser scanning point clouds. In: Koziel, S., Leifsson, L., Yang, X.-S. (Eds.), *Solving Computationally Expensive Engineering Problems: Methods and Applications*. Springer, New York, USA, pp. 269–292.
- Weimann, M., Jutzi, B., Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3d point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-5/W2*, pp. 313–318.

- Weinmann, M., Jutzi, B., Mallet, C., 2014. Semantic 3d scene interpretation: a framework combining optimal neighborhood size selection with relevant features. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3*, pp. 181–188.
- Weinmann, M., Urban, S., Hinz, S., Jutzi, B., Mallet, C., 2015. Distinctive 2d and 3d features for automated large-scale scene analysis in urban areas. *Comput. Graph. XXX, XXX-XXX*.
- West, K.F., Webb, B.N., Lersch, J.R., Pothier, S., Triscari, J.M., Iverson, A.E., 2004. Context-driven automated target detection in 3-d data. In: *Proceedings of SPIE 5426, Automatic Target Recognition XIV*, SPIE, Orlando, USA, 12 April, pp. 133–143.
- Wurm, K.M., Kretschmar, H., Kümmerle, R., Stachniss, C., Burgard, W., 2014. Identifying vegetation from laser data in structured outdoor environments. *Robot. Autonom. Syst.* 62 (5), 675–684.
- Xiong, X., Munoz, D., Bagnell, J.A., Hebert, M., 2011. 3-d scene analysis via sequenced predictions over points and regions. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, Shanghai, China, 9–13 May, pp. 2609–2616.
- Xu, S., Vosselman, G., Oude Elberink, S., 2014. Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS J. Photogr. Remote Sens.* 88, 1–15.
- Yu, L., Liu, H., 2003. Feature selection for high-dimensional data: a fast correlation-based filter solution. In: *Proceedings of the International Conference on Machine Learning*. AAAI Press, Washington, USA, 21–24 August, pp. 856–863.
- Yu, T.H., Woodford, O.J., Cipolla, R., 2013. A performance evaluation of volumetric 3d interest point detectors. *Int. J. Comput. Vis.* 102 (1–3), 180–197.
- Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., Liu, H., 2010. Advancing feature selection research – ASU feature selection repository. Technical Report, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, USA.
- Zhou, Q.Y., Neumann, U., 2013. Complete residential urban area reconstruction from dense aerial lidar point clouds. *Graph. Models* 75 (3), 118–125.
- Zhou, L., Vosselman, G., 2012. Mapping curbstones in airborne and mobile laser scanning data. *Int. J. Appl. Earth Observ. Geoinform.* 18 (1), 293–304.



Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on Processing Large Geospatial Data

Distinctive 2D and 3D features for automated large-scale scene analysis in urban areas

M. Weinmann^{a,*}, S. Urban^a, S. Hinz^a, B. Jutzi^a, C. Mallet^b^a Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Englerstr. 7, 76131 Karlsruhe, Germany^b Université Paris-Est, IGN, SRIG, MATIS, 73 Avenue de Paris, 94160 Saint-Mandé, France

ARTICLE INFO

Article history:

Received 1 September 2014

Received in revised form

23 December 2014

Accepted 29 January 2015

Available online 7 February 2015

Keywords:

3D scene analysis

Point cloud

Feature

Classification

Large-scale

Urban

ABSTRACT

We propose a new methodology for large-scale urban 3D scene analysis in terms of automatically assigning 3D points the respective semantic labels. The methodology focuses on simplicity and reproducibility of the involved components as well as performance in terms of accuracy and computational efficiency. Exploiting a variety of low-level 2D and 3D geometric features, we further improve their distinctiveness by involving individual neighborhoods of optimal size. Due to the use of individual neighborhoods, the methodology is not tailored to a specific dataset, but in principle designed to process point clouds with a few millions of 3D points. Consequently, an extension has to be introduced for analyzing huge 3D point clouds with possibly billions of points for a whole city. For this purpose, we propose an extension which is based on an appropriate partitioning of the scene and thus allows a successive processing in a reasonable time without affecting the quality of the classification results. We demonstrate the performance of our methodology on two labeled benchmark datasets with respect to robustness, efficiency, and scalability.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The automated analysis of 3D point clouds has become a topic of great importance in photogrammetry, remote sensing, computer vision and robotics. One avenue of research directly addresses the analysis of urban environments, where recent investigations focus on 3D reconstruction [1–3], consolidation of imperfect scan data [4,5], object detection [6–9], extraction of roads and curbstones or road markings [10–12], urban accessibility analysis [13], recognition of power-line objects [14], extraction of building structures [15], vegetation mapping [16], large-scale city modeling [17], semantic perception for ground robotics [18] and semantization of complex 3D scenes [19]. A common task for many of these different applications consists of point cloud classification [20,21], where each 3D point is assigned a specific (semantic) class label.

Addressing the task of urban point cloud classification – where the spatial 3D data may be collected via dense matching as well as airborne, terrestrial and/or mobile laser scanning – we face a variety of challenges arising from the complexity of respective 3D scenes caused by an irregular sampling and very different types of objects. Since the results of urban 3D scene analysis may vary from

one dataset to another, publicly available standard datasets are desirable in order to compare the performance of different methodologies. Consequently, there has been a steadily increasing availability of 3D point cloud datasets in recent years [22]. However, urban point clouds with respective point-wise manual annotations in terms of semantic class labels are still rarely available, although this represents a prerequisite for supervised point cloud classification. One of the most widely used datasets is the Oakland 3D Point Cloud Dataset [23] which, however, only contains approximately 1.6 million labeled 3D points. Hence, this dataset is not tailored to designing large-scale processing pipelines.

Due to the recent technological advancements, it is meanwhile possible to collect geospatial data in a fast and efficient way via terrestrial and mobile laser scanning. In order to foster research in advanced 3D point cloud processing, two labeled point cloud datasets representing densely sampled urban environments with a significantly higher number of 3D points have been presented recently [24,25]. These can be considered as a first step towards large geospatial datasets in terms of city-scale or even larger. The availability of such datasets is important for comparing large-scale processing workflows which is the core issue of a recent benchmark [25] and addressed in this paper.

In our work, we consider each individual 3D point and its local 3D neighborhood for extracting respective geometric features. By exploiting a fully generic approach for optimizing the neighborhood

* Corresponding author. Tel.: +49 721 608 47302.

E-mail address: martin.weinmann@kit.edu (M. Weinmann).

size, our approach is generally applicable and not tailored to a specific dataset. Furthermore, our approach represents a basic requirement for either smooth labeling techniques or methods involving contextual information, since both of them are based on the results of individual point classification. In summary, our contributions extend [26] and consist of

- a new methodology for large-scale urban 3D point cloud classification,
- an in-depth analysis of a powerful strategy for recovering individual 3D neighborhoods of optimal size,
- efficient feature extraction and classification, and
- an extension towards data-intensive processing.

In the following, we first reflect related work in Section 2 and provide an up-to-date view of approaches for processing 3D point cloud data in order to efficiently obtain significant information contained in the data. Subsequently, in Section 3, we present a methodology which is closely linked to recent investigations on 3D scene analysis involving optimal neighborhoods and different classifiers [27,28]. Based on these investigations with a very detailed evaluation, we can directly select the most appropriate solution with respect to urban 3D scene analysis. The considered criteria address feasibility in terms of simplicity and reproducibility of the involved components as well as performance in terms of accuracy and computational effort. We further introduce an increase in efficiency resulting from efficient neighborhood recovery. In order to extend the applicability of the selected methodology towards huge datasets, in Section 4, an extension towards large-scale urban point cloud classification is presented which does not affect the quality of the results, but allows the successive processing of huge point clouds in a reasonable time. Afterwards, the datasets involved in our experiments and the experimental results are presented in Sections 5 and 6. The derived results are subsequently discussed in Section 7. Finally, in Section 8, concluding remarks are provided and suggestions for future work are outlined.

2. Related work

Although modern devices nowadays allow the acquisition of additional information such as echo-based features or full-waveform features [29,30] which may alleviate 3D scene analysis, we focus on the use of geometric features as most of the available point cloud datasets only contain spatial 3D information. Other features may however easily be appended to the feature vectors defined in the scope of our work. In the following, we first present fundamental concepts for defining appropriate features for 3D scene analysis. Subsequently, we discuss approaches for (i) optimizing the derived feature vectors by involving feature relevance or (ii) describing the local 3D structure either at a single, but optimized neighborhood or at multiple neighborhoods of different size. Finally, we briefly reflect related work on classification methods for 3D scene analysis.

2.1. Feature design

A crucial issue for 3D scene analysis consists of designing appropriate features. For this purpose, numerous histogram-based methods have been proposed which accumulate information about the spatial 3D geometry into a histogram according to a specific quantized domain [31]. A still popular approach has been presented with Spin Images [32], where 2D histograms are derived by spinning a 2D plane patch around the surface normal and counting the number of points falling into each bin of the 2D

patch. Powerful alternatives have recently been presented with Point Feature Histograms (PFHs) [33] and their modification denoted as Fast Point Feature Histograms (FPFHs) [34]. Considering a point \mathbf{X} and all points within its local neighborhood, these approaches first assign the respective surface normal to \mathbf{X} and then sample geometric relations between the nearest neighbors in terms of angular variations and point distances into histograms. A different strategy has been presented with shape distributions [35] which are based on the idea of randomly sampling simple geometric measures such as distances or angles in order to obtain a descriptor characterizing the neighborhood around a point \mathbf{X} [36]. Furthermore, a combination of histograms with signatures has been presented with the Signature of Histograms of Orientations (SHOT) descriptor [31] in order to achieve a better balance between descriptiveness and robustness. The performance of different histogram descriptors has been compared in [37]. However, for all these approaches, single entries of the derived feature vectors are hardly interpretable.

Alternatively, the local 3D structure can be described by deriving the 3D covariance matrix from the spatial coordinates of a point \mathbf{X} and its neighbors. Based on the respective eigenvalues, a direct scene analysis may be conducted [38], or a set of features may be defined [39] which encapsulate geometric information about the local 3D structure. In particular the latter approach is nowadays commonly applied in lidar data processing, and the respective features or feature subsets are typically complemented with other geometric features [40,23,41,27,42,43]. A specific advantage consists of the fact that the respective entries in the feature vector are interpretable as they address local 3D shape primitives. In contrast to 3D covariance matrices encoding the relationships among points within a local neighborhood, covariance matrices of higher dimension have been used to combine multiple features such as angular measures and point distances to a compact representation [44]. Further information such as radiometric information may also be taken into account in this representation.

2.2. Feature relevance assessment

For compensating a lack of knowledge, often as many features as possible are extracted and involved in the classification process, although some of these features may be more and others less suitable. Consequently, investigations focusing on feature selection have also been introduced for 3D point cloud processing in order to improve the classification accuracy while simultaneously reducing both computational effort and memory consumption. Respective approaches allow a ranking of single features according to their relevance and the selection of a subset of the best-ranked features. Whereas the ranking can be obtained by involving a classifier [29,30,45], classifier-independent approaches are in the focus of filter-based feature selection which offers both simplicity and efficiency. Herein, univariate filter-based feature selection methods rely on a score function which simply evaluates feature-class relations based on the training data. A general relevance metric composed of several score functions addressing different intrinsic properties of the given training data has recently been proposed [27]. In contrast, multivariate filter-based feature selection methods rely on both feature-class and feature-feature relations in order to discriminate between relevant, irrelevant and redundant features [28].

2.3. Single-scale vs. multi-scale representation

The heuristic determination of local 3D neighborhoods is conducted either with respect to the absolute size [38,46] or with respect to the scale parameter [27]. However, in order to avoid heuristically determining a suitable neighborhood size – which

may even be specific for each dataset and which may not be identical across all 3D points in consideration – there have been few attempts to automatically derive individual neighborhoods of optimal size. Respective approaches rely on the local surface variation [47,48], a combination involving curvature, point density and noise of normal vector estimation [49,40], dimensionality features [41] or the measure of eigenentropy [28]. The need for involving such techniques becomes for instance apparent when analyzing the behavior of features derived from the 3D structure tensor and shape distribution features across different scales [36] or when observing a significant improvement in comparison to neighborhoods of fixed scale [28].

Instead of focusing on the concept of optimal neighborhoods, a consideration of features at multiple scales may be applied. For instance, it has been proposed to calculate features at different scales and involve a training procedure in order to define which combination of scales allows the best separation of different classes [50]. Further approaches even extract features based on different entities such as points and regions [51,52].

2.4. Individual vs. contextual classification

When selecting an appropriate classifier, we may follow the strategy of individual point classification by exploiting respective feature vectors, where Support Vector Machines [53], Random Forests [29], AdaBoost [54] or classical Maximum Likelihood (ML) classifiers exploiting Gaussian Mixture Models (GMMs) [40] represent the most commonly applied approaches for point cloud classification. Alternatively, contextual learning approaches may be applied which address the idea that semantic labels of nearby 3D points tend to be correlated and hence involve relationships among 3D points within a local neighborhood in addition to the respective feature vectors. Respective approaches applied for point cloud classification are represented by Associative and non-Associative Markov Networks [23,55,56], Conditional Random Fields [21], multi-stage inference procedures focusing on point cloud statistics and relational information over different scales [51], and spatial inference machines modeling mid- and long-range dependencies inherent in the data [57].

When applying contextual learning approaches, it has to be taken into account that the relationships are inferred from the training data. Hereby, the local neighborhood is typically different from the neighborhood used for feature extraction. Furthermore, since the training data is limited, exact inference is computationally intractable and therefore either approximate inference techniques or smoothing techniques are commonly applied. The selection of an approximate inference technique remains challenging as there is no indication towards an optimal inference strategy, and such techniques quickly reach their limitations if the considered neighborhood becomes too large. In contrast, smoothing techniques enforce the desirable smooth labeling of nearby 3D points and may thus provide a significant improvement with respect to classification accuracy [58]. Such smoothing techniques, however, exploit either

the estimated probability of a 3D point belonging to each of the defined classes or the direct assignment of the respective label, and thus the results of an individual point classification.

3. Methodology

For 3D scene analysis in terms of uniquely assigning each 3D point a semantic label, we propose a fully automatic and generic methodology which consists of three successive steps (Fig. 1). In the first step, each 3D point is characterized with an individual local 3D neighborhood of optimal size (Section 3.1). This allows an extraction of highly distinctive features which is pursued in the second step, where various geometric 3D and 2D features are taken into consideration (Section 3.2). Finally, in the third step, the distinctive features and a given set of training examples are provided to a supervised classification scheme (Section 3.3). The main focus of our investigations is put on feature design in terms of deriving distinctive geometric features from individual neighborhoods of optimal size.

3.1. Neighborhood selection

Considering a point \mathbf{X} in a point cloud \mathcal{P} , the respective neighborhood selection generally involves (i) a suitable neighborhood definition, (ii) an efficient recovery of the local neighborhood and (iii) an optimal parameterization of the neighborhood in terms of neighborhood size. These aspects are addressed in the following subsections.

3.1.1. Neighborhood definitions

In general, very different approaches may be applied in order to define the local 3D neighborhood \mathcal{N} of a given 3D point $\mathbf{X} \in \mathbb{R}^3$. For instance, a spherical neighborhood definition may be applied, where the local neighborhood is formed by all 3D points in a sphere with a fixed radius [59]. An alternative consists of applying a cylindrical neighborhood definition, where the local neighborhood is formed by all those 3D points whose 2D projections onto the ground plane are within a circle with a fixed radius [60]. A further neighborhood definition involves a fixed number of k closest 3D points for a given query point [61], which results in spherical neighborhoods of variable absolute size. Note that all these neighborhood definitions rely on the specification of one free scale parameter.

Since we want to account for more flexibility in case of varying point density and thereby avoid including a-priori knowledge on the scene, we employ a neighborhood definition based on the k closest neighbors of a given 3D point. Consequently, the nearest neighbors have to be recovered and an appropriate k has to be selected.

3.1.2. Exact vs. approximate nearest neighbors

As a consequence of the selected neighborhood definition, a computationally quite expensive part consists of the calculation of

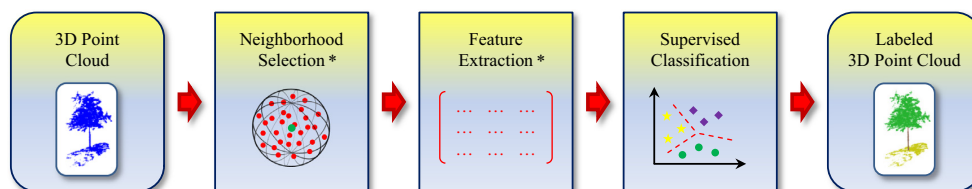


Fig. 1. The proposed methodology: after calculating an individual local 3D neighborhood of optimal size for each 3D point, highly distinctive 3D and 2D features are extracted and provided to a supervised classification scheme in order to obtain a semantically labeled 3D point cloud (*The respective implementation (Matlab, C++ and binaries) is released with this paper and available at <http://www.ipf.kit.edu/code.php>).

the nearest neighbors for each 3D point. This nearest neighbor search can formally be described as follows: given a point set $\mathcal{P} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$ in a 3-dimensional Euclidean vector space, those points $\mathbf{X} \in \mathcal{P}$ that are nearest to a given query point \mathbf{X}_Q should be recovered efficiently. The commonly used approach for nearest neighbor search is based on a Kd -tree [62] which represents a compact, hierarchical data structure for point sets sampled from a K -dimensional manifold. A point in a Kd -tree with N points can thus be localized with an average complexity of $O(\log N)$ and, in the worst case, with a complexity of $O(N)$ [63]. In order to further increase efficiency, an approximate nearest neighbor search has been proposed [64] which can be much faster than the exact nearest neighbor search with only little loss in accuracy since non-optimal neighbors may be returned. For details on these approaches, we refer to an extensive survey on data structures [65]. Addressing the criteria of query time and accuracy, a powerful approach with public availability and fully automatic parameter selection has been presented in the Fast Library for Approximate Nearest Neighbors (FLANN) [66] which is based on either searching hierarchical K -means trees with a priority search order or using multiple randomized Kd -trees. Hence, we apply the FLANN for nearest neighborhood search.

3.1.3. Optimal parameterization

Besides an efficient nearest neighbor search, it is desirable to automatically find the optimal parameterization for the selected neighborhood definition which is based on the k closest neighbors in our case. When addressing this issue, we may also take into account that the optimal choice of the parameter k (which is also commonly referred to as scale) may vary within a dataset since k certainly depends on the respective 3D structures and thus also on the respective class label.

Seminal work addressing the selection of an optimal value for the scale parameter k and thus the selection of the optimal neighborhood size is based on fundamental geometric properties of the point cloud data. For instance, the local surface variation (i.e. the change of curvature) may be exploited since a critical neighborhood size is indicated by a significant change of curvature when successively increasing the neighborhood by adding the next closest 3D point [47,48]. Furthermore, an iterative scheme involving curvature, point density and noise of normal vector estimation has been proposed [49,40]. Whereas these approaches are tailored for smoothly varying surfaces, they may face severe challenges when considering surfaces acquired with lidar systems or dense matching. Consequently, we focus on the use of multiple low-level geometric features which adequately capture the variability of natural environments.

Our approach is inspired by dimensionality based scale selection [41], where optimal neighborhood size selection is based on the idea that the optimal neighborhood size favors one dimensionality the most. More specifically, for describing the local 3D structure around a 3D point $\mathbf{X} = \mathbf{X}_0$, the dimensionality features of linearity L_λ , planarity P_λ and scattering S_λ are derived from the set of neighboring 3D points \mathbf{X}_i with $i = 1, \dots, k$ by considering the respective 3D covariance matrix

$$\mathbf{C} = \frac{1}{k+1} \sum_{i=0}^k (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \quad (1)$$

with

$$\bar{\mathbf{X}} = \frac{1}{k+1} \sum_{i=0}^k \mathbf{X}_i \quad (2)$$

which is also known as the 3D structure tensor. This 3D structure tensor \mathbf{C} represents a symmetric positive-definite matrix. Consequently, its eigenvalues exist, are non-negative and correspond to

an orthogonal system of eigenvectors. For the sake of generality, we assume that there might not necessarily be a preferred variation with respect to the eigenvectors. This results in the general case of a structure tensor with rank 3 as well as eigenvalues λ_1, λ_2 and λ_3 with $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$. Based on these eigenvalues, the dimensionality features are defined as

$$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1} \quad P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1} \quad S_\lambda = \frac{\lambda_3}{\lambda_1} \quad (3)$$

and, as these features sum up to 1, they may be considered as the probabilities of a 3D point to be labeled as 1D, 2D or 3D structure [41]. Favoring one dimensionality the most thus corresponds to minimizing a measure given by the Shannon entropy [67] as

$$E_D = -L_\lambda \ln(L_\lambda) - P_\lambda \ln(P_\lambda) - S_\lambda \ln(S_\lambda) \quad (4)$$

across different scales k , and the optimal neighborhood size corresponds to the respective k with the minimal Shannon entropy. Instead of directly varying the scale parameter k as later tested in [28], however, the respective radius has been taken into account in [41]. Sampling the interval $[r_1, r_2]$ between specified radii r_1 and r_2 into 16 scales, where the radii are not linearly increased since the radius of interest is usually closer to r_1 , the optimal neighborhood size corresponds to the radius yielding the minimal Shannon entropy. However, the two radii r_1 and r_2 depend on various characteristics of the given data and are therefore specific for each dataset [41].

In order to avoid strong assumptions on the presence of specific geometric structures in the scene and to get rid of heuristic parameter selection, a more general solution for optimal neighborhood size selection has been proposed very recently [28] which has proven to outperform dimensionality based scale selection. Instead of exploiting the three dimensionality features, this approach directly exploits the eigenvalues of the 3D structure tensor which correspond to the principal components and thus span a 3D covariance ellipsoid. Normalizing the three eigenvalues λ_1, λ_2 and λ_3 by their sum Σ_λ yields normalized eigenvalues e_1, e_2 and e_3 summing up to 1. Thus, in analogy to the dimensionality based scale selection, the measure of eigenentropy E_λ given by the Shannon entropy according to

$$E_\lambda = -e_1 \ln(e_1) - e_2 \ln(e_2) - e_3 \ln(e_3) \quad (5)$$

is calculated and minimized across different scales k which, in turn, relates to minimizing the disorder of points within a 3D covariance ellipsoid. The optimal neighborhood size finally corresponds to the respective k with the minimal eigenentropy. For our experiments, we consider relevant statistics to start with $k_1 = 10$ – which is in accordance to [41] – and successively increase the scale parameter k with a step size of $\Delta k = 1$ up to an upper bound of $k_2 = 100$ as proposed in [28], which already represents a relatively high number. Note that k_2 can be arbitrary, and hence we will later focus on this issue in the experimental results.

The resulting optimal neighborhood size thus depends on contextual information preserved in the spatial arrangement of neighboring 3D points and may even be different for each individual 3D point. Even though optimal neighborhood size selection causes a higher computational effort with respect to both processing time and memory consumption, it should be taken into consideration since the classification accuracy is significantly improved according to recent investigations [28].

3.2. Feature extraction

Since many of the publicly available 3D point cloud datasets only contain information about the spatial 3D geometry in terms of XYZ coordinates, we focus on the use of geometric features. Such

geometric features typically rely on a local 3D neighborhood which has already been derived in the previous step. Involving the optimal neighborhood size for a respective 3D point, we may assume that highly distinctive geometric features can be derived from the set of 3D points within the neighborhood. Due to the high point density of recently published lidar point cloud datasets, the selected scale k tends to correspond to a relatively small absolute size, and the respective local 3D structure can therefore only be described with low-level geometric features. Such features show a specific behavior for planar patches, ridges, edges and vertices. By involving a 2D projection, we may also account for structures with larger extent, e.g. in the vertical direction.

In order to define adequate low-level geometric features, we follow the strategy involving a variety of geometric 3D and 2D features [27], but we additionally increase their distinctiveness by taking into account the optimal neighborhood size for each individual 3D point [28]. In total, a set of 21 distinctive low-level geometric features is thus calculated for each 3D point. For the sake of clarity, we briefly describe the involved 3D and 2D features in the following subsections. The respective code (Matlab, C++ and binaries) is released with this paper.

3.2.1. 3D features

Obviously, a variety of 3D features can directly be derived by describing basic geometric properties of the considered 3D neighborhood such as the absolute height Z of the considered 3D point \mathbf{X} , the radius r_{k-NN} of the sphere encompassing the local 3D neighborhood, the local point density D defined as number of points per unit volume and the verticality V which is based on the vertical component of the local normal vector. Hereby, the local normal vector is related to the eigenvector corresponding to the smallest eigenvalue of the respective 3D structure tensor. Furthermore, all 3D points within the neighborhood may be considered in order to calculate the maximum height difference ΔZ_{k-NN} and the standard deviation $\sigma_{Z,k-NN}$ of height values.

Additionally, we take into account that the 3D structure tensor encodes the general distribution of 3D points within the local neighborhood. Consequently, the normalized eigenvalues e_1 , e_2 and e_3 of the 3D structure tensor may also be exploited to define local 3D shape features. Besides the aforementioned dimensionality features of linearity L_λ , planarity P_λ and scattering S_λ , further features are represented by omnivariance O_λ , anisotropy A_λ , eigenentropy E_λ , the sum Σ_λ of eigenvalues and the local surface variation C_λ .

3.2.2. 2D features

Taking into account that urban areas are typically characterized by an aggregation of man-made structures, specific geometric relations in terms of symmetry and orthogonality are likely to occur. In particular, we may face a large number of structures which are oriented perpendicular to a horizontally oriented plane, e.g. building façades, traffic signs or curbstone edges. Consequently, a projection of the 3D point cloud onto a horizontally oriented plane might reveal additional information and possibly also clear evidence about the presence of specific structures in the observed scene. Respective features which are based on this 2D projection can easily be defined as 2D properties of the neighborhood such as the radius $r_{k-NN,2D}$ and the local point density D_{2D} .

Furthermore, the coordinates resulting from the 2D projection of all 3D points within the neighborhood may be exploited to derive the 2D covariance matrix also known as the 2D structure tensor. In analogy to the 3D case, the respective two eigenvalues may be used to define characteristic features such as the sum $\Sigma_{\lambda,2D}$ of the eigenvalues and their ratio $R_{\lambda,2D}$.

Interestingly, the aforementioned 2D features are based on spherical neighborhoods of relatively small absolute metric size, whereas man-made structures tend to provide a similar behavior across different height levels, i.e. across several meters of height. Consequently, a discretization of the 2D projection in terms of a 2D accumulation map [68] with discrete, quadratic bins (here with a side length of 0.25 m) may provide further interesting properties about local 3D structures which are not yet covered by the already defined features. Respective features based on the accumulation map have been proposed with the number M of points falling in the respective bin as well as the maximum height difference ΔZ and standard deviation σ_Z of height values within the respective bin. Particularly, the feature M provides clear evidence on the existence of building façades and, if the point density is sufficiently high, also on the existence of curbstone edges. The detection of such man-made structures, in turn, is important for urban accessibility analysis which is one of the main intentions of current research.

3.3. Supervised classification

For classification, we apply a standard supervised classification scheme involving a set of training examples which, in turn, encapsulate feature vectors as well as the respective class labels. An adequate choice among a variety of classification strategies and respective approaches, however, should directly address applicability, reproducibility, and scalability in order to facilitate 3D scene analysis in large-scale urban environments. In the following, we first motivate our choice for the classifier and subsequently address the issue of how to obtain suitable training data.

3.3.1. Classifier selection

Since we focus on the applicability and reproducibility of all involved components, the involved classification strategy should be easy-to-use without crucial parameter selection, and respective implementations should be available in different software packages. For this reason, we focus on individual point classification based on a set of derived features. With the intention of processing huge point clouds, it is mandatory to apply a powerful but still computationally efficient classifier. In particular for scalability towards huge datasets, combining a set of weak classifiers such as decision trees via bootstrap aggregating which is commonly referred to as bagging [69] has proven to be successful. Using bootstrapped replica of the training data, i.e. subsets of the complete training data which are randomly drawn with replacement [70], diversity is obtained by training a weak learner of the same type for each subset of the training data.

The most popular example for bagging is represented by Random Forests [71] which represent a modern discriminative method and provide efficiency in case of a large amount of input data. Efficiency in this context covers simplicity and a high degree of parallelization which results in a fast classification scheme. Additionally, robustness to outliers, noise and missing data is provided. More specifically, a Random Forest represents an ensemble of randomly trained decision trees and thus aggregates hypotheses derived via decision trees which, in turn, are trained over different distributions of the training data. Thereby, each decision tree is constructed based on a top-down strategy successively selecting the variable which best splits the respective training data. Consequently, a split function and a stopping criterion have to be defined. Since, in the training phase, individual decision trees are assumed to be trained on randomly selected subsets of the given training data, the trees may be expected to be randomly different from one another which results in a decorrelation between individual tree predictions and thus improved

generalization and robustness [72]. Once a Random Forest is trained, the classification phase consists of letting each decision tree vote for a single class and assigning the respective label according to the majority vote of all decision trees. The parameters to be specified are the number of involved decision trees and the tree depth. In the experiments, we use a Random Forest with 100 trees and a tree depth of $\lfloor \sqrt{d} \rfloor$, where d equals the number of extracted features, i.e. $d=21$.

3.3.2. Training data

Given a set of training examples, we have to take into account that an unbalanced distribution of training examples across all classes may have a detrimental effect on the training process [72]. Consequently, we involve a class re-balancing in terms of randomly selecting the same number of training examples for each class. Alternatively, the known prior class distribution of the training set could be used for weighting the contribution of each class.

4. Extension towards large-scale urban point cloud classification

Recent investigations clearly show that the use of individual neighborhoods with optimal size has a significant, beneficial impact on the classification results [28]. However, the additional calculations cause a drastic increase in computational effort. Consequently, the described methodology is suited to process point clouds containing only up to a few millions of 3D points. When considering huge point clouds at city scale with possibly billions of points – which is the aim of recent effort in order to obtain an adequate 3D model of a whole city like Paris – an extension of the presented methodology has to be introduced.

The extension described in this paper does not affect the quality of 3D scene analysis, but only the scalability of the methodology in order to process larger datasets. Specifically, it focuses on successively processing a huge point cloud by applying a 2D sliding window function which is shifted within a horizontally oriented plane in discrete steps and involves a small padding region in order to avoid discontinuities at its borders. The size of the window should be chosen in a way that, for each step, a still reasonable number of 3D points is in consideration since the window size is practically limited depending on the available memory size. The partial results are subsequently merged together to obtain the results for the full scene. Accordingly, the approach represents a partitioning of the scene into subparts which, in turn, are extended by small padding regions at the borders and can be processed in parallel. For the sake of simplicity, we focus on two specific scenarios and select the one which is suited best with respect to the given data:

- For the more general scenario, we propose the use of a tiling approach. Taking a defined area within a horizontally oriented plane, e.g. a small quadratic area of $10\text{ m} \times 10\text{ m}$, allows a successive processing of data and thus also the analysis of huge point clouds at city-scale, where directly applying the methodology is intractable due to the computational burden with respect to computational effort and memory consumption.
- Without loss of generality, we may also take into account that the recently published benchmark datasets describe straight street sections with a length of approximately 160–200 m [24,25]. Consequently, the tiling approach can be substituted by a slicing approach, where the slices have a specified width, e.g. a width of 10 m, along the street direction and infinite extent along the two perpendicular directions.

Note that for both partitioning schemes, those 3D points within the small padding around the considered tile or slice are also used if they are within the neighborhood of other 3D points within the considered part of the scene in order to avoid artifacts at boundaries between tiles or slices. Due to the high point density of recent point cloud datasets, a padding with a width of 0.50 m is considered to be sufficient. As alternative to scene partitioning, streaming methods could be applied [73].

Consequently, we may repeat the proposed workflow consisting of (i) optimal neighborhood selection, (ii) feature extraction and (iii) classification independently for each tile or for each slice. Thus, large-scale urban 3D scene analysis is composed of successive steps which exploit the same methodology and could be parallelized if respective hardware is available.

5. Datasets

Since we focus on the issue of urban 3D point cloud classification and want to facilitate an objective comparison to other methodologies, we consider two publicly available and labeled 3D point cloud datasets representing densely sampled urban environments. Both datasets have been acquired in the city of Paris, France, via mobile laser scanning (MLS) systems:

*Paris-rue-Madame database*¹ [24]: This 3D point cloud dataset has been acquired with the mobile laser scanning system L3D2 [74] equipped with a Velodyne HDL32. It contains 20 million points corresponding to a digitized street section with a length of approximately 160 m. A respective annotation has been conducted in a manually assisted way and includes both point-wise labels (26 different classes) and segmented objects (642 objects in total). This annotation relies on an initial segmentation based on elevation images [9] which is followed by a manual refinement [24].

Paris-rue-Cassette database [25]: This point cloud dataset has been acquired with the mobile laser scanning system called Stereopolis II [75] in January 2013. This system captures the local 3D geometry of the scene with two plane sweep lidars (Riegl LMS-Q120i) placed on each side of the vehicle in order to mainly observe the building façades with a centimeter accuracy and a 3D lidar (Velodyne HDL-64E) to observe the bottom part in between. The dataset contains 12 million points corresponding to a digitized street section with a length of approximately 200 m, and a manually assisted annotation is available which includes both point-wise labels and segmented objects. The respective annotation is based on recovering a regular 2D topology for the point cloud stream during data acquisition and an offline human interaction via a graph editing tool based on standard 2D image segmentation techniques [76]. A further extension in the form of 10 different zones with a total number of about 100 million points has been released in the scope of a recent contest [25].

6. Experimental results

In the following, we first provide an in-depth analysis addressing the results of optimal neighborhood size selection for both involved datasets. Afterwards, we focus on the respective results for individual point classification and, finally, we demonstrate the performance of the new approach with respect to computational effort.

¹ *Paris-rue-Madame database: MINES ParisTech 3D mobile laser scanner dataset from Madame street in Paris.* ©2014 MINES ParisTech. MINES ParisTech created this special set of 3D MLS data for the purpose of detection-segmentation-classification research activities, but does not endorse the way they are used in this project or the conclusions put forward. The database is publicly available at <http://cmm.ensmp.fr/~serna/rueMadameDataset.html> (last access: 30 August 2014).

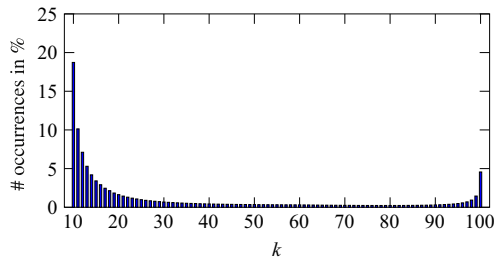


Fig. 2. Distribution of the assigned optimal neighborhood size k for all 3D points in the Paris-rue-Madame database.

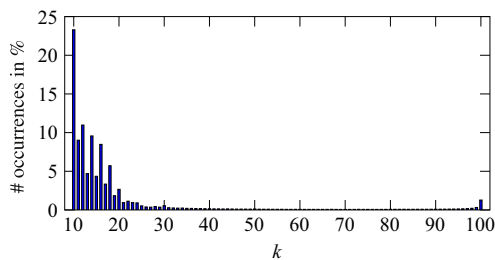


Fig. 3. Distribution of the assigned optimal neighborhood size k for all 3D points in the Paris-rue-Cassette database.

6.1. Optimal neighborhood size selection

In order to obtain more insights into the process of optimal neighborhood size selection, we first focus on a detailed analysis of the respective results. Note that each 3D point is assigned an individual value for k and that all integer values in $[k_1, k_2]$ with $k_1 = 10$ and $k_2 = 100$ are taken into consideration. Since the upper boundary k_2 has been selected for reasons of computational costs and may principally be set to an arbitrary value, we first consider the distribution of the parameter k across all 3D points. For this purpose, the respective distributions of the assigned optimal neighborhood size across all 3D points obtained for the Paris-rue-Madame database with 20 million 3D points and the Paris-rue-Cassette database with 12 million 3D points are visualized as histograms in Figs. 2 and 3. These figures clearly reveal a trend towards small values of k . Since the last bin in the histograms ($k_2 = 100$) is likely to also represent those 3D points which might have a higher value than $k_2 = 100$, a small increase can be observed. However, the percentage of 3D points which are assigned an optimal neighborhood with less than k_2 neighbors is 95.44% and 98.72% for the Paris-rue-Madame database and the Paris-rue-Cassette database, respectively. This shows that our selection of k_2 is appropriate.

A qualitative visualization of the distributions of the assigned optimal neighborhood size across all 3D points of both involved datasets is depicted in Figs. 4 and 5. For the Paris-rue-Cassette database, a significantly smoother behavior can be observed.

6.2. Individual point classification

In order to assign a semantic label to each individual 3D point, we exploit only those 3D points with labels corresponding to the most dominant semantic classes *façade*, *ground*, *cars*, *motorcycles/2 wheelers*, *traffic signs/road inventory*, *pedestrians* and *vegetation* (Table 1), which represent a fraction of 99.81% of the Paris-rue-Madame database and 99.56% of the Paris-rue-Cassette database. All 3D points belonging to the other classes are removed since the respective number of samples per class is not considered to be

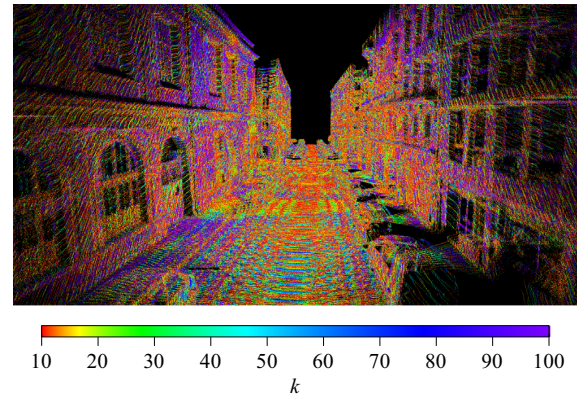


Fig. 4. Qualitative distribution of the assigned optimal neighborhood size k for all 3D points in the Paris-rue-Madame database.

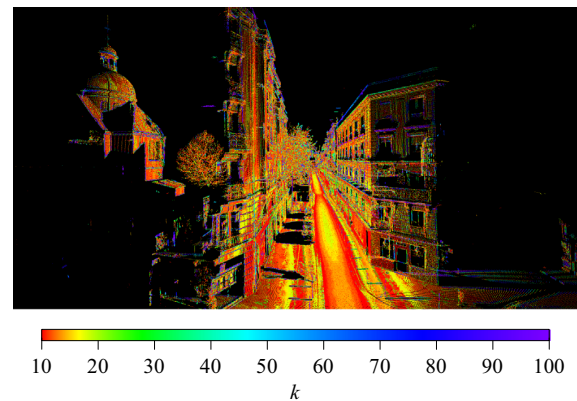


Fig. 5. Qualitative distribution of the assigned optimal neighborhood size k for all 3D points in the Paris-rue-Cassette database.

representative. For training, we randomly select a small, balanced training set \mathcal{X} with 1,000 training examples per class, and the remaining data is used as test set \mathcal{Y} .

In order to allow a comparison to recent work, we design the evaluation scheme in analogy to [28]. Thus, our evaluation is based on (i) overall accuracy which indicates the performance of the classifier on the test set, (ii) recall which represents a measure of completeness or quantity, (iii) precision which represents a measure of correctness or quality, (iv) F_1 -score which combines recall and precision with equal weights, (v) mean class recall which represents an averaged measure of completeness/quantity across all classes and (vi) a visual inspection of the derived results.

Evaluating our approach on both datasets, the overall accuracy is 88.82% for the Paris-rue-Madame database and 89.60% for the Paris-rue-Cassette database. The resulting recall and precision values as well as the corresponding F_1 -scores are provided in Tables 2 and 3. Accordingly, mean class recall values of 83.53% and 81.78% are obtained for the Paris-rue-Madame database and the Paris-rue-Cassette database, respectively. Finally, a visual impression on the quality of the derived results for individual point classification is depicted in Figs. 6 and 7.

6.3. Computational effort

The experiments have been conducted on an Intel Core i7-3820 with 3.6 GHz and 64GB RAM. We use the proposed slicing approach,

Table 1

Number of points in the most dominant classes. These classes cover 99.81% of the Paris-rue-Madame database and 99.56% of the Paris-rue-Cassette database.

Class	Paris-rue-Madame database [24]	Paris-rue-Cassette database [25]
Façade	9,978,435	7,026,016
Ground	8,024,295	4,228,639
Cars	1,835,383	367,271
Motorcycles/ 2 wheelers	98,867	39,331
Traffic signs/ road inventory	15,480	45,105
Pedestrians	10,048	22,999
Vegetation	–	211,131

Table 2

Recall, precision and F_1 -scores for the Paris-rue-Madame database.

Paris-rue-Madame	Recall	Precision	F_1
Façade	0.9527	0.9620	0.9573
Ground	0.8650	0.9782	0.9182
Cars	0.6476	0.7948	0.7137
Motorcycles	0.7198	0.0980	0.1725
Traffic signs	0.9485	0.0491	0.0934
Pedestrians	0.8780	0.0163	0.0320

Table 3

Recall, precision and F_1 -scores for the Paris-rue-Cassette database.

Paris-rue-Cassette	Recall	Precision	F_1
Façade	0.8721	0.9928	0.9285
Ground	0.9646	0.9924	0.9783
Cars	0.6112	0.6767	0.6423
2 wheelers	0.8285	0.1774	0.2923
Road inventory	0.7657	0.1495	0.2501
Pedestrians	0.8225	0.0924	0.1661
Vegetation	0.8602	0.2566	0.3953

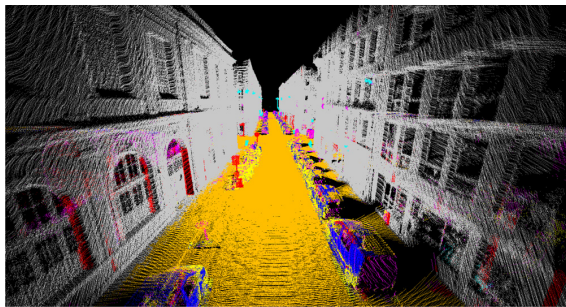


Fig. 6. Paris-rue-Madame database: classified point cloud with assigned semantic labels (façade: gray, ground: brown, cars: blue, motorcycles: yellow, traffic signs: red, pedestrians: pink). The points represented in cyan are those points which are not considered as the respective classes are not covered representatively. The noisy appearance results from individual point classification. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

where the slices have a width of 10 m and the padding has a width of 0.50 m. Whereas the prototype released with [28] is based on a full and straightforward Matlab implementation, our investigations revealed that a significant speedup can be achieved in two ways. Firstly, for the considered small point sets formed by up to $k_2 = 100$ neighboring 3D points, a considerable speedup in the calculation of the respective 3D covariance matrices results from simply replacing

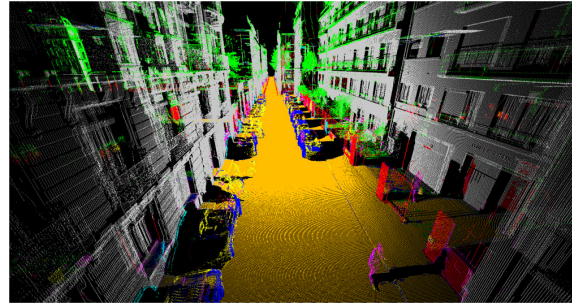


Fig. 7. Paris-rue-Cassette database: classified point cloud with assigned semantic labels (façade: gray, ground: brown, cars: blue, 2 wheelers: yellow, road inventory: red, pedestrians: pink, vegetation: green). The points represented in cyan are those points which are not considered as the respective classes are not covered representatively. The noisy appearance results from individual point classification. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Table 4

Computational effort for processing the Paris-rue-Cassette database: the required processing times t_1 for optimal neighborhood size selection, t_2 for feature extraction, t_3 for training on the small training set and t_4 for testing on the respective test set are listed for different approaches. Note that t_1 and t_2 correspond to a successive processing of all slices, and that t_3 and t_4 do not change since they are not affected by our optimization.

Time	Prototype [28]	Optimized Matlab version	Proposed approach
t_1	27.45 h	10.90 h	2.11 h
t_2	11.84 h	10.75 h	4.28 h
t_3	~1–2 s	~1–2 s	~1–2 s
t_4	~90 s	~90 s	~90 s

the internal Matlab function `cov` with the respective vectorized straightforward implementation. Secondly, our new optimized approach can be used which has been implemented in C++ and, for comparison, we used the respective binaries in the test environment in Matlab. We consider both approaches. For the example involving the Paris-rue-Cassette database, the respective speedup achieved with the latter implementation clearly becomes visible in Table 4, where the processing times for the different subtasks are listed. Since the training phase will not change with larger datasets, the respective classification will only remain a question of computational and not human effort.

Taking a tile of defined size as a reference area would even allow us to extrapolate the resulting computational effort for data processing (which also accounts for those points in the padding). Thus, we would even be able to extrapolate the computational effort to full cities.

7. Discussion

In the experiments, it becomes apparent that particularly for the smaller classes a decrease in performance can be observed (Tables 2 and 3). This might indicate that those classes are still not covered representatively for the complexity of urban 3D scenes. However, the derived mean class recall values for both datasets indicate that completeness/quantity across all classes is relatively high compared to other approaches focusing on individual point classification [27,28]. Thus, the methodology is less prone to overfitting. Consequently, the decrease in performance might mainly arise from the similarity of local 3D structures belonging to respective classes. Additionally, for the Paris-rue-Cassette database, we can observe that vegetation is detected at the balconies

which is in contradiction to the reference labels, but maybe not always in contradiction to the real scene (Fig. 7). Note that the Paris-rue-Madame database contains more noise than the Paris-rue-Cassette database (Figs. 4 and 5) which might be a further reason for relatively low precision values obtained for the respective smaller classes (Table 2). In order to increase the precision values, introducing further features and/or multi-scale considerations seem to be necessary. Furthermore, due to the individual point classification, a noisy labeling can be expected which indeed can be observed in Figs. 6 and 7.

Concerning the computational complexity, we may consider the computation times required for processing different numbers of points. In Figs. 8 and 9, this is done for the separate slices (with the respective padding) as well as separately for optimal neighborhood size selection and feature extraction. It becomes apparent that optimal neighborhood size selection has a linear complexity for increasing numbers of considered 3D points, whereas feature

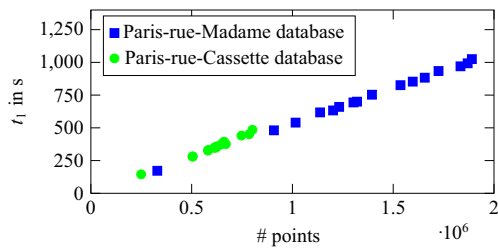


Fig. 8. Required computation times per slice for optimal neighborhood size selection: a linear behavior can be observed for increasing numbers of points.

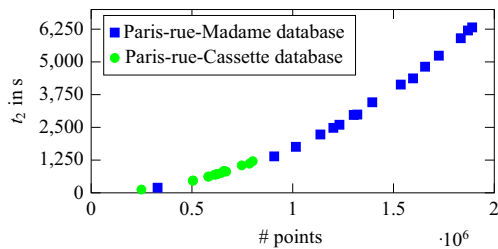


Fig. 9. Required computation times per slice for feature extraction: a non-linear behavior can be observed for increasing numbers of points.

extraction shows a non-linear dependency. In this case, we have a superposition of (i) a linear behavior for calculating 3D features in terms of basic geometric properties or eigenvalue-based features, (ii) a linear behavior for 2D features in terms of basic geometric properties or eigenvalue-based features and (iii) a non-linear behavior for 2D features based on the accumulation map.

The most crucial issue of the whole methodology remains an appropriate selection of the scale parameter k . The motivation of the applied approach is to avoid the use of empiric or heuristic knowledge on the scene with respect to neighborhood size and to obtain an automated, appropriate selection instead. Consequently, the approach is generally applicable and not tailored to a specific dataset. The consideration of individual neighborhoods even accounts for the idea that an optimal neighborhood size depends on the respective 3D structure and thus varies within a dataset. In order to provide further insights in addition to the clear trend of the scale parameter towards smaller values (Figs. 2 and 3), the behavior for the different classes is visualized in Fig. 10. Even though the analysis per class reveals a slight difference between the different classes, there is no clear indication of a characteristic which is specific for a certain class. Furthermore, we may state that the behavior of individual neighborhoods across a dataset indicates the quality of a dataset, since a much smoother behavior can be observed for the dataset with less noise (Figs. 4 and 5).

8. Conclusions

In this paper, we have presented a methodology for automated 3D scene analysis and its extension towards huge point clouds. The methodology generally requires a higher computational effort due to the consideration of individual 3D neighborhoods of optimal size which, in turn, is justified as it significantly improves the classification results in comparison to state-of-the-art approaches [28] and furthermore avoids human interaction guided by empiric or heuristic knowledge. Specifically, involving such optimal neighborhoods for feature extraction results in distinctive low-level geometric 3D and 2D features as important prerequisites for obtaining appropriate classification results. The further extension towards data-intensive processing via scene partitioning overcomes the limitation with respect to the computational burden and also allows large-scale 3D scene analysis. For two recently published point cloud datasets captured in urban areas, the derived results clearly reveal the potential of our methodology. For future work, we plan to involve spatial smoothing techniques,

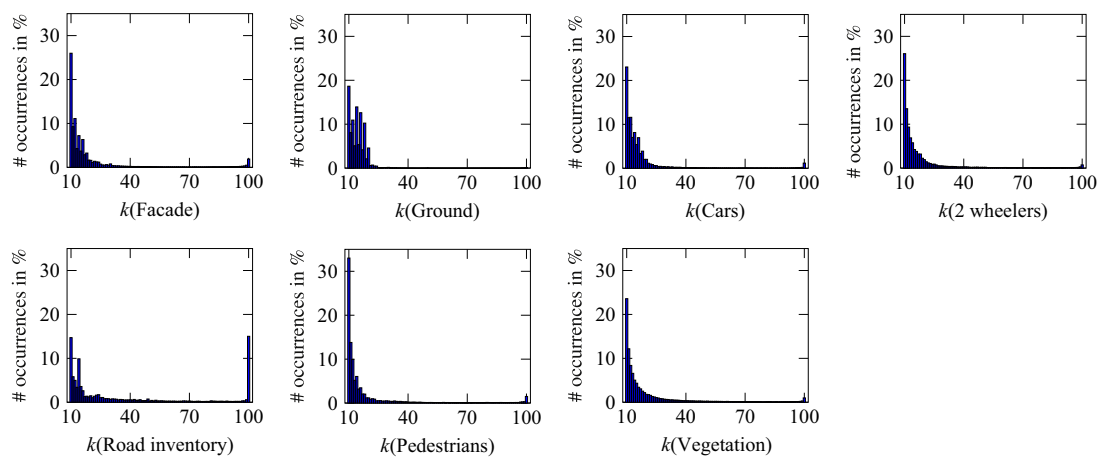


Fig. 10. Distribution of the assigned optimal neighborhood size k for different classes of the Paris-rue-Cassette database.

since neighboring 3D points tend to have correlated labels. Furthermore, an extended analysis up to object level would be desirable.

Acknowledgements

The project was partially supported by KIT-GRACE, the Graduate School for Climate and Environment at the Karlsruhe Institute of Technology (KIT), and by the FP7 project IQmulus (FP7-ICT-2011-318787). Furthermore, a funding for a stay abroad of the first author was provided by the Karlsruhe House of Young Scientists (KHYS) at KIT in order to support the collaboration.

References

- Musiński P, Wonka P, Aliaga DG, Wimmer M, van Gool L, Purgathofer W. A survey of urban reconstruction. *Comput Graph Forum* 2013;32(6):146–77.
- Zhou QY, Neumann U. Complete residential urban area reconstruction from dense aerial lidar point clouds. *Graph Model* 2013;75(3):118–25.
- Lafarge F, Alliez P. Surface reconstruction through point set structuring. *Comput Graph Forum* 2013;32(2):225–34.
- Zheng Q, Sharf A, Wan G, Li Y, Mitra NJ, Cohen-Or D, et al. Non-local scan consolidation for 3d urban scenes. *ACM Trans Graph* 2010;29(4):1–9 [Article no. 94].
- Ceylan D, Mitra NJ, Zheng Y, Pauly M. Coupled structure-from-motion and 3d symmetry detection for urban facades. *ACM Trans Graph* 2014;33(1):2:1–15.
- Pu S, Rutzinger M, Vosselman G, Oude Elberink S. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS J Photogramm Remote Sens* 2011;66(6):528–39.
- Velizhev A, Shapovalov R, Schindler K. Implicit shape models for object detection in 3d point clouds. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2012;1-3:179–84.
- Bremer M, Wichmann V, Rutzinger M. Eigenvalue and graph-based object extraction from mobile laser scanning point clouds. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2013;11-5/W2:55–60.
- Serna A, Marcotegui B. Detection, segmentation and classification of 3d urban objects using mathematical morphology and supervised learning. *ISPRS J Photogramm Remote Sens* 2014;93:243–55.
- Boyko A, Funkhouser T. Extracting roads from dense point clouds in large scale urban environment. *ISPRS J Photogramm Remote Sens* 2011;66(6):S02–12.
- Zhou L, Vosselman G. Mapping curbstones in airborne and mobile laser scanning data. *Int J Appl Earth Obs Geoinf* 2012;18(1):293–304.
- Guan H, Li J, Yu Y, Wang C, Chapman M, Yang B. Using mobile laser scanning data for automated extraction of road markings. *ISPRS J Photogramm Remote Sens* 2014;87:93–107.
- Serna A, Marcotegui B. Urban accessibility diagnosis from mobile laser scanning data. *ISPRS J Photogramm Remote Sens* 2013;84:23–32.
- Kim HB, Sohn G. Random forests based multiple classifier system for power-line scene classification. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2011; XXXVIII-5/W12:253–8.
- Vanegas CA, Aliaga DG, Benes B. Automatic extraction of Manhattan-world building masses from 3d laser range scans. *IEEE Trans Vis Comput Graph* 2012;18(10):1627–37.
- Wurm KM, Kretschmar H, Kümmerle R, Stachniss C, Burgard W. Identifying vegetation from laser data in structured outdoor environments. *Robot Auton Syst* 2014;62(5):675–84.
- Lafarge F, Mallet C. Creating large-scale city models from 3d-point clouds: a robust approach with hybrid representation. *Int J Comput Vis* 2012;99(1):69–85.
- Hebert M, Bagnell JA, Bajracharya M, Daniilidis K, Matthies LH, Mianzo L, et al. Semantic perception for ground robotics. In: Proceedings of SPIE 8387, unmanned systems technology XIV; 2012. p. 83870Y: 1–12.
- Boulch A, Houllier S, Marlet R, Tournaire O. Semantizing complex 3d scenes using constrained attribute grammars. *Comput Graph Forum* 2013;32(5):33–42.
- Hu H, Munoz D, Bagnell JA, Hebert M. Efficient 3-d scene analysis from streaming data. In: Proceedings of the IEEE international conference on robotics and automation; 2013. p. 2297–304.
- Niemeyer J, Rottensteiner F, Soergel U. Contextual classification of lidar data and building object detection in urban areas. *ISPRS J Photogramm Remote Sens* 2014;87:152–65.
- Nüchter A, Lingemann K. Robotic 3D scan repository. Jacobs University Bremen gGmbH and University of Osnabrück; 2011. <http://kos.informatik.uni-osnabrueck.de/3Dscans/> (last access: 30 August 2014).
- Munoz D, Bagnell JA, Vandapel N, Hebert M. Contextual classification with functional max-margin Markov networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2009. p. 975–82.
- Serna A, Marcotegui B, Goulette F, Deschaud JE. Paris-rue-Madame database: a 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. In: Proceedings of the international conference on pattern recognition applications and methods; 2014. p. 819–24.
- Papadimitris N, Vallet B, Marcotegui B, Serna A. IQmulus & TerraMobilita contest – analysis of mobile laser scans (MLS) in dense urban environments. MATIS Laboratory, French National Mapping Agency (IGN) and Center for Mathematical Morphology (CMM), MINES ParisTech; 2014. <http://data.ign.fr/benchmarks/UrbanAnalysis/> (last access: 30 August 2014).
- Weinmann M, Jutzi B, Mallet C. Describing Paris: automated 3d scene analysis via distinctive low-level geometric features. In: Proceedings of the IQmulus workshop on processing large geospatial data; 2014a. p. 1–8.
- Weinmann M, Jutzi B, Mallet C. Feature relevance assessment for the semantic interpretation of 3d point cloud data. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2013;11-5/W2:313–8.
- Weinmann M, Jutzi B, Mallet C. Semantic 3d scene interpretation: a framework combining optimal neighborhood size selection with relevant features. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2014;11-3:181–8.
- Chehata N, Guo L, Mallet C. Airborne lidar feature selection for urban classification using random forests. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2009;XXXVIII-3/W8:207–12.
- Mallet C, Bretar F, Roux M, Soergel U, Heipke C. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS J Photogramm Remote Sens* 2011;66(6):S71–84.
- Tombari F, Salti S, Di Stefano L. Unique signatures of histograms for local surface description. In: Proceedings of the European conference on computer vision; 2010. p. 356–69.
- Johnson AE, Hebert M. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans Pattern Anal Mach Intell* 1999;21(5):433–49.
- Rusu RB, Marton ZC, Blodow N, Beetz M. Persistent point feature histograms for 3d point clouds. In: Proceedings of the international conference on intelligent autonomous systems; 2008. p. 119–28.
- Rusu RB, Blodow N, Beetz M. Fast Point Feature Histograms (FPFH) for 3d registration. In: Proceedings of the IEEE international conference on robotics and automation; 2009. p. 3212–17.
- Osada R, Funkhouser T, Chazelle B, Dobkin D. Shape distributions. *ACM Trans Graph* 2002;21(4):807–32.
- Blomley R, Weinmann M, Leitloff J, Jutzi B. Shape distribution features for point cloud analysis – a geometric histogram approach on multiple scales. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2014;11-3:9–16.
- Behley J, Steinhege V, Cremers AB. Performance of histogram descriptors for the classification of 3d laser range data in urban environments. In: Proceedings of the IEEE international conference on robotics and automation; 2012. p. 4391–8.
- Jutzi B, Gross H. Nearest neighbour classification on laser point clouds to gain object structures from buildings. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2009;XXXVIII-1-4-7/W5.
- West KF, Webb BN, Lersch JR, Pothier S, Triscari JM, Iverson AE. Context-driven automated target detection in 3-d data. In: Proceedings of SPIE 5426; 2004. p. 133–43.
- Lalonde JF, Unnikrishnan R, Vandapel N, Hebert M. Scale selection for classification of point-sampled 3d surfaces. In: Proceedings of the international conference on 3-D digital imaging and modeling; 2005. p. 285–92.
- Demantké J, Mallet C, David N, Vallet B. Dimensionality based scale selection in 3d lidar point clouds. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2011; XXXVIII-5/W12:97–102.
- Guo B, Huang X, Zhang F, Sohn G. Classification of airborne laser scanning data using JointBoost. *ISPRS J Photogramm Remote Sens* 2014;92:124–36.
- Waldhauser C, Hochreiter R, Otepka J, Pfeifer N, Ghuffar S, Korzenjowska K, et al. Automated classification of airborne laser scanning point clouds. In: Kozel S, Leifsson L, Yang XS, editors. Solving computationally expensive engineering problems. New York, USA: Springer; 2014. p. 269–92.
- Fehr D, Cherian A, Sivalingam R, Nickolay S, Morellas V, Papanikolopoulos N. Compact covariance descriptors in 3d point clouds for object recognition. In: Proceedings of the IEEE international conference on robotics and automation; 2012. p. 1793–8.
- Khoshelham K, Oude Elberink SJ. Role of dimensionality reduction in segment-based classification of damaged building roofs in airborne laser scanning data. In: Proceedings of the international conference on geographic object based image analysis; 2012. p. 372–7.
- Niemeyer J, Rottensteiner F, Soergel U. Conditional random fields for lidar point cloud classification in complex urban areas. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2012;1-3:263–8.
- Pauly M, Keiser R, Gross M. Multi-scale feature extraction on point-sampled surfaces. *Comput Graph Forum* 2003;22(3):281–9.
- Belton D, Lichti DD. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2006;XXXVI-5:44–9.
- Mitra NJ, Nguyen A. Estimating surface normals in noisy point cloud data. In: Proceedings of the annual symposium on computational geometry; 2003. p. 322–8.
- Brodu N, Lague D. 3d terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology. *ISPRS J Photogramm Remote Sens* 2012;68:121–34.
- Xiong X, Munoz D, Bagnell JA, Hebert M. 3-d scene analysis via sequenced predictions over points and regions. In: Proceedings of the IEEE international conference on robotics and automation; 2011. p. 2609–16.

- [52] Xu S, Vosselman G, Oude Elberink S. Multiple-entropy based classification of airborne laser scanning data in urban areas. *ISPRS J Photogramm Remote Sens* 2014;88:1–15.
- [53] Secord J, Zakhor A. Tree detection in urban regions using aerial lidar and image data. *IEEE Geosci Remote Sens Lett* 2007;4(2):196–200.
- [54] Lodha SK, Fitzpatrick DM, Helmbold DP. Aerial lidar data classification using AdaBoost. In: Proceedings of the international conference on 3-D digital imaging and modeling; 2007. p. 435–42.
- [55] Shapovalov R, Velizhev A, Barinova O. Non-associative Markov networks for 3d point cloud classification. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2010;XXXVIII-3A:103–8.
- [56] Najafi M, Taghavi Namin S, Salzmann M, Petersson L. Non-associative higher-order Markov networks for point cloud classification. In: Proceedings of the European conference on computer vision; 2014. p. 500–15.
- [57] Shapovalov R, Vetrov D, Kohli P. Spatial inference machines. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2013. p. 2985–92.
- [58] Schindler K. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Trans Geosci Remote Sens* 2012;50(11):4534–45.
- [59] Lee I, Schenk T. Perceptual organization of 3d surface points. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2002;XXXIV-3A:193–8.
- [60] Filin S, Pfeifer N. Neighborhood systems for airborne laser data. *Photogramm Eng Remote Sens* 2005;71(6):743–55.
- [61] Linsen L, Prautzsch H. Local versus global triangulations. In: Proceedings of eurographics; 2001. p. 257–63.
- [62] Friedman JH, Bentley JL, Finkel RA. An algorithm for finding best matches in logarithmic expected time. *ACM Trans Math Softw* 1977;3(3):209–26.
- [63] Vosselman G, Klein R. Visualisation and structuring of point clouds. In: Vosselman G, Maas HG, editors. Airborne and terrestrial laser scanning. Dunbeath, UK: Whittles Publishing; 2010. p. 45–81.
- [64] Arya S, Mount DM, Netanyahu NS, Silverman R, Wu AY. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *J ACM* 1998;45(6):891–923.
- [65] Samet H. Foundations of multidimensional and metric data structures. St. Louis, USA: Morgan Kaufmann; 2006.
- [66] Muja M, Lowe DG. Fast approximate nearest neighbors with automatic algorithm configuration. In: Proceedings of the international conference on computer vision theory and applications, vol. 1; 2009. p. 331–40.
- [67] Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;27(3):379–423.
- [68] Monnier F, Vallet B, Soheilian B. Trees detection from laser point clouds acquired in dense urban areas by a mobile mapping system. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2012;1-3:245–50.
- [69] Breiman L. Bagging predictors. *Mach Learn* 1996;24(2):123–40.
- [70] Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat* 1979;7(1):1–26.
- [71] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32.
- [72] Criminisi A, Shotton J. Decision forests for computer vision and medical image analysis. *Advances in computer vision and pattern recognition*. London, UK: Springer; 2013.
- [73] Zhou QY, Neumann U. A streaming framework for seamless building reconstruction from large-scale aerial lidar data. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2009. p. 2759–66.
- [74] Goulette F, Nashashibi F, Abuhadrous I, Ammoun S, Lurgeau C. An integrated on-board laser range sensing system for on-the-way city and road modelling. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2006;XXXVI-1.
- [75] Paparoditis N, Papelard JP, Cannelle B, Devaux A, Soheilian B, David N, et al. Stereopolis II: a multi-purpose and multi-sensor 3d mobile mapping system for street visualisation and 3d metrology. *Rev Fr Photogramm Télédélect* 2012;200:69–79.
- [76] Brédif M, Vallet B, Serna A, Marcotegui B, Paparoditis N. TerraMobilita/IQmulus urban point cloud classification benchmark. In: Proceedings of the IQmulus workshop on processing large geospatial data; 2014. p. 1–6.

SEMANTIC 3D SCENE INTERPRETATION: A FRAMEWORK COMBINING OPTIMAL NEIGHBORHOOD SIZE SELECTION WITH RELEVANT FEATURES

Martin Weinmann^a, Boris Jutzi^a, Clément Mallet^b

^a Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Englerstr. 7, 76131 Karlsruhe, Germany - {martin.weinmann, boris.jutzi}@kit.edu

^b IGN, MATIS lab., Université Paris Est
73 avenue de Paris, 94160 Saint-Mandé, France - clement.mallet@ign.fr

Commission III, WG III/2

KEY WORDS: LIDAR, Point Cloud, Neighborhood Size Selection, Features, Feature Selection, Classification, Interpretation

ABSTRACT:

3D scene analysis by automatically assigning 3D points a semantic label has become an issue of major interest in recent years. Whereas the tasks of feature extraction and classification have been in the focus of research, the idea of using only relevant and more distinctive features extracted from optimal 3D neighborhoods has only rarely been addressed in 3D lidar data processing. In this paper, we focus on the interleaved issue of extracting relevant, but not redundant features and increasing their distinctiveness by considering the respective optimal 3D neighborhood of each individual 3D point. We present a new, fully automatic and versatile framework consisting of four successive steps: (i) optimal neighborhood size selection, (ii) feature extraction, (iii) feature selection, and (iv) classification. In a detailed evaluation which involves 5 different neighborhood definitions, 21 features, 6 approaches for feature subset selection and 2 different classifiers, we demonstrate that optimal neighborhoods for individual 3D points significantly improve the results of scene interpretation and that the selection of adequate feature subsets may even further increase the quality of the derived results.

1 INTRODUCTION

The automatic interpretation of 3D point clouds represents a fundamental issue in photogrammetry, remote sensing and computer vision. Nowadays, different subtopics are in the focus of research such as point cloud classification (Hu et al., 2013; Niemeyer et al., 2014; Xu et al., 2014), object recognition (Pu et al., 2011; Velizhev et al., 2012), creation of large-scale city models (Lafarge and Mallet, 2012) or urban accessibility analysis (Serna and Marcotegui, 2013). For all of them, it is important to cope with the complexity of 3D scenes caused by the irregular sampling and very different types of objects as well as the computational burden arising from both large 3D point clouds and a variety of available features.

For scene interpretation in terms of uniquely assigning each 3D point a semantic label (e.g. *ground*, *building* or *vegetation*), the straightforward approach is to extract respective geometric features from its local 3D structure. Thus, the features rely on a local 3D neighborhood which is typically chosen as spherical neighborhood with fixed radius (Lee and Schenk, 2002), cylindrical neighborhood with fixed radius (Filin and Pfeifer, 2005) or spherical neighborhood formed by a fixed number of the k closest 3D points (Linsen and Prautzsch, 2001). Once features have been calculated, the classification of each 3D point may be conducted via standard supervised learning approaches such as Gaussian Mixture Models (Lalonde et al., 2005), Support Vector Machines (Secord and Zakhor, 2007), AdaBoost (Lodha et al., 2007), a cascade of binary classifiers (Carlberg et al., 2009), Random Forests (Chehata et al., 2009) and Bayesian Discriminant Classifiers (Khoshelham and Oude Elberink, 2012). In contrast, contextual learning approaches also involve relationships among 3D points in a local neighborhood¹ which have to be inferred from the training data. Respective methods for classifying

¹This local neighborhood is typically different from the one used for feature extraction.

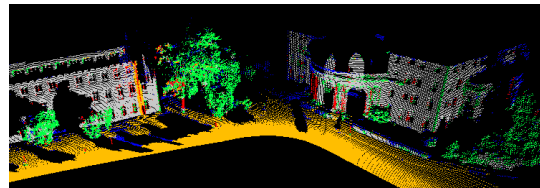


Figure 1: 3D point cloud with assigned labels (wire: blue, pole/trunk: red, façade: gray, ground: brown, vegetation: green).

point cloud data have been proposed with Associative and non-Associative Markov Networks (Munoz et al., 2009a; Shapovalov et al., 2010), Conditional Random Fields (Niemeyer et al., 2012), multi-stage inference procedures focusing on point cloud statistics and relational information over different scales (Xiong et al., 2011), and spatial inference machines modeling mid- and long-range dependencies inherent in the data (Shapovalov et al., 2013).

Since the semantic labels of nearby 3D points tend to be correlated (Figure 1), involving a smooth labeling is often desirable. However, exact inference is computationally intractable when applying contextual learning approaches. Instead, either approximate inference techniques or smoothing techniques are commonly applied. Approximate inference techniques remain challenging as there is no indication towards an optimal inference strategy, and they quickly reach their limitations if the considered neighborhood is becoming too large. In contrast, smoothing techniques may provide a significant improvement concerning classification accuracy (Schindler, 2012). All of these techniques exploit either the estimated probability of a 3D point belonging to each of the defined classes or the direct assignment of the respective label, and thus the results of a classification for individual 3D points. Consequently, it seems desirable to investigate sources for potential improvements with respect to classification accuracy.

One potential improvement may address the design of utilized features as, despite the different neighborhood definitions, the pa-

parameterization of the neighborhood is still typically selected with respect to empirical a priori knowledge on the scene and identical for all 3D points. This raises the question about estimating the optimal neighborhood for each individual 3D point and thus increasing the distinctiveness of derived features. Respective approaches addressing this issue are based on local surface variation (Pauly et al., 2003; Belton and Lichti, 2006), iterative schemes relating neighborhood size to curvature, point density and noise of normal estimation (Mitra and Nguyen, 2003; Lalonde et al., 2005), or dimensionality-based scale selection (Demantké et al., 2011). Instead of mainly focusing on optimal neighborhoods, further approaches extract features based on different entities such as points and regions (Xiong et al., 2011; Xu et al., 2014). Alternatively, it would be possible to calculate features at different scales and later use a training procedure to define which combination of scales allows the best separation of different classes (Brodu and Lague, 2012).

Considering the variety of features which have been proposed for classifying 3D points, it may further be expected that there are more and less suitable features among them. For compensating lack of knowledge, however, often all extracted features are included in the classification process, and a respective feature selection has only rarely been applied in 3D point cloud processing. The main idea of such a feature selection is to improve the classification accuracy while simultaneously reducing both computational effort and memory consumption (Guyon and Elisseeff, 2003; Liu et al., 2010). Respective approaches allow to assess the relevance/importance of single features, rank them according to their relevance and select a subset of the best-ranked features (Chehata et al., 2009; Mallet et al., 2011; Khoshelham and Oude Elberink, 2012; Weinmann et al., 2013).

In this paper, we use state-of-the-art approaches for classifying 3D points and focus on the interleaved issue of deriving an optimal subset of relevant, but not redundant, features extracted from individual neighborhoods with optimal size. In comparison to seminal work addressing optimal neighborhood size selection (Pauly et al., 2003; Mitra and Nguyen, 2003; Demantké et al., 2011), we directly assess the order/disorder of 3D points in the local neighborhood from the eigenvalues of the 3D structure tensor. In comparison to recent work on feature selection for 3D lidar data processing (Mallet et al., 2011; Weinmann et al., 2013), we exploit entropy-based measures for (i) determining the optimal neighborhood size for each 3D point and (ii) removing irrelevant and redundant features in order to derive an adequate feature subset. Both of these issues are crucial for the whole processing chain, and it is therefore of great importance to avoid parameters or thresholds which are explicitly selected by human interaction based on empiric or heuristic knowledge.

In summary, the main contribution of our work is a fully automatic versatile framework which is based on

- determining the optimal neighborhood size for each individual 3D point by considering the order/disorder of 3D points within a covariance ellipsoid,
- extracting optimized 3D and 2D features from the derived optimal neighborhoods in order to optimally describe the local structure for each 3D point,
- selecting a compact and robust feature subset by addressing different intrinsic properties of the given training data via multivariate filter-based feature selection (based on both feature-class and feature-feature relations) in order to remove feature redundancy, and
- improving the classification accuracy by exploiting the derived feature subsets and state-of-the-art classifiers.

Our framework is generally applicable for interpreting 3D point cloud data acquired via airborne laser scanning (ALS), terrestrial laser scanning (TLS), mobile laser scanning (MLS), range imaging by 3D cameras or 3D reconstruction from images. While the selected feature subset may vary with respect to different datasets, the beneficial impact of both optimal neighborhood size selection and feature selection remains. Further extensions of the framework by involving additional features such as color/intensity or full-waveform features can easily be taken into account.

The paper is organized as follows. In Section 2, we explain the single components of our framework in detail. Subsequently, in Section 3, we evaluate the proposed methodology on MLS data acquired within an urban environment. The derived results are discussed in Section 4. Finally, in Section 5, concluding remarks are provided, and suggestions for future work are outlined.

2 METHODOLOGY

For semantically interpreting 3D point clouds, we propose a new methodology which involves neighborhood selection with optimal neighborhood size for each individual 3D point (Section 2.1), 3D and 2D feature extraction (Section 2.2), feature subset selection via feature-class and feature-feature correlation (Section 2.3), and supervised classification of 3D point cloud data (Section 2.4). A visual representation of the whole framework and its components is provided in Figure 2.

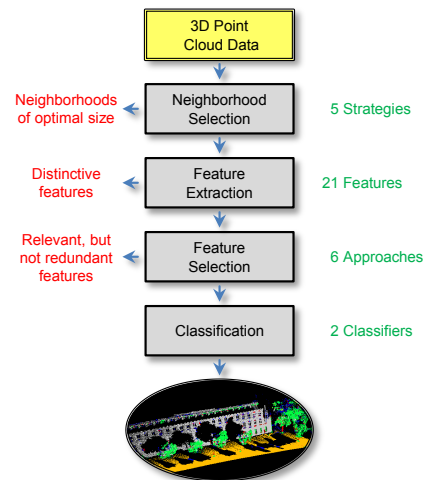


Figure 2: The proposed framework: the contributions are highlighted in red, and the quantity of attributes/approaches used for evaluation is indicated in green.

2.1 Neighborhood Selection

In general, we may face a varying point density in the captured 3D point cloud data. Since we do not want to assume a priori knowledge on the scene, we exploit the spherical neighborhood definition based on a 3D point and its k closest 3D points (Linsen and Prutzsch, 2001), which allows more flexibility with respect to the geometric size of the neighborhood. In order to avoid heuristically selecting a certain value for the parameter k , we focus on automatically estimating the optimal value for k .

Assuming a point cloud formed by a total number of N 3D points and a given value $k \in \mathbb{N}$, we may consider each individual 3D point $\mathbf{X} = (X, Y, Z)^T \in \mathbb{R}^3$ and the respective k neighbors

defining its scale. For describing the local 3D structure around \mathbf{X} , the respective 3D covariance matrix also known as 3D structure tensor $\mathbf{S} \in \mathbb{R}^{3 \times 3}$ is derived which is a symmetric positive-definite matrix. Thus, its three eigenvalues $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ exist, are non-negative and correspond to an orthogonal system of eigenvectors. Since there may not necessarily be a preferred variation with respect to the eigenvectors, we consider the general case based on a structure tensor with rank 3. Hence, it follows that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ holds for each 3D point \mathbf{X} .

From the eigenvalues of the 3D structure tensor, the surface variation C_λ (i.e. the change of curvature) with

$$C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \quad (1)$$

can be estimated. For an increasing neighborhood size, the heuristic search for locations with significant increase of C_λ allows to find the critical neighborhood size and thus to select a respective value for k (Pauly et al., 2003). This procedure is motivated by the fact that occurring jumps indicate strong deviations in the normal direction. As alternative, it has been proposed to select the neighborhood size according to a consistent curvature level (Belton and Lichti, 2006).

Further investigations focus on extracting the dimensionality features of linearity L_λ , planarity P_λ and scattering S_λ according to

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

which represent 1D, 2D and 3D features. As these features sum up to 1, they may be considered as the probabilities of a 3D point to be labeled as 1D, 2D or 3D structure (Demantké et al., 2011). Accordingly, a measure E_{dim} of unpredictability given by the Shannon entropy (Shannon, 1948) as

$$E_{\text{dim}} = -L_\lambda \ln(L_\lambda) - P_\lambda \ln(P_\lambda) - S_\lambda \ln(S_\lambda) \quad (3)$$

can be minimized across different scales k to find the optimal neighborhood size which favors one dimensionality the most. For this purpose, the radius has been taken into account, and the interval $[r_{\min}, r_{\max}]$ has been sampled in 16 scales, where the radii are not linearly increased since the radius of interest is usually closer to r_{\min} . The values r_{\min} and r_{\max} depend on various characteristics of the given data and are therefore specific for each dataset. However, the results are based on the assumption of particular shapes being present in the observed scene.

In order to avoid assumptions on the scene, we propose a more general solution to optimal neighborhood size selection. Since the eigenvalues correspond to the principal components, they span a 3D covariance ellipsoid. Consequently, we may normalize the three eigenvalues by their sum Σ_λ and consider the measure of eigenentropy E_λ given by the Shannon entropy according to

$$E_\lambda = -e_1 \ln(e_1) - e_2 \ln(e_2) - e_3 \ln(e_3) \quad (4)$$

where the e_i with $e_i = \lambda_i / \Sigma_\lambda$ for $i \in \{1, 2, 3\}$ represent the normalized eigenvalues summing up to 1. The eigenentropy thus provides a measure of the order/disorder of 3D points within the covariance ellipsoid². Hence, we propose to select the parameter k by minimizing the eigenentropy E_λ over varying values for k . For this purpose, we consider relevant statistics to start with $k_{\min} = 10$ samples which is in accordance to similar investigations (Demantké et al., 2011). As maximum, we select a

²Note that the occurrence of eigenvalues identical to zero has to be avoided by adding an infinitesimal small value ϵ .

relatively high number of $k_{\max} = 100$ samples, and all integer values in $[k_{\min}, k_{\max}]$ are taken into consideration.

2.2 Feature Extraction

For feature extraction, we follow the strategy of deriving a variety of both 3D and 2D features (Weinmann et al., 2013), but we optimize their distinctiveness by taking into account the optimal neighborhood size of each individual 3D point. Based on the normalized eigenvalues e_1, e_2 and e_3 of the 3D structure tensor \mathbf{S} , we extract a feature set consisting of 8 eigenvalue-based features for each 3D point \mathbf{X} (Table 1). Additionally, we derive 6 further 3D features for characterizing the local neighborhood: absolute height Z , radius $r_{k\text{-NN}}$ of the spherical neighborhood, local point density D , verticality V which is derived from the vertical component of the normal vector, and maximum height difference $\Delta Z_{k\text{-NN}}$ as well as height variance $\sigma_{Z,k\text{-NN}}$ within the local neighborhood.

Linearity:	$L_\lambda = \frac{e_1 - e_2}{e_1}$
Planarity:	$P_\lambda = \frac{e_2 - e_3}{e_1}$
Scattering:	$S_\lambda = \frac{e_3}{e_1}$
Omnivariance:	$O_\lambda = \sqrt[3]{e_1 e_2 e_3}$
Anisotropy:	$A_\lambda = \frac{e_1 - e_3}{e_1}$
Eigenentropy:	$E_\lambda = -\sum_{i=1}^3 e_i \ln(e_i)$
Sum of eigenvalues:	$\Sigma_\lambda = e_1 + e_2 + e_3$
Change of curvature:	$C_\lambda = \frac{e_3}{e_1 + e_2 + e_3}$

Table 1: Eigenvalue-based 3D features.

Finally, we consider 7 features arising from the 2D projection of the 3D point cloud data onto a horizontally oriented plane. Four of them are directly derived: radius $r_{k\text{-NN},2D}$, local point density D_{2D} and sum $\Sigma_{\lambda,2D}$ as well as ratio $R_{\lambda,2D}$ of eigenvalues. The other three features are derived via the construction of a 2D accumulation map with discrete, quadratic bins of side length 0.25 m as number M of points, maximum height difference ΔZ and height variance σ_Z within the respective bin.

2.3 Feature Selection

The definition of adequate feature vectors remains a common and crucial issue for classification problems. Hence, the interest in feature selection techniques emerged for finding compact and robust subsets of relevant and informative features in order to gain predictive accuracy, improve computational efficiency with respect to both time and memory consumption, and retain meaningful features (Guyon and Elisseeff, 2003; Liu et al., 2010). By definition, a feature is statistically relevant if its removal from a feature set will reduce the prediction power. In general, feature selection methods can be categorized into filter-based methods, wrapper-based methods and embedded methods. As both wrapper-based and embedded feature selection methods involve a classifier, they generally yield a better performance than filter-based methods. In particular, embedded methods provide the capability of dealing with exhaustive feature sets as input and letting the classifier internally select a suitable feature subset during the training phase (Chehata et al., 2009; Tokarczyk et al., 2013).

However, they face a relatively high computational effort and provide feature subsets which are only optimized with respect to the applied classifier. Hence, we focus on a filter-based method.

Due to their simplicity and efficiency, such filter-based methods are commonly applied. These methods are classifier-independent and only exploit a score function directly based on the training data. Univariate filter-based feature selection methods rely on a score function which evaluates feature-class relations and thus the relation between the values of each single feature across all observations and the respective label vector. In general, the score function may address different intrinsic properties of the given training data such as distance, information, dependency or consistency. Accordingly, a variety of possible score functions addressing a specific intrinsic property (Guyon and Elisseeff, 2003; Zhao et al., 2010) as well as a general relevance metric addressing different intrinsic properties (Weinmann et al., 2013) have been proposed. Multivariate filter-based feature selection methods rely on both feature-class and feature-feature relations in order to discriminate between relevant, irrelevant and redundant features.

Defining random variables X for the feature values and C for the classes, we can apply the general definition of the Shannon entropy $E(X)$ indicating the distribution of feature values x_a as

$$E(X) = - \sum_a P(x_a) \ln P(x_a) \quad (5)$$

and the Shannon entropy $E(C)$ indicating the distribution of (semantic) classes c_b as

$$E(C) = - \sum_b P(c_b) \ln P(c_b) \quad (6)$$

respectively. The joint Shannon entropy results in

$$E(X, C) = - \sum_{a,b} P(x_a, c_b) \ln P(x_a, c_b) \quad (7)$$

and can be used for deriving the mutual information

$$MI(X, C) = E(X) + E(C) - E(X, C) \quad (8)$$

$$= E(X) - E(X|C) \quad (9)$$

$$= E(C) - E(C|X) \quad (10)$$

$$= IG(X|C) \quad (11)$$

$$= IG(C|X) \quad (12)$$

which represents a symmetrical measure defined as *information gain* (Quinlan, 1986). Thus, the amount of information gained about C after observing X is equal to the amount of information gained about X after observing C . Following the definition, a feature X is regarded as more correlated to the classes C than a feature Y if $IG(C|X) > IG(C|Y)$. For feature selection, information gain is evaluated independently for each feature and features with a high information gain are considered as relevant. Consequently, those features with the highest values may be selected as relevant features. Information gain can also be derived via the conditional entropy, e.g. via $E(X|C)$ which quantifies the remaining uncertainty in X given that the value of the random variable C is known.

However, information gain is biased in favor of features with greater numbers of values since these appear to gain more information than others, even if they are not more informative (Hall, 1999). The bias can be compensated by considering the measure

$$SU(X, C) = 2 \frac{MI(X, C)}{E(X) + E(C)} \quad (13)$$

defined as *symmetrical uncertainty* (Press et al., 1988) with values in $[0, 1]$. Information gain and symmetrical uncertainty however are only measures for ranking features according to their relevance to the class and do not eliminate redundant features.

In order to remove redundancy, Correlation-based Feature Selection (CFS) has been proposed (Hall, 1999). Considering a subset of n features and taking the symmetrical uncertainty as correlation measure, we may define $\bar{\rho}_{XC}$ as average correlation between features and classes as well as $\bar{\rho}_{XX}$ as average correlation between different features. The relevance R of the feature subset results in

$$R(X_{1\dots n}, C) = \frac{n\bar{\rho}_{XC}}{\sqrt{n + n(n-1)\bar{\rho}_{XX}}} \quad (14)$$

which can be maximized by searching the feature subset space (Hall, 1999), i.e. by iteratively adding a feature to the feature subset (forward selection) or removing a feature from the feature subset (backward elimination) until R converges to a stable value.

For comparison only, we also consider feature selection exploiting a Fast Correlation-Based Filter (FCBF) (Yu and Liu, 2003) which involves heuristics and thus does not meet our intention of a fully generic methodology. For deciding whether features are relevant to the class or not, a typical feature ranking based on symmetrical uncertainty is conducted in order to determine the feature-class correlation. If the symmetrical uncertainty is above a certain threshold, the respective feature is considered to be relevant. For deciding whether a relevant feature is redundant or not, the symmetrical uncertainty among features is compared to the symmetrical uncertainty between features and classes in order to remove redundant features and only keep predominant features.

2.4 Classification

Based on given training data, a supervised classification of individual 3D points can be conducted by using the training data to train a classifier which afterwards should be able to generalize to new, unseen data. Introducing a formal description, the training set $\mathcal{X} = \{(\mathbf{x}_i, l_i)\}$ with $i = 1, \dots, N_X$ consists of N_X training examples. Each training example encapsulates a feature vector $\mathbf{x}_i \in \mathbb{R}^d$ in a d -dimensional feature space and the respective class label $l_i \in \{1, \dots, N_C\}$, where N_C represents the number of classes. In contrast, the test set $\mathcal{Y} = \{\mathbf{x}_j\}$ with $j = 1, \dots, N_Y$ only consists of N_Y feature vectors $\mathbf{x}_j \in \mathbb{R}^d$. If available, the respective class labels may be used for evaluation. For multi-class classification, we apply different classifiers. Following recent work on smooth image labeling (Schindler, 2012), we apply a classical (Gaussian) maximum-likelihood (ML) classifier as well as Random Forest (RF) classifier as representative of modern discriminative methods.

The classical ML classifier represents a simple generative model – the Gaussian Mixture Model (GMM) – which is based on the assumption that the classes can be represented by different Gaussian distributions. Hence, in the training phase, a multivariate Gaussian distribution is fitted to the given training data. For each new feature vector, the probability of belonging to the different classes is evaluated and the class with maximum probability is assigned. Since the decision boundary between any two classes in such a model represents a quadratic function (Schindler, 2012), the resulting classifier is also referred to as Quadratic Discriminant Analysis (QDA) classifier.

A Random Forest (Breiman, 2001) is an ensemble of randomly trained decision trees. In the training phase, individual trees are trained on randomly selected feature subsets of the given training

data. Thus, the trees are all randomly different from one another which results in a de-correlation between individual tree predictions and thus improved generalization and robustness (Criminisi and Shotton, 2013). For a new feature vector, each tree votes for a single class and a respective label is subsequently assigned according to the majority vote of all trees. We use a RF classifier with 100 trees and a tree depth of $\lfloor \sqrt{d} \rfloor$, where d is the dimension of the feature space.

Since we may often face an unbalanced distribution of training examples per class in the training set, which may have a detrimental effect on the training process (Criminisi and Shotton, 2013), we apply a class re-balancing which consists of resampling the training data in order to obtain a uniform distribution of randomly selected training examples per class. The alternative would be to exploit the known prior class distribution of the training set for weighting the contribution of each class.

3 EXPERIMENTAL RESULTS

We demonstrate the performance of the proposed methodology for two publicly available MLS benchmark datasets which are described in Section 3.1. The conducted experiments are outlined in Section 3.2. A detailed evaluation and a comparison of single approaches are presented in Section 3.3.

3.1 Datasets

For our experiments, we use the Oakland 3D Point Cloud Dataset³ (Munoz et al., 2009a) which is a labeled benchmark MLS dataset representing an urban environment. The dataset has been acquired with a mobile platform equipped with side looking SICK LMS laser scanners used in push-broom mode. A separation into training set \mathcal{X} , validation set \mathcal{V} and test set \mathcal{Y} is provided, and each 3D point is assigned one of the five semantic labels *wire*, *pole/trunk*, *façade*, *ground* and *vegetation*. After class re-balancing, the reduced training set encapsulates 1,000 training examples per class. The test set contains 1.3 million 3D points.

Additionally, we apply our framework on the Paris-rue-Madame database⁴ (Serna et al., 2014) acquired in the city of Paris, France. The point cloud data consists of 20 million 3D points and corresponds to a street section with a length of approximately 160 m. For data acquisition, the Mobile Laser Scanning (MLS) system L3D2 (Goulette et al., 2006) equipped with a Velodyne HDL32 was used, and annotation has been conducted in a manually assisted way. Since the annotation includes both point labels and segmented objects, the database contains 642 objects which are in turn categorized in 26 classes. We exploit the point labels of the six dominant semantic classes *façade*, *ground*, *cars*, *motorcycles*, *traffic signs* and *pedestrians*. All 3D points belonging to the remaining classes are removed since the number of samples per class is less than 0.05% of the complete dataset. For class re-balancing, we take into account that the smallest of the selected classes comprises little more than 10,000 points. In order to provide a higher ratio between training and testing samples across all

³The Oakland 3D Point Cloud Dataset is available online at http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/ (last access: 30 March 2014).

⁴Paris-rue-Madame database: MINES ParisTech 3D mobile laser scanner dataset from Madame street in Paris. ©2014 MINES ParisTech. MINES ParisTech created this special set of 3D MLS data for the purpose of detection-segmentation-classification research activities, but does not endorse the way they are used in this project or the conclusions put forward. The database is publicly available at <http://cmm.ensmp.fr/~serna/rueMadameDataset.html> (last access: 30 March 2014).

classes, we randomly select a training set \mathcal{X} with 1,000 training examples per class, and the remaining data is used as test set \mathcal{Y} .

3.2 Experiments

In the experiments, we first consider the impact of five different neighborhood definitions on the classification results:

- the neighborhood \mathcal{N}_{10} formed by the 10 nearest neighbors,
- the neighborhood \mathcal{N}_{50} formed by the 50 nearest neighbors,
- the neighborhood \mathcal{N}_{100} formed by the 100 nearest neighbors,
- the optimal neighborhood $\mathcal{N}_{\text{opt,dim}}$ for each individual 3D point when considering dimensionality features, and
- the optimal neighborhood $\mathcal{N}_{\text{opt},\lambda}$ for each individual 3D point when considering our proposed approach⁵.

The latter two definitions involving optimal neighborhoods are based on varying the scale parameter k between $k_{\min} = 10$ and $k_{\max} = 100$ with a step size of $\Delta k = 1$, and selecting the value with minimum Shannon entropy of the respective criterion. Subsequently, we focus on testing six different feature sets for each neighborhood definition:

- the whole feature set \mathcal{S}_{all} with all 21 features,
- the feature subset \mathcal{S}_{dim} covering the three dimensionality features L_λ , P_λ and S_λ ,
- the feature subset $\mathcal{S}_{\lambda,3D}$ covering the 8 eigenvalue-based 3D features,
- the feature subset \mathcal{S}_5 consisting of the five features $R_{\lambda,2D}$, V , C_λ , $\Delta Z_{k\text{-NN}}$ and $\sigma_{Z,k\text{-NN}}$ proposed in recent investigations (Weinmann et al., 2013),
- the feature subset \mathcal{S}_{CFS} derived via Correlation-based Feature Selection, and
- the feature subset $\mathcal{S}_{\text{FCBF}}$ derived via the Fast Correlation-Based Filter.

The latter three feature subsets are based on either explicitly or implicitly assessing feature relevance. In case of combining feature subsets with RF-based classification, the tree depth of the Random Forest is determined as $\max\{\lfloor \sqrt{d} \rfloor, 3\}$, since at least 3 features are required for separating 5 or 6 classes. Note that the full feature set only has to be calculated and stored for the training data, whereas a smaller feature subset automatically selected during the training phase has to be calculated for the test data.

All implementation and processing was done in Matlab. In the following, the main focus is put on the impact of both optimal neighborhood size selection and feature selection on the classification results. We may expect that (i) optimal neighborhoods for individual 3D points significantly improve the classification results and (ii) feature subsets selected according to feature relevance measures provide an increase in classification accuracy.

3.3 Results and Evaluation

For evaluation, we consider five commonly used measures: (i) precision which represents a measure of exactness or quality, (ii) recall which represents a measure of completeness or quantity, (iii) F_1 -score which combines precision and recall with equal weights, (iv) overall accuracy (OA) which reflects the overall performance of the respective classifier on the test set, and (v) mean class recall (MCR) which reflects the capability of the respective classifier to detect instances of different classes. Since the results for classification may slightly vary for different runs, the mean

⁵The code is publicly available at <http://www.ipf.kit.edu/code.php>

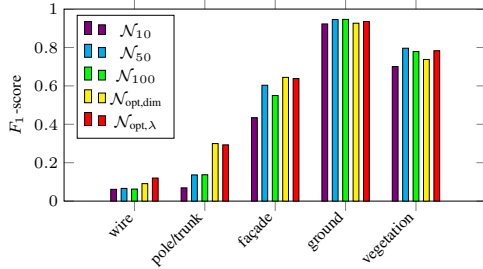


Figure 3: F_1 -scores for QDA-based classification.

values across 20 runs are used in the following in order to allow for more objective conclusions. Additionally, we consider that, for CFS and FCBF, the derived feature subsets may vary due to the random selection of training data, and hence determine them as the most often occurring feature subsets over 20 runs.

First, we test our framework on the Oakland 3D Point Cloud Dataset. Since the upper boundary $k = 100$ has been selected for reasons of computational costs, we have to take into account that it is likely to also represent 3D points which might favor a higher value. Accordingly, we consider the percentage of 3D points which are assigned neighborhoods with $k < 100$ neighbors which is 98.12% and 98.08% for $\mathcal{N}_{opt,dim}$ and $\mathcal{N}_{opt,\lambda}$. For QDA-based classification based on all 21 features, the derived recall and precision values for different neighborhood definitions are provided in Table 2 and Table 3, and the respective F_1 -scores are visualized in Figure 3. The recall and precision values when using a RF classifier are provided in Table 4 and Table 5, and the respective F_1 -scores are visualized in Figure 4. For both classifiers, it becomes visible that introducing an optimal neighborhood size for each individual 3D point has a beneficial impact on both recall and precision values, and consequently also on the F_1 -score. Exemplary results for RF-based classification using $\mathcal{N}_{opt,\lambda}$ and all 21 features are illustrated in Figure 1 and Figure 5.

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.662	0.522	0.434	0.882	0.616
\mathcal{N}_{50}	0.667	0.507	0.473	0.916	0.788
\mathcal{N}_{100}	0.606	0.417	0.472	0.916	0.767
$\mathcal{N}_{opt,dim}$	0.754	0.750	0.543	0.890	0.778
$\mathcal{N}_{opt,\lambda}$	0.791	0.765	0.519	0.906	0.829

Table 2: Recall values for QDA-based classification using all features and different neighborhood definitions.

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.032	0.037	0.614	0.967	0.812
\mathcal{N}_{50}	0.035	0.079	0.832	0.977	0.805
\mathcal{N}_{100}	0.033	0.082	0.659	0.979	0.791
$\mathcal{N}_{opt,dim}$	0.048	0.187	0.793	0.966	0.701
$\mathcal{N}_{opt,\lambda}$	0.065	0.181	0.829	0.966	0.742

Table 3: Precision values for QDA-based classification using all features and different neighborhood definitions.

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.705	0.684	0.503	0.981	0.668
\mathcal{N}_{50}	0.578	0.617	0.679	0.988	0.779
\mathcal{N}_{100}	0.513	0.579	0.631	0.987	0.724
$\mathcal{N}_{opt,dim}$	0.850	0.791	0.659	0.985	0.794
$\mathcal{N}_{opt,\lambda}$	0.862	0.798	0.672	0.985	0.809

Table 4: Recall values for RF-based classification using all features and different neighborhood definitions.

If, besides the neighborhood definitions, the different feature sets are also taken into account, we get a total number of 30 possible combinations. For each combination, the resulting overall accuracy and mean class recall value are provided in Table 6 and

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.054	0.079	0.786	0.970	0.946
\mathcal{N}_{50}	0.048	0.196	0.845	0.979	0.942
\mathcal{N}_{100}	0.041	0.134	0.742	0.980	0.938
$\mathcal{N}_{opt,dim}$	0.080	0.219	0.832	0.976	0.950
$\mathcal{N}_{opt,\lambda}$	0.091	0.236	0.846	0.972	0.959

Table 5: Precision values for RF-based classification using all features and different neighborhood definitions.

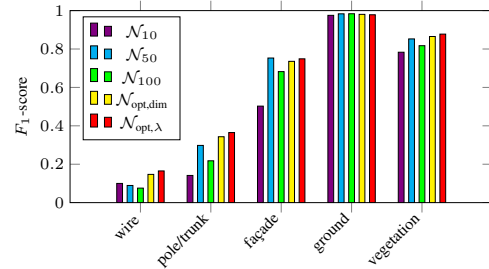


Figure 4: F_1 -scores for RF-based classification.

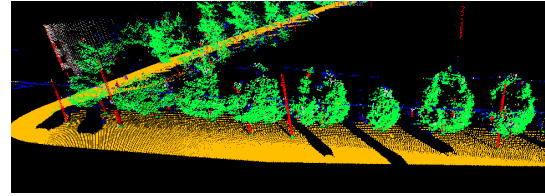


Figure 5: 3D point cloud with semantic labels assigned by the RF classifier (wire: blue, pole/trunk: red, façade: gray, ground: brown, vegetation: green).

Table 7 for QDA-based classification. The respective values for RF-based classification are provided in Table 8 and Table 9. Here, \mathcal{S}_{CFS} contains between 12 and 14 features, whereas \mathcal{S}_{FCBF} contains between 6 and 8 features. For both subsets, the respective features are distributed across all types of 3D and 2D features. The derived results clearly reveal that the feature subset \mathcal{S}_{dim} is not sufficient for obtaining adequate classification results. In contrast, using the feature subsets \mathcal{S}_5 , \mathcal{S}_{CFS} and \mathcal{S}_{FCBF} which are all based on feature relevance assessment yields classification results of better quality and, in particular when using a RF classifier, partially even a higher quality than the full feature set \mathcal{S}_{all} .

Oakland	\mathcal{S}_{all}	\mathcal{S}_{dim}	$\mathcal{S}_{\lambda,3D}$	\mathcal{S}_5	\mathcal{S}_{CFS}	\mathcal{S}_{FCBF}
\mathcal{N}_{10}	0.788	0.689	0.741	0.867	0.667	0.678
\mathcal{N}_{50}	0.850	0.771	0.822	0.927	0.725	0.762
\mathcal{N}_{100}	0.845	0.758	0.823	0.924	0.713	0.903
$\mathcal{N}_{opt,dim}$	0.837	0.371	0.798	0.910	0.715	0.687
$\mathcal{N}_{opt,\lambda}$	0.857	0.480	0.801	0.920	0.851	0.723

Table 6: Overall accuracy for QDA-based classification using different neighborhood definitions and different feature sets.

Oakland	\mathcal{S}_{all}	\mathcal{S}_{dim}	$\mathcal{S}_{\lambda,3D}$	\mathcal{S}_5	\mathcal{S}_{CFS}	\mathcal{S}_{FCBF}
\mathcal{N}_{10}	0.623	0.365	0.454	0.583	0.570	0.618
\mathcal{N}_{50}	0.670	0.509	0.588	0.673	0.633	0.699
\mathcal{N}_{100}	0.636	0.474	0.555	0.668	0.600	0.708
$\mathcal{N}_{opt,dim}$	0.743	0.440	0.561	0.666	0.694	0.703
$\mathcal{N}_{opt,\lambda}$	0.762	0.477	0.576	0.704	0.755	0.739

Table 7: Mean class recall values for QDA-based classification using different neighborhood definitions and different feature sets.

Since the RF classifier in combination with our approach for optimal neighborhood size selection ($\mathcal{N}_{opt,\lambda}$) yields high values for both overall accuracy and mean class recall, we select this combination for a test on the Paris-rue-Madame database. The obtained

Oakland	S_{all}	S_{dim}	$S_{\lambda,3D}$	S_5	S_{CFS}	S_{FCBF}
N_{10}	0.875	0.579	0.742	0.887	0.873	0.857
N_{50}	0.917	0.734	0.805	0.912	0.916	0.924
N_{100}	0.901	0.728	0.814	0.901	0.902	0.920
$N_{opt,dim}$	0.918	0.696	0.773	0.915	0.918	0.907
$N_{opt,\lambda}$	0.922	0.628	0.851	0.911	0.924	0.923

Table 8: Overall accuracy for RF-based classification using different neighborhood definitions and different feature sets.

Oakland	S_{all}	S_{dim}	$S_{\lambda,3D}$	S_5	S_{CFS}	S_{FCBF}
N_{10}	0.708	0.483	0.598	0.686	0.699	0.702
N_{50}	0.728	0.544	0.642	0.655	0.728	0.742
N_{100}	0.687	0.504	0.612	0.638	0.693	0.697
$N_{opt,dim}$	0.816	0.615	0.676	0.754	0.812	0.808
$N_{opt,\lambda}$	0.825	0.596	0.692	0.759	0.827	0.825

Table 9: Mean class recall values for RF-based classification using different neighborhood definitions and different feature sets.

recall and precision values using the feature sets S_{all} and S_{CFS} are provided in Table 10 as well as the resulting F_1 -scores. Based on the full feature set S_{all} , the RF classifier provides an overall accuracy of 90.1% and a mean class recall of 77.6%, whereas based on the feature subset S_{CFS} , a slight improvement to an overall accuracy of 90.5% and a mean class recall of 77.8% can be observed. A visualization for RF-based classification using $N_{opt,\lambda}$ and all 21 features is provided in Figure 6.

Paris	R	P	F_1	R	P	F_1
façade	0.957	0.962	0.960	0.958	0.964	0.961
ground	0.902	0.964	0.932	0.911	0.960	0.935
cars	0.606	0.755	0.672	0.603	0.768	0.676
motorcycles	0.639	0.123	0.206	0.657	0.136	0.225
traffic signs	0.974	0.055	0.105	0.978	0.058	0.109
pedestrians	0.575	0.019	0.036	0.559	0.020	0.038

Table 10: Recall (R), precision (P) and F_1 -score for RF-based classification involving all 21 features (left) and only the features in S_{CFS} (right).

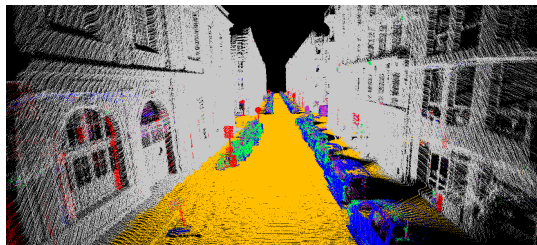


Figure 6: 3D point cloud with semantic labels assigned by the RF classifier (façade: gray, ground: brown, cars: blue, motorcycles: green, traffic signs: red, pedestrians: pink).

4 DISCUSSION

Certainly, a huge advantage of the proposed methodology is that it avoids the use of empiric or heuristic a priori knowledge on the scene with respect to neighborhood size. For the sake of generality, involving such data-dependent knowledge should not be an option and the optimal neighborhood of each individual 3D point should be considered instead. This is in accordance with the idea that the optimal neighborhood size may not be the same for different classes and furthermore depend on the respective point density. In the provided Tables 2-5, the class-specific classification results clearly reveal that the suitability of all three neighborhood definitions based on a fixed scale parameter may vary from one class to the other. Instead, the approaches based on optimal neighborhood size selection address this issue and hence provide a significant improvement in recall and precision, and thus also in the F_1 -score over all classes (Figure 3 and Figure 4).

In particular, the detailed evaluation provides a clear evidence that the proposed approach for optimal neighborhood size selection is beneficial in comparison to the other neighborhood definitions, since it often yields a significant improvement with respect to performance and behaves close to the best performance otherwise. A strong indicator for the quality of the derived results has been defined by the mean class recall, as only a high overall accuracy may not be sufficient for analyzing the derived results. For the Oakland 3D Point Cloud Dataset, for instance, we have an unbalanced test set and an overall accuracy of 70.5% can be obtained if only the instances of the class *ground* are correctly classified. This clear trend to overfitting becomes visible when considering the respective mean class recall of only 20.0%.

In comparison to other recent investigations based on a fixed scale parameter k (Weinmann et al., 2013), the recall values are significantly increased, and a slight improvement with respect to the precision values can be observed. Even in comparison to investigations involving approaches of contextual learning (Munoz et al., 2009b), our methodology yields higher precision values with approximately the same recall values over all classes.

Considering the different feature sets (Tables 6-9), it becomes visible that the feature subset S_{dim} of the three dimensionality features L_λ , P_λ and S_λ is not sufficient for 3D scene interpretation. This might be due to ambiguities, since the classes *wire* and *pole/trunk* provide a linear behavior, whereas the classes *façade* and *ground* provide a planar behavior. This can only be adequately handled by considering additional features. Even when only using the feature subset $S_{\lambda,3D}$ of the eigenvalue-based 3D features, the results are significantly worse than when using the full feature set S_{all} . In contrast, the feature subsets derived via the three approaches for feature selection provide a performance close to the full feature set S_{all} or even better. In particular, the feature subset S_{CFS} derived via Correlation-based Feature Selection provides a good performance without being based on manually selected parameters such as the feature subset S_{FCBF} derived via the Fast Correlation-Based Filter.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have addressed the interleaved issue of optimally describing 3D structures by geometrical features and selecting the *best* features among them as input for classification. We have presented a new, fully automatic and versatile framework for semantic 3D scene interpretation. The framework involves optimal neighborhood size selection which is based on minimizing the measure of eigenentropy over varying scales in order to derive optimized features with higher distinctiveness in the subsequent step of feature extraction. Further applying the measure of entropy for feature selection, irrelevant and redundant features are recognized based on a relatively small training set and, consequently, these features do not have to be calculated and stored for the test set. In a detailed evaluation, we have demonstrated the significant and beneficial impact of optimal neighborhood size selection, and that the selection of adequate feature subsets may even further increase the quality of 3D scene interpretation.

For future work, we plan to address the step from individual 3D point classification to a spatially smooth labeling of nearby 3D points. This could be based on probabilistic relaxation or smooth labeling techniques adapted from image processing.

ACKNOWLEDGEMENTS

For collaboration, a stay abroad of the first author was supported by Karlsruhe House of Young Scientists (KHYS).

REFERENCES

- Belton, D. and Lichti, D. D., 2006. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, Part 5, pp. 44–49.
- Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5–32.
- Brodu, N. and Lague, D., 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, pp. 121–134.
- Carlberg, M., Gao, P., Chen, G. and Zakhor, A., 2009. Classifying urban landscape in aerial lidar using 3D shape analysis. *Proceedings of the IEEE International Conference on Image Processing*, pp. 1701–1704.
- Chehata, N., Guo, L. and Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 3/W8, pp. 207–212.
- Criminisi, A. and Shotton, J., 2013. *Decision forests for computer vision and medical image analysis*. Advances in Computer Vision and Pattern Recognition, Springer, London, UK.
- Demantké, J., Mallet, C., David, N. and Vallet, B., 2011. Dimensionality based scale selection in 3D lidar point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 5/W12, pp. 97–102.
- Filin, S. and Pfeifer, N., 2005. Neighborhood systems for airborne laser data. *Photogrammetric Engineering & Remote Sensing*, 71(6), pp. 743–755.
- Goulette, F., Nashashibi, F., Abuhadrous, I., Ammoun, S. and Laurgeau, C., 2006. An integrated on-board laser range sensing system for on-the-way city and road modelling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, Part 1.
- Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, pp. 1157–1182.
- Hall, M. A., 1999. Correlation-based feature subset selection for machine learning. PhD thesis, Department of Computer Science, University of Waikato, New Zealand.
- Hu, H., Munoz, D., Bagnell, J. A. and Hebert, M., 2013. Efficient 3-D scene analysis from streaming data. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2297–2304.
- Khoshelham, K. and Oude Elberink, S. J., 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*, pp. 372–377.
- Lafarge, F. and Mallet, C., 2012. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *International Journal of Computer Vision*, 99(1), pp. 69–85.
- Lalonde, J.-F., Unnikrishnan, R., Vandapel, N. and Hebert, M., 2005. Scale selection for classification of point-sampled 3D surfaces. *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*, pp. 285–292.
- Lee, I. and Schenk, T., 2002. Perceptual organization of 3D surface points. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, Part 3A, pp. 193–198.
- Linsen, L. and Prautsch, H., 2001. Natural terrain classification using three-dimensional lidar data for ground robot mobility. *Proceedings of Eurographics*, pp. 257–263.
- Liu, H., Motoda, H., Setiono, R. and Zhao, Z., 2010. Feature selection: an ever evolving frontier in data mining. *Proceedings of the Fourth International Workshop on Feature Selection in Data Mining*, pp. 4–13.
- Lodha, S. K., Fitzpatrick, D. M. and Helmbold, D. P., 2007. Aerial lidar data classification using AdaBoost. *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*, pp. 435–442.
- Mallet, C., Bretar, F., Roux, M., Soergel, U. and Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), pp. S71–S84.
- Mitra, N. J. and Nguyen, A., 2003. Estimating surface normals in noisy point cloud data. *Proceedings of the Annual Symposium on Computational Geometry*, pp. 322–328.
- Munoz, D., Bagnell, J. A., Vandapel, N. and Hebert, M., 2009a. Contextual classification with functional max-margin Markov networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 975–982.
- Munoz, D., Vandapel, N. and Hebert, M., 2009b. Onboard contextual classification of 3-D point clouds with learned high-order Markov random fields. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2009–2016.
- Niemeyer, J., Rottensteiner, F. and Soergel, U., 2012. Conditional random fields for lidar point cloud classification in complex urban areas. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, pp. 263–268.
- Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp. 152–165.
- Pauly, M., Keiser, R. and Gross, M., 2003. Multi-scale feature extraction on point-sampled surfaces. *Computer Graphics Forum*, 22(3), pp. 81–89.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T., 1988. *Numerical recipes in C*. Cambridge University Press, Cambridge, UK.
- Pu, S., Rutzinger, M., Vosselman, G. and Oude Elberink, S., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), pp. S28–S39.
- Quinlan, J. R., 1986. Induction of decision trees. *Machine Learning*, 1(1), pp. 81–106.
- Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11), pp. 4534–4545.
- Secord, J. and Zakhor, A., 2007. Tree detection in urban regions using aerial lidar and image data. *IEEE Geoscience and Remote Sensing Letters*, 4(2), pp. 196–200.
- Serna, A. and Marcotegui, B., 2013. Urban accessibility diagnosis from mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84, pp. 23–32.
- Serna, A., Marcotegui, B., Goulette, F. and Deschaud, J.-E., 2014. Paris-rue-Madame database: a 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, pp. 819–824.
- Shannon, C. E., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), pp. 379–423.
- Shapovalov, R., Velizhev, A. and Barinova, O., 2010. Non-associative Markov networks for 3D point cloud classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 3A, pp. 103–108.
- Shapovalov, R., Vetrov, D. and Kohli, P., 2013. Spatial inference machines. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2985–2992.
- Tokarczyk, P., Wegner, J. D., Walk, S. and Schindler, K., 2013. Beyond hand-crafted features in remote sensing. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3/W1, pp. 35–40.
- Velizhev, A., Shapovalov, R. and Schindler, K., 2012. Implicit shape models for object detection in 3D point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, pp. 179–184.
- Weinmann, M., Jutzi, B. and Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-5/W2, pp. 313–318.
- Xiong, X., Munoz, D., Bagnell, J. A. and Hebert, M., 2011. 3-D scene analysis via sequenced predictions over points and regions. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2609–2616.
- Xu, S., Vosselman, G. and Oude Elberink, S., 2014. Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, pp. 1–15.
- Yu, L. and Liu, H., 2003. Feature selection for high-dimensional data: a fast correlation-based filter solution. *Proceedings of the International Conference on Machine Learning*, pp. 856–863.
- Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A. and Liu, H., 2010. Advancing feature selection research - ASU feature selection repository. Technical Report, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University.

D. Publikationen zu 2.4 Objektrekonstruktion

- [D1] Blomley R, Weinmann M, Leitloff J, **Jutzi B** (2014) Shape Distribution Features for Point Cloud Analysis - A Geometric Histogram Approach on Multiple Scales. In: Schindler K, Paparoditis N (Eds) ISPRS Technical Commission III Symposium. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3, 2014: 9-16
- [D2] Braun AC, Weidner U, **Jutzi B**, Hinz S (2012) Kernel Composition and the One-against-one cascade for integrating model knowledge into SVM classification. In: Heipke C, Jacobsen K, Rottensteiner F, Müller S, Sörgel U (Eds) PFG Photogrammetrie - Fernerkundung - Geoinformation. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung 2012 (4): 371-384
- [D3] **Jutzi B**, Neulist J, Stilla U (2005) Sub-pixel edge localization based on laser waveform analysis. In: Vosselman G, Brenner C (Eds) Laserscanning 2005. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W19): 109-114
- [D4] Kirchhof M, **Jutzi B**, Stilla U (2008) Iterative processing of laser scanning data by full waveform analysis in close neighborhood. In: Lichti D, Pfeifer N, Maas HG (Eds) ISPRS Journal of Photogrammetry & Remote Sensing 63 (1): 99-114 doi:10.1016/j.isprsjprs.2007.08.006
- [D5] Stilla U, Yao W, **Jutzi B** (2007) Full waveform stacking of weak laser pulses by exploiting neighbourhood relation. In: Stilla U, Mayer H, Rottensteiner F, Heipke C, Hinz S (Eds) Photogrammetric Image Analysis PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W49A): 25-30
- [D6] Weinmann M, **Jutzi B** (2012) A step towards dynamic scene analysis with active multi-view range imaging systems. In: Shortis M, Paparoditis N, Mallet C (Eds) XXII ISPRS Congress: Imaging a sustainable future. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 39 (B3): 433-438 doi:10.5194/isprsarchives-XXXIX-B3-433-2012

SUB-PIXEL EDGE LOCALIZATION BASED ON LASER WAVEFORM ANALYSIS

B. Jutzi^a, J. Neulist^a, U. Stilla^b

^a FGAN-FOM Research Institute for Optronics and Pattern Recognition, 76275 Ettlingen, Germany - {jutzi,neulist}@fom.fgan.de

^b Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, 80290 Muenchen, Germany - stilla@bv.tum.de

Commission III, Working Group III/3

KEY WORDS: Urban, Analysis, Laser scanning, LIDAR, Edge, Sub-pixel precision.

ABSTRACT:

Laser range data is of high interest in photogrammetry. However, compared to passive imagery, it usually has a lower image resolution, making geometry extraction and modeling challenging. A method to overcome this handicap using a laser scanner capable of full waveform analysis is proposed. The recorded pulse waveform is analyzed to find range and intensity values for all of the surfaces in the beam footprint. Sub-pixel edge localization is implemented and tested for straight edges of man-made objects, allowing a much higher precision of geometry extraction in urban areas. The results show that it is possible to find edges with an accuracy of at least one tenth of a pixel.

1. INTRODUCTION

The automatic generation of 3-d models for a description of man-made objects, like buildings, is of great interest in photogrammetric research. In photogrammetry, a spatial surface is classically measured by triangulation of corresponding image points from two or more pictures of the surface. The points are manually chosen or automatically detected by analyzing image structures. Besides this indirect measurement using object characteristics dependant on natural illumination, active laser scanner systems allow a direct and illumination-independent measurement of range. Laser scanners capture the range of 3-d objects in a fast, contactless and accurate way. Overviews for laser scanning systems are given in (Huising & Pereira, 1998; Wehr & Lohr, 1999; Baltsavias, 1999).

For the task of automatic model generation, a precise measurement of the edges and vertices of regularly shaped objects is paramount. Often, the spatial resolution of laser scanners used in urban surveying is not sufficient for this. In this case, an approach to locate the edges with sub-pixel accuracy is desirable. To achieve this, as much information as possible should be gained per pixel¹ to offset the low number of pixels in the image.

Current pulsed laser scanner systems for topographic mapping are based on time-of-flight ranging techniques to determine the range of the illuminated object. The signal analysis to determine the elapsed time between the emitted and backscattered laser pulses typically operates by analogous threshold detection. Some systems capture multiple reflections caused by objects which are smaller than the footprint located in different ranges. Such systems usually record the first and the last backscattered laser pulse (Baltsavias, 1999).

First pulse as well as last pulse exploitation is used for different applications like urban planning or forestry surveying. While first pulse registration is the optimum choice to measure the hull of partially penetrable objects (e.g. canopy of trees), last pulse registration should be chosen to measure non-penetrable surfaces (e.g. ground surface). Due to multiple pulse reflection at the boundary of buildings and the processing by first or last

pulse mode, building areas dilate or erode. For visualizing the various sizes of the building footprints in first and last pulse images a difference image can be calculated. The actual building edges are then expected within the bright areas in this image (Figure 1).

Apart from the range measurement of laser scanner systems some systems deliver a single reflectance value derived from the intensity or the power of the backscattered laser light. The intensity is determined by the signal maximum and the power by signal integration of the measured laser light and gives radiometric information about the surveyed area. This intensity (power) value can be used for separating segments of artificial objects from vegetation (Hug & Wehr, 1997; Maas, 2001).

One step further, the complete waveform of the recorded signal might be of interest, because it includes the backscattering characteristic of the illuminated field. Investigations on analyzing the waveform were done to explore the vegetation concerning the bio mass, foliage or density (e.g. trees, bushes, and ground). Recent laser scanner system developments provide commercial airborne laser scanning systems that allow capturing the waveform: RIEGL LMS-Q560, LITEMAPPER 5600, OPTECH ALTM 3100, TOPEYE II. To interpret the backscattered waveform, a better understanding of the physical principles is important and has to be developed. The physical measurement process and the influence of the surface on the emitted waveform are discussed by (Jutzi & Stilla, 2002; Wagner *et al.*, 2004).

By analyzing the backscattered waveform for the received pulses it is possible to determine specific surface properties for each received pulse. Typical surface properties of interest can be distance, roughness, reflectance and number of surface responses. The estimates of these properties can be used for further processing. Vosselman (2002) suggested considering the reflectance strength of the laser beam response to estimate and improve the accuracy of reflectance edge positions. Besides the edge positions of planimetric offsets, the estimation of edges in different heights is investigated by considering the reflectance strength to increase the accuracy of boundaries at plane surfaces, where the height data captured with laser scanner systems generally suffers by unresolved ambiguity.

In this paper we describe investigations for a detailed analysis of laser pulses. In Section 2, our method for generation of

¹ Note that in this paper the label *pixel* describes the conical region in space illuminated by a single laser beam. During visualization, this region is compressed into a single pixel of a displayed image, hence the name.

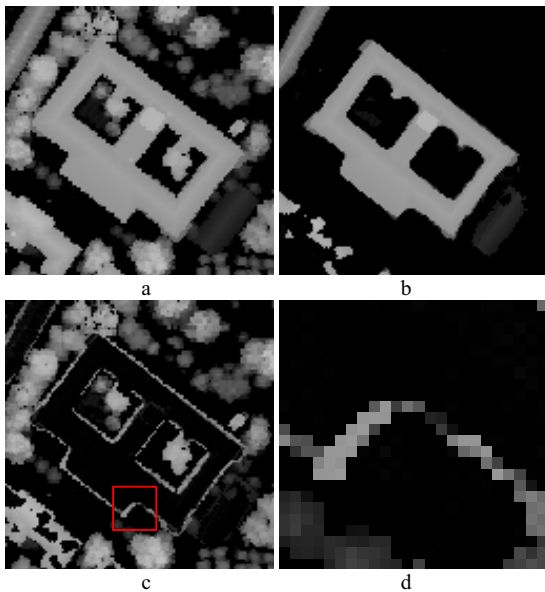


Figure 1. Sections of an urban scene (Test area Karlsruhe, Germany).

- a) elevation images captured by first pulse mode,
- b) elevation images captured by last pulse mode,
- c) difference image of first and last pulse mode,
- d) section of the difference image (building boundary).

synthetic test data is discussed. A description of the general waveform analysis, image segmentation and surface boundary extraction can be found in Section 3. The actual sub-pixel edge localization algorithms are developed in Section 4. Section 5 presents results and a discussion of the merits and flaws of the method.

2. DATA GENERATION

To simulate the temporal waveform of the backscattered pulses a *scene model* (i) and a *sensor model* (ii) is required (Jutzi & Stilla, 2004).

2.1 Scene modeling

2.1.1 Scene representation

For a 3-d scene representation, our simulation setup considers geometric and radiometric features of the illuminated surface in the form of a 3-d object model with homogeneous surface reflectance.

2.1.2 Sampling

The object model with homogeneous surface reflectance is then sampled at a higher spatial resolution than the scanning grid we simulate and process, to enable us to simulate the spatial distribution of the laser beam. Considering the position and orientation of the sensor system we receive high-resolution range and intensity images (45-fold oversampling, i.e. one image pixel is produced by averaging over 45x45 high-resolution sub-pixels. The oversampling window size does not have any practical relevance if it is sufficiently large to not induce errors in a higher magnitude as those incurred by our discretized beam profile). Depending on the predetermined position and orientation of the sensor system, various range images can be captured.

2.2 Sensor modeling

The sensor modeling takes into account the specific properties of the sensing process: the position and orientation of the sensor, the laser pulse description, scanning and the receiver properties.

2.2.1 Orientation

To simulate varying perspectives, a description of the extrinsic orientation of the laser scanning system with the help of a GPS/INS system is used.

2.2.2 Laser pulse description

The transmitted laser pulse of the system is characterized by specific pulse properties (Jutzi et al., 2002). We assume a Gaussian pulse energy distribution in both space and time, the spatial distribution thus being radially symmetric. With real data, the sampled actual pulse distribution depending on the used laser type can be used to model the edge appearance in the image (q.v. Section 4.1).

2.2.3 Scanning

Depending on the scan pattern of the laser scanner system, the grid spacing of the scanning, and the divergence of the laser beam, a sub-area of the high-resolution range image is processed. By convolving this sub-area with the temporal waveform of the laser pulse, we receive a high-resolution intensity cube. Furthermore, the corresponding sub-area of the high-resolution intensity image is weighted with the spatial energy distribution of the laser beam, where the grid spacing is taken to be 6σ of the spatial beam energy distribution (i.e. the grid lines are at $\pm 3\sigma$ relative to the beam center) to take into account the amount of backscattered laser light for each reflectance value. Then we have a description of the backscattered laser beam with a higher spatial resolution than necessary for processing.

2.2.4 Receiver

By focusing the beam with its specific properties on the detector of the receiver, the spatial resolution is reduced and this is simulated with a spatial undersampling of the sub-areas.

Finally we receive an intensity cube spaced with the scanning width of the simulated laser scanner system and containing the temporal description of the backscattered signal. Because each intensity value in the sub-area is processed by undersampling, multiple reflections can be observed in the backscattered signal.

3. DATA ANALYSIS

Algorithms are developed and evaluated with simulated signals of synthetic objects. First, a signal preprocessing of the intensity cube with a matched filter is implemented to improve the detection rate. These results are used to analyze the waveform of each pulse for gaining the surface properties: range, reflectance and number of peaks. Then the surface properties are processed with a region based segmentation algorithm. By the use of images the region boundary pixels derived from multiple reflections at the same spatial position are shared by separate regions.

3.1 Pulse property extraction

Depending on the size of the observed surface geometry in relation to the laser beam (footprint and wavelength) different properties can be extracted (Jutzi & Stilla, 2003). In this paper we focus on the pulse properties average time value, maximum intensity and number of peaks.

- The average time value is processed to determine the distance from the system to the illuminated surface.
- The maximum intensity value is computed to get a description for the reflectance strength of the illuminated area.
- Multiple peaks in one signal indicate multiple surfaces at differing ranges illuminated by the beam. Therefore, they are clues to object boundaries.

For determining the property values for each pulse of the whole waveform the intensity cube is processed in different ways. First, the pulse has to be detected in the signal profile by using a matched filter. Then, a neighborhood area of interest in the temporal waveform is selected for temporal signal analysis.

For obtaining surface characteristics, each waveform of the cube is analyzed for pulse property values. For pulse detection it is necessary to separate each single pulse from the background noise. The number of detected pulses depends critically on this separation method. Therefore the signal background noise is estimated, and where the intensity of the waveform is above three times the noise standard deviation for a duration of at least 5 ns (full-width-half-maximum of the pulse), a pulse is assumed to have been found and a waveform interval including the pulse is accepted for further processing.

Typical surface features we wish to extract from a waveform are *range*, *roughness*, and *reflectance*. The corresponding pulse properties of these surface features are: *time*, *width* and *intensity*. Because of the strong fluctuations of the waveform, extracting the relevant properties of the waveform can be difficult. Therefore, the recorded waveform is approximated by a Gaussian to get a parametric description. Fitting a Gaussian to the complete waveform instead of quantizing a single value of the waveform has the advantage of decreasing the influence of noise and fluctuation. To solve the Gaussian mixture problem, the Gauss-Newton method (Hartley & Zisserman, 2000) with iterative parameter estimation is used. The estimated parameters for pulse properties are the averaged time value τ , standard deviation σ and maximum intensity a :

$$w(t - \tau) = \frac{a}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t - \tau)^2}{2\sigma^2}\right) \quad (1)$$

To start the iteration, we use the actual parameter values (time at pulse maximum, width of signal at half pulse height, and pulse maximum) of the original waveform.

The averaged time value τ of the estimated waveform is used to exploit the temporal form of the received pulses. The averaged range value r can then easily be determined by

$$r = \frac{\tau c}{2} \quad (2)$$

where c is the speed of light.

3.2 Segmentation

General approaches for segmenting laser range data, as those described by Besl (1988), usually do not take into account the additional information acquired by full waveform processing and therefore have to be expanded upon.

By using waveform processing, we are not only generating a range image, but in fact a whole set of data for each pixel. For the purposes of this paper, the features of particular importance

will be: range for each return pulse, intensity for each return pulse, and number of return pulses. The number of return pulses is used as a clue to region boundaries, while range and intensity further facilitate identifying homogeneous regions inside these boundaries. Without these multiple pulse clues, the region boundaries can not reliably be pinpointed with pixel accuracy. Since the sub-pixel localization scheme works on the intensity of pixels partially covering a surface, this pixel precision is necessary for the accuracy of the resulting edges.

Proceeding from these boundary clues, an iterative region growing algorithm examines the range properties of all pulses in the spatial neighborhood: if the range difference of a pulse and the proofing pulse is below a given threshold, then the pulse is connected and grouped as a new element to this region.

The segmentation leads to a description of image pixels as *region interior* or *region boundary*. The region interior is characterized by single reflections and fills up the region to the boundary. For each homogeneous region found in this manner, the average return pulse power P_0 inside this region is calculated and stored. The region boundary pixels are connected in a 4-neighbor fashion, i.e. each boundary pixel has at least one neighbor in horizontal or vertical direction which also belongs to the boundary.

4. BOUNDARY REFINEMENT

To achieve a higher precision for object localization and reconstruction, it is desirable to further refine these measurements (Figure 2a). Standard sub-pixel edge localization approaches use the intensity of grayscale images to obtain improved edge information. The intensity values acquired may form a grayscale image and thus permit the application of these algorithms to our data. But we will go one step further since the full waveform laser data has several advantages over passive optical images for this purpose.

For precise edge localization, it is important to know the properties of the data acquisition unit very well and be able to model the effects of a beam being only partially reflected by a given surface. This modeling will be explained in the first subsection. In the second subsection, we will show how this model enables us to determine the sub-pixel location of an edge in each pixel. The third subsection will examine a straightforward approach to determine edge direction using neighborhood information. Then we will detail our proposed scheme to use the complete edge information to fortify the edge estimate in the fourth subsection. The last subsection will deal with vertices and the problem of their precise localization in the image.

For the purposes of this paper, we will call the measured edge pixels of the image *boundary pixels*, or *corners*, if they do not belong to a straight edge. Furthermore, let the true geometry be denoted by *edges* and *vertices*, to clear up the description of our approach.

4.1 Modeling the rasterized edge intensity profile

As shown in the introduction to this Section, it is necessary to be aware of the meaning of the intensity values acquired alongside the range measurements. Therefore we will examine the results of a beam hitting a homogeneous surface perpendicular to the beam propagation direction assuming uniform reflectance for the surface.

An analysis of the spatial beam profile (Jutzi *et al.*, 2002) has shown that it can be approximated by a radially symmetric

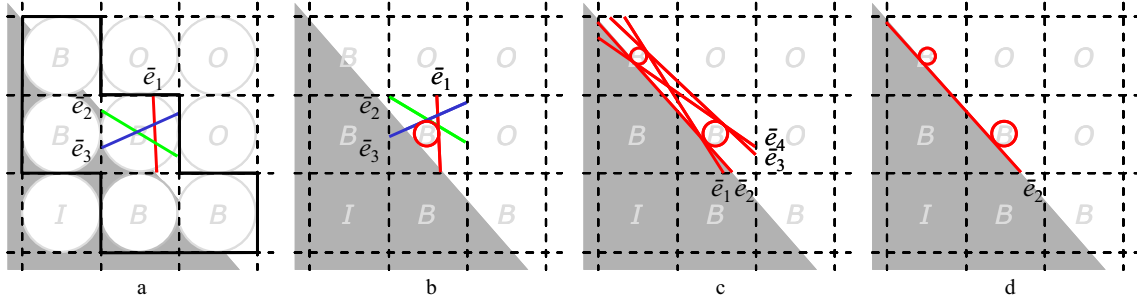


Figure 2. Edge estimation (Pixels labelled *I* are inner region pixels, *O* are outside pixels, and *B* are boundary pixels): a) sample of edges \bar{e}_1 , \bar{e}_2 and \bar{e}_3 , b) estimated edges by considering the distance from the beam center d , c) possible edges tangential to two circles, d) final edge by using the signs of d_i

Gaussian distribution for the sake of simplicity. For the case of the beam partially hitting the surface, we assume the object edge to be straight. In reality, this is not a very strong demand, since the edge has to be essentially straight only for the extent of the beam. Furthermore, we let the edge be parallel to the y -coordinate axis. Because of the radial symmetry of the beam profile, the results can then be generalized to arbitrary edge orientations. If we let d the distance of the surface edge to the beam center, σ the standard deviation of the beam profile and P_0 the average beam power inside the region (q.v. Section 3.2), we find the reflected beam power to be

$$P_B(d) = \frac{P_0}{\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^d e^{-\frac{x^2+y^2}{\sigma^2}} dx dy = \frac{P_0}{\sqrt{\pi}\sigma} \int_{-\infty}^d e^{-\frac{x^2}{\sigma^2}} dx \quad (3)$$

This integral can be described by the complementary error function

$$P_B(d) = \frac{P_0}{2} \operatorname{erfc}\left(-\frac{d}{\sigma}\right) \quad (4)$$

Figure 3a shows a plot of the beam intensity versus the edge offset. Here, the grid spacing is $2f$ and the standard deviation of the Gaussian used to model the beam profile is $f/3$.

4.2 Sub-pixel edge localization in each pixel

Looking at a boundary pixel, we can easily acquire the edge distance from the beam center d_s by inverting the above relationship (using standard numerical procedures) and applying it to the measured return pulse power P_B . However, it is impossible to estimate the edge orientation using only a single pixel (Figure 2b). Therefore we now have a circle with radius d_s around the beam center and know that the edge has to be a tangent to this circle (Figure 3b).

To determine the orientation of this tangent, and along with it the orientation of the edge, we have to use neighborhood information to further restrict the edge hypotheses.

4.3 Estimating tangents in 2 neighborhood edge pixels

A very simple approach consists of using a neighborhood boundary pixel to reduce the problem's degrees of freedom. If we have hints that both of these pixels belong to the same straight object edge, we can use their combined information to estimate the edge. We are looking for a line that is tangential to two circles (radii d_1 and d_2 , respectively). This problem generally has four different solutions (Figure 2c), mathematically. From the intensity response, we know which side of the beam center the tangent passes through, since we

actually get a signed result for d_1 and d_2 , corresponding to the measured intensity being smaller or larger than 50% of the intensity inside the region (Figure 3a & b). If the signs of these two radii are different, the correct edge solution is one of the two tangents crossing between the circles, else one of the outer tangents.

From the segmentation step it is already known which side is the inner and which is the outer side of the boundary, i.e. we know where the boundary is connected to the region. Again, watching the signs of d_1 and d_2 , we know which of the two remaining solutions to choose (Figure 2d, in this case both of the signs are negative).

However, this straightforward approach is very sensitive to noisy data, since it does not use an over-determined system of equations. We will be applying our method to urban areas, where we assume much longer edges to be present. This full edge information can be used to gain higher precision results.

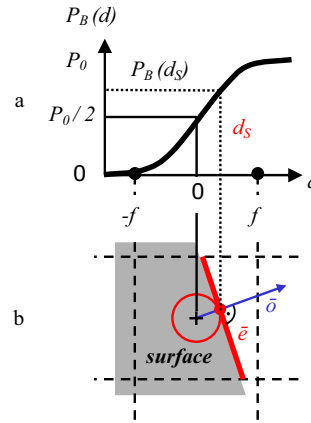


Figure 3. a) Edge \bar{e} with distance d_s , b) Integration over the spatial beam profile (the grid spacing is $2f$)

4.4 Complete edge localization

To use the complete information available for any given edge, we first have to determine the set of boundary pixels belonging to that edge. In the description of the segmentation algorithm, we already explained how to find region boundaries (Section 3.3). This boundary is transformed into a polygon, at first taking each pixel as a vertex (solid line in Figure 4). If there are any pixels in this list occurring more than once, all of their instances are removed. We do this because we assumed each

pixel to belong to only one edge, therefore mixed pixels are not allowed to appear in our calculation.

This vertex list is then pruned by removing all vertices not significantly affecting the polygon contour. This is achieved by tentatively removing each vertex in turn from the contour (dashed line) and calculating the distance of the new polygon (dotted line) to all boundary pixels along this edge. If the largest of these distances g is smaller than one half pixel, the vertex is removed permanently. This step is repeated until no more vertices can be removed.

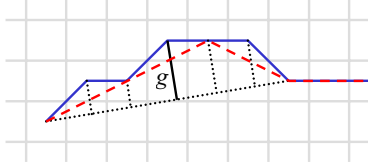


Figure 4. Region boundary simplification

The last polygon found in this fashion is now a pixel-precise estimate of the region boundary, composed of a minimal number of vertices and edges of maximal length. To increase the precision of our estimates for each of these edges, we select the set of all boundary pixels along this edge, leaving out the vertices themselves because of our assumption of pixels fully belonging to straight edges.

For this set of k points,

$$(\vec{p}_i)_{i \in \{1,2,3,\dots,k\}} \subset \mathfrak{R}^2 \quad (5)$$

we calculate the edge distances d_i from the pixel intensity and solve the optimization problem

$$J = \frac{1}{2k} \sum_{i=1}^k (\vec{n}\vec{p}_i - c - d_i)^2 \quad (6)$$

for the edge normal \vec{n} and the edge offset to the coordinate origin c .

As we see, the functional connection between P_B and d does not appear in the formula. Therefore the above problem is a fairly standard optimization problem. Going back to the image, if we replace each intensity value P_B by the associated value d , Equation 6 corresponds to a standard edge localization problem in grayscale images. However, we now have the advantage of knowing our edge models exactly and can present a finer solution than those typically found for passive imagery. Parameterizing the edge normal \vec{n} by its direction ($\vec{n} = (\cos \varphi, \sin \varphi)$), we have a simple two-dimensional optimization problem, though it is nonlinear in φ . We solve for φ using the trust region approach by Coleman & Li (1996). The edge offset to the origin c can then be estimated by

$$\hat{c} = \frac{1}{k} \sum_{i=1}^k (\vec{n}\vec{p}_i - d_i) =: \vec{n}\vec{p} - \vec{d} \quad (7)$$

which is very simple to solve.

4.5 Estimating vertices

Due to the modeling approach, our edge model is correct only for straight edges and incorrect for vertices or curves. Therefore we have been explicitly leaving out corner pixels in the edge localization step. To determine the vertices of the depicted geometry, we calculate the intersections of every pair of neighboring edges.

We chose this approach since the connection of vertex location with the measured intensity is very ambiguous. Furthermore, corners are usually darker and thus more strongly affected by the detector noise. Tying their information in to our optimization problem would complicate matters, remove the independence of neighboring edge localization problems, and does not promise much gain.

5. EXPERIMENTAL VALIDATION

To test the algorithms, we created several images with various resolutions, by the method described in Section 2. The example presented here has a size of ten by ten pixels. In all of the images, the solid line is the ground truth geometry used for image generation, whereas the dashed line shows the result of the individual processing steps.

The range image depicted in Figure 5a is actually a range segmentation result, i.e. the white pixels designate returns from the examined surface. Figure 5b shows the intensity measurement at the range of the surface. The image was brightened to enhance the visibility of the overlay lines – the darkest gray actually stands for zero measured intensity.

The boundary segmentation algorithm presented in Section 3.3 results in the boundary polygon shown in Figure 5c. The boundary simplification method (Section 4.4) reduces this polygon to the correct five vertices, resulting in Figure 5d. These two images are typical results using only pixel-precise edge localization (for example, if no intensity information is available).

The next image (Figure 5e) shows the corner and boundary masks used for the sub-pixel edge localization algorithm. The black pixel [3|2] is a boundary pixel, but it was not used for boundary simplification and edge localization, since it belongs to two of our edges (compare Figure 5c). The image in Figure 5f shows the result of the sub-pixel edge localization.

Figure 6a and b show two tests with noisy images. For both of these images, it has been assumed that the proper region boundary pixels can still be extracted from the waveform information. For Figure 6a, Gaussian noise with standard deviation 0.1 has been applied to the image. Figure 6b shows the limits of the algorithm at a noise level of 0.3.

Orientation error [°]	A	B	C	D	E
Noiseless case	0.00	0.64	1.21	0.04	0.00
Noise 0.1	2.33	1.86	0.19	0.13	2.77
Noise 0.3	12.6	0.34	39.7	15.9	1.84

Table 1. Edge orientation errors in degrees; Edge A is the horizontal edge at the top of the image, the remaining edges follow clockwise.

Localization error [pixels]	A	B	C	D	E
Noiseless case	0.09	0.05	0.06	0.05	0.03
Noise 0.1	0.24	0.20	0.13	0.12	0.21
Noise 0.3	0.72	0.30	1.04	0.68	0.09

Table 2. Edge localization errors in pixels; The localization error measures the maximum distance of the vertices of the edge estimate to the straight line extension of the ground truth edge.

The results of the edge localization are very accurate, despite the low resolution. The vertex positions are generally precise up to about a tenth to a twentieth of a pixel. Tables 1 and 2 give a detailed result of the errors acquired for the edges. While the performance of the algorithm degrades with noisy data, it is

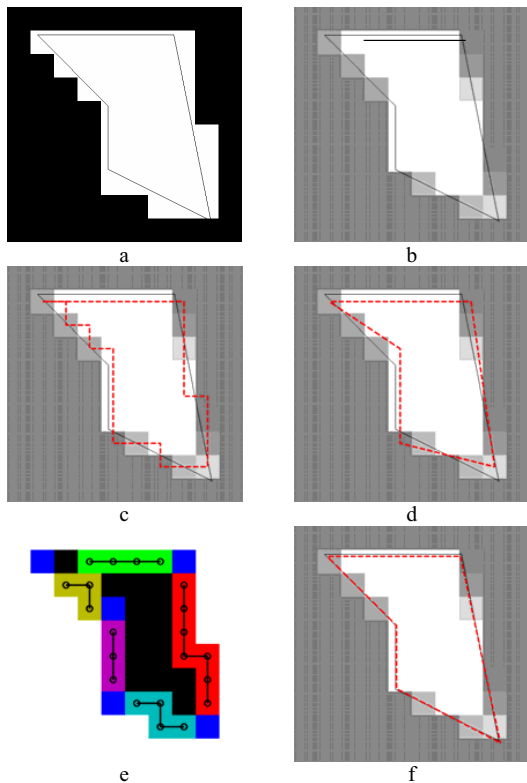


Figure 5. 10x10 source images and results:
 a) Range image with overlaid ground truth geometry
 b) Brightened intensity image with ground truth
 c) Boundary extraction result
 d) Boundary simplification result (dashed) versus ground truth (solid)
 e) Boundary and corner masks
 f) Result of sub-pixel edge localization

evident that longer edges serve to stabilize the results. Edge B, being the longest edge in the test image, gets quite acceptable results even at very high noise levels.

In this work we only considered plane surfaces, fully ignoring the range information except for segmentation purposes. The problems of edge localization in lateral and in range direction are independent of each other and can therefore be tackled separately. Especially in the case of roofs, where the ridge can sometimes not be extracted in the intensity image, a further examination of the geometry in range direction is important.

Also, adjoining surfaces sharing a common edge should be investigated. In this case, we might want to determine one edge using the information from both surfaces instead of two different edges. A typical example for this is a building roof edge, which is usually exactly above an edge between the wall of the building and the ground.

6. CONCLUSION

We have presented a scheme to extract the geometry of man-made objects from laser scanning images under the consideration of the intensity value for each received laser pulse. We have shown that using a laser scanner capable of full waveform processing, edge localization precision can be increased by a factor of at least ten. For the actual sub-pixel

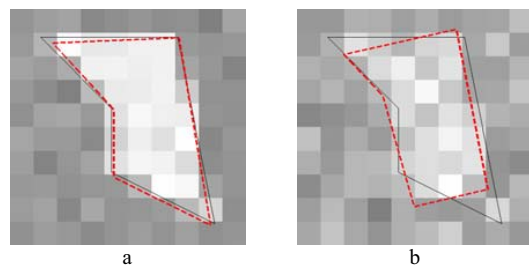


Figure 6. Noisy source images and results:
 a) Noise standard deviation 0.1
 b) Noise standard deviation 0.3

localization algorithm the knowledge of a first pulse intensity image would be sufficient. However, the additional information leads to a much more stable segmentation and consequently higher precision edges. The data generation and analysis we carried out are general investigations for a laser system which records the waveform of laser pulses. The method remains to be tested with real data, and expanded to handle more complex geometries (e.g. vehicles, buildings).

REFERENCES

- Baltsavias EP (1999) Airborne laser scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 164-198.
- Besl PJ (1988) *Surfaces in Range Image Understanding*. New York: Springer-Verlag.
- Coleman TF, Li Y (1996) An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds, *SIAM Journal on Optimization*, Vol. 6, pp.418-445
- Hartley R, Zisserman A (2000) *Multiple View Geometry in Computer Vision*. Proc. Cambridge University Press, Cambridge.
- Hug C, Wehr A (1997) Detecting and identifying topographic objects in laser altimeter data. *ISPRS, International Archives of Photogrammetry & Remote Sensing*, Vol. 32, Part 3-4W2: 19-26.
- Huisjen EJ, Gomes Pereira LM (1998) Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of Photogrammetry & Remote Sensing* 53: 245-261.
- Jutzi B, Eberle B, Stilla U (2002) Estimation and measurement of backscattered signals from pulsed laser radar. In: Serpico SB (ed) (2003) *Image and Signal Processing for Remote Sensing VIII*, SPIE Proc. Vol. 4885: 256-267.
- Jutzi B, Stilla U (2003) Analysis of laser pulses for gaining surface features of urban objects. 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and data fusion on urban areas, URBAN 2003: 13-17.
- Jutzi B, Stilla U (2004) Extraction of features from objects in urban areas using space-time analysis of recorded laser pulses. In: Altan MO (ed) *International Archives of Photogrammetry and Remote Sensing*. Vol. 35, Part B2, 1-6.
- Maas HG (2001) The suitability of airborne laser scanner data for automatic 3D object reconstruction. In: Baltsavias EP, Gruen A, Van Gool L (eds) *Automatic Extraction of Man-Made Objects From Aerial and Space Images (III)*, Lisse: Balkema.
- Vosselman G (2002) On estimation of planimetric offsets in laser altimetry data. Vol. XXXIV, *International Archives of Photogrammetry and Remote Sensing*: 375-380.
- Wagner W, Ullrich A, Melzer T, Briese C, Kraus K (2004) From single-pulse to full-waveform airborne laser scanners: Potential and practical challenges. In: Altan MO (ed) *International Archives of Photogrammetry and Remote Sensing*. Vol 35, Part B3, 201-206.
- Wehr A, Lohr U (1999) Airborne laser scanning – an introduction and overview. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 68-82.



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

ISPRS Journal of Photogrammetry & Remote Sensing 63 (2008) 99–114

PHOTOGRAMMETRY
& REMOTE SENSING

www.elsevier.com/locate/isprsjprs

Iterative processing of laser scanning data by full waveform analysis

Michael Kirchhof ^{a,*}, Boris Jutzi ^a, Uwe Stilla ^b

^a *FGAN-FOM Research Institute for Optronics and Pattern Recognition, 76275 Ettlingen, Germany*

^b *Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, 80290 Muenchen, Germany*

Received 8 December 2006; received in revised form 22 August 2007; accepted 23 August 2007

Available online 24 October 2007

Abstract

The latest developments in laser scanners allow capturing and digitizing of the full waveform of the backscattered pulse. The waveform can be analyzed for measurement features such as range, reflectance values and spreading of the pulse. These features are used to distinguish between locally planar surface elements and partly penetrable objects caused by partial occlusions. This pre-segmentation and the derived range values are used to automatically generate surface primitives in the form of planes. This allows refining each range value taking the surface geometry in a close neighborhood into account. To refine the modeling of the surface, partly occluded surface areas are extended by prediction of the expected range values. This prediction is further improved by considering the surface slope for the estimated received waveform. Then the point cloud associated with the surface is enhanced by additional range values that were missed in the first processing step due to weak signal response. This procedure is repeated several times until all useful range values are considered to estimate the surface.

© 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Keywords: Laser scanning; Waveform analysis; Feature extraction

1. Introduction

The generation of accurate 3D models from laser scanning data for a description of man-made objects is of great interest in photogrammetric research. Detailed description of objects such as buildings requires sampling the surface as a point cloud as dense and complete as possible. However, depending on the scene and the point of view of the laser scanning system, foreground objects like trees in the line of sight of the laser beam interrupt the uniform sampling of the surfaces. These gaps in the sampling of the surface are called partly occluded regions. [Maas \(2000\)](#) included a discussion about partly occluded

regions in his work on TIN (Triangulated Irregular Network) structures and least-squares matching. A different strategy to handle gaps in the point cloud is the application of morphological operations ([Gorte and Pfeifer, 2004](#)). Morphological operations can also be used in the case that no reflections of the region can be measured due to total occlusion by an impervious object. However, vegetation often shows a semi-penetrable property.

State of the art laser scanning systems allow recording the first and last pulse or a given number of pulses. While first pulse registration is the optimum choice to measure the hull of partially penetrable objects, e.g. canopy of trees, last pulse registration should be chosen to measure impenetrable surfaces, e.g. ground surface below vegetation for airborne applications or a building behind vegetation for terrestrial applications.

* Corresponding author. Tel.: +49 7243 992 213; fax: +49 7243 992 299.

E-mail address: kirchhof@fom.fgan.de (M. Kirchhof).

Recent developments in commercial airborne laser scanning systems have led to systems such as OPTECH ALTM 3100, TOPEYE MK II, and TOPOSYS HARRIER 56 (based on RIEGL LMS-Q560) that allow capturing the waveform. These systems are specified to operate with a transmitted pulse width of 4–10 ns at full-width-at-half-maximum (FWHM) and allow digitization and acquisition of the waveform with approximately 0.5–1 GSample/s. Detailed overviews for laser scanning systems are given in (Huising and Gomes Pereira, 1998; Wehr and Lohr, 1999; Baltsavias, 1999).

In contrast to airborne systems, the prototype of the terrestrial laser scanning system ECHIDNA (Lovell et al., 2003) allows to capture the waveform of the backscattered pulse. Terrestrial applications often show a distribution of objects that range from several meters in the foreground to hundreds of meters in the background. Multiple reflections of objects within the beam corridor can be extracted from the full waveform data (Reitberger et al., 2006; Ullrich and Reichert, 2005).

Assuming that waveforms are sampled with sufficiently high frequency, techniques of digital signal processing can be applied. Comparing the transmitted and the received waveform by the cross-correlation function can improve the range estimation. The argument of the maximum of the cross-correlation function estimates the range value with higher reliability and accuracy than the amplitude of the received waveform alone. Details of the improvement can be found in Hofton and Blair (2002), Jutzi and Stilla (2005), and Thiel et al. (2005).

Different surfaces have to be analyzed for different applications. For example for urban objects, it is relevant to deal with objects at different elevations (Brenner et al., 2001). In rural environments it is relevant to deal with randomly distributed natural objects (Reitberger et al., 2006). The impact of the scene on the received waveform has been discussed using standard examples (Jutzi et al., 2002; Wagner et al., 2004; Jutzi and Stilla, 2006).

The focus of this work is reconstruction of man-made objects that can be approximated by planar surfaces. A robust algorithm for finding planes can be implemented using the RANSAC algorithm (Fischler and Bolles, 1981; Hartley and Zisserman, 2000). Many authors have used RANSAC in 2D for extraction of low parameterized transformation models (Workshop 25 Years of RANSAC, 2006). However RANSAC can also be used to fit planes (Brenner et al., 2001) or cylinders (Beder and Förstner, 2006) to 3D point clouds. While Brenner et al. (2001) exploit well defined neighborhood relations to bound the search region for plane primitives, Beder and Förstner (2006) make no assumptions about the search regions. In both approaches a coarse estimation of the variance of

point distances to the fitted model is required. The membership of points to the fitted model is determined by a constant threshold. Local distortions of the point distribution caused by details of the object that are not represented by the model at the given level of detail (LOD) lead to exclusion of the points related to these details. This disadvantage can be overcome by an improved RANSAC based on automatic threshold detection. Speaking precisely, it is possible to capture all points of the object even when the current model does not fit the corresponding point cloud exactly at the current LOD.

In this paper, we propose a method for iterative knowledge-based processing of terrestrial laser scanning data by full waveform analysis. The approach that we present improves the point density and accuracy of the range values. Furthermore, it allows closing gaps in partly occluded surface regions by knowledge-based search.

In Section 2, the surface response of a planar surface with slope is derived. Using this surface response, the corresponding range value is estimated and the correlation function is analyzed. The derived property values are used in a pre-segmentation to separate locally planar surface elements from partly penetrable objects. The result of this process is used to extract surface primitives by RANSAC with automatic threshold selection. In an iterative processing step the slope of the surface is used to improve the accuracy of the range value by increasing the cross-correlation between the received waveform and the expected surface response. This procedure requires the assumption that the surface is locally planar. In Section 3, outdoor experiments with an experimental laser scanning system capturing an urban scene are described. The results of the iterative processing are presented in Section 4. In Section 5 the improvement of the point clouds and the pre-segmentation are discussed. Section 6 completes the contribution with a summary of the advantages of our approach and directions of further research.

2. Methods

The following section describes the whole processing chain. An overview of the processing chain is depicted in Fig. 1. First we introduce the concept of estimating the surface response in Section 2.1. The general interaction between a surface and a laser pulse is analyzed. The output of the analysis is the reduction of the 3D surface characteristic to a 1D range dependent signal. In the following this signal will be called the surface response. A theoretical derivation of the surface response as a function of the slope of a planar surface is given in Section 2.2. In Section 2.3, we introduce a

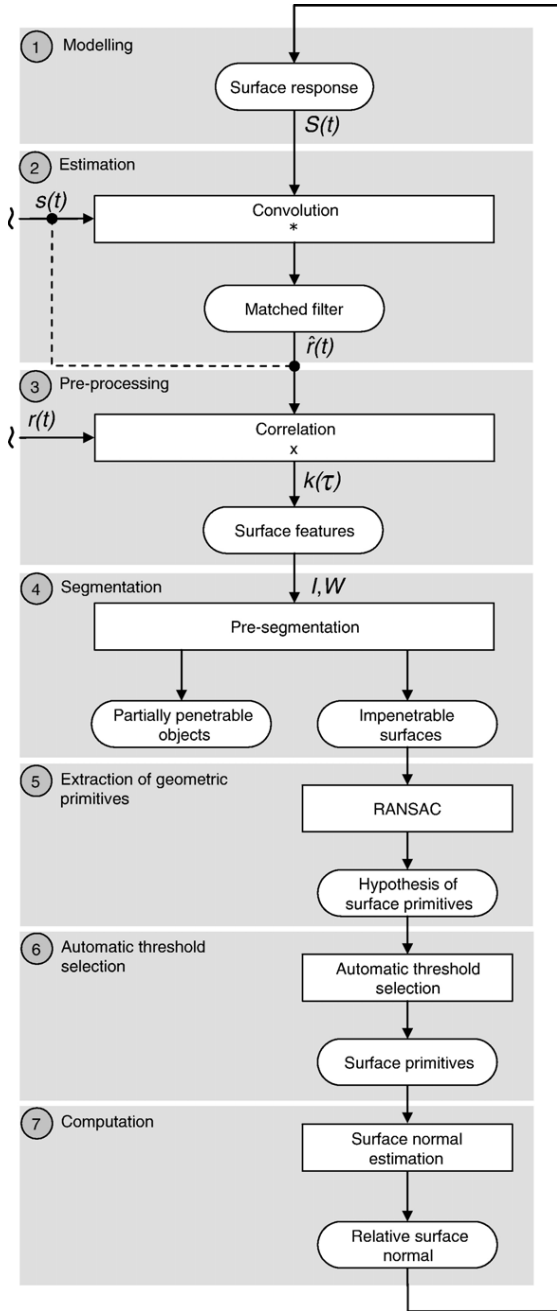


Fig. 1. Flowchart of the processing procedure.

matched filter approach that is used to estimate the range value and additional features of the response. A method for pre-segmenting partly penetrable objects is described in Section 2.4. The result is a set of 3D points that are expected to be located on one surface of a building, e.g. facades or roofs. Here we assume that these surfaces are locally planar. A method for detecting these surface primitives using RANSAC is presented in Section

2.5. Section 2.6 addresses the problem of missing knowledge about the distribution of the 3D points located on a surface primitive. Section 2.7 closes the processing loop and gives initial estimates of the surface response for the matched filter in the next iteration loop.

2.1. Estimating the surface response

The received waveform of a laser pulse depends on the transmitted waveform $s(t)$, the impulse response $h(t)$ of the receiver unit, the spatial beam profile of the used laser $P(x,y)$, and the illuminated surface $S(x,y,z)$. Specifically the received waveform $r(x,y,z,t)$ can be expressed by a convolution of these terms,

$$r(x,y,z,t) = s(t)*h(t)*P(x,y)*S(x,y,z), \quad (1)$$

where $(*)$ denotes the convolution operation. The impulse response consists mainly of the receiving properties of the photodiode and amplifier that are used. The spatial beam profile typically takes the shape of a Gaussian or uniform distribution and the surface characteristics can be described by the geometry of the surface and reflectance properties (mixture of diffuse and specular). We assume a receiver unit consisting of an ideal photodiode and amplifier with an infinite bandwidth and a linear frequency characteristic. The 3D surface characteristic can thus be reduced to the range-dependent 1D surface response $S(z)$.

An analytical description of the surface response for a sloped plane surface is derived in the following section.

2.2. Plane surface with slope

We assume the beam profile of our laser to be uniform (top-hat form), because measurements of the beam profile in the near field have shown that a uniform distribution fits our data best (Jutzi and Stilla, 2006). This is in contrast to the commonly used Gaussian distribution for the beam profile. The divergence of the beam is denoted by Θ . The range from the origin to the idealized plane surface A is denoted by r_{0A} . The angle between the normal \vec{n} of the surface and the optical axis at the centre of the beam is φ (Fig. 2b).

We use a Cartesian coordinate system for which the z -axis points in the direction of propagation of the beam and x is chosen such that \vec{n} coincides with the x - z plane. This implies that small variations in the y direction have negligible influence on the range value z . The general form of the beam cone is

$$x^2 + y^2 \leq \left(z \tan \frac{\Theta}{2} \right)^2. \quad (2)$$

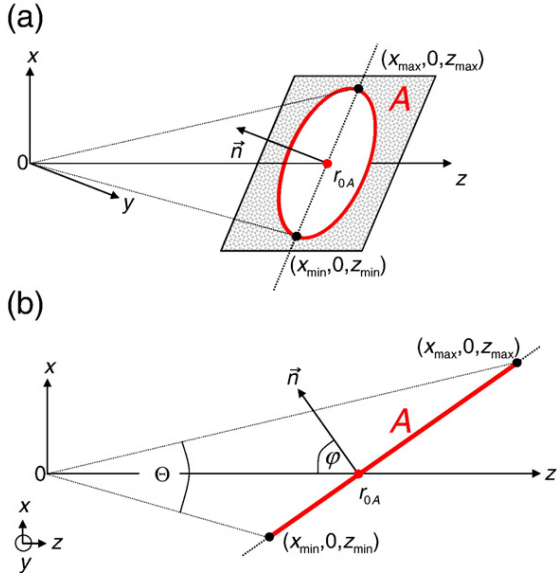


Fig. 2. Schematic description of the projected beam (footprint) at a plane surface. a) Oblique view, b) side view.

The illuminated surface area can be parameterized by Inequality (2) combined with

$$x \sin \varphi + z \cos \varphi = r_{0A} \cos \varphi. \quad (3)$$

Eq. (3) represents the orientation of the surface (Fig. 2). With

$$x = (r_{0A} - z) \cot \varphi \quad (4)$$

and Inequality (2) y is bounded by $y_{\min} = -\left(\left(z \tan \frac{\Theta}{2}\right)^2 - \left((r_{0A} - z) \cot \varphi\right)^2\right)^{0.5}$ and $y_{\max} = \left(\left(z \tan \frac{\Theta}{2}\right)^2 - \left((r_{0A} - z) \cot \varphi\right)^2\right)^{0.5}$.

The surface response $S(z)$ is directly related to the area of the illuminated surface at depth z . The choice of the coordinate system ensures that the range value z is independent of y . Therefore the surface response is given by

$$S(z) = \frac{y_{\max} - y_{\min}}{z} = 2 \sqrt{\left(z \tan \frac{\Theta}{2}\right)^2 - \left((r_{0A} - z) \cot \varphi\right)^2}. \quad (5)$$

As can be seen in Fig. 2a the limits for the range z are given by the equation $y_{\min} = y_{\max} = 0$. It follows that the range z is only valid within the range interval

$$z \in [z_{\min}; z_{\max}] = \left[\frac{r_{0A} \cot \varphi}{\cot \varphi + \tan \frac{\Theta}{2}}; \frac{r_{0A} \cot \varphi}{\cot \varphi - \tan \frac{\Theta}{2}} \right]. \quad (6)$$

Substituting $z = ct/2$, where c is the speed of light and t is the travel time, we obtain a temporal description

of the surface response $S(t)$. If the range to the surface is large compared to the beam diameter, then the surface response is symmetric.

2.3. Estimating the range value by processing the correlation coefficient

First the transmitted waveform $s(t)$ and the received waveform $r(t)$ have to be measured with the receiver unit of the laser system (Fig. 1(2) and (3)). Then, by the use of the transmitted waveform $s(t)$ and the modeled surface response $S(t)$ (Fig. 1(1)) of the known surface for each position on the surface, the estimated received waveform $\hat{r}(t)$ (Fig. 1(3)) is calculated by a convolution to derive the matched filter. If the surface is unknown, a sub-optimal matched filter based on the transmitted waveform $s(t)$ has to be used (dotted line). If the surface is known, the estimated received waveform $\hat{r}(t)$ is compared with the measured waveform $r(t)$ by determining the normalized cross-correlation function $k(\tau)$. The maximum coefficient of the normalized cross-correlation function yields accurate range (Fig. 1(3)).

2.3.1. Matched filter approach

The data analysis starts with the detection of the backscattered pulses in the temporal signal. This signal is usually disturbed by various noise components: background radiation, amplifier noise, photo detector noise, etc. Detecting the received waveform of the backscattered pulse in noisy data and extracting the associated travel time is a well-known problem and is discussed in detail in radar techniques (Skolnik, 1980) and system theory (Papoulis, 1984). In radar techniques, these problems are tackled by a matched filter.

To improve the range accuracy and the signal-to-noise ratio (SNR), the matched filter for the waveform of the backscattered pulse has to be determined. In practice, it is difficult to determine the optimal matched filter. In cases where no optimal matched filter is available, sub-optimal filters may be used at the cost of decreasing the SNR.

In the simplest case, the received waveform is the uniformly attenuated (isotropic attenuation by reflection or transmission of the pulse) copy of the transmitted waveform. In this case, the transmitted waveform of the emitted pulse is the best choice for the matched filter. This is assumed in the first iteration step, where the surface is unknown. In practice, the received waveform is influenced by the interaction between the laser beam and the illuminated surface.

If the surface is known by prior knowledge or rough estimation from previous iteration, the estimated received waveform can be determined (Section 2.1). This estimated

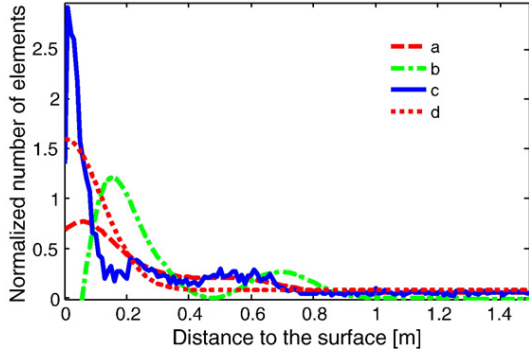


Fig. 3. Example for the automatic threshold detection. a) Smoothed histogram function: $g_{\text{hist}}(\underline{x})$, b) first derivation of the smoothed histogram function: $g_{\text{hist}}'(\underline{x})$, c) normalized histogram function $f_{\text{hist}}(\underline{x})$, and d) mixture fit function $f(\underline{x})$.

received waveform can be used to improve the matched filter. To be precise, the estimated received waveform is the best matched filter if the surface properties are exactly known.

Let us assume that the noise components of the system mentioned above are sufficiently described by white noise with a constant factor N . Let the signal energy of the pulse be E . The maximum SNR occurs if the signal and the filter match. In this case, the associated travel time t of the delayed pulse is τ and the SNR is described by

$$\text{SNR}(\tau) = \frac{2E}{N}. \tag{7}$$

An interesting fact of this postulate is that the maximum of the instantaneous SNR depends only on the signal energy of the emitted pulse and the noise.

Generally, if the surface is unknown, the matched filter is computed by the normalized cross-correlation function R_{sr} between the transmitted waveform $s(t)$ of the emitted pulse and the received waveform $r(t)$ of the backscattered pulse (Fig. 1(2), dotted line). Assuming zero-mean waveforms, we obtain the output signal R_{sr} with a local maximum I_{sr} at the delay time τ via

$$R_{sr}(\tau) = \frac{\int_{t=-\infty}^{\infty} s(t) \cdot r(t + \tau) dt}{\sqrt{\int_{t=-\infty}^{\infty} s^2(t) dt \cdot \int_{t=-\infty}^{\infty} r^2(t) dt}}. \tag{8}$$

If the surface is known, the estimated received waveform $\hat{r}(t)$ is substituted for the transmitted waveform $s(t)$ and the normalized cross-correlation function $R_{\hat{r}r}$ is calculated in an analogous manner. The local maximum of $R_{\hat{r}r}$ is denoted by $I_{\hat{r}r}$.

2.3.2. Processing the range value

The output signal R_{sr} with improved SNR is analyzed by searching for the local maximum to determine the travel time of the pulse. By using the correlation signal to calculate the travel time t , higher accuracy is reached than by operating on the waveform alone (Jutzi and Stilla, 2005). Instead of using a single value of the waveform, the shape of the waveform can be used to increase the accuracy. This is because the specific pulse properties (e.g., asymmetric shape, intensity fluctuations) are taken into account and so less time-varying influence (jitter) is expected on range estimation.

2.4. Pre-segmentation

Pre-segmentation (Fig. 1(4)) is used to compute an initial cue for partially penetrable or partially illuminated objects and impenetrable surfaces. These surfaces are the focus of interest in this contribution.

Partially penetrable objects like vegetation spread the received waveform (Wagner et al., 2004). It is straightforward to show that this spreading propagates to the width W of the cross-correlation function. Since energy is conserved, the spreading also leads to decreased signal amplitude. It is again straightforward that the decreased signal amplitude leads to a decreased maximum I of the

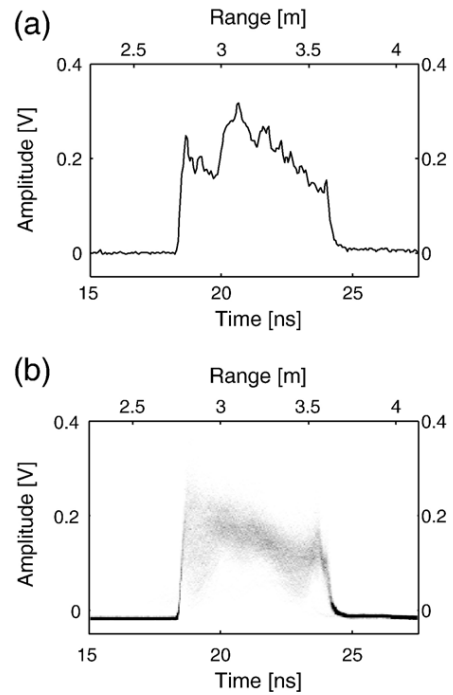


Fig. 4. Transmitted waveform. a) Single example, b) adaptive overlay of 500 waveforms.

cross-correlation function. Thus a decreased cross-correlation maximum I and increased width W of the cross-correlation function will indicate partially penetrable objects. From these two measures we can compose the feature

$$F = (1 - I)W. \quad (9)$$

The new feature F has the physical interpretation of lost energy. An energy difference occurs if a surface is only partially illuminated or the object is partially penetrable. Using a single threshold we use the new feature F to separate points located on impenetrable surfaces from partially illuminated or partially penetrable objects. This threshold is derived manually from the histogram of the feature F .

As it will be shown in the next section it is not important to detect every 3D point located on the building's surface by the pre-segmentation with a single

threshold. It is more important to exclude the vegetation from further processing to reduce ambiguities, because, for example, a virtual plane could be fitted through a large subset of 3d-points located on aligned trees. However this plane could have no physical reality and would therefore belong to no surface primitive.

2.5. Extraction of geometric primitives

The segmentation of Section 2.4 produces a set of 3D points that are expected to be located on a surface primitive of a building (Fig. 1(4)). From this set we generate hypotheses for possible planes in the scene (Fig. 1(5)) based on the random sample consensus algorithm RANSAC (Fischler and Bolles, 1981). RANSAC is an architecture for hypothesis generation and testing. Hypotheses are generated from a minimal set of

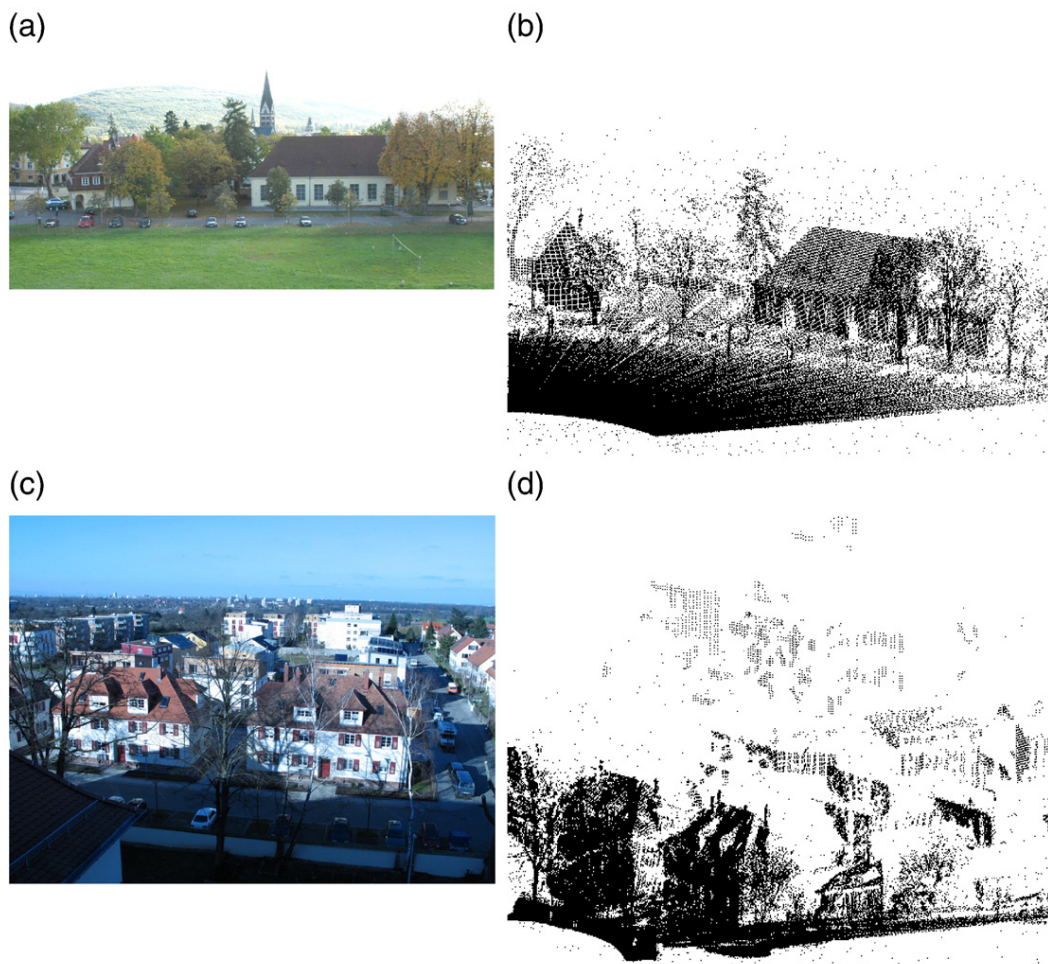


Fig. 5. a) Photo of the test SCENE_A with different urban objects, b) complete 3D point cloud of SCENE_A (APR), c) photo of SCENE_B, d) complete 3D point cloud of SCENE_B.

correspondences. At least three 3D points are needed to compute a unique plane. The hypotheses are computed from a homogeneous representation of the 3D points \vec{x}_i by the formula:

$$B \begin{bmatrix} \vec{n} \\ d \end{bmatrix} = \vec{0}, \text{ with } B = \left[\begin{bmatrix} \vec{x}_1 \\ 1 \end{bmatrix}, \begin{bmatrix} \vec{x}_2 \\ 1 \end{bmatrix}, \dots, \begin{bmatrix} \vec{x}_i \\ 1 \end{bmatrix} \right]^T, \quad (10)$$

or $\begin{bmatrix} \vec{n} \\ d \end{bmatrix} = \underset{\|x\|=1}{\operatorname{argmin}} (\|Bx\|).$

The second formulation has to be used in the over determined case. Here $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_i\}$ can be any arbitrary subset of the 3D point cloud with cardinality $i > 2$. Following the exact RANSAC idea, we use $i = 3$ for the hypothesis generation combined with least squares optimization (guided matching). The optimization is computed from the subset of 3D points that are initially layered as inliers to the current hypothesis. The vector $[\vec{n}^T \ d]^T$ is the hypothesis of the plane. The first three components \vec{n} describe the orientation of the plane while d represents the distance to the origin if \vec{n} is a unit vector. With this parameterization, the geometric distance between the plane and a 3D point \vec{x}_i is given by $|\vec{n}^T d| |[\vec{x}_i^T \ 1]^T|$. Eq. (10) for $[\vec{n}^T \ d]^T$ is linear and can be solved directly using singular value decomposition.

Suitable hypotheses are then evaluated by the number of inliers in the dataset. An initial threshold is set (0.5 m) and each point is initially called an inlier if the geometric distance between the plane and the point is smaller than this threshold. The hypothesis with the largest number of inliers is our solution. The procedure of generating and testing is stopped if the probability that at least one hypothesis was computed only from inliers is greater than $p = 95\%$. The number N_H of needed hypotheses is computed from the formula (Hartley and Zisserman, 2000):

$$N_H = \frac{\log(1 - p)}{\log\left(1 - \left(\frac{\#\text{inlier}}{\#\text{points}}\right)^3\right)}. \quad (11)$$

Here #inlier denotes the size of the currently largest inlier set and #points the size of the 3D point set. The

Table 1
SCENE_A: Surfaces areas and distances between surfaces and the sensor

	SCENE_A FACADE1	SCENE_A ROOF1	SCENE_A FACADE2	SCENE_A ROOF2
Estimated area of the surface [m ²]	232	221	132	60
Distance to center of gravity [m]	115	118	110	112

Table 2
SCENE_B: Surfaces areas and distances between surfaces and the sensor

SCENE_B surface number 1–4	1	2	3	4
Estimated area of the surface [m ²]	59	42	143	42
Distance to center of gravity [m]	113	92	58	98

Table 3
SCENE_B: Surfaces areas and distances between surfaces and the sensor

SCENE_B surface number 5–8	5	6	7	8
Estimated area of the surface [m ²]	134	129	93	16
Distance to center of gravity [m]	63	59	63	93

number $\left(\frac{\#\text{inlier}}{\#\text{points}}\right)$ is the current estimate for the probability that a 3D point lies on the largest remaining plane.

2.6. Automatic threshold selection

The computation of the planar surface primitives (Fig. 1(5)) is based on an initial threshold (0.5 m) for the distance of points to the plane. The threshold can be chosen arbitrarily between the accuracy of the range values and the extension of the smallest plane to detect. This is sufficient to determine a suitable hypothesis for the plane but now we have to decide which of the points are parts of the plane and which are not. This is achieved by an automatic threshold detection algorithm (Fig. 1 (6)).

We assume that the distribution of the distances ρ between the estimated plane and the 3D points in the interval $[0; \rho_{\max}]$ is a mixture composed of a narrow Gaussian with mean $\mu = 0$ for the inliers and a uniform distribution for the outliers. Here ρ_{\max} is an arbitrary limit for the range of interest greater than 2 m. The function

$$f(\rho) = ae^{-\left(\frac{\rho}{b}\right)^2} + c \text{ for } \rho \in [0; \rho_{\max}] \quad (12)$$

is fitted to the histogram function $f_{\text{hist}}(\rho)$ by nonlinear optimization using the Levenberg–Marquardt algorithm (Marquardt, 1963). The function $f_{\text{hist}}(\rho)$ is constructed from piecewise linear interpolation of the histogramR, normalized to $\int_{\mathbb{R}} f_{\text{hist}}(\rho) d\rho = 1$ to be independent of the number of bins. The residuals $f(\rho) - f_{\text{hist}}(\rho)$ are weighted by the term $\left(3 - \frac{\rho}{\rho_{\max}}\right)$ representing the increasing uncertainty of our assumption with increasing range to the plane. Then the resulting equation $f_{\text{cost}}(\rho, a, b, c) = \sum_{i=1}^{N_B} \left(ae^{-\left(\frac{\rho_i}{b}\right)^2} + c - f_{\text{hist}}(\rho_i)\right)^2 \left(3 - \frac{\rho_i}{\rho_{\max}}\right)$ is minimized, where ρ_i are the bins of the histogram and N_B is the number of bins. Let $g_{\text{hist}}(\rho)$ be the convolution of $f_{\text{hist}}(\rho)$ with a Gaussian

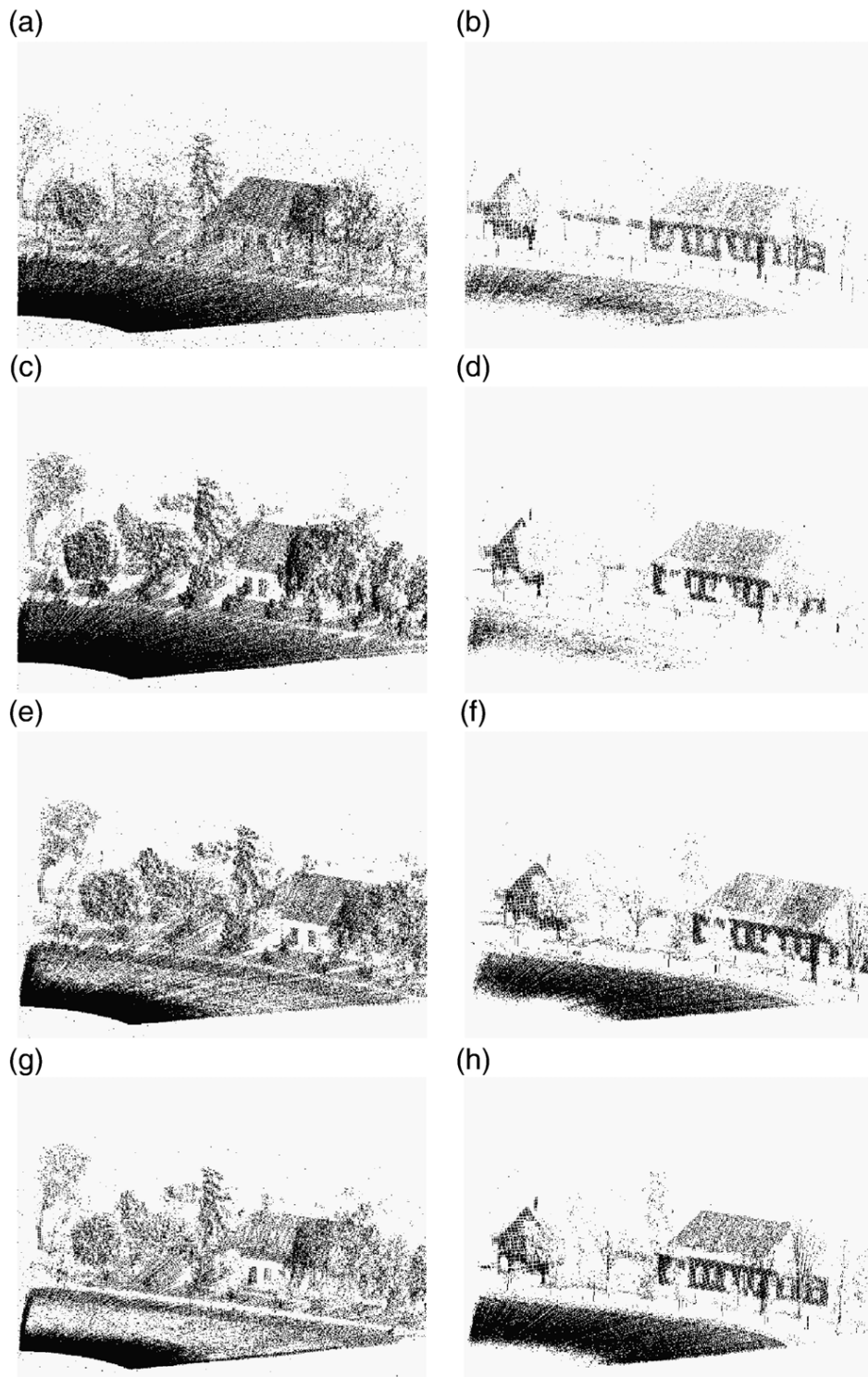


Fig. 6. Oblique view of the pre-segmentation results: Partly penetrable objects (left side) and 3D points corresponding to planar surfaces (right side). a) and b) APR, c) and d) MAY, e) and f) OCT, g) and h) NOV.



Fig. 7. Pre-segmentation result of the points labeled impervious without the main plane of the scene (APR).

function with mean $\mu=0$ cm and a standard deviation of $\sigma=0.1$ cm. Then the optimization is initialized by

$$\begin{aligned} a_0 &= \max (\{f_{\text{hist}}(\rho) | x \in [0; \rho_{\text{max}}]\}), \\ b_0 &= \arg \max_{\rho \in [0; \rho_{\text{max}}]} \left(\frac{\partial g_{\text{hist}}(\rho)}{\partial \rho} \right), \text{ and} \\ c_0 &= \int_{\rho=0}^{\rho_{\text{max}}} \frac{f_{\text{hist}}(\rho)}{\rho_{\text{max}}} d\rho. \end{aligned} \tag{13}$$

A typical example for the automatic threshold detection is depicted in Fig. 3. The solid curve shows

the piecewise linear approximation of the histogram $f_{\text{hist}}(\rho)$, the smoothed histogram function $g_{\text{hist}}(\rho)$ is depicted by a dashed curve, the dash-dot curve shows the first derivative of $g_{\text{hist}}(\rho)$ and the resulting mixture fit function $f(\rho)$ is drawn as a dotted curve.

The histogram is computed with 1000 bins. Because of the normalization used for the histogram, the result $f(\rho)$ is independent of the number of bins but depends on quantization effects. These effects are suppressed by low pass filter, namely the convolution with the Gaussian function above.

2.7. Computation of the relative surface normal and the iterative loop

The surface primitives and the corresponding point sets computed by RANSAC combined with the automatic threshold detection give us an initial guess for the scene geometry (Fig. 1 (5) and (6)). Since high level modeling methods require dense point clouds, we use this initial cue to improve our results. The convex hull of a point set is increased within the surface primitive to estimate the search region for new points in the scene. The process is similar to a morphological dilation within the plane in 3D. The range value z_0 of a beam hitting the

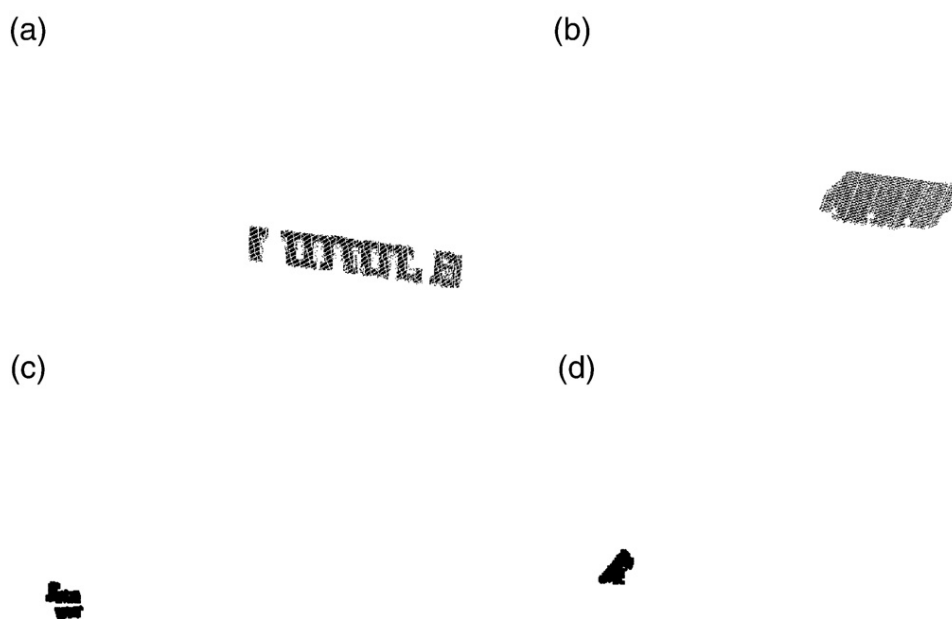


Fig. 8. Examples of four different surface primitives. a) SCENE_A {FAÇADE}, b) SCENE_A {ROOF}, c) SCENE_A façade of the second building, d) SCENE_A roof of the second building.

geometric primitive within the increased convex hull is computed by the formula

$$z_0 = -d/\vec{v}^T \vec{n}. \quad (14)$$

Again the vector $[\vec{n}^T \ d]^T$ represents the geometric primitive as described in Section 2.5 and \vec{v} is the normalized direction of the laser beam.

The computed angle between the normal vector of the surface and the laser beam is used to calculate the expected surface response (Fig. 1 (7) and (1)). This surface response is used to increase the cross correlation function between the expected and the measured waveform. Since we have already computed the direction vector of the laser beam \vec{v} , the angle

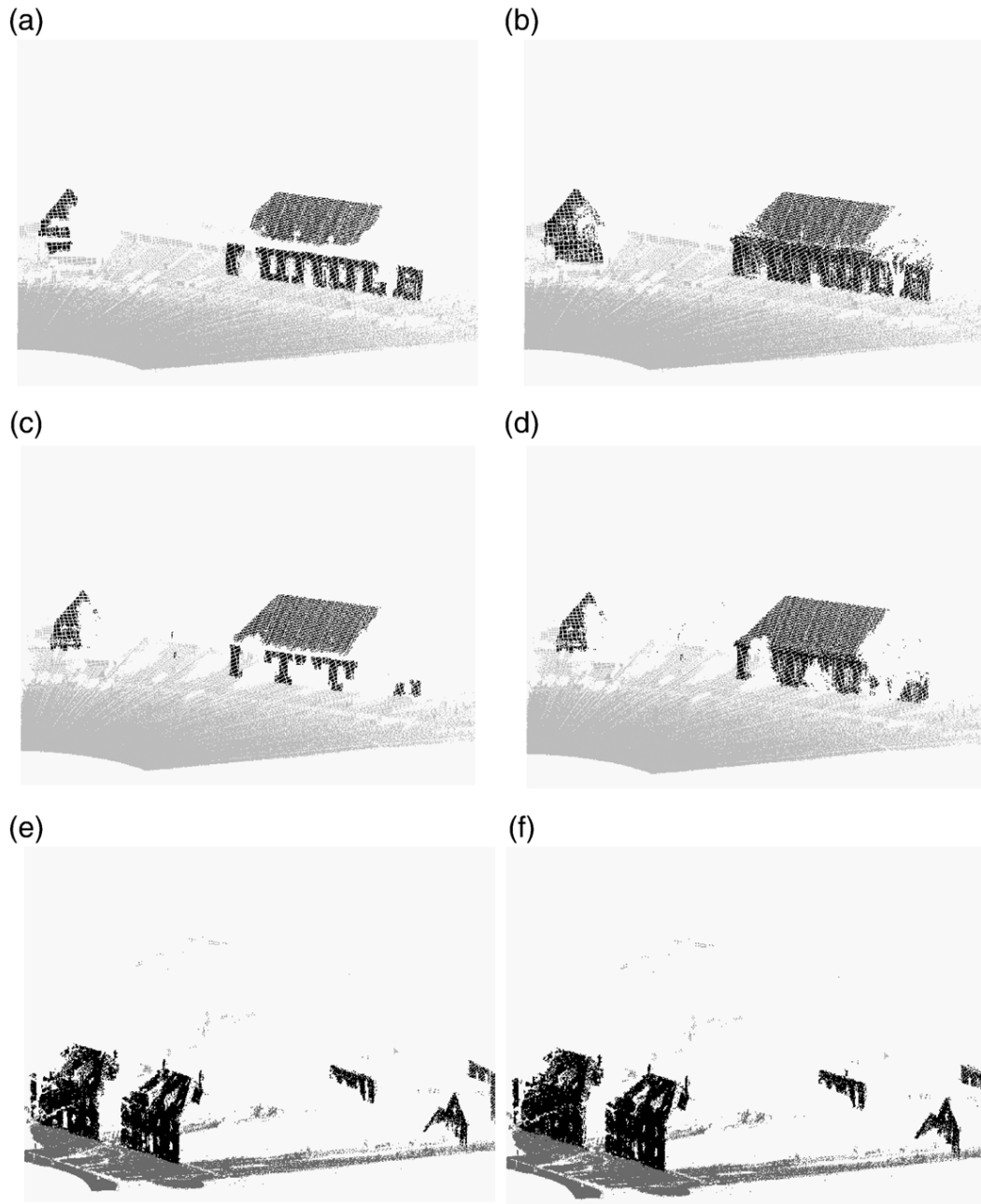


Fig. 9. Complete set of surface primitives detected in the SCENE_A (APR). a) Before iteration, b) after iteration. Complete set of surface primitives detected in the SCENE_A (MAY). c) Before iteration, d) after iteration. The complete set of evaluated surface primitives in SCENE_B. e) Before iteration, f) after iteration.

between the laser beam and the surface normal \vec{n} is given by

$$\hat{\varphi} = \arccos(\vec{v}^T \vec{n}). \quad (15)$$

This can be used with Eq. (5) and z_0 from Eq. (14) to compute the resulting surface response

$$S(z, \hat{\varphi}) = 2\sqrt{\left(z \tan \frac{\Theta}{2}\right)^2 - ((z_0 - z) \cot \hat{\varphi})^2}. \quad (16)$$

The estimated surface response $S(z, \hat{\varphi})$ can now be used as prior knowledge in the matched filter approach described in Section 2.3.1.

3. Experiments

3.1. Laser scanning system

An experimental setup that allows capturing the waveform was built to explore the capabilities of the laser scanning system described in Section 2. The laser scanning system has three main components: an emitter unit, a motion control unit, and a receiver unit.

We use a laser pulse system with a pulse duration of 5 ns at full-width-at-half-maximum (FWHM) and a high repetition rate (42 kHz). The peak power of the laser is up to 10 kW. The multi-mode erbium fiber laser operates at a wavelength of 1550 nm with a beam divergence of 1 mrad. The system uses a photodiode to pump the multi-mode fiber cavity and a fiber amplifier. The transmitted waveform shows strong random modulation for each emitted pulse. In Fig. 4 a single example of the transmitted waveform and an adaptive overlay of 500 transmitted waveforms is depicted to make the strong random modulation clear.

For the 2D scanning process, a moving mirror is used for an elevation scan with ± 15 degrees in vertical direction (320 raster steps) and a rotation stage for an azimuth scan with 360 degree rotation in the horizontal direction (variable number and spacing of the raster steps). For the experiments, we selected 60° (600 raster

steps) for equiangular spacing of approximately 0.1° in vertical and horizontal direction.

For our investigations we use a receiver with a bandwidth of 250 MHz containing a PIN photodiode sensitive at wavelengths of around 900 to 1700 nm. Furthermore, we use an A/D converter with 20 GSamples/s. The A/D conversion and digital recording is accomplished by using a digital memory oscilloscope (Le Croy – Wavemaster 8600A), where the bandwidth of the oscilloscope is limited to 6 GHz.

3.2. Data acquisition

We used four datasets of SCENE_A (Fig. 5a, b) and one dataset of SCENE_B (Fig. 5 c, d) to assess the algorithms and to demonstrate the performance of our methodology. The datasets were taken in April (APR), May (MAY), October (OCT), and November (NOV) for SCENE_A and in February for SCENE_B. The sensor platform were placed at a height of 15 m for SCENE_A and 23 m for SCENE_B. Objects in the scene are buildings, streets, vehicles, parking lots, meadow, bushes, and trees. The scenes do not include coniferous trees but do include deciduous trees (birches and chestnuts). Some objects are partly occluded and the materials have different reflectance properties.

Since the evaluation would not be reliable on small surface patches (Fig. 10), we focus on two large surfaces in SCENE_A: SCENE_A{ROOF} and SCENE_A{FAÇADE}. The area size of the evaluated surfaces and their distances (center of gravity) to the sensor are depicted in Tables 1, 2 and 3.

The pre-segmentation is influenced by the foliage of the trees. Fig. 6 shows an oblique view of the partial penetrable 3D point cloud (left side) and the pre-segmentation results (right side) of all SCENE_A datasets. The evaluation shows that pre-segmentation is much easier if the trees are leafy. This was expected since leafy trees spread the waveform. For clarity the pre-segmentation result of the APR scene is shown without the main scene plane in Fig. 7. The detected

Table 4
Results of the evaluation methods for the surface primitive SCENE_A{FAÇADE}

SCENE_A{FAÇADE}	APR	MAY	OCT	NOV
Median angle $\hat{\varphi}$ between surface and beam	10.4°	8.7°	8.3°	8.5°
μ mean of the improvements Φ_i	$[10^{-3}] 0.752$	$[10^{-3}] 0.637$	$[10^{-3}] 0.537$	$[10^{-3}] 0.705$
σ standard deviation of the improvements Φ_i	$[10^{-3}] 0.629$	$[10^{-3}] 0.479$	$[10^{-3}] 0.384$	$[10^{-3}] 0.583$
Probability $P(\mu > 0)$ (Eq. (17))	90%	90%	92%	89%
# points on plane without iteration	4038	1830	3207	4102
# points on plane after iteration	6013	3641	4398	4588

Table 5
Results of the evaluation methods for the surface primitive SCENE_A {ROOF}

SCENE_A {ROOF}	Apr	May	Oct	Nov
Median angle $\hat{\varphi}$ between surface and beam	37.2°	40.2°	40.55°	40.3°
μ mean of the improvements Φ_i	[10 ⁻³] 6.9	[10 ⁻³] 9.9	[10 ⁻³] 7.4	[10 ⁻³] 6.6
σ standard deviation of the improvements Φ_i	[10 ⁻³] 4.2	[10 ⁻³] 3.2	[10 ⁻³] 3.9	[10 ⁻³] 3.5
Probability $P(\mu > 0)$ (Eq. (17))	94.5%	99%	97%	98%
# points on plane without iteration	4701	4615	5565	2814
# points on plane after iteration	5543	5259	6037	4025

surface primitives SCENE_A {ROOF} and SCENE_A {FAÇADE} of the two buildings are shown in Fig. 8. Beside SCENE_A, a second urban scene was processed. SCENE_B is more complex, with many occlusions and objects of various sizes with dimensions of 15–400 m.

4. Results

We evaluate our results quantitatively with the evaluation methods described in Section 4.1. A visual impression of the results is presented in Fig. 9. All 3D points located on the evaluated surfaces and the main scene plane are shown before and after the iteration procedure. The ground surface is only shown for visualization purposes. In Section 4.3, we focus on the influence of the surface response on the maximum of the cross-correlation function.

4.1. Evaluation methods

We evaluate our results using two different measures. The first measure quantifies the increase of the cross-

correlation value computed in the matched filter by using the estimated surface normal. This evaluation also tests our theory of the surface response. If the surface response is modeled correctly, the use of the surface normal should increase the cross-correlation value between the estimated received waveform and the measured received waveform. The evaluation is conducted using the 3D points located on the detected planes. If the estimation is not improved, the “improvements” $\Phi_i = I_{ir} - I_{ST}$ will be normally distributed with zero mean. The improvements Φ_i are defined by the local maxima of the cross-correlation functions described in Section 2.3.1. The first measure is the mean μ of the improvements Φ_i for each surface.

We now assume that the improvements Φ_i are normally distributed and we compute their standard deviation σ . Combining μ and σ into the cumulative probability density function $\Phi_{\mu\sigma}$ for the normal distribution, we can compute the probability of the hypothesis that the improvements Φ_i have a positive mean $\mu > 0$ by

$$P(\mu > 0) = 1 - \Phi_{\mu\sigma}(0) = 1 - \int_{-\infty}^0 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (17)$$

The second measure is the power of the 3D point set (number of elements) located on the estimated planes. Tables 4–6 show the power of the 3D point sets before and after iteration. The additional points were not detected in the initial processing step.

4.2. Numerical results

Tables 4–6 are computed with the method described in Section 4.1. In SCENE_A we focus on the surfaces SCENE_A {FAÇADE} and SCENE_A {ROOF}. They provide reliable results because of their large size. The dependence of the reliability of the results on the size can be seen in Fig. 10 and is further analyzed in Section 4.3. Table 6 shows the results for eight

Table 6
Results of the evaluation methods for the SCENE_B

SCENE_B	1	2	3	4	5	6	7	8
Median angle $\hat{\varphi}$	6.8°	7.8°	15.9°	16.3°	27.8°	35.8°	37.5°	38.4°
μ mean	[10 ⁻³] 0.40	[10 ⁻³] 0.36	[10 ⁻³] 0.49	[10 ⁻³] 0.96	[10 ⁻³] 1.3	[10 ⁻³] 1.8	[10 ⁻³] 2.3	[10 ⁻³] 4.1
σ standard deviation	[10 ⁻³] 0.17	[10 ⁻³] 0.12	[10 ⁻³] 0.22	[10 ⁻³] 0.48	[10 ⁻³] 0.9	[10 ⁻³] 1.2	[10 ⁻³] 1.6	[10 ⁻³] 3.2
Probability $P(\mu > 0)$	99%	99.8%	98%	98%	93%	92%	92%	90%
# without iteration	682	1306	10912	1114	7839	6364	4219	371
# after iteration	889	1494	13502	1390	9976	7986	5342	525

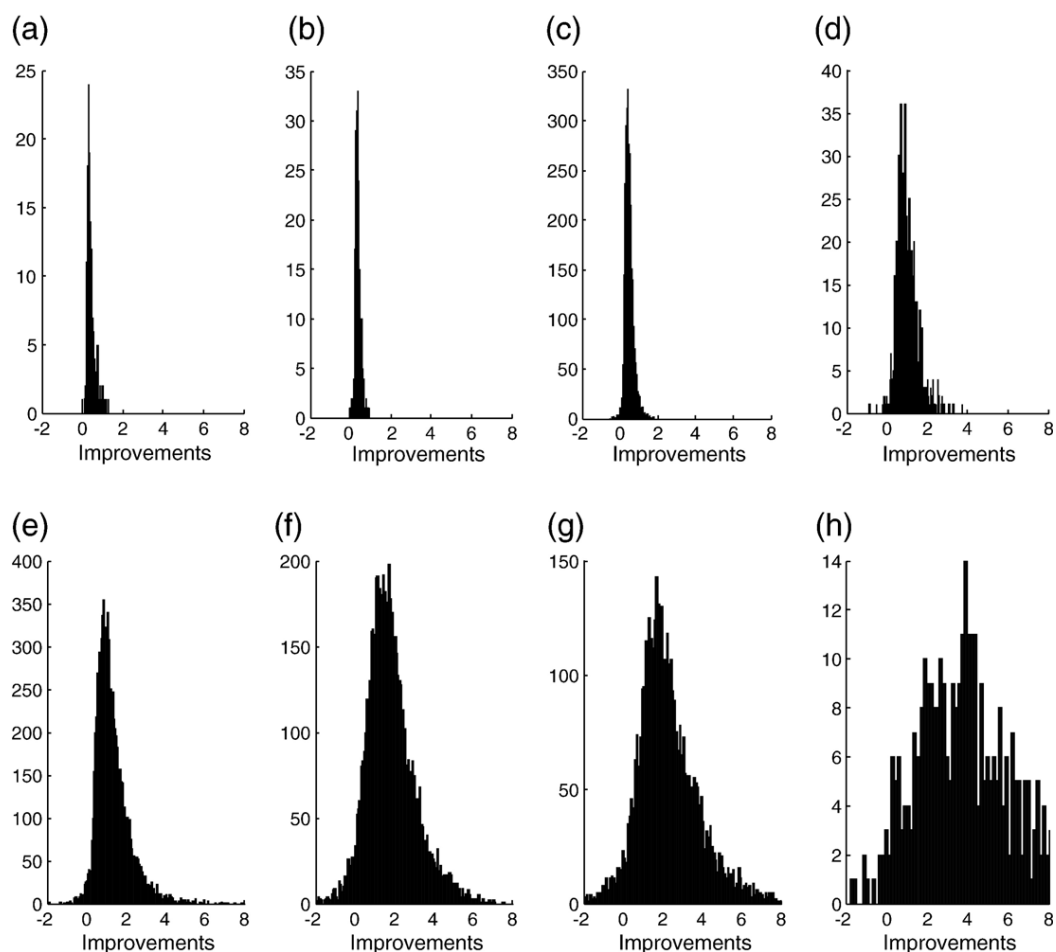


Fig. 10. Histograms of the cross-correlation improvements Φ_i for all evaluated surfaces of SCENE_B with the angles a) 7°, b) 8°, c) 16°, d) 16.5°, e) 28°, f) 36°, g) 37°, and h) 38°.

randomly chosen surfaces of SCENE_B. Tables 4–6 show the results of the evaluation methods described in Section 4.1.

4.3. Improvements depending on the surface response

Fig. 10 shows the 8 histograms of the improvements Φ_i evaluated for the surfaces of SCENE_B. The improvements Φ_i are approximately normally distributed. Mean and standard deviation rise with increasing surface slope.

Surfaces elements with a large area (Fig. 10 c, e, f, g) provide many 3D points that can be analyzed. Therefore the histogram is similar to a continuous distribution of the improvements Φ_i . In addition to this it is known that the standard deviation of the mean estimation decreases with $\frac{1}{\sqrt{n-1}}$ where n is the number of measurements that contribute to the mean.

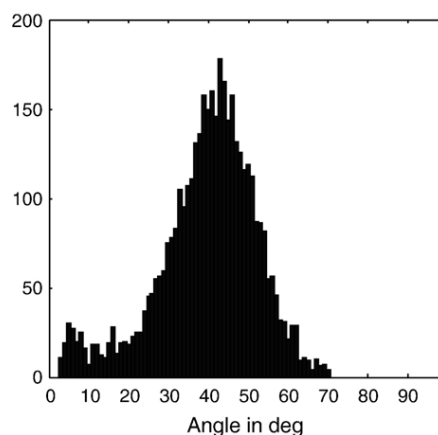


Fig. 11. Histogram of the angles estimated by the surface responses (SCENE_A{ROOF}).

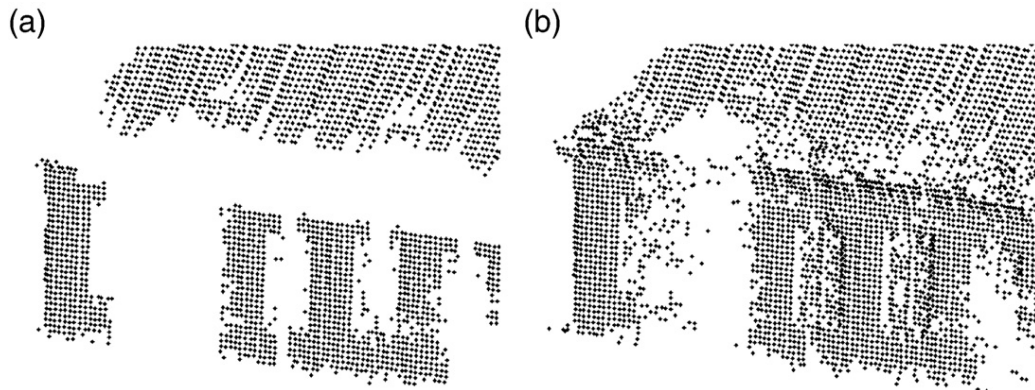


Fig. 12. Enlargement of the surface boundaries. a) Before iteration data was incomplete at interesting regions like windows, b) after iteration an improvement is visible with additional points at interesting regions.

To test the validity of the results we conducted another experiment: Each range value of SCENE_A {ROOF} taken in November was reprocessed with varying assumed surface slope $\hat{\varphi}$ in the interval $[1^\circ; 90^\circ]$ and the angle with the maximal cross-correlation was recorded. The results are shown in Fig. 11. The mean surface angle was about 40° as can be seen in Table 5.

5. Discussion

The experiments show that the use of the estimated surface response from the surface normal can increase the response of the matched filter for feature extraction. They also show that the methodology still holds in real life scenarios with low SNR. Tables 4, 5, and 6 show that the mean μ of the improvements Φ_i in row 3 is correlated with the surface slope $\hat{\varphi}$ in row 2. But the standard deviation σ of the improvements Φ_i in row 4 is correlated with the slope $\hat{\varphi}$ as well. More precisely, if we can achieve larger improvements Φ_i by considering the estimated slope $\hat{\varphi}$, then the variations of the improvements Φ_i increase (Fig. 10).

As mentioned in Section 2, an increased surface slope φ has the effect of spreading the waveform, which decreases the amplitude of the waveform due to conservation of the energy. Therefore the estimated range value becomes less accurate, independent of the detection method used, for increasing the surface slope. Nevertheless the algorithm proposed improves the detection rate as demonstrated in rows 6 and 7 of Tables 4–6.

This results show that the surface slope is relevant for almost every application. For example, actual sensors like RIEGL LMS-Q560 have a maximum field of view of 60° . Even if we assuming further a planar scene, the range variation increases with the angle between the

direction of propagation of the beam and the surface normal and becomes at least 30° .

In Section 4.1 we assumed that the improvements Φ_i can be approximated by a normal distribution. This was required to compute our evaluation measure, the probability $P(\mu > 0)$. All histograms in Fig. 10 are obviously close to a normal distribution which matches our assumption. Due to noise and numerical effects, negative improvements Φ_i can be observed.

The set of points associated with a surface primitive can be increased between by 10% and 100%. A huge number of the additional 3D points are located near the boundaries of the surface, which are the edges of the object (Fig. 12). If the surfaces are curved, the 3D points in these regions describe the edges more accurately than the intersections of the computed planes. Again, we would like to point out that we see our contribution as an input for high-level modeling which requires as few gaps as possible. For example, a method based on 3D morphological operations, such as that of Gorte and Pfeifer (2004), would benefit from additional points in partly occluded regions.

Fig. 9 shows the point sets belonging to the detected planar surface primitives. Beside these the remaining points, which were initially detected, might correspond to partially penetrable objects like vegetation (e.g., foliage of trees) but could although correspond to man-made objects (e.g., fences or street lamps).

The segmentation of vegetation and man-made objects is required for many applications in urban environments (Maas and Vosselman, 1999). The application of full-waveform laser scanning systems in forestry environments has been demonstrated (Wynne and Nelson, 2006). In addition to laser scanning data, additional data like close infrared thermal images are typically used (Haala and Brenner, 1999) for segmentation.

The comparison of the pre-segmentation results of an identical scene at different times (Fig. 9) shows that the denser vegetation in summer makes the pre-segmentation more accurate. This might be caused by the foliage which leads to a spreading of the received waveform. This effect can be seen in Fig. 6c and d. For the same reason, we could detect only a few 3D points occluded by trees in summer. The effect is visible in Fig. 9c and d when compared with Fig. 6c. The trees become almost impervious in leaf-on situations. No reflections can be measured from the occluded regions.

6. Conclusion

The iterative processing of full waveform laser data increases the set of 3D points associated with each surface. Additional 3D points can be detected in partially occluded and partly illuminated regions. A rough pre-segmentation can be achieved by a single threshold using a combination of the width and the amplitude of the cross correlation function. The use of prior knowledge of the surface normal improves the range and correlation value. Vice versa the surface response can be used to compute the surface normal without prior knowledge but at the cost of high computational effort. Local plane fitting with RANSAC is a useful method to compute this surface normal. The method presented requires full waveform laser data but is not restricted to terrestrial or airborne laser scanning systems.

Additional work could be done on analyzing the improvements of the cross-correlation maximum Φ_i depending on the surface slope and reflectance properties. The next step for a more detailed description of the scene is to fit quadrics (e.g., cylinders, spheres, ellipsoids, hyperboloids) to the point clouds. Note here that planes are degenerate versions of a quadric and are therefore also described by the same mathematical model. This means that planes and curved surfaces can be determined using the same model without detection of planarity or curvature. Further, the performance of the RANSAC approach can be improved by using sets of points in the hypothesis generation step that are neighbored in a general way, e.g. the distances between pairs of points in the generation set is bounded.

References

- Baltsavias, E.P., 1999. Airborne laser scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 164–198.
- Beder, C., Förstner, W., 2006. Direct solutions for computing cylinders from minimal sets of 3D points. In: Leonardis, A., Bischof, H., Pinz, A. (Eds.), *Computer Vision—ECCV 2006: 9th European Conference on Computer Vision, Part I. Lecture Notes in Computer Science*, vol. 3951. Springer, pp. 135–146.
- Brenner, C., Haala, N., Fritsch, D., 2001. Towards fully automated 3D city model generation. In: Baltsavias, E., Grün, A., Van Gool, L. (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*. Balkema Publishers, Rotterdam, pp. 47–57.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communication of the Association for Computing Machinery* 24 (6), 381–395.
- Gorte, B., Pfeifer, N., 2004. Structuring laser-scanned trees using 3D mathematical morphology. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B5), 929–933.
- Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 130–137.
- Hartley, R., Zisserman, A., 2000. *Multiple View Geometry*. Cambridge University Press, Cambridge, UK.
- Hofton, M.A., Blair, J.B., 2002. Laser altimeter return pulse correlation: A method for detecting surface topographic change. *Journal of Geodynamics* 34 (3), 491–502.
- Huising, E.J., Gomes Pereira, L.M., 1998. Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (5), 245–261.
- Jutzi, B., Eberle, B., Stilla, U., 2002. Estimation and measurement of backscattered signals from pulsed laser radar. In: Serpico, S.B. (Ed.), *Image and Signal Processing for Remote Sensing VIII*. SPIE Proceedings, vol. 4885, pp. 256–267.
- Jutzi, B., Stilla, U., 2005. Measuring and processing the waveform of laser pulses. In: Grün, A., Kahmen, H. (Eds.), *Optical 3-D Measurement Techniques VII*, vol. I, pp. 194–203.
- Jutzi, B., Stilla, U., 2006. Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS Journal of Photogrammetry and Remote Sensing* 61 (2), 95–107.
- Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., Coops, N.C., 2003. Using airborne and ground based ranging Lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing* 29 (5), 607–622.
- Maas, H.-G., 2000. Least-squares matching with airborne laserscanning data in a TIN structure. *International Archives of Photogrammetry and Remote Sensing* 33 (Part 3A), 548–555.
- Maas, H.-G., Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 153–163.
- Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 11 (2), 431–441.
- Papoulis, A., 1984. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Tokyo.
- Reitberger, J., Krzystek, P., Stilla, U., 2006. Analysis of full waveform LIDAR data for tree species classification. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), 228–233.
- Skolnik, M.I., 1980. *Introduction to radar systems*, Second Edition. McGraw-Hill International Editions, New York.
- Thiel, K.H., Wehr, A., Hug, C., 2005. A new algorithm for processing fullwave laser scanner data. *EARSeL 3D—Remote Sensing Workshop*. (on CD-ROM).
- Ullrich, A., Reichert, R., 2005. High resolution laser scanner with waveform digitization for subsequent full waveform analysis. In:

- Kamerman, G.W. (Ed.), *Laser Radar Technology and Applications X*. SPIE Proceedings, vol. 5791, pp. 82–88.
- Wagner, W., Ullrich, A., Melzer, T., Briese, C., Kraus, K., 2004. From single-pulse to full-waveform airborne laser scanners: potential and practical challenges. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3)*, 201–206.
- Wehr, A., Lohr, U., 1999. Airborne laser scanning— an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing 54 (2–3)*, 68–82.
- Workshop 25 Years RANSAC, (conjunction with CVPR'06), 2006, ISBN 0-7695-2646-2, IEEE Computer Society, Los Alamitos USA.
- Wynne, R.H., Nelson, R.F. (Eds.), 2006. *SilviScan special issue— Lidar applications in forest assessment and inventory*. *Photogrammetric Engineering and Remote Sensing*, vol. 72 (12).

A STEP TOWARDS DYNAMIC SCENE ANALYSIS WITH ACTIVE MULTI-VIEW RANGE IMAGING SYSTEMS

Martin Weinmann and Boris Jutzi

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Kaiserstr. 12, 76128 Karlsruhe, Germany
{martin.weinmann, boris.jutzi}@kit.edu

Commission III, WG III/5

KEY WORDS: LIDAR, Multisensor, Point Cloud, Imagery, Automation, Close Range, Dynamic

ABSTRACT:

Obtaining an appropriate 3D description of the local environment remains a challenging task in photogrammetric research. As terrestrial laser scanners (TLSs) perform a highly accurate, but time-dependent spatial scanning of the local environment, they are only suited for capturing static scenes. In contrast, new types of active sensors provide the possibility of simultaneously capturing range and intensity information by images with a single measurement, and the high frame rate also allows for capturing dynamic scenes. However, due to the limited field of view, one observation is not sufficient to obtain a full scene coverage and therefore, typically, multiple observations are collected from different locations. This can be achieved by either placing several fixed sensors at different known locations or by using a moving sensor. In the latter case, the relation between different observations has to be estimated by using information extracted from the captured data and then, a limited field of view may lead to problems if there are too many moving objects within it. Hence, a moving sensor platform with multiple and coupled sensor devices offers the advantages of an extended field of view which results in a stabilized pose estimation, an improved registration of the recorded point clouds and an improved reconstruction of the scene. In this paper, a new experimental setup for investigating the potentials of such multi-view range imaging systems is presented which consists of a moving cable car equipped with two synchronized range imaging devices. The presented setup allows for monitoring in low altitudes and it is suitable for getting dynamic observations which might arise from moving cars or from moving pedestrians. Relying on both 3D geometry and 2D imagery, a reliable and fully automatic approach for co-registration of captured point cloud data is presented which is essential for a high quality of all subsequent tasks. The approach involves using sparse point clouds as well as a new measure derived from the respective point quality. Additionally, an extension of this approach is presented for detecting special objects and, finally, decoupling sensor and object motion in order to improve the registration process. The results indicate that the proposed setup offers new possibilities for applications such as surveillance, scene reconstruction or scene interpretation.

1 INTRODUCTION

An appropriate 3D description of the local environment is represented in the form of point clouds consisting of a large number of measured 3D points and, optionally, different attributes for each point. Such point clouds can directly be acquired with different scanning devices such as terrestrial laser scanners (TLSs), time-of-flight (ToF) cameras or devices based on the use of structured light. However, a single scan often is not sufficient and hence, multiple scans have to be acquired from different locations in order to get a full scene coverage. As each captured point cloud represents 3D information about the local area only with respect to a local coordinate frame, a basic task for many applications consists of a point cloud registration. This process serves for estimating the transformation parameters between different point clouds and transforming all point clouds into a common coordinate frame. Existing techniques for point cloud registration rely on

- 3D geometry,
- 3D geometry and the respective 2D representation as range image and
- 3D geometry and the corresponding 2D representation of intensity values.

Standard approaches involving only the spatial 3D information for calculating the transformation parameters between two partially overlapping point clouds are based on the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) and its variants

(Rusinkiewicz and Levoy, 2001). Iteratively minimizing the difference between two point clouds however shows a high computational effort for large numbers of points. Hence, other registration approaches are based on information extracted from the point clouds. This information may for instance be derived from the distribution of the points within each point cloud by using the normal distributions transform (NDT) either on 2D scan slices (Brenner et al., 2008) or in 3D (Magnusson et al., 2007). If the presence of regular surfaces can be assumed in the local environment, various types of geometric features are likely to occur, e.g. planes, spheres and cylinders. These features can directly be extracted from the point clouds and strongly support the registration process (Brenner et al., 2008; Pathak et al., 2010; Rabbani et al., 2007). In cluttered scenes, descriptors representing local surface patches are more appropriate. Such descriptors may be derived from geometric curvature or normal vectors of the local surface (Bae and Lichti, 2008).

As the scans are acquired on a regular grid resulting from a cylindrical or spherical projection, the spatial 3D information can also be represented as range image. This range image provides additional features such as distinctive feature points which strongly support the registration process (Barnea and Filin, 2008; Steder et al., 2010).

Currently, most of the scanning devices can not only capture 3D information but also either co-registered camera images or panoramic reflectance images representing the respective energy of the backscattered laser light. The additional information typ-

ically is represented as intensity image. This intensity image might provide a higher level of distinctiveness than shape features (Seo et al., 2005) and thus information about the local environment which is not represented in the range measurements. Hence, the registration process can efficiently be supported by using reliable feature correspondences between the respective intensity images. Although different kinds of features can be used for this purpose, most of the current approaches are based on the use of feature points or keypoints as these tend to yield the most robust results for registration without assuming the presence of regular surfaces in the scene. Distinctive feature points simplify the detection of point correspondences and for this reason, SIFT features are commonly used. These features are extracted from the co-registered camera images (Al-Manasir and Fraser, 2006; Barnea and Filin, 2007) or from the reflectance images (Wang and Brenner, 2008; Kang et al., 2009). For all point correspondences, the respective 2D feature points are projected into 3D space using the spatial information. This yields a much smaller set of 3D points for the registration process and thus a much faster estimation of the transformation parameters between two point clouds. Furthermore, additional constraints considering the reliability of the point correspondences (Weinmann et al., 2011; Weinmann and Jutzi, 2011) allow for increasing the accuracy of the registration results.

Once 2D/2D correspondences are detected between images of different scans, the respective 3D/3D correspondences can be derived. Thus knowledge about the closest neighbor is available and the computationally expensive ICP algorithm can be replaced by a least squares adjustment. Least squares methods involving all points of a scan have been used for 3D surface matching (Gruen and Akca, 2005), but since a large overlap between the point clouds is required which can not always be assumed, typically sparse 3D point clouds consisting of a very small subset of points are derived from the original 3D point clouds (Al-Manasir and Fraser, 2006; Kang et al., 2009). To further exclude unreliable 3D/3D correspondences, filtering schemes based on the RANSAC algorithm (Fischler and Bolles, 1981) have been proposed in order to estimate the rigid transformation aligning two point clouds (Seo et al., 2005; Böhm and Becker, 2007; Barnea and Filin, 2007).

For dynamic environments, terrestrial laser scanners which perform a time-dependent spatial scanning of the scene are not suited. Furthermore, due to the background illumination, monitoring outdoor environments remains challenging with devices based on structured light such as the Microsoft Kinect device which uses random dot patterns of projected infrared points for getting reliable and dense close-range measurements in real-time. Hence, this paper is focused on airborne scene monitoring with range imaging devices mounted on a sensor platform. Although the captured point clouds are corrupted with noise and the field of view is very limited, a fast, but still reliable approach for point cloud registration is presented. The approach involves an initial camera calibration for increased accuracy of the respective 3D point clouds and the extraction of distinctive 2D features. The detection of 2D/2D correspondences between two successive frames and the subsequent projection of the respective 2D points into 3D space yields 3D/3D correspondences. Using such sparse point clouds significantly increases the performance of the registration process, but the influence of outliers has to be considered. Hence, a new weighting scheme derived from the respective point quality is introduced for adapting the influence of each 3D/3D correspondence on a weighted rigid transformation. Additionally, an extension of this approach is presented which is based on the already detected features and focuses on a decoupling of sensor and object motion.

The remainder of this paper is organized as follows. In Section 2, the proposed methodology for successive pairwise registration in dynamic environments is described as well as a simple extension for decoupling sensor and object motion. The configuration of the sensor platform is outlined in Section 3. Subsequently, the performance of the presented approach is tested in Section 4. The derived results are discussed in Section 5. Finally, in Section 6, the content of the entire paper is concluded and suggestions for future work are outlined.

2 METHODOLOGY

The proposed methodology provides fast algorithms which are essential for time-critical surveillance applications and should be capable for a real-time implementation on graphic processors. After data acquisition (Section 2.1), a preprocessing has to be carried out in order to get the respective 3D point cloud (Section 2.2). However, the point cloud is corrupted with noise and hence, a quality measure is calculated for each point of the point cloud (Section 2.3). Subsequently extracting distinctive features from 2D images allows for detecting reliable 2D/2D correspondences between different frames (Section 2.4), and projecting the respective 2D points into 3D space yields 3D/3D correspondences of which each 3D point is assigned a value for the respective point quality (Section 2.5). The point cloud registration is then carried out by estimating the rigid transformation between two sparse point clouds where the weights of the 3D/3D correspondences are derived from the point quality of the respective 3D points (Section 2.6). Finally, a feature-based method for object detection and segmentation is introduced (Section 2.7) which can be applied for decoupling sensor and object motion.

2.1 Data Acquisition

In contrast to the classical stereo observation techniques with passive sensors, where data from at least two different viewpoints has to be captured, the monostatic sensor configuration of the PMD[vision] CamCube 2.0 preserves information without the need of a co-registration of the captured data. A PMD[vision] CamCube 2.0 simultaneously captures various types of data, i.e. geometric and radiometric information, by images with a single shot. The images have a size of 204×204 pixels which corresponds to a field of view of $40^\circ \times 40^\circ$. Thus, the device provides measurements with an angular resolution of approximately 0.2° . For each pixel, three features are measured, namely the respective range R , the active intensity I_a and the passive intensity I_p . The active intensity depends on the illumination emitted by the sensor, whereas the passive intensity depends on the background illumination arising from the sun or other external light sources. As a single frame consisting of a range image \mathbf{I}_R , an active intensity image \mathbf{I}_a and a passive intensity image \mathbf{I}_p can be updated with high frame rates of more than 25 releases per second, this device is well-suited for capturing dynamic scenes.

2.2 Preprocessing

In a first step, the intensity information of each frame, i.e. \mathbf{I}_a and \mathbf{I}_p , has to be adapted. This is achieved by applying a histogram normalization of the form

$$I_n = \frac{I - I_{min}}{I_{max} - I_{min}} \cdot 255 \quad (1)$$

which adapts the intensity information I of each pixel to the interval $[0, 255]$. The modified frames thus consist of a normalized active intensity image $\mathbf{I}_{n,a}$, a normalized passive intensity image $\mathbf{I}_{n,p}$ and the range image \mathbf{I}_R which are illustrated in Figure 1.

For all subsequent tasks, it is essential to get the 3D information as accurate as possible. Due to radial lens distortion and decentring distortion, however, the image coordinates have to be adapted in order to be able to appropriately capture a scene. Hence, a camera calibration is carried out for the used devices. This yields a corrected grid of image coordinates with the principal point as origin of the new 2D coordinate frame. For each point $\mathbf{x} = (x, y)$ on the new grid, the respective 3D information in the local coordinate frame can then be derived from the measured range value R with

$$R = \sqrt{X^2 + Y^2 + Z^2} \quad (2)$$

and a substitution of X and Y with

$$X = \frac{x}{f_x} \cdot Z \quad \text{and} \quad Y = \frac{y}{f_y} \cdot Z \quad (3)$$

where f_x and f_y are the focal lengths in x - and y -direction. Solving for the depth Z along the optical axis yields

$$Z = \frac{R}{\sqrt{\left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2 + 1}} \quad (4)$$

and thus, the 3D point $\mathbf{X} = (X, Y, Z)$ corresponding to the 2D point $\mathbf{x} = (x, y)$ has been calculated. Consequently, the undistortion of the 2D grid and the projection of all points onto the new grid lead to the respective point cloud data.



Figure 1: Image representation of normalized active intensity, normalized passive intensity and range data.

2.3 Point Quality Assessment

For further calculations, it is feasible to derive a measure which describes the quality of each 3D point. Those points which arise from objects in the scene will probably provide a smooth surface, whereas points corresponding to the sky or points along edges of the objects might be very noisy. Hence, for each point on the regular 2D grid, the standard deviation σ of all range values within a 3×3 neighborhood is calculated and used as a measure describing the reliability of the range information of the center point. This yields a 2D confidence map according to which the influence of a special point on subsequent tasks can be weighted. For the example depicted in Figure 1, the corresponding confidence map is shown in Figure 2.

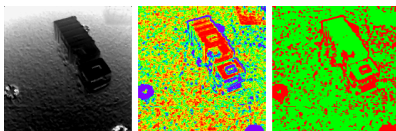


Figure 2: Range image, confidence map (pseudo-color representation where reliable points are marked in red and unreliable ones in blue) and thresholded confidence map (green: $\sigma \leq 0.05$ m).

2.4 2D Feature Extraction

As each frame consists of range and image data acquired on a regular grid, the alignment of two point clouds is based on using both

kinds of information. However, instead of using the whole 3D information available which results in a high computational effort, the intensity information is used to derive a much smaller set of 3D points. Hence, distinctive 2D features are extracted from the intensity information which later have to be projected into 3D space. For this purpose, the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is carried out on the normalized active intensity image as well as on the normalized passive intensity image. This yields distinctive keypoints and the respective local descriptors which are invariant to image scaling and image rotation, and robust with respect to image noise, changes in illumination and small changes in viewpoint. The vector representation of these descriptors allows for deriving correspondences between different images by considering the ratio

$$r = \frac{d(N_1)}{d(N_2)} \quad (5)$$

where $d(N_i)$ with $i = 1, 2$ denotes the Euclidean distance of a descriptor belonging to a keypoint in one image to the i -th nearest neighbor in the other image. This ratio $r \in [0, 1]$ describes the distinctiveness of a keypoint. Distinctive keypoints arise from low values and hence, the ratio r has to be below a certain threshold t_{des} . Typical values for this threshold are between 0.6 and 0.8. This procedure yields n_a correspondences between the normalized active intensity images of the two frames and n_p correspondences between the normalized passive intensity images. For the registration process, it is not necessary to distinguish between the two types of correspondences as only the spatial relations are of interest. Hence, a total number of $n = n_a + n_p$ correspondences is utilized for subsequent tasks.

2.5 Point Projection

In contrast to the measured range and intensity data which are only available on a regular grid, the location of SIFT features is determined with subpixel accuracy. Hence, an interpolation has to be carried out in order to obtain the respective 3D information as well as the respective range reliability. For this purpose, a bilinear interpolation is used. Assuming a total number of m SIFT features extracted from an image, this yields a set of samples s_i with $i = 1, \dots, m$ which are described by a 2D location \mathbf{x}_i , a 3D location \mathbf{X}_i and a quality measure σ_i . Compared to the original point cloud, the derived 3D points \mathbf{X}_i represent a much smaller point cloud where each point is assigned a quality measure σ_i .

Extending this on two frames with m_1 and m_2 SIFT features, between which $n \leq \min\{m_1, m_2\}$ correspondences have been detected, yields additional constraints. From the set of all n correspondences, it is now possible to derive subsets of

- 2D/2D correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ which can be used for image-based techniques, e.g. using the fundamental matrix (Hartley and Zisserman, 2008),
- 3D/3D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ which can be used for techniques based on the 3D geometry such as the ICP algorithm (Besl and McKay, 1992) and approaches estimating a rigid or non-rigid transformation, or
- 3D/2D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{x}'_i$ which can be used for hybrid techniques such as the methods presented in (Weinmann et al., 2011) and (Weinmann and Jutzi, 2011) which involve the EPnP algorithm (Moreno-Noguer et al., 2007).

The additional parameters σ_i can also be included for weighting the influence of each correspondence on any of the algorithms described above.

2.6 Point Cloud Registration

The spatial relation between two point clouds with n 3D/3D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ with $\mathbf{X}_i, \mathbf{X}'_i \in \mathbb{R}^3$ can formally be described as

$$\mathbf{X}'_i = \mathbf{R}\mathbf{X}_i + \mathbf{t} \quad (6)$$

where $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ represents a rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ represents a translation vector. A fully automatic estimation of the transformation parameters can be derived from minimizing the error between the point clouds. Including a weighting $w_i \in \mathbb{R}$ for each 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ yields an energy function E with

$$E = \sum_{i=1}^n w_i \|\mathbf{X}'_i - (\mathbf{R}\mathbf{X}_i + \mathbf{t})\|^2 \quad (7)$$

for the registration process. For minimizing this energy function E , the registration is carried out by estimating the rigid transformation from all 3D/3D correspondences and the weights are derived from a histogram-based approach. This approach is initialized by dividing the interval $[0m, 1m]$ into $n_b = 100$ bins of equal size. For all detected correspondences, the calculated quality measures σ_i and σ'_i assigned to the 3D points \mathbf{X}_i and \mathbf{X}'_i are mapped to the respective bins b_j and b'_j . Points with standard deviations greater than 1 m are mapped to the last bin. The occurrence of mappings to the different bins is stored in histograms $\mathbf{h} = [h_j]_{j=1, \dots, 100}$ and $\mathbf{h}' = [h'_j]_{j=1, \dots, 100}$. Subsequently, cumulative histograms

$$\mathbf{h}_c = \left[\sum_{j=1}^i h_j \right]_{i=1, \dots, 100} \quad \text{and} \quad \mathbf{h}'_c = \left[\sum_{j=1}^i h'_j \right]_{i=1, \dots, 100}$$

are derived. The entries of the cumulative histograms reach from 0 to the number n of detected correspondences. As points with a low standard deviation are more reliable, they should be assigned a higher weight. For this reason, the inverse cumulative histograms

$$\mathbf{h}_{c,inv} = \left[n - \sum_{j=1}^i h_j \right]_{i=1, \dots, 100} \quad (8)$$

and

$$\mathbf{h}'_{c,inv} = \left[n - \sum_{j=1}^i h'_j \right]_{i=1, \dots, 100} \quad (9)$$

are calculated. Finally, the weight w_i of a 3D/3D correspondence $\mathbf{X}_i \leftrightarrow \mathbf{X}'_i$ is set to

$$w_i = \min\{\mathbf{h}_{c,inv}(\sigma_i), \mathbf{h}'_{c,inv}(\sigma'_i)\} \quad (10)$$

where σ_i and σ'_i are considered as quality measures for the respective 3D points \mathbf{X}_i and \mathbf{X}'_i . Estimating the transformation parameters can thus be carried out for both range imaging devices separately. However, as the relative orientation between the devices is already known from a priori measurements and both devices are running synchronized, the rigid transformation can be estimated from the respective correspondences detected by both devices between successive frames. Combining information from both devices corresponds to extending the field of view and this yields more reliable results for the registration process. The extension can be expressed by transforming the projected 3D points \mathbf{X}_i which are related to the respective camera coordinate frame (superscript c) into the body frame (superscript b) of the sensor platform according to

$$\mathbf{X}_i^b = \mathbf{R}_c^b \cdot \mathbf{X}_i^c + \mathbf{t}_c^b \quad (11)$$

where \mathbf{R}_c^b describes the rotation and \mathbf{t}_c^b denotes the translation between the respective coordinate frames. For this, it is assumed that the origin of the body frame is in the center between both range imaging devices.

2.7 Object Detection and Segmentation

As 2D SIFT features have already been calculated for the registration process, they can also be utilized for detecting special objects in the scene. This allows for calculating the coarse area of an object and for automatically selecting features which should not be included in the registration process as they arise from objects which are likely to be dynamic. These features have to be treated in a different way as the static background being relevant for registration. For this purpose, image representations of several objects have to be stored in a database before starting the surveillance application. One of these images contains a template for the object present in the scene, but from a different measurement campaign at a different place and at a different season. Due to a similar altitude, the active intensity images show a very similar appearance. Comparing the detected SIFT features of the normalized active intensity image to the object templates in the database during the flight yields a maximum similarity to the correct template. Defining a spatial transformation based on the SIFT locations as control points, the template is transformed. The respective area of the transformed template is then assumed to cover the detected object. This procedure allows for detecting both static and moving objects in the scene as well as for decoupling sensor and object motion. Hence, the presented approach for registration also remains reliable in case of dynamic environments if representative objects are already known.

3 ACTIVE MULTI-VIEW RANGE IMAGING SYSTEMS

The proposed concept focuses on airborne scene monitoring with range imaging devices. For simulating a future operational system involving such range imaging devices fairly realistically, a scaled test scenario has been set up. However, due to the large payload of several kilograms for the whole system, mounting the required components for data acquisition and data storage on an unmanned aerial vehicle (UAV) still is impracticable. Hence, in order to investigate the potentials of active multi-view range imaging systems, a cable car moving along a rope is used as sensor platform which is shown in Figure 3. The components mounted on this platform consist of

- two range imaging devices (PMD[vision] CamCube 2.0) for recording the data,
- a notebook with a solid state hard disk for efficiently storing the recorded data and
- a 12 V battery with 6.5 Ah for independent power supply.

As the relative orientation of the two range imaging devices can easily be changed, the system allows for variable multi-view options with respect to parallel, convergent or divergent data acquisition geometries.

However, due to the relatively large influence of noise effects arising from the large amount of ambient radiation in comparison to the emitted radiation as well as from multipath scattering, the utilized devices only have a limited absolute range accuracy of a few centimeters and noisy point clouds can be expected. Furthermore, due to the measurement principle of such time-of-flight cameras, the non-ambiguous range R_n with

$$R_n = \frac{1}{2} \frac{c}{f_m} \quad (12)$$

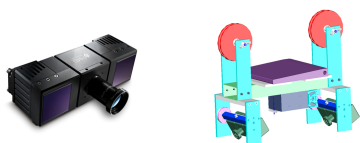


Figure 3: PMD[vision] CamCube 2.0 and model of a cable car equipped with two range imaging devices.

depends on the modulation frequency f_m , where c_0 denotes the speed of light. A modulation frequency of 20 MHz thus corresponds to a non-ambiguous range of 7.5 m. In order to overcome this range measurement restriction, image- or hardware-based unwrapping procedures have been introduced (Jutzi, 2009; Jutzi, 2012). When dealing with multiple range imaging devices, it also has to be taken into account that these may influence each other and that interferences are likely to occur. This can be overcome by choosing different modulation frequencies.

4 EXPERIMENTAL RESULTS

The estimation of the flight trajectory of a sensor platform requires the definition of a global world coordinate frame. This world coordinate frame is assumed to equal the local coordinate frame of the sensor platform at the beginning. The local coordinate frame has a fixed orientation with respect to the sensor platform. It is oriented with the X -direction in forward direction tangential to the rope, the Y -direction to the right and the Z -direction downwards. For evaluating the proposed methodology, a successive pairwise registration is performed. The threshold for the matching of 2D features is selected as $t_{des} = 0.7$. The resulting 2D/2D correspondences are projected into 3D space which yields 3D/3D correspondences. Including the weights in the estimation of the rigid transformation yields position estimates and, finally, an estimated trajectory which is shown in Figure 4 in nadir view and in Figure 5 from the side. The green and blue points describe thinned point clouds captured with both range imaging devices and transformed to the global world coordinate frame.

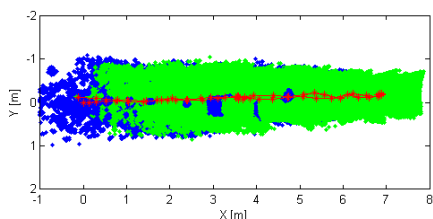


Figure 4: Projection of the estimated trajectory and thinned point cloud data onto the XY -plane.

A limitation of the experimental setup seems to be the fact that no reference values are available for checking the deviation of the position estimates from the real positions. However, due to the relative orientation of the sensor platform to the rope, the projection of the real trajectory onto the XY -plane should approximately be a straight line. Additionally, the length of the real trajectory projected onto the ground plane can be estimated from aerial images or simply be measured. Here, the distance Δ_{ground} between the projections of the end points onto the ground plane has been measured as well as the difference $\Delta_{altitude}$ between maximum and minimum altitude. From the measured values of $\Delta_{ground} = 7$ m and $\Delta_{altitude} = 1.25$ m, a total distance of

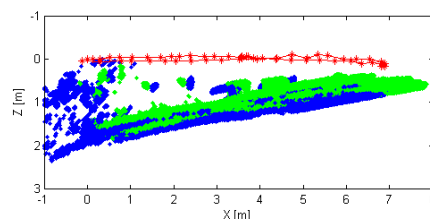


Figure 5: Projection of the estimated trajectory and thinned point cloud data onto the XZ -plane.

approximately 7.11 m can be assumed. A comparison between the start position and the point with the maximum distance on the estimated trajectory results in a distance of 6.90 m. As a consequence, the estimated trajectory can be assumed to be of relatively high quality. The results for a subsequent object detection and segmentation is illustrated for an example frame in Figure 6.

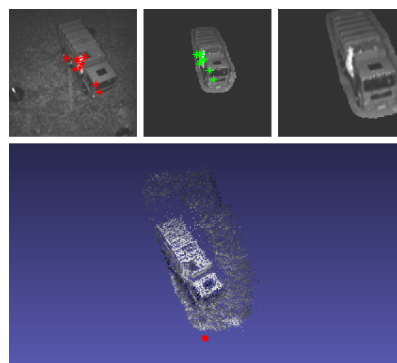


Figure 6: SIFT-based object detection and segmentation: normalized active intensity image, template and transformed template (upper row, from left to right). The corresponding point cloud for the area of the transformed template and the sensor position (red dot) are shown below.

5 DISCUSSION

The presented methodology is well-suited for dynamic environments. Instead of considering the whole point clouds, the problem of registration is reduced on sparse point clouds of physically almost identical 3D points. Due to this fact and the non-iterative processing scheme, the proposed algorithm for point cloud registration is very fast which is required for monitoring in such demanding environments. Although the current Matlab implementation is not fully optimized with respect to parallelization of tasks, a total time of approximately 1.63 s is required for pre-processing, point quality assessment, feature extraction and point projection. Further 0.46 s are required for feature matching, calculation of weights and point cloud registration. This can significantly be reduced with a GPU-implementation of SIFT, as the calculation of SIFT features already takes approximately 1.54 s.

Furthermore, the simple estimation of a rigid transformation is not sufficient, as used 3D/3D correspondences have the same weight, even if the uncertainty of the respective 3D points is very high or if outlier correspondences not fitting to the transformation have been detected. Hence, a quality measure for 3D/3D correspondences has been introduced which is based on the quality of the respective 3D points. This quality measure is used for

weighting the influence of each 3D/3D correspondence on the estimation of the rigid transformation. As most of the 3D points of a frame are assigned a higher quality, the introduced weights of 3D/3D correspondences with low quality are approximately 0. Consequently, the presented approach shows similar characteristics as a RANSAC-based approach, but it is faster and a deterministic solution for the transformation parameters is calculated.

6 CONCLUSIONS AND FUTURE WORK

In this paper, an experimental setup involving a moving sensor platform with multiple and coupled sensor devices for monitoring in low altitudes has been presented. For successive pairwise registration of the measured point clouds, a fast and reliable image-based approach has been presented which can also cope with dynamic environments. The concept is based on the extraction of distinctive 2D features from the image representation of measured intensity information and the projection into 3D space with respect to the measured range information. Detected 2D/2D correspondences between two frames, which have a high reliability, thus yield sparse 3D point clouds of 3D/3D correspondences. For increased robustness, the influence of each 3D/3D correspondence is weighted with a new measure derived from the quality of the respective 3D points. Finally, the point cloud registration is carried out by estimating the rigid transformation between two sparse point clouds which involves the calculated weights. As demonstrated, this approach can easily be extended towards using the already detected features for object detection and, even further, decoupling sensor and object motion which significantly improves the registration process in dynamic environments. The results indicate that the presented concept of active multi-view range imaging strongly supports navigation, point cloud registration and scene analysis.

The presented methodology can further be extended towards the detection, the segmentation and the recognition of multiple static or moving objects. Furthermore, a tracking method for estimating the trajectory of a moving object could be introduced as well as a model for further stabilizing the estimated trajectory of the sensor platform. Hence, active multi-view range imaging systems have a high potential for future research on dynamic scene analysis.

ACKNOWLEDGEMENTS

The authors would like to thank Michael Weinmann (Institute of Computer Science II, University of Bonn) for helpful discussions and Peter Runge (Geodetic Institute, Karlsruhe Institute of Technology) for constructing the sensor platform. Further thanks go to André Dittrich and Annette Schmitt for assistance during the measurement campaign.

REFERENCES

Al-Manasir, K. and Fraser, C. S., 2006. Registration of terrestrial laser scanner data using imagery. *The Photogrammetric Record* 21(115), pp. 255–268.

Bae, K.-H. and Lichti, D. D., 2008. A method for automated registration of unorganised point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63(1), pp. 36–54.

Barnea, S. and Filin, S., 2007. Registration of terrestrial laser scans via image based features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), pp. 32–37.

Barnea, S. and Filin, S., 2008. Keypoint based autonomous registration of terrestrial laser point-clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63(1), pp. 19–35.

Besl, P. J. and McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), pp. 239–256.

Böhm, J. and Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. *Optical 3-D Measurement Techniques VIII*, pp. 338–344.

Brenner, C., Dold, C. and Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 63(1), pp. 4–18.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), pp. 381–395.

Gruen, A. and Akca, D., 2005. Least squares 3D surface and curve matching. *ISPRS Journal of Photogrammetry and Remote Sensing* 59(3), pp. 151–174.

Hartley, R. I. and Zisserman, A., 2008. *Multiple view geometry in computer vision*. University Press, Cambridge.

Jutzi, B., 2009. Investigations on ambiguity unwrapping of range images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 3 / W8), pp. 265–270.

Jutzi, B., 2012. Extending the range measurement capabilities of modulated range imaging devices by time-frequency multiplexing. *AVN - Allgemeine Vermessungs-Nachrichten* 2 / 2012.

Kang, Z., Li, J., Zhang, L., Zhao, Q. and Zlatanova, S., 2009. Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. *Sensors* 9(4), pp. 2621–2646.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.

Magnusson, M., Lilienthal, A. and Duckett, T., 2007. Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics* 24(10), pp. 803–827.

Moreno-Noguer, F., Lepetit, V. and Fua, P., 2007. Accurate non-iterative $O(n)$ solution to the PnP problem. *IEEE 11th International Conference on Computer Vision*, pp. 1–8.

Pathak, K., Birk, A., Vaskevicius, N. and Poppinga, J., 2010. Fast registration based on noisy planes with unknown correspondences for 3-D mapping. *IEEE Transactions on Robotics* 26(3), pp. 424–441.

Rabbani, T., Dijkman, S., van den Heuvel, F. and Vosselman, G., 2007. An integrated approach for modelling and global registration of point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 61(6), pp. 355–370.

Rusinkiewicz, S. and Levoy, M., 2001. Efficient variants of the ICP algorithm. *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling*, pp. 145–152.

Seo, J. K., Sharp, G. C. and Lee, S. W., 2005. Range data registration using photometric features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, pp. 1140–1145.

Steder, B., Grisetti, G. and Burgard, W., 2010. Robust place recognition for 3D range data based on point features. *IEEE International Conference on Robotics and Automation*, pp. 1400–1405.

Wang, Z. and Brenner, C., 2008. Point based registration of terrestrial laser data using intensity and geometry features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (Part B5), pp. 583–589.

Weinmann, Ma. and Jutzi, B., 2011. Fully automatic image-based registration of unorganized TLS data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 5 / W12).

Weinmann, Ma., Weinmann, Mi., Hinz, S. and Jutzi, B., 2011. Fast and automatic image-based registration of TLS data. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(6), pp. S62–S70.

Z. Nicht referenzierte Publikationen des Verfassers

- [Z1] Braun AC, Weidner U, **Jutzi B**, Hinz S (2011) Integrating external knowledge into SVM classification - Fusing hyperspectral and laserscanning data by kernel composition. In: Heipke C, Jacobsen K, Rottensteiner F, Müller S, Sörgel U (Eds) High-resolution earth imaging for geospatial information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 4 / W19) (on CD)
- [Z2] Centeno J, **Jutzi B** (2010) Evaluation of a range imaging sensor concerning resolution and illumination. ISPRS TCI Symposium 2010. Symposium of ISPRS Commission I: Image Data Acquisition - Sensors & Platforms. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 1) (on CD)
- [Z3] Centeno J, **Jutzi B**, De Oliveira A (2010) Noise reduction for range imaging devices. In: Mayer M, Krueger CP, Heck B (Eds) Highly Precise Positioning and Height Determination using GPS. Karlsruhe: KIT Scientific Publishing, KIT Scientific Reports 7604: 53-57
- [Z4] De Oliveira A, Centeno J, **Jutzi B** (2010) Tratamento de ruído em imagens de distância obtidas com a câmara PMD[VISION]®CAMCUBE 2.0. SIMGEO - III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, Recife-PE, 27-30 de Julho de 2010: 001-007
- [Z5] Đuričić A, **Jutzi B** (2013) Supporting UAVs in low visibility conditions by multiple-pulse laser scanning devices. In: Heipke C, Jacobsen K, Rottensteiner F, Sörgel U (Eds) High-resolution earth imaging for geospatial information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W1, 2013: 93-98
- [Z6] Đuričić A, Weinmann M, **Jutzi B** (2013) Potentials of Small, Lightweight, and Low Cost Multi-Echo Laser Scanners for Detecting Grape Berries. In: Remondino F, Menna F (Eds) ISPRS Technical Commission V Symposium. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5, 2014: 211-216
- [Z7] Heipke C, Sörgel U, Rottensteiner F, **Jutzi B** (Eds) (2015) Theme Issue: High-Resolution Earth Imaging for Geospatial Information. ISPRS Journal of Photogrammetry & Remote Sensing 100 (2015): 1-128
- [Z8] Hinz S, **Jutzi B**, Kron A, Leitloff J, Musall M, Nestmann F, Brockmann H (2015) Ableitung von Flussvorlandrauheiten aus Laserdaten für die

hydrodynamisch-numerische Modellierung. Wasserstraßenbezogene geodätische Anwendungen und Produkte der Fernerkundung, BfG Schriftenreihe Veranstaltungen 2/2015: 28-34

- [Z9] **Jutzi B** (2010) Architectural applications - a challenge for active sensing. In: Amorim AL (Ed) Documentação do Patrimônio Arquitetônico Vol. 8, No 2 (2009)
- [Z10] **Jutzi B** (2011) Grundlagen der elektrooptischen Lasermesstechnik. Optronische Informationssysteme. In: Schmidtke H (Ed) Handbuch der Ergonomie (HdE). Zweite, überarbeitete und erweiterte Auflage. BWB (Hrsg), Koblenz (Nachf. auf CD-ROM)
- [Z11] **Jutzi B** (2011) Potenziale des Full-Waveform-Laserscanning zur Erfassung von Flussvorlandrauheiten. Zeitgemäße Erfassung und Bereitstellung von Geobasisdaten für die WSV, BfG Schriftenreihe Veranstaltungen 3/2011: 118-124
- [Z12] **Jutzi B**, Gabler R, Jäger K (2001) Stereo vision for small targets in IR image sequences. In: Drummond OE (Ed) Signal and data processing of small targets 2001. The International Symposium on Optical Science and Technology. The International Society for Optics and Photonics (SPIE) Proceedings. Vol. 4473: 361-370
- [Z13] **Jutzi B**, Neulist J, Stilla U (2005) High-Resolution waveform acquisition and analysis for pulsed laser. In: Heipke C, Jacobsen K, Gerke M (Eds) High-resolution Earth Imaging for Geospatial Information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 1 W3) (on CD)
- [Z14] **Jutzi B**, Stilla U (2003) Analysis of laser pulses for gaining surface features of urban objects. 2nd IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and data fusion on urban areas, URBAN 2003. IEEE: 13-17 [ISBN 0-7803-7719-2]
- [Z15] **Jutzi B**, Stilla U (2006) Characteristics of the measurement unit of a full-waveform laser system. Symposium of ISPRS Commission I: From Sensors to Imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 1/A) (on CD)
- [Z16] **Jutzi B**, Stilla U (2005) Erfassung und Analyse der zeitlichen Signalform bei gepulsten Lasersystemen. In: Luhmann T (Hrsg) Photogrammetrie - Laser-Scanning - Optische 3D-Messtechnik. Beiträge der Oldenburger 3D-Tage 2005. Herbert Wichmann Verlag, Heidelberg: 203-213
- [Z17] **Jutzi B**, Stilla U (2004) Extraction of features from objects in urban areas using space-time analysis of recorded laser pulses. In: Altan MO (Ed) XXth ISPRS Congress: Geo-Imagery Bridging Continents. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B2): 1-6

- [Z18] **Jutzi B**, Stilla U (2005) Measuring and processing the waveform of laser pulses. In: Gruen A, Kahmen H (Eds) Optical 3D Measurement Techniques VII. Vol. I: 194-203
- [Z19] **Jutzi B**, Stilla U (2005) Waveform analysis of laser pulses for gaining range accuracy. ISPRS WGI/2 Workshop on 3D Mapping from InSAR and LiDAR (on CD)
- [Z20] **Jutzi B**, Stilla U (2005) Waveform processing of laser pulses for reconstruction of surfaces in urban areas. In: Moeller M, Wentz E (Eds) 3rd International Symposium: Remote sensing and data fusion on urban areas, URBAN 2005. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 8 W27) (on CD)
- [Z21] **Jutzi B**, Thiele A, Meyer F, Hinz S (2010) Relations between SAR tomography and full-waveform LIDAR for structural analysis of forested areas. Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium: 3267-3270
- [Z22] Kron A, **Jutzi B**, Leitloff J, Musall M, Nestmann F, Hinz S, Brockmann H (2015) Ableitung von Vorlandrauheiten für Strömungssimulationen auf der Basis von Full-Waveform-Airborne-Laserscannerdaten (FW-ALS). Dresdner Wasserbauliche Mitteilungen - Messen und Überwachen im Wasserbau und am Gewässer. Heft 53: 161-171 [ISSN 0949-5061]
- [Z23] Stilla U, Rottensteiner F, Mayer H, **Jutzi B**, Butenuth M (Eds) (2011) Photogrammetric Image Analysis. Lecture Notes in Computer Science - LNCS 6952, Springer: Heidelberg doi:10.1007/978-3-642-24393-6
- [Z24] Stilla U, Rottensteiner F, Mayer H, **Jutzi B**, Butenuth M (Eds) (2011) Photogrammetric Image Analysis PIA11. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W22) (CD only)
- [Z25] Stilla U, Rottensteiner F, Mayer H, **Jutzi B**, Schmitt M (Eds) (2012) Special Issue: Photogrammetric Image Analysis. PFG Photogrammetrie - Fernerkundung - Geoinformation. Stuttgart: Schweizerbartsche Verlagsbuchhandlung 2012 (5)
- [Z26] Toth C, Holm T, **Jutzi B** (Eds) (2014) ISPRS Technical Commission I Symposium. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1, 2014
- [Z27] Toth C, **Jutzi B** (Eds) (2014) ISPRS Technical Commission I Symposium. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-1, 2014
- [Z28] Weinmann M, **Jutzi B**, Mallet C (2014) 3D Scene Analysis Based on Optimal Neighborhoods and Relevant Features. International Computer Vision Summer School 2014 - From Fundamentals to Applications: 76

- [Z29] Weinmann M, **Jutzi B**, Mallet C (2014) Describing Paris: Automated 3D Scene Analysis via Distinctive Low-Level Geometric Features. Proceedings of the IQmulus Workshop on Processing Large Geospatial Data: 1-8

Literaturverzeichnis

- [1] Bae KH, Lichti DD (2008) A method for automated registration of unorganised point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1): 36-54
- [2] Bai X, Zhou F, Xue B (2011) Fusion of infrared and visual images through region extraction by using multi scale center-surround top-hat transform. *Optics Express* 19 (9): 8444-8457
- [3] Barnea S, Filin S (2008) Keypoint based autonomous registration of terrestrial laser point-clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1): 19-35
- [4] Belton D, Lichti D (2006) Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (5): 44-49
- [5] Besl PJ, McKay ND (1992) A Method for Registration of 3-D Shapes. *IEEE Transaction on Pattern Analysis and Machine Intelligence* (Los Alamitos, CA, USA: IEEE Computer Society Press) 14 (2): 239-256. doi:10.1109/34.121791
- [D1] Blomley R, Weinmann M, Leitloff J, **Jutzi B** (2014) Shape Distribution Features for Point Cloud Analysis - A Geometric Histogram Approach on Multiple Scales. In: Schindler K, Paparoditis N (Eds) *ISPRS Technical Commission III Symposium*. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3*, 2014: 9-16
- [6] Böhm J, Becker S (2007) Automatic marker-free registration of terrestrial laser scans using reflectance features. In: Gruen A, Kahmen H (Eds) *Optical 3-D Measurement Techniques VIII*: 338-344
- [7] Bouguet JY (2010) Camera calibration toolbox for Matlab [online]. Computer Vision Research Group, Department of Electrical Engineering, California Institute of Technology, Pasadena, USA. Available from: http://www.vision.caltech.edu/bouguetj/calib_doc/ [Accessed 31 December 2014]
- [8] Boulch A, Houllier S, Marlet R, Tournaire O (2013) Semantizing complex 3d scenes using constrained attribute grammars. *Computer Graphics Forum* 32 (5): 33-42

- [B1] Bradley PE, **Jutzi B** (2011) Improved feature detection in fused intensity-range images with Complex SIFT (CSIFT). *Remote Sensing - Open Access Journal* 2011, 3 (9): 2076-2088 doi:10.3390/rs3092076
- [D2] Braun AC, Weidner U, **Jutzi B**, Hinz S (2012) Kernel Composition and the One-against-one cascade for integrating model knowledge into SVM classification. In: Heipke C, Jacobsen K, Rottensteiner F, Müller S, Sörgel U (Eds) *PFG Photogrammetrie - Fernerkundung - Geoinformation*. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung 2012 (4): 371-384
- [9] Breiman L (2001) Random forests. *Machine Learning* 45 (1): 5-32
- [10] Brenner C, Dold C, Ripperda N (2008) Coarse orientation of terrestrial laser scans in urban environments. *ISPRS Journal of Photogrammetry & Remote Sensing* 63 (1): 4-18
- [A1] Bretar F, Chauve A, Mallet C, **Jutzi B** (2008) Managing full waveform LiDAR data: A challenging task for the forthcoming years. In: Chen J, Jiang J, Baudoin A (Eds) *XXIth ISPRS Congress: Silk Road for Information from Imagery*. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (Part B1): 415-420
- [11] Brown DC (1971) Close-range camera calibration. *Photogrammetric Engineering* 37 (8): 855-866
- [12] Chehata N, Guo L, Mallet C (2009) Airborne lidar feature selection for urban classification using random forests. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38-3/W8: 207-212
- [13] Chen S, Leung H (2009) An EM-CI based approach to fusion of IR and visual images. *Proceedings of the 12th International Conference on Information Fusion*: 1325-1330
- [14] Cortes C, Vapnik V (1995) Support-vector networks. *Machine Learning* 20 (3): 273-297
- [15] Cover T, Hart P (1967) Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13 (1): 21-27
- [16] Dal Mutto C, Zanuttigh P, Cortelazzo GM (2012) *Time-of-Flight Cameras and Microsoft Kinect - A user perspective on technology and applications*. SpringerBriefs in Electrical and Computer Engineering, Springer ISBN 978-1-4614-3806-9
- [17] Demantké J, Mallet C, David N, Vallet B (2011) Dimensionality based scale selection in 3d lidar point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (5/W12): 97-102
- [18] Demantké J, Vallet B, Paparoditis N (2012) Streamed vertical rectangle detection in terrestrial laser scans for facade database production. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* Vol. I-3: 99-104

- [19] Dold C (2005) Extended Gaussian images for the registration of terrestrial scan data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*
- [20] Dold C, Brenner C (2004) Automatic matching of terrestrial scan data as a basis for the generation of detailed 3D city models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B3): 1091-1096
- [21] Dold C, Brenner C (2006) Registration of terrestrial laser scanning data using planar patches and image data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 5): 78-83
- [22] Droschel D, Holz D, Behnke S (2010) Probabilistic Phase Unwrapping for Time-of-Flight Cameras. In *Proceedings of the joint conference of the 41st International Symposium on Robotics (ISR 2010) and the 6th German Conference on Robotics (ROBOTIK 2010)*, Munich, Germany: 318-324
- [23] Durrant-Whyte H, Bailey T (2006) Simultaneous localization and mapping: Part I. *IEEE Robotics & Automation Magazine* 13 (2): 99-110 doi:10.1109/MRA.2006.1638022
- [24] Filin S, Pfeifer N (2005) Neighborhood systems for airborne laser data. *Photogrammetric Engineering & Remote Sensing* 71 (6): 743-755
- [25] Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24 (6): 381-395
- [26] Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Ann. Eugen* 7 (2): 179-188
- [27] Förstner W, Gülch E (1987) A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. *Proceedings of the ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data*: 281-305
- [28] Frey BJ, Koetter R, Petrovic N (2001) Very loopy belief propagation for unwrapping phase images. In *Advances in Neural Information Processing Systems* 14: 737-743
- [29] Fröhlich C (1996) *Aktive Erzeugung korrespondierender Tiefen- und Reflektivitätsbilder und ihre Nutzung zur Umgebungserfassung*. Technische Universität München, Dissertation
- [30] Ghiglia DC, Pritt MD (1998) *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. John Wiley & Sons: New York
- [31] Gini C (1912) *Variabilita e mutabilita*. *Memorie di metodologia statistica*
- [32] Goldstein RM, Zebker HA, Werner CL (1988) Satellite radar interferometry: two-dimensional phase unwrapping. *Radio Science* 23: 713-720

- [C1] Gross H, **Jutzi B**, Thoennesen U (2009) Classification of Elevation Data based on analytical versus trained Feature Values to determine Object Boundaries. In: Seyfert E (Ed) Geoinformatik und Erdbeobachtung: 29. Wissenschaftlich-Technische Jahrestagung der DGPF, 2009 (18): 315-326
- [A2] Gross H, **Jutzi B**, Thoennesen U (2008) Intensity normalization by incidence angle and range of full-waveform LiDAR data. Chen J, Jiang J, Nayak S (Eds) XXIth ISPRS Congress: Silk Road for Information from Imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (Part B4): 405-412
- [C2] Gross H, **Jutzi B**, Thoennesen U (2007) Segmentation of tree regions using data of a full-waveform laser. In: Stilla U, Mayer H, Rottensteiner F, Heipke C, Hinz S (Eds) Photogrammetric Image Analysis PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W49A): 57-62
- [33] Guo B, Huang X, Zhang F, Sohn G (2014) Classification of airborne laser scanning data using JointBoost. ISPRS Journal of Photogrammetry & Remote Sensing 92: 124-136
- [34] Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. Journal of Machine Learning Research 3 (2003): 1157-1182
- [35] Hall MA (1999) Correlation-based feature subset selection for machine learning. Ph.D. thesis, Department of Computer Science, University of Waikato, New Zealand
- [36] Hansard M, Lee S, Choi O, Horaud R (2013) Time-of-Flight Cameras: Principles, Methods and Applications. SpringerBriefs Series in Computer Science, Springer ISBN 978-1-4471-4658-2
- [37] Hanssen RF (2001) Radar Interferometry: Data Interpretation and Error Analysis. Kluwer Academic Publishers, Dordrecht
- [38] Hartley RI, Zisserman A (2008) Multiple view geometry in computer vision. University Press, Cambridge, UK
- [39] Hebel M (2012) Änderungsdetektion in urbanen Gebieten durch objektbasierte Analyse und schritthaltenden Vergleich von Multi-Aspekt ALS-Daten. Dissertation, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften (DGK), Reihe C, Nr. 690
- [40] Heikkilä J, Silvén O (1997) A four-step camera calibration procedure with implicit image correction. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 17-19 June, San Juan: 1106-1112
- [B2] Hinz S, Weinmann M, Runge P, **Jutzi B** (2011) Potentials of image based active ranging to capture dynamic scenes. In: Heipke C, Jacobsen K,

- Rottensteiner F, Müller S, Sörgel U (Eds) High-resolution earth imaging for geospatial information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 4 / W19) (on CD)
- [B3] Hoegner L, Hanel A, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2014) Towards People Detection from Fused Time-Of-Flight and Thermal Infrared Images. In: Schindler K, Paparoditis N (Eds) ISPRS Technical Commission III Symposium. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-3, 2014: 121-126
- [41] Hoegner L, Kumke H, Meng L, Stilla U (2007) Automatic extraction of textures from infrared image sequences and database integration for 3D building models. PFG Photogrammetrie, Fernerkundung, Geoinformation 2007 (6): 459-468
- [B4] Hoegner L, Roth L, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2013) Fusion von Time-of-Flight-Entfernungsdaten und thermalen IR-Bildern. AVN - Allgemeine Vermessungs-Nachrichten, Ausgabe 5/2014: 192-197
- [B5] Hoegner L, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2013) Co-registration of Time-of-Flight (TOF) camera generated 3d point clouds and thermal infrared images (IR). In: Seyfert E (Ed) 33. Wissenschaftlich-Technische Jahrestagung der DGPF, 2013 (22): 481-488
- [B6] Hoegner L, Weinmann M, **Jutzi B**, Hinz S, Stilla U (2012) Synchrone Koregistrierung von 3d Punktwolken und thermischen Infrarotbildern. In: Luhmann T (Hrsg) Optische 3D-Messtechnik - Photogrammetrie - Laser-Scanning. Beiträge der 12. Oldenburger 3D-Tage 2013. Herbert Wichmann Verlag, Heidelberg
- [42] Höfle B, Pfeifer N (2007) Correction of laser scanning intensity data: Data and model-driven approaches. ISPRS Journal of Photogrammetry & Remote Sensing 62 (6): 415-433
- [43] Iwaszczuk D, Hoegner L, Stilla U (2011) Detection of windows in IR building textures using masked correlation. In: Stilla U, Rottensteiner F, Mayer H, **Jutzi B**, Butenuth M (Eds), Photogrammetric Image Analysis, ISPRS Conference - Proceedings. Lecture Notes in Computer Science, Vol. 6952, Springer, Heidelberg, Germany: 133-146
- [44] John GH, Langley P (1995) Estimating continuous distributions in Bayesian classifiers. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence: 338-345
- [A3] **Jutzi B** (2007) Analyse der zeitlichen Signalform von rückgestreuten Laserpulsen. Dissertation, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften (DGK), Reihe C, Nr. 611
- [A4] **Jutzi B** (2010) Extending the range measurement capabilities of modulated range imaging devices by time-frequency-multiplexing. AVN - Allgemeine Vermessungs-Nachrichten, Ausgabe 2/2012: 54-62

- [A5] **Jutzi B** (2009) Investigations on ambiguity unwrapping of range images. In: Bretar F, Pierrot-Deseilligny M, Vosselman G (Eds) Laserscanning 2009. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 3 / W8): 265-270
- [A6] **Jutzi B**, Eberle B, Stilla U (2002) Estimation and measurement of backscattered signals from pulsed laser radar. In: Serpico SB (Ed) Image and signal processing for remote sensing VIII. The International Society for Optics and Photonics (SPIE) Proceedings. Vol. 4885: 256-267
- [A7] **Jutzi B**, Gross H (2010) Investigations on surface reflection models for intensity normalization in airborne laser scanning (ALS) data. In: Heipke C, Jacobsen K, Müller S, Sörgel U (Eds) Journal of Photogrammetric Engineering & Remote Sensing (PE&RS), Vol. 76, No. 9, September 2010: 1051-1060
- [C3] **Jutzi B**, Gross H (2009) Nearest neighbour classification on Laser point clouds to gain object structures from buildings. In: Heipke C, Jacobsen K, Müller S, Sörgel U (Eds) High-resolution Earth Imaging for Geospatial Information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 1-4-7/W5) (on CD)
- [A8] **Jutzi B**, Gross H (2009) Normalization of lidar intensity data based on range and surface incidence angle. In: Bretar F, Pierrot-Deseilligny M, Vosselman G (Eds) Laserscanning 2009. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 3 / W8): 213-218
- [D3] **Jutzi B**, Neulist J, Stilla U (2005) Sub-pixel edge localization based on laser waveform analysis. In: Vosselman G, Brenner C (Eds) Laserscanning 2005. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W19): 109-114
- [A9] **Jutzi B**, Stilla U (2007) Characteristics of the measurement unit of a full-waveform laser system. *Revue Française de Photogrammétrie et de Télédétection* 182 (2006-2): 17-22
- [A10] **Jutzi B**, Stilla U (2003) Laser pulse analysis for reconstruction and classification of urban objects. In: Ebner H, Heipke C, Mayer H, Pakzad K (Eds) Photogrammetric Image Analysis PIA03. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3 / W8): 151-156
- [A11] **Jutzi B**, Stilla U (2006) Precise range estimation on known surfaces by analysis of full-waveform laser. In: Förstner W, Steffen R (Eds) Symposium of ISPRS Commission III: Photogrammetric Computer Vision PCV06. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3): 234-239
- [A12] **Jutzi B**, Stilla U (2006) Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS Journal of Photogrammetry & Remote*

Sensing 61 (2): 95-107 doi:10.1016/j.isprsjprs.2006.09.001 U.V. Helava Award 2006

- [A13] **Jutzi B**, Stilla U (2007) Simulation and analysis of full-waveform laser data of urban objects. Remote sensing and data fusion on urban areas, URBAN 2007. IEEE 07EX1577 (on CD) [ISBN 1-4244-0712-5]
- [B7] **Jutzi B**, Weinmann M, Meidow J (2013) Improved UAV-borne 3D mapping by fusing optical and laserscanner data. In: Grenzdörffer G, Bill R (Eds) UAV-g2013. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W2, 2013: 223-228
- [B8] **Jutzi B**, Weinmann M, Meidow J (2013) Weighted Data Fusion for UAV-borne 3D Mapping with Camera and Line Laserscanner. International Journal of Image and Data Fusion 5 (3): 226-243 doi:10.1080/19479832.2014.889228
- [45] Kaasalainen S, Hyyppä J, Litkey P, Hyyppä H, Ahokas E, Kukko A, Kaartinen H (2007) Radiometric calibration of ALS intensity. In: Rönnholm P, Hyyppä H, Hyyppä J (Eds) Proceedings of Laserscanning 2007. International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences 36 (Part 3-W52): 201-205
- [46] Kaasalainen S, Krooks A, Kukko A, Kaartinen H (2009) Radiometric calibration of terrestrial laser scanners with external reference targets. Remote Sensing 2009 1 (3): 144-158 doi:10.3390/rs1030144
- [47] Kamerman GW (1993) Laser radar. In: Fox CS (Ed), Active Electro-Optical Systems, The Infrared & Electro-Optical Systems Handbook, SPIE Optical Engineering Press, Michigan
- [48] Kang Z, Li J, Zhang L, Zhao Q, Zlatanova S (2009) Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. Sensors 9 (4): 2621-2646
- [49] Kassir A, Peynot T (2010) Reliable automatic camera-laser calibration. Proceedings of the Australasian conference on robotics and automation (ACRA 2010), 1-3 December, Brisbane
- [50] Khoshelham K, Oude Elberink SJ (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: 372-377
- [D4] Kirchhof M, **Jutzi B**, Stilla U (2008) Iterative processing of laser scanning data by full waveform analysis in close neighborhood. In: Lichti D, Pfeifer N, Maas HG (Eds) ISPRS Journal of Photogrammetry & Remote Sensing 63 (1): 99-114 doi:10.1016/j.isprsjprs.2007.08.006

- [51] Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In: Proceedings of the European Conference on Machine Learning. Springer, Catania, Italy, 6-8 April: 171-182
- [52] Kukko A, Kaasalainen S, Litkey P (2007) Effect of incidence angle on laser scanner intensity and surface data. *Applied Optics* 47 (7): 986-992
- [53] Lafferty JD, McCallum A, Pereira FCN (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proceedings of the International Conference on Machine Learning: 282-289
- [54] Lagüela S, González-Jorge H, Armesto J, Arias P (2011) Calibration and verification of thermographic cameras for geometric measurements. *Infrared Physics & Technology* 54: 92-99
- [55] Lagüela S, Martínez J, Armesto J, Arias P (2011) Energy efficiency studies through 3D laser scanning and thermographic technologies. *Energy and Buildings* 43: 1216-1221
- [56] Lari Z, Habib A (2012) Alternative methodologies for estimation of local point density index: Moving towards adaptive lidar data processing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39 (Part B3): 127-132
- [57] Lee I, Schenk T (2002) Perceptual organization of 3d surface points. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (3A): 193-198
- [58] Linsen L, Prautzsch H (2001) Natural terrain classification using three-dimensional ladar data for ground robot mobility. *Proceedings of Eurographics*: 257-263
- [59] Lowe D (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60 (2): 91-110
- [60] Luhmann T, Ohm J, Piechel J, Roelfs T (2010) Geometric calibration of thermographic cameras. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38-5: 411-416
- [61] Mallet C, Bretar F (2009) Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry & Remote Sensing* 64 (1): 1-16
- [62] Mallet C, Bretar F, Roux M, Soergel U, Heipke C (2011) Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry & Remote Sensing* 66 (6): S71-S84
- [63] Mallet C, Bretar F, Soergel U (2008) Analysis of Full-Waveform Lidar Data for Classification of Urban areas. *Photogrammetrie - Fernerkundung - Geo-information (PFG)* 2008 (5): 337-349
- [64] Markov S, Birk A (2007) Detecting humans in 2D thermal images by generating 3D models. In: Hertzberg J, Beetz M, Englert R (Eds) *KI 2007: Advances*

- in Artificial Intelligence. Lecture Notes in Artificial Intelligence, Vol. 4667, Springer, Heidelberg, Germany: 293-307
- [65] Monnier F, Vallet B, Scheilian B (2012) Trees detection from laser point clouds acquired in dense urban areas by a mobile mapping system. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3 2012: 245-250
 - [66] Moreno-Noguer F, Lepetit V, Fua P (2007) Accurate non-iterative $O(n)$ solution to the PnP problem. Proceedings of the International Conference on Computer Vision: 1-8
 - [67] Munoz D, Bagnell JA, Vandapel N, Hebert M (2009) Contextual classification with functional max-margin Markov networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Miami, USA, 20-25 June: 975-982
 - [68] Niemeyer J, Rottensteiner F, Soergel U (2012) Conditional random fields for lidar point cloud classification in complex urban areas. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3 2012: 263-268
 - [69] Niemeyer J, Rottensteiner F, Soergel U (2014) Contextual classification of lidar data and building object detection in urban areas. ISPRS Journal of Photogrammetry & Remote Sensing 87: 152-65
 - [70] Osada R, Funkhouser T, Chazelle B, Dobkin D (2002) Shape distributions. ACM Transactions on Graphics 21(4): 807-832
 - [71] Pauly M, Keiser R, Gross M (2003) Multi-scale feature extraction on point-sampled surfaces. Computer Graphics Forum 22 (3): 281-289
 - [72] Pearson K (1896) Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. Philosophical Transactions of the Royal Society of London A 187: 253-318
 - [73] Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8): 1226-1238
 - [74] Pfeifer N, Dorninger P, Haring A, Fan H (2007) Investigating terrestrial laser scanning intensity data: Quality and functional relations. In: Gruen A, Kahmen H (Eds) Proceedings of the International Conference on Optical 3D Measurement Techniques VIII, Zürich, Switzerland, ISBN 3-906467-67-8: 328-337
 - [75] Phong BT (1975) Illumination for computer generated pictures. Communications of the ACM 18 (6): 311-317
 - [76] Pirotti F (2011) Analysis of full-waveform LiDAR data for forestry applications: A review of investigations and methods. Journal of Biogeosciences and Forestry 2011 (4): 100-106 doi:10.3832/ifor0562-004

- [77] Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical Recipes in C. Cambridge University Press, Cambridge, UK
- [78] Quinlan JR (1986) Induction of decision trees. *Machine Learning* 1: 81-106
- [79] Remondino F, Stoppa D (Eds) (2013) TOF Range-Imaging Cameras. Springer-Verlag Berlin Heidelberg doi:10.1007/978-3-642-27523-4
- [80] Rusinkiewicz S, Levoy M (2001) Efficient variants of the ICP algorithm. In: Proceedings of the Third International Conference on 3D Digital Imaging and Modeling (3DIM 2001), 28 May-1 June 2001, Quebec City, Canada. IEEE Computer Society, Los Alamitos, CA: 145-152
- [81] Rusu RB, Blodow N, Beetz M (2009) Fast point feature histograms (FPFH) for 3D registration. Proceedings of the IEEE International Conference on Robotics and Automation: 1848-1853
- [82] Schmidt A, Niemeyer J, Rottensteiner F, Soergel U (2014): Contextual Classification of Full Waveform Lidar Data in the Wadden Sea. *IEEE Geoscience and Remote Sensing Letters* 11(9): 1614-1618
- [83] Seo JK, Sharp GC, Lee SW (2005) Range data registration using photometric features. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2: 1140-1145
- [84] Shan J, Toth CK (Eds) (2008) Topographic Laser Ranging and Scanning: Principles and Processing. Boca Raton, FL: Taylor & Francis
- [85] Shannon CE (1948) A mathematical theory of communication. *The Bell System Technical Journal* 27 (3): 379-423
- [86] Steger C (2001) Similarity measures for occlusion, clutter, and illumination invariant object recognition. In: Radig B, Florczyk S (Eds) Pattern Recognition, DAGM 2001. Lecture Notes in Computer Science Vol. 2191, Springer: 148-154
- [87] Steinvall O (2000) Effects of Target Shape and Reflection on Laser Radar Cross Sections. *Applied Optics* 39 (24): 4381-4391
- [A14] Stilla U, **Jutzi B** (2008) Waveform analysis for small-footprint pulsed laser systems. In: Shan J, Toth CK (Eds) Topographic Laser Ranging and Scanning: Principles and Processing. CRC Press, Boca Raton: 215-234
- [A15] Stilla U, **Jutzi B**, Reitberger J, Yao W, Krzystek P (2009) Full Waveform Laserscanning - Auswertemethoden und Anwendungen. *Terrestrisches Laserscanning (TLS2009)*, Schriftenreihe des DVW, Band 60: 49-67 (eingeladener Beitrag)
- [D5] Stilla U, Yao W, **Jutzi B** (2007) Full waveform stacking of weak laser pulses by exploiting neighbourhood relation. In: Stilla U, Mayer H, Rottensteiner F, Heipke C, Hinz S (Eds) Photogrammetric Image Analysis PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W49A): 25-30

- [88] Tombari F, Salti S, Di Stefano L (2013) Performance evaluation of 3D key-point detectors. *International Journal of Computer Vision* 102: 198-220
- [89] Tombari F, Salti S, Di Stefano L (2010) Unique signatures of histograms for local surface description. In: Daniilidis K, Maragos P, Paragios N (Eds) *ECCV 2010 Part III. Lecture Notes in Computer Science Vol. 6313*, Springer, Heidelberg, Germany: 356-369
- [90] Ulrich M (2003) Hierarchical real-time recognition of compound objects in images. Dissertation, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften (DGK), Reihe C, Nr. 568
- [91] Vallet B, Xiao W, Brédif M (2015) Extracting mobile objects in images using a velodyne LiDAR point cloud. In: Stilla U, Heipke C (Eds) *Photogrammetric Image Analysis PIA15 + High-Resolution Earth Imaging for Geospatial Information HRIGI15. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W4, 2015: 247-253*
- [92] van Rijsbergen CJ (1979) *Information retrieval*. Butterworths, London
- [93] Vidas S, Moghadam P, Bosse M (2013) 3D thermal mapping of building interiors using an RGB-D and thermal camera. *Proceedings of the IEEE International Conference on Robotics and Automation: 2311-2318*
- [94] von Hansen W (2006) Robust automatic marker-free registration of terrestrial scan data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3): 105-110
- [95] Vosselman G, Maas HG (Eds) (2010) *Airborne and Terrestrial Laser Scanning*. Whittles Publishing, CRC Press
- [96] Wagner W, Ullrich A, Briese C (2003) Der Laserstrahl und seine Interaktion mit der Erdoberfläche. *Österreichische Zeitschrift für Vermessung & Geoinformation, VGI 4/2003: 223-235*
- [97] Waldhauser C, Hochreiter R, Otepka J, Pfeifer N, Ghuffar S, Korzeniowska K, Wagner G (2014) Automated classification of airborne laser scanning point clouds. In: Koziel S, Leifsson L, Yang XS (Eds) *Solving Computationally Expensive Engineering Problems: Methods and Applications*. Springer, New York, USA: 269-292
- [98] Wang Z, Brenner C (2008) Point based registration of terrestrial laser data using intensity and geometry features. *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences* 37 (Part B5): 583-589
- [99] Wanga C, Li Q, Liu Y, Wu G, Liu P, Ding X (2015) A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry. *ISPRS Journal of Photogrammetry & Remote Sensing* 101: 22-35
- [100] Weinmann M (2015) *Reconstruction and Analysis of 3D Scenes: From Irregularly Distributed 3D Points to Object Classes*. Dissertation, Springer

- [B9] Weinmann M, Dittrich A, Hinz S, **Jutzi B** (2013) Automatic feature-based point cloud registration for a moving sensor platform. In: Heipke C, Jacobsen K, Rottensteiner F, Sörgel U (Eds) High-resolution earth imaging for geospatial information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W1, 2013: 373-378
- [B10] Weinmann M, Hoegner L, Leitloff J, Stilla U, Hinz S, **Jutzi B** (2012) Fusing passive and active sensed images to gain infrared-textured 3D models. In: Shortis M, El-Sheimy N (Eds) XXII ISPRS Congress: Imaging a sustainable future. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 39 (B1): 71-76 doi:10.5194/isprsarchives-XXXIX-B1-71-2012
- [D6] Weinmann M, **Jutzi B** (2012) A step towards dynamic scene analysis with active multi-view range imaging systems. In: Shortis M, Paparoditis N, Mallet C (Eds) XXII ISPRS Congress: Imaging a sustainable future. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 39 (B3): 433-438 doi:10.5194/isprsarchives-XXXIX-B3-433-2012
- [B11] Weinmann M, **Jutzi B** (2013) Fast and Accurate Point Cloud Registration by Exploiting Inverse Cumulative Histograms (ICHs). Joint Urban Remote Sensing Event, JURSE 2013. IEEE CFP13RSD (on USB) [ISBN 978-1-4799-0212-5]: 218-221
- [B12] Weinmann M, **Jutzi B** (2011) Fully automatic image-based registration of unorganized TLS data. In: Lichti DD, Habib AF (Eds) Laserscanning 2011. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 5 / W12) (on CD)
- [C4] Weinmann M, **Jutzi B**, Hinz S, Mallet C (2015) Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry & Remote Sensing 105: 286-304 doi:10.1016/j.isprsjprs.2015.01.016
- [C6] Weinmann M, **Jutzi B**, Mallet C (2013) Feature relevance assessment for the semantic interpretation of 3D point cloud data. In: Scaioni M, Lindenberg RC, Oude Elberink S, Schneider D, Pirotti F (Eds) ISPRS Workshop Laserscanning 2013. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-5/W2, 2013: 313-318
- [C7] Weinmann M, **Jutzi B**, Mallet C (2014) Semantic 3D Scene Interpretation: A Framework Combining Optimal Neighborhood Size Selection with Relevant Features. In: Schindler K, Paparoditis N (Eds) ISPRS Technical Commission III Symposium. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3, 2014: 181-188 PCV 2014 Best Paper - Honourable Mention
- [B13] Weinmann M, Leitloff J, Hoegner L, **Jutzi B**, Stilla U, Hinz S (2014) Thermal 3D Mapping for Object Detection in Dynamic Scenes. In: Toth CK,

- Jutzi, B (Eds) ISPRS Technical Commission I Symposium. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-1, 2014: 53-60
- [C8] Weinmann M, Mallet C, Hinz S, **Jutzi B** (2015) Efficient interpretation of 3D point clouds by assessing feature relevance. AVN - Allgemeine Vermessungs-Nachrichten
- [C9] Weinmann M, Schmidt A, Mallet C, Hinz S, Rottensteiner F, **Jutzi B** (2015) Contextual classification of point cloud data by exploiting individual 3D neighborhoods. In: Stilla U, Heipke C (Eds) Photogrammetric Image Analysis PIA15 + High-Resolution Earth Imaging for Geospatial Information HRIGI15 - Joint ISPRS conference. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W4, 2015: 271-278
- [C5] Weinmann M, Urban S, Hinz S, **Jutzi B**, Mallet C (2015) Distinctive 2D and 3D Features for Automated Large-Scale Scene Analysis in Urban Areas. Special Section on Processing Large Geospatial Data. Computers & Graphics 49: 47-57 doi:10.1016/j.cag.2015.01.006
- [B14] Weinmann M(artin), Weinmann M(ichael), Hinz S, **Jutzi B** (2011) Fast and automatic image-based registration of TLS data. In: Bretar F, Wagner W, Paparoditis N (Eds) ISPRS Journal of Photogrammetry & Remote Sensing 66 (6): 62-70 doi:10.1016/j.isprsjprs.2011.09.010
- [B15] Weinmann M, Wursthorn S, **Jutzi B** (2011) Semi-automatic image-based co-registration of range imaging data with different characteristics. In: Stilla U, Rottensteiner F, Mayer H, Jutzi B, Butenuth M (Eds) Photogrammetric Image Analysis PIA11. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W22): 119-124
- [101] West KF, Webb BN, Lersch JR, Pothier S, Triscari JM, Iverson AE (2004) Context-driven automated target detection in 3d data. In: Sadjadi FA (Ed) Automatic Target Recognition XIV. Proceedings of SPIE Vol. 5426: 133-143
- [102] Wiener N (1949) Extrapolation, Interpolation, and Smoothing of Stationary Time Series. MIT Press, Cambridge, MA.
- [103] Yu L, Liu H (2003) Feature selection for high-dimensional data: a fast correlation-based filter solution. Proceedings of the International Conference on Machine Learning. AAAI Press, Washington, USA, 21-24 August: 856-863
- [104] Zhang Q, Pless R (2004) Extrinsic calibration of a camera and laser range finder (improves camera calibration). Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, 28 September-2 October, Sendai 3: 2301-2306

Danksagung

Die Habilitationsschrift entstand während meiner Zeit am Institut für Photogrammetrie und Fernerkundung (IPF) des Karlsruher Institut für Technologie (KIT).

Mein besonderer Dank gilt meinem Fachmentor Prof. Dr.-Ing. habil. Stefan Hinz, der mich bestärkte eine Habilitation anzugehen und mir stets vollstes Vertrauen bei meinen Forschungsaktivitäten entgegenbrachte. Ebenso bedanke ich mich bei Prof. Dr.-Ing. habil. Richard Bamler von der Technische Universität München (TUM) und Prof. Dr. sc. techn. habil. Hans-Gerd Maas von der Technische Universität Dresden für die Übernahme des Fachmentorats.

Zudem danke ich meinen Kollegen, insbesondere Jens Leitloff, Steffen Urban, Uwe Weidner, Martin Weinmann und Sven Wursthorn, die Diskussionen mit Ihnen waren und sind immer sehr anregend und bereichernd.

Allen Studierenden die mit ihren studentischen Arbeiten zu der Habilitation beigetragen haben spreche ich meinen Dank aus.

Ein besonderer Dank für die erfolgreiche Zusammenarbeit gilt meinen Koautoren: Rosmarie Blomley, Patrick Bradley, Andreas Braun, Frédéric Bretar, Herbert Brockmann, Jorge Centeno, Adrien Chauve, André Dittrich, Ana Đuričić, Bernd Eberle, Richard Gabler, Hermann Gross, Alexander Hanel, Stefan Hinz, Ludwig Hoegner, Aarne Hovi, Klaus Jäger, Michael Kirchhof, Ilkka Korpela, Andreas Kron, Peter Krzystek, Jens Leitloff, Clément Mallet, Jochen Meidow, Franz Meyer, Mark Mussall, Franz Nestmann, Jörg Neulist, Andrey Oliveira, Josef Reitberger, Lukas Roth, Franz Rottensteiner, Peter Runge, Alena Schmidt, Uwe Stilla, Antje Thiele, Ulrich Thoennessen, Charles Toth, Steffen Urban, Uwe Weidner, Martin Weinmann, Michael Weinmann, Sven Wursthorn und Wei Yao.

Meiner Familie, insbesondere meiner Frau Pamela, meinen Kindern Lewin und Quentin danke ich für die Unterstützung und das Verständnis, dass ich stets meinen persönlichen Zielen nachgehen kann.