Vorankündigung der gemeinsamen Tagung 2014

**Geoinformationen öffnen das
Tor zur Welt**

**62. Deutscher Kartographentag der DGfK
34. Wissenschaftlich-Technische Jahrestagung der DGPF
Geoinformatik 2014 der GFGi
Jahrestagung 2014 des GiN**

Gemeinsam veranstaltet mit der

HCU | **HafenCity Universität**
      Hamburg

26. – 28. März 2014

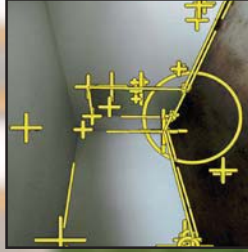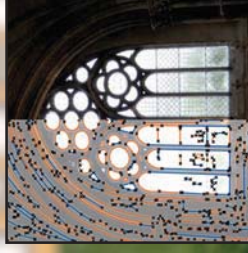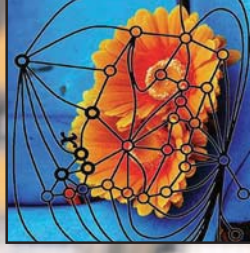| Termine:		
	30.9.2013	Beginn Teilnehmerregistrierung
	15.11.2013	Schlusstermin für Einreichung von Abstracts
	20.12.2013	Mitteilung an Autoren über Annahme oder Ablehnung der Beiträge
	15.1.2014	Ende Frühbucherrabatt
	1.2.2014	Annahmeschluss für Beiträge zur Übernahme in die Tagungs-DVD

Weitere Informationen siehe www.dgpf.de

Die Tagung wird organisiert durch

Geomatik
Tagungs-GmbH

Zum Titelbild

Die vier Kacheln auf dem Hintergrund des Hauptgebäudes der Universität Bonn (Photo von FRANK LUERWEG/Uni Bonn) zeigen Arbeiten am dortigen Institut für Photogrammetrie.

Links oben: Unbemannte Plattformen mit Kameras und anderen Sensoren ausgestattet ermöglichen die Anwendung photogrammetrischer Verfahren durch Nichtexperten. Die Autonomie solcher Fluggeräte ist jedoch noch Forschungsgebiet. Innerhalb der DFG-Forschergruppe 'Mapping on Demand' befasst sich das Institut mit der visuellen Ortung auf der Basis eines Kamerasystems mit Fisheye-Objektiven. In der Figur sieht man die geometrische Konfiguration eines Systems mit zwei omnidirektionalen Kameras zu zwei Zeitpunkten, die sowohl nahe Punkte wie Punkte im Unendlichen, also Raumrichtungen, beobachten und zur Eigenbewegungsbestimmung verwenden. Der Beitrag von JOHANNES SCHNEIDER behandelt die Ortung und die Kalibrierung eines solchen Kamerasystems.

Rechts oben: Verfahren zur automatischen Bildinterpretation sind entscheidend für die Effizienzsteigerung bei der Bereitstellung von Geodaten für Geoinformationssysteme. Die Entwicklung solcher Verfahren ist wegen der Komplexität der Objekte langsam. Die objektbezogene Klassifikation wird seit einigen Jahren durch so genannte graphische Modelle, speziell Markoff-Felder, unterstützt, die die Nachbarschaften zur Verbesserung der Klassifikation verwenden. In der Figur sieht man die Zentren der Objekte (nicht deren Grenzen) und ihre gegenseitigen Nachbarschaften. Der Beitrag von FALKO SCHINDLER zeigt effiziente Verfahren zur Partitionierung von Bildern und Punktwolken in Regionen. Der Beitrag von WOLFGANG FÖRSTNER diskutiert die Bedeutung graphischer Modelle für klassische und zukünftige Aufgaben in der Photogrammetrie.

Rechts unten: Die automatische Bildinterpretation künstlicher Objekte wie Gebäuden kann sich auf die Form der für sie typischen Grenzen stützen. Neben geraden Linien handelt es sich oft um Kreise, etwa bei Fenster- oder Torbögen, Konturen von Kugeln oder, wie etwa bei der Kontur von Zylindern, um Folgen von Geraden und Kreisbögen. Kreise erscheinen wegen der allgemeinen Orientierung der Kamera zum Objekt im Bild als Kegelschnitte, meist als Ellipsen. Eine zuverlässige und genaue Extraktion und Partitionierung von Bildkanten in eine Folge von Geraden- und Ellipsensegmenten (blau und orange im linken Bildteil), wie sie der Beitrag von SUSANNE WENZEL präsentiert, stellt daher einen wichtigen Schritt zur Vereinfachung der sich anschließenden Interpretation dar.

Links unten: Die Orientierung von Kameras als zentraler Aufgabe der Photogrammetrie soll unter Umständen in Umgebungen erfolgen, wo die Zahl der markanten Punkte außerordentlich gering ist, wie etwa in frisch zu beziehenden Wohnungen. Wenn man auf künstliche Signale verzichten und die Bildorientierung automatisieren will, ist es sinnvoll, nicht nur andere Objektmerkmale wie gerade Kanten und Regionen, sondern vor allem auch ihre gegenseitig räumliche Lage zu verwenden. Bei nicht zu stark in der Tiefe gegliederten Objekten kann man annehmen, dass Inzidenzen und Rechts-Links-Beziehungen invariant gegenüber der Orientierung der Kamera sind. Wie der Beitrag von TIMO DICKSCHEID zeigt, lassen sich Bildmerkmale von Eckpunkten (Kreuze mit ihrer Größe), Kanten (Linien) und Regionen (durch Kreise dargestellt) einerseits und ihre gegenseitigen Beziehung andererseits in einem Markoff-Feld, dessen Parameter aus Trainingsdaten lernbar sind, integrieren.

WOLFGANG FÖRSTNER, Bonn

PFG – Jahrgang 2013, Heft 4
Inhaltsverzeichnis

Editorial

Originalbeiträge

Beiträge aus Wissenschaft und Praxis

Mitteilungen

Zusammenfassungen der „Originalbeiträge" und der „Beiträge aus Wissenschaft und Praxis"
(deutsch und englisch) sind auch verfügbar unter www.dgpf.de/neu/pfg/ausgaben.htm

Editorial: Schwerpunktheft zur Photogrammetrie an der Universität Bonn

Auch dieses Heft der PFG spiegelt die Breite unseres Themenspektrums wider – Photogrammetrie, Fernerkundung und Geoinformation. Ich freue mich sehr, dass es diesmal gelungen ist, anlässlich der Pensionierung von Prof. Dr.-Ing. Dr. h.c. mult. Wolfgang Förstner ein Schwerpunktheft zu gestalten, das Förstners langjährige Arbeiten, vor allem in seiner Bonner Zeit, mit der Veröffentlichung von jüngst erreichten Ergebnissen attraktiv illustriert. Beteiligt haben sich viele seiner Schüler und er selbst. Ich danke herzlich für die Wahl der PFG zur Veröffentlichung der Artikel und für die gute Zusammenarbeit in der Vorbereitung des Heftes. Den inhaltlichen Schwerpunkt des Heftes wird Wolfgang Förstner anschließend selbst vorstellen.

In letzter Zeit besaßen fast alle Ausgaben der PFG ein Schwerpunktthema. Wir erinnern uns zum Beispiel an hochaufgelöste Satellitenbilder (PFG 4/12) und photogrammetrische Bildanalyse (PFG 5/12). Beide Hefte waren aus Workshops entstanden, die in Hannover und München stattgefunden hatten. Keines der Schwerpunkthefte enthielt aber ausschließlich Artikel zum Thema. Viele Leserinnen und Leser der PFG erwarten in jedem Heft ein möglichst breites Spektrum. In diesem Heft wird der vom Bonner Institut gesetzte Schwerpunkt „Photogrammetrische Forschung" durch drei Beiträge zur Fernerkundung und einen zur Geoinformation ergänzt.

Suchenwirth et al. haben einen neuen Ansatz zur Analyse von Fernerkundungsszenen mit Hilfe zusätzlicher geographischer Daten und des maschinellen Lernens entwickelt. Mit diesem Ansatz kann die Speicherkapazität von Auenwäldern für organischen Kohlenstoff großräumig bestimmt werden. Das Verfahren wird am Beispiel des österreichischen Nationalparks Donau-Auen unterhalb von Wien erprobt.

Gnyp et al. bilden eine deutsch-chinesische Autorengruppe. Sie entwickelten ein Verfahren zur Abschätzung der Biomasse auf der Basis von Reflexionsmessungen am Bestand. Dabei werden die Kombination verschiedener spektraler Bänder und mehrere Indizes miteinander verglichen und bewertet.

Bannehr et al. versuchen Oberflächen in städtischen Gebieten, vor allem Dächer, mit Hilfe der Kombination von Messdaten aus verschiedenen Fernerkundungssensoren genauer zu bestimmen als bisher möglich. Dabei kommen eine Thermalkamera, eine Hyperspektralkamera, ein Laserscanner und eine rgb-Kamera zum Einsatz. Für die Klassifizierung werden Support-Vector-Machines benutzt. Das Testgebiet ist die Innenstadt von Oldenburg.

Stoter et al. berichten schließlich über den aktuellen Stand des Niederländischen 3D-Geodatenstandards. Nach seiner Implementierung auf der Basis von CityGML sind nun die Spezifikation, Beispieldatensätze und Werkzeuge zur Daten-Validierung verfügbar und werden vorgestellt. Im nächsten Entwicklungsschritt soll das Verfahren zur Laufendhaltung definiert werden.

Wolfgang Kresse

www.schweizerbart.de

Photogrammetrische Forschung – Eine Zwischenbilanz aus Bonner Sicht

Wolfgang Förstner, Bonn

*Die Schwierigkeit liegt nicht darin,
die neuen Ideen zu finden,
sondern die alten loszuwerden.*
John M. Kaynes

So wie ein runder Geburtstag im Jahre 2000 Anlass für einen Zwischenbericht in dieser Zeitschrift über die Forschung nach 10 Jahren Leitung des Instituts für Photogrammetrie war, so gibt mir meine Pensionierung die Gelegenheit, die Entwicklung der letzten Jahre, wie sie sich aus meiner Sicht darstellt, zu skizzieren, sie mit repräsentativen Arbeiten aus meiner Arbeitsgruppe in diesem Heft vorzustellen. Selbst bei dieser Beschränkung lässt sich die Dynamik in unserem in Breite und Tiefe wachsenden Forschungsfeld beleuchten.

Während wir mit einer Promotion zur strukturerhaltenden Generalisierung von Digitalen Geländemodellen (Braunmandl 2002) innerhalb des Sonderforschungsbereichs 350 ,*Wechselwirkungen kontinentaler Stoffsysteme und ihre Modellierung*' unsere Arbeiten zur Qualität von Geodaten abschlossen, hat uns unser zentrales Aufgabengebiet, die Automation der dreidimensionalen Gebäudeerfassung, bis heute nicht losgelassen. In den 90er-Jahren stützten sich unsere Arbeiten noch auf starke geometrische Modelle. Bei ihnen ergab sich die dreidimensionale Form durch Aggregation von Gebäudeteilen. Sie waren daher sowohl für die automatische wie die manuelle Erfassung geeignet und erforderten daher zur empirischen Prüfung eigene Konzepte (Ragia 2001). Sie konnten jedoch die große Vielfalt realer Objekte nicht beschreiben und verlangten einen neuen Zugang zum Problem der Gebäudemodellierung. Dieser sollte und musste die alten Konzepte grundsätzlich ablösen, ein in seinen Folgen bis heute nicht abgeschlossener Schritt.

Zunächst reduzierten wir die Rigidität der Modelle, die sich durch die Festlegung auf manuell erstellte Gebäudeteile ergab und fokussierten auf polyederförmige Oberflächenbeschreibungen. Schindler & Förstner (2011), durch die Rekonstruktion von Polyedern aus Punktwolken motiviert, stellen in diesem Heft ein flexibles Segmentierungsverfahren vor, das geometrische und radiometrische Information integrieren kann. Die Zuwendung zu Polyedern als elementares Oberflächenmodell von Gebäuden und seiner Erscheinungsform in Bildern erforderte eine fundierte Modellierung der Unsicherheit der mit projektiver Geometrie beschreibbaren geometrischen Grundelemente und ihrer Transformationen. Grundlegend dazu waren die Arbeiten von Heuel (2004) zum unsicheren geometrischen Schlussfolgern mit der erforderlichen Erweiterung des Kalmanfilter-Updates bei impliziten Funktionen, siehe der Beitrag von Steffen. Appel (2004) nutzte die projektive Geometrie zur Integration vom Bildern und Plänen. Die Arbeit von Beder (2006) zur Gruppierung unsicherer geometrischer Elemente zum Zwecke der Gebäuderekonstruktion zeigt auch die Limitierungen rein geometrischer Verfahren zur Rekonstruktion von Polyedern, legt aber auch die Basis für allgemeinere Gruppierungsaufgaben, siehe der Beitrag von Wenzel.

Die Flexibilisierung der Gebäudemodelle erlaubte jedoch kaum eine Interpretation der Bilddaten, in dem Sinne, dass die resultierenden Objekte klassifiziert und ihre gegenseitigen Beziehungen identifiziert wurden. Mitentscheidend für die Arbeitsrichtung in den

letzten Jahren war die Frage eines Gutachters bei der Beurteilung unserer Arbeiten zur Gebäudeerfassung im Jahr 1995: „Wenn Sie die Vielfalt der Formen und Strukturen realer Gebäude bei der automatischen Bildauswertung nutzen wollen, können Sie nicht alle auftretenden Formen manuell modellieren. M.E. benötigen Sie dazu Methoden des maschinellen Lernens." Wir mussten beginnen zu lernen wie Rechner statistische Modelle lernen können.

Mit dem von der EU geförderten Projekt *E-Training for Interpreting Images of Man-Made Scenes* (eTRIMS) begannen wir 2006 zusammen mit Kollegen aus der Computer Vision (Radim Sara, CMP, Prag), der Mustererkennung (Maria Petrou, UCL, London) und der Künstlichen Intelligenz (Bernd Neumann, KOGS, Hamburg) am Problem der Interpretation von Fassadenbildern das Problem der Bildinterpretation neu aufzugreifen. Zentral in diesem Projekt war eine Steuerung der Interpretation durch ein Fassadenmodell, das Partonomien, Taxonomien und relationale Modelle zwischen den Fassadenteilen enthielt und mit den Bildregionen interagierte. Wir befassten uns in diesem Projekt mit so genannten graphischen Modellen, auf Graphen gestützte stochastische Modelle für Objektteile und ihre Beziehungen, als Schnittstelle zwischen dem Fassadenmodell und den Bildregionen. Hierbei ging es um die simultane Klassifikation vieler, hierarchisch angeordneter Bildregionen, was durch bedingte Bayesnetze oder bedingte Markoff-Zufallsfelder realisierbar ist (Drauschke 2011, Yang 2011). Dazu kamen statistisch begründete Verfahren zur Schätzung, d.h. Lernen der freien Parameter des Klassifikationsmodells: Korč (2012) zeigt in seiner Arbeit, dass diese Parameterschätzung auf ein konvexes Optimierungsproblem führt, wenn man ein so genanntes logistisches Modell für die Bestimmung der a posteriori Wahrscheinlichkeiten verwendet. Dieses Klassifikationsmodell hat sich als außerordentlich flexibel herausgestellt, da es bei gleicher Qualität wie Standardverfahren auf natürliche Weise mehrere Klassen verarbeiten kann, zuverlässige a posteriori Wahrscheinlichkeiten liefert und sich wirksam zu einem inkrementelles Lernverfahren verallgemei-

nern lässt, wie die Arbeit von Roscher (2012) zeigt.

Graphische Modelle haben ein außerordentlich breites Anwendungsfeld, nicht nur in der Geodäsie und Photogrammetrie, siehe der Beitrag von Förstner. Sie haben daher einige unser Arbeiten beeinflusst: bereits sehr früh zur Klassifikation von Laserdaten (Brunn 2000), später zur Klassifikation des Gesundheitszustands von Pflanzen (Bauer 2011). Dickscheid (2010) modelliert das Problem der Zuordnung von Bildpunkten und -kanten über ein Markoff-Zufallsfeld, und kann so auch Nachbarschaftsrelationen unter den Bildmerkmalen integrieren, siehe seinen Beitrag in diesem Heft.

Parallel zu diesen Arbeiten konnten wir, u.a. in Industriekooperationen unsere Arbeiten zur Bildfolgenanalyse fortsetzen. Dabei ging es um die Echtzeitauswertung von Stereobildfolgen für Fahrerassistenzsysteme (Barth 2010, Siegemund 2013). Die Verfügbarkeit leichter unbemannter Flugsysteme ist nicht nur hoch attraktiv in der Lehre, sondern stellt wegen der Gewichtsrestriktionen besondere Anforderungen an die Auswertetechnik bei der Bestimmung der Eigenbewegung aus Bildfolgen einzelner Videokameras (Steffen 2009). Im Jahr 2011 konnten wir eine DFG-Forschergruppe zum Thema Mapping on Demand einwerben. Zusammen mit der Robotik (Behnke, Bonn), der Computer Graphik (Klein, Bonn), der Computer Vision (Cremers, München), der Ingenieurgeodäsie (Kuhlmann, Bonn) und der Geoinformation (Plümer, Bonn) adressieren wir das Problem der autonomen Navigation eines mit Kameras, Ultraschallsensoren, Laserabtaster und GPS/INS ausgestatteten fliegenden Roboters und die durch den Benutzer beauftragte Erfassung der Umgebung. In diesem Projekt sind wir für die visuelle Ortung aus synchron aufgenommenen Bildfolgen von Gruppen von Fisheye-Kameras zuständig, siehe der Beitrag von Schneider. Damit rückt eines der klassischen photogrammetrischen Probleme, das der Kamerasystemkalibrierung (Mitschke 2002) bzw. der Approximation der Aufnahmegeometrie durch ein Zentralprojektionsmodell (Wolff 2006) erneut in den Vordergrund.

Wir freuen uns über die Anerkennung, die wir durch den DAGM-Preis 2001 und 2010,

den Commerzbank-Dissertationspreis 2005, den best-paper-award bei der Konferenz Photogrammetric Computer Vision 2011 und den DGPF-Preis 2012 erfahren konnten, siehe Luxen & Förstner (2001), Barth et al. (2010), Appel (2004), Schindler & Förstner (2011) und Schneider (2011).

Der Lösung der 1990 avisierten Aufgabe der automatischen Gebäudeerfassung sind wir bis heute einige Schritte näher gekommen. Was ich damals nicht sah, sind die heute bei weitem nicht ausgeschöpften Methoden und Möglichkeiten, insbesondere der Mustererkennung und des maschinellen Lernens, vor allem graphische Modelle und Grammatiken, um der Komplexität der Gebäudeerfassung Herr zu werden. Ebenso wenig war abzusehen, dass der damals als Kern der Photogrammetrie bezeichnete und zwischenzeitlich durch die Bildinterpretation scheinbar verdrängte Bereich der Kalibrierung und Orientierung durch autonome fahrende und fliegende Plattformen und durch die außerordentliche Vielzahl von Nicht-Standard-Kameramodellen Anforderungen an die Abbildungskonzepte, ihre Integration in Echtzeitsysteme und ihre Vermittlung in der Lehre stellt. Die Photogrammetrische Forschung integriert zunehmend messtechnische und interpretatorische Fragestellungen, bindet so Themen der Ingenieurgeodäsie und der Fernerkundung ein und kann daher auf eine spannende Zukunft blicken.

Ich möchte mich an dieser Stelle bei allen, Forschungsförderern, Firmen, der Universität und der Fakultät, die die Arbeit an unserem Institut finanziell ermöglicht haben, bedanken. Der die ganzen Jahre kooperative und unterstützende Geist in der Bonner Fachgruppe Geodäsie und Geoinformation hat mir die Arbeit leicht gemacht und mich getragen. Mein besonderer Dank gilt meinen Mitarbeitern. Sie haben nicht nur Hervorragendes geleistet, vielmehr haben sie, mit zunehmendem Altersunterschied zu mir, mehr und mehr die Innovation am Institut getragen – was könnte hoffnungsvoller stimmen, wenn man die auf uns

zukommenden lohnenden Herausforderungen geodätisch-photogrammetrischer Forschung betrachtet.

Ich wünsche dem Leser eine interessante Lektüre.

Wer seine Ziele erreicht,
hat sie zu niedrig gewählt.
Herbert von Karajan

References

Appel, M., 2004: From Images and Technical Drawings to 3D Models: A Novel Approach to As-Built Reconstruction. – PhD thesis, Institute of Photogrammetry, University of Bonn, ibidem-Verlag, Stuttgart.

Barth, A., 2010: Vehicle Tracking and Motion Estimation Based on Stereo Vision Sequences. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.unibonn.de/2010/2356/2356.htm.

Barth, A., Siegemund, J., Meissner, A., Franke, U. & Förstner, W., 2010: Probabilistic Multi-Class Scene Flow Segmentation for Traffic Scenes. – Goesele, M., Roth, S., Kuijper, A., Schiele, B. & Schindler, K. (eds.): Pattern Recognition (Symposium of DAGM) '10, Springer.

Bauer, S.D., 2011: Automatische Detektion von Krankheiten auf Blättern von Nutzpflanzen. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.unibonn.de/2011/2744/2744.htm.

Beder, C., 2006: Gruppierung unsicherer orientierter projektiver geometrischer Elemente mit Anwendung in der automatischen Gebäuderekonstruktion. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.unibonn.de/2007/0935/0935.htm.

Braunmandl, A., 2002: Geometrische Generalisierung von Digitalen Höhenmodellen. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.unibonn.de/2002/0280/0280.htm.

Brunn, A., 2000: Semantik-basierte Gebäudeerfassung mit verkoppelten Markoff-Zufallsfeldern. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.uni-bonn.de/2000/0139/0139.htm.

Dickscheid, T., 2010: Robust Wide-Baseline Stereo Matching for Sparsely Textured Scenes. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.unibonn.de/2011/2603/2603.htm.

Drauschke, M., 2011: Ein hierarchischer Ansatz zur Interpretation von Gebäudeaufnahmen. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.unibonn. de/2012/2871/2871.htm.

Heuel, S., 2004: Statistical Reasoning in Uncertain Projective Geometry for Polyhedral Object Reconstruction. – LNCS, Springer.

Korč, F., 2012: Tractable Learning for a Class of Global Discriminative Models for Context Sensitive Image Interpretation. – PhD thesis, Department of Photogrammetry, University of Bonn, http://hss.ulb.uni-bonn. de/2012/3010/3010.htm.

Luxen, M. & Förstner, W., 2001: Optimal Camera Orientation from Points and Straight Lines. – Radig, B. & Florczyk, S. (eds.): Mustererkennung 2001, 23. DAGM-Symposium, LNCS **2191:** 84–91, Springer, München.

Mitschke, M., 2002: Methods for Geometric Calibration of X-ray Imaging Systems. – PhD thesis, Institute of Photogrammetry, University of Bonn, ibidem-Verlag, Stuttgart.

Ragia, L., 2001: Ein Modell für die Qualität räumlicher Daten zur Bewertung der photogrammetrischen Gebäudeerfassung. – PhD thesis, Institute of Photogrammetry, University of Bonn. GMD Series **14**.

Roscher, R., 2012: Sequential Learning using Incremental Import Vector Machines for Semantic Segmentation. – PhD thesis, University Bonn.

Schneider, J., 2011: Lösung von Orientierungsaufgaben der Photogrammetrie mit konvexer Optimierung. – Master thesis, University of Bonn, http://hss.ulb.uni-bonn.de/2012/3009/3009.htm.

Schindler, F. & Förstner, W., 2011: Fast marching for robust surface segmentation. – Photogrammetric Image Analysis, LNCS **6952:** 147–158.

Schneider, J., 2011: Lösung von Orientierungsaufgaben der Photogrammetrie mit konvexer Optimierung, Master thesis. University of Bonn.

Siegemund, J., 2013: Street Surfaces and Boundaries from Depth Image Sequences Using Probabilistic Models. – PhD thesis. University Bonn. submitted.

Steffen, R., 2009: Visual SLAM from Image Sequences Acquired by Unmanned Aerial Vehicles. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.unibonn.de/ 2009/1971/1971.htm.

Wolff, K., 2006: Zur Approximation allgemeiner optischer Abbildungsmodelle und deren Anwendung auf eine geometrisch basierte Mehrbildzuordnung am Beispiel einer Mehrmedienabbildung. – PhD thesis, Institute of Photogrammetry, University of Bonn, http://hss.ulb.uni-bonn. de/2007/1099/1099.htm.

Yang, M.Y., 2011: Hierarchical and Spatial Structures for Interpreting Images of Man-made Scenes Using Graphical Models. – PhD thesis, Institute of Photogrammetry, University of Bonn, http:// hss.ulb.uni-bonn.de/2012/2765/2765.htm.

Address of the Author:

Prof. Dr.-Ing. Dr. h. c. mult. Wolfgang Förstner, Josef-Schell-Str. 34, D-53121 Bonn, e-mail: wf@ ipb.uni-bonn.de

Article

Graphical Models in Geodesy and Photogrammetry

WOLFGANG FÖRSTNER, Bonn

Summary: The paper gives an introduction into graphical models and their use in specifying stochastic models in geodesy and photogrammetry. Basic task in adjustment theory can intuitively be described and analysed using graphical models. The paper shows that geodetic networks and bundle adjustments can be interpreted as graphical models, both as Bayesian networks or as conditional random fields. Especially hidden Markov random fields and conditional random fields are demonstrated to be versatile models for parameter estimation and classification.

Zusammenfassung: *Graphische Modelle in Geodäsie und Photogrammetrie*. Der Beitrag gibt eine Einführung in Graphische Modelle und ihren Einsatz zur Erstellung probabilistischer Modelle in der Geodäsie und der Photogrammetrie. Grundaufgaben der Ausgleichungsrechnung lassen sich intuitiv beschreiben und analysieren. Der Beitrag zeigt, wie geodätische/photogrammetrische Netze als Bayesnetze oder Markoff-Zufallsfelder interpretiert werden können. Besonders bedingte Zufallsfelder erweisen sich als flexibel für die Modellierung und die Optimierung von Parameterschätz- und Klassifikationsaufgaben.

1 Introduction

For more than 50 years geodetic and photogrammetric networks are classical tools for point positioning and orientation determination. They are characterized by a sparse link between observed data, image coordinates, distances, angles, height differences, or GPS-coordinates, on one hand and unknown parameters, mostly coordinates but also orientation parameters or additional parameters on the other hand, to capture various systematic effects. The work horse for determining the unknown parameters in a statistically optimal manner is the classical adjustment theory, including its Bayesian variants which allows for including prior information or sequential estimation.

In the last decades so-called graphical models, especially Markov random fields, have found their way into photogrammetric research, mainly for image interpretation. In contrast to simple pixel-wise or image region-wise classifiers these models allow for statistical modelling the spatial neighbourhood between pixels or image regions, and lead to an increase in classification performance.

The tools for finding optimal classifications based on these graphical models are in no way related to methods for solving large equation systems. Even more, in most cases only approximate solutions can be found and the statistical properties of the results are difficult to characterize.

Therefore the question arises: How is the relation between the current methodology and the new one? Do they address different problems? Is there an overlap? Is the current methodology a special case of the new one? The answer is clear: Graphical models are a real generalization of the well-known tools from statistics and adjustment theory, in the way they are trained and used in geodesy and photogrammetry. If a new methodology comes up, which claims to be a true generalization of the current one, it has to prove (a) that the current methods can be derived from the new ones by specialization and (b) that there is a relevant potential for successfully solving new problems, which cannot

be tackled by the current methodology. The paper mainly addresses the first point. The second point is sketched and proven in some of this issue's articles.

Graphical models are probabilistic models on graphs, the nodes representing single or aggregated random variables, and the edges representing probabilistic relationships between the random variables. In case these relations are directed, the networks fall into the category of Bayesian networks, otherwise they are so-called Markov random fields. In the above mentioned application of interpreting images, the random variables in the graphical models are discrete, explaining the conceptually different procedures for finding optimal solutions. However, graphical models are not restricted to model and analyse situations with discrete random variables, i.e. classification tasks, but also can handle problems with continuous variables. A prominent example is the classical Kalman filter, see Fig. 1. It is a graphical model. It con-
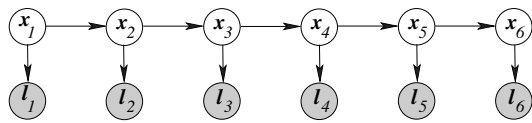
Fig. 1: A hidden Markov-chain as a special continuous Bayesian network used as basic model in Kalman filtering. The joint probability of all nodes can be factorized into $p(\boldsymbol{l}, \boldsymbol{x}) = p(x_1) \prod_{t=2}^{6} p(x_t|x_{t-1}) \prod_{t_1}^{6} p(l_t|x_t)$, each factor depending only on one or two variables. This eases learning and reasoning. The naming *Markov-chain* indicates, that (1) the time series follows the Markov-property, i.e. future states do only depend on the present, not on past states, and (2) the structure of the underlying graph of the unknown states (white) is a chain, i.e. only has nodes with maximum two neighbours.

tains two types of nodes $\boldsymbol{x}_t$ and $\boldsymbol{l}_t$. The nodes $\boldsymbol{x}_t$ represent the unknown state, shown in white, varying over time, building a so-called hidden Markov chain, as the new state only depends on the previous state, and not on older states. The nodes $\boldsymbol{l}_t$ represent the observations at time t. Here, we assume observed values are available, therefore the nodes are shown in grey.

The methods for optimal estimation, prediction and filtering of such a Bayesian network are well known, not only in geodesy and photogrammetry.

This paper wants to uncover more correspondences between graphical models and geodetic and photogrammetric parameter estimation problems, namely showing the close relation between geodetic networks and photogrammetric blocks and so-called conditional random fields, which have shown their power in image interpretation during the last years.

The paper does not go into the details of graphical models, which are documented in quite a number of lectures, e.g. BILMES (2000), MURPHY (1998) and books, e.g. PEARL (1988b), LI (2000), BISHOP (2006), WINKLER (2006).

The paper is organized as follows. We first give a short introduction into graphical models, relating the concepts to basic statistical tasks in geodetic education. Using several examples, we show the close relation between adjustment theory and Bayesian networks. Using a simple four-node network we on one hand demonstrate the versatility of graphical models to describe the probabilistic models for geodetic networks and for image interpretation, on the other hand uncover the intimate link between independence relations in Markov random fields and the sparsity of classical normal equation matrices.

2 Graphical Models

2.1 *Motivation*

The complexity of probabilistic models increases exponentially with the number of variables if no structure is imposed. Take as an example a small binary image of 326×119 pixels as shown in Fig. 2. In order to describe the joint

Fig. 2: Binary image with 326×119 pixels.

probability of all $N = 326 \times 119 = 38794$ pixels, one would need an enormous number of $2^N - 1 = 2^{38794} - 1 \approx 10^{11678}$ probabilities $P(x_1, ..., x_{38794})$. The number of high resolution colour images is even larger. In order to get an impression of the size of this large number

10^{11678}, one can relate it either to the comparably really microscopic number of 10^{78} atoms in the universe, or to the a bit larger number 10^{130} of possible images taken at any micrometer within the universe at any microsecond of its estimated lifetime in one of 100 million directions.

Probabilistic models with a large number of variables usually reveal an internal structure, which therefore needs to be exploited. The structure on one hand results from the object modelled. As an example take the model of time-dependent processes, which refers to the states in a sequence of points in time, see Fig. 1, or take the stochastic model of human body configuration in image analysis, which is related to the pose of connected limbs.

As geodesists and photogrammetrists we usually work with a simplified assumption about the distribution, a Gaussian distribution. As a consequence the number of parameters necessary to specify the distribution only increases quadratically with the number of parameters, which for a million variables would require the specification of $\approx 10^{12}$ entries in the covariance matrix. The ability to approximate covariance matrices using covariance functions of large point clouds suggests that in some applications much less parameters are necessary to arrive at an adequate stochastical model. Another geodetic approach to reduce the complexity of a stochastical model is the concept of primary errors, which allows to explain correlations in high dimensional covariance matrices using only a few causing effects.

2.2 *Definition and Types*

A graphical model is a probabilistic model where the dependence structure between the variables is described by a graph $G(\mathcal{N}, \mathcal{E})$. The nodes $n \in \mathcal{N}$ represent stochastical variables, the edges $e = (n_i, n_j) \in \mathcal{E}$ represent probabilistic relations.

A node may represent a single random variable, say $\underline{x}$, following some distribution, e.g. specified by $\underline{x} \sim p_x(x)$. It may be a set or vector $\underline{\boldsymbol{x}}$ of variables or even a more complex structure of random variables. Random variables are indicated by an underscore. The index x in p_x indicates the name of the variable the density function refers to, which we omit, in

case there is no confusion. The graphical presentation shows the name, say x, and possibly the type of the random variable, e.g. $\boldsymbol{x}$ if it is a vector. The random variables may be discrete or continuous, or mixed. The distribution of the variables may be completely unknown, partially or fully known. This flexibility of course requires to specify the content of the nodes in the legend of the graphical model. In the following we will assume the nodes to either represent a vector of continuous random variables, e.g. standing for measurements or parameters useful in parameter estimation problems, or a discrete random variable, e.g. standing for a class name out of a given set of possible classes useful in classification or interpretation problems.

The variable of a node may be either unknown, then the node is drawn with a white background, or it is observed, equivalent to having a sample value of the underlying distribution, then it is drawn with a grey background.

The edges of the graph are either directed or undirected. Directed edges are used, in case one wants to specify conditional probabilities.

Fig. 3 shows several graphical models with two nodes. For directed networks we show various variants with observed nodes and repetitions.
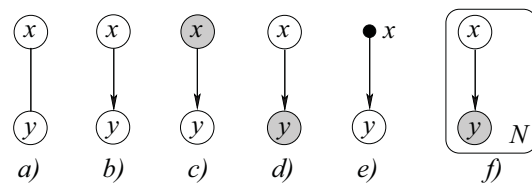
Fig. 3: Graphical models with two nodes. White nodes: unobserved. Grey nodes: observed. Black node: fixed value. Undirected edge: $p(x, y)$ is part of the model. Directed edge: $p(y|x)$ is part of the model. The rounded box indicates that the two-node network exists N-times.

The essential idea of graphical models is that the joint probability distribution $p(\boldsymbol{x}) = p(x_1, x_2, ..., x_n, ..., x_N)$ can be written as a product

$$p(x_1, x_2, ..., x_n, ..., x_N) = \frac{1}{Z} \prod_i f_i(\mathcal{X}_i) \quad (1)$$

of functions f_i of small subsets $\mathcal{X}_i = \{x_{i_1}, ..., x_{i_{N_i}}\}$ of variables, and these sets $\mathcal{X}_i$ can be seen in the graph. The functions either

result from some statistical knowledge or just can be chosen such that large values of f_i support the joint probability $p(\boldsymbol{x})$. The constant Z, the so-called partition function, sometimes is necessary to guarantee that the resulting probability density fulfills $\int p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 1$.

In the following we show that classical models in geodesy and photogrammetry can be modelled using graphical models.

2.3 *Bayesian Nets*

2.3.1 The model for collocation

In time series often a noisy signal is observed, and the original signal is to be recovered. The problem is known as collocation in geodesy and photogrammetry, see MORITZ (1978, section 4), KRAUS (1972). The signal, say $\underline{\boldsymbol{x}}$, is assumed to be a random vector. Its covariance structure usually is described using covariance functions, a modelling tool also having found its place in pattern recognition via so-called Gaussian processes, see RASMUSSEN & WILLIAMS (2005). The signal is contaminated by noise, say $\underline{\boldsymbol{n}}$. The noise is assumed to be independent of the signal and again, in case it is vector valued, may be correlated. The resulting observed signal, say $\boldsymbol{y}$, is a function of the unknown random signal $\boldsymbol{x}$ and the noise $\boldsymbol{n}$, say by addition. The Bayesian network in Fig. 4 is a graphical model for this situation. Observe, the model only specifies the principal relationship. The joint probability of all three variables can
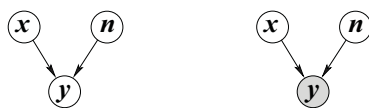
Fig. 4: A model for filtering a noisy signal. The signal x is contaminated by noise n, independent on x. The contaminated signal is y, depending on both x and n. Left: model prior to an observation. Right: The situation where the contaminated signal is observed: In case certain characteristics of the signal x and the noise n are known, both can be recovered from the observed signal y.

easily derived from the graph. The nodes with no parents, $\boldsymbol{x}$ and $\boldsymbol{n}$ are assumed to be independent and follow some distribution, the nodes $\boldsymbol{y}$ is depending on these two nodes, and the conditional probability $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{n})$ is assumed to be

part of the model. Thus, the joint probability of all three variables is

$$p(\boldsymbol{x}, \boldsymbol{n}, \boldsymbol{y}) = p(\boldsymbol{x})\, p(\boldsymbol{n})\, p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{n}) \qquad (2)$$

indicating no specific density $p(\boldsymbol{x})$ and $p(\boldsymbol{n})$ for the independent variables nor for the conditional probability $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{n})$. Thus, when specifying the model, there is no need to specify the densities.

In case all variables are zero mean Gaussian distributed and the contamination model is additive, i.e. $\underline{\boldsymbol{y}} = \underline{\boldsymbol{x}} + \underline{\boldsymbol{n}}$, the covariance matrix of the joint vector $[\underline{\boldsymbol{x}}, \underline{\boldsymbol{n}}, \underline{\boldsymbol{y}}]^\mathsf{T}$ would be

$$\mathbb{D}\left(\begin{bmatrix} \underline{\boldsymbol{x}} \\ \underline{\boldsymbol{n}} \\ \underline{\boldsymbol{y}} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{xx} & 0 & \Sigma_{xx} \\ 0 & \Sigma_{nn} & \Sigma_{nn} \\ \Sigma_{xx} & \Sigma_{nn} & \Sigma_{xx} + \Sigma_{nn} \end{bmatrix} \tag{3}$$

which follows from variance propagation. Given an observed value $\boldsymbol{y}$ for the contaminated signal, using Bayesian theorem in the form $p(\boldsymbol{x}, \boldsymbol{n}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{n})p(\boldsymbol{x})p(\boldsymbol{n})/p(\boldsymbol{y})$ one in the general setting can derive the density $p(\boldsymbol{x}|\boldsymbol{y})$ of the signal given $\boldsymbol{y}$ and the density $p(\boldsymbol{n}|\boldsymbol{y})$ of the noise given $\boldsymbol{y}$.

In the case of normally distributed variables, with

$$\Sigma_{yy} = \Sigma_{xx} + \Sigma_{nn} \tag{4}$$

we obtain the classical result

$$\mathbb{E}\left(\begin{bmatrix} \underline{\boldsymbol{x}}|\boldsymbol{y} \\ \underline{\boldsymbol{n}}|\boldsymbol{y} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{xx} \\ \Sigma_{nn} \end{bmatrix} \Sigma_{yy}^{-1}\boldsymbol{y} \tag{5}$$

and

$$\mathbb{D}\left(\begin{bmatrix} \underline{\boldsymbol{x}}|\boldsymbol{y} \\ \underline{\boldsymbol{n}}|\boldsymbol{y} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{xx} \\ \Sigma_{nn} \end{bmatrix} \Sigma_{yy}^{-1} \left[\Sigma_{xx}|\Sigma_{nn}\right] \tag{6}$$

indicating the derived signal and noise variables $\underline{\boldsymbol{x}}|\boldsymbol{y}$ and $\underline{\boldsymbol{n}}|\boldsymbol{y}$ are 100 % correlated.

This result can be directly transferred into a rule interpreting the independence relations in a Bayesian network. In a three-node network of the type in Fig. 4 left, the parent nodes $\boldsymbol{x}$ and $\boldsymbol{n}$ are independent of $\boldsymbol{y}$, whereas in the right network the parent nodes $\boldsymbol{x}$ and $\boldsymbol{n}$ are conditionally dependent given $\boldsymbol{y}$. This result is independent of the type of density functions involved. The conditional dependency can be used to infer from a subsequent observation of one of the

two causing variables to the other variable – a special case of PEARL's (1988a) so-called *explaining away*. Similar interpretation rules exist for chains and can be used to analyse more complex networks.

2.3.2 The concept of primary errors

A well-established geodetic principle to obtain a probabilistic model for the observations with only a few parameters is the concept of primary errors (PELZER 1974): Correlations in the high-dimensional covariance matrix of observations are explained by a few, say K unknown effects $\underline{p}_k$, which, besides individual random perturbations $\underline{e}$, influence all N variables $\underline{l}_n$. This also is the basis for modelling systematic image errors using additional parameters in bundle adjustment. With the, in general unknown, mean values $\underline{y}_n$ the general model reads as

$$p(l_n|y_n, \boldsymbol{p}) = \prod_n p(y_n) \prod_k p(l_n|p_k)p(p_k)$$

(7)

or in an additive setting

$$\underline{l}_n = \underline{y}_n + \sum_{k=1}^{K} h_{nk}\underline{p}_k \qquad (8)$$

the random variables $\underline{y}_n \sim \mathcal{N}(\mu_n, \sigma_{e_n}^2)$ and $\underline{p}_k \sim \mathcal{N}(0, \sigma_{p_k}^2)$ being statistically independent. Variance propagation yields the full rank covariance matrix

$$\Sigma_{ll} = \mathrm{Diag}([\sigma_{y_n}^2]) + H \, \mathrm{Diag}([\sigma_{p_k}^2])H^{\mathsf{T}} \quad (9)$$

with $H = [h_{nk}]$. The number of parameters for specifying this model is in the order of $O(NK)$, making the model specification efficient if $K \ll N$. The random variable $\underline{l}$ can easily be described by the graphical model in Fig. 5.

The exchangeability of the functional (8) and the stochastical model, using Σ_{ll} from (9), has been discussed in the early days of aerial triangulation, see ACKERMANN (1965) and SCHILCHER (1980).

The graphical model makes the difference of modelling systematic errors, as done in bundle adjustment, transparent: Systematic errors usually are modelled with the second term $\sum_{k=1}^{K} h_{nk}\underline{p}_k$, then the causing effects are made
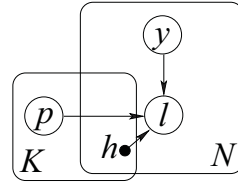
Fig. 5: Graphical model for describing the uncertainty of the N correlated random variables $\underline{l}_n$ using the primary error concept: The individual elements $\underline{l}_n$ depend on individual random variables $\underline{y}_n$ and common systematic effects $\underline{p}_k$ weighted by h_{nk}. The rounded boxes indicate that the internal structure, namely $\underline{l}$ depending on $\underline{y}$ and h on one hand and p and h on the other hand, are repeated N and $\bar{K}$ times respectively as made explicit in (8).

explicit. But they also can be modelled using a fully correlated observation vector with covariance matrix Σ_{ll} in (9), see SCHILCHER (1980), then not making the causing effects explicit. Modelling systematic errors in the functional model or the stochastical model only is equivalent, if the expected values $\mathbb{E}(\underline{p}_k) = 0$ for the systematic effects are zero and constant over time. Obviously, modelling the systematic errors in the functional model not only requires less parameters to be specified compared to modelling them in the stochastical model, but also allows to model time dependent effects, see SCHROTH (1986) and to estimate their mean values and thus *learn* the parameters from data within a self-calibration.

2.3.3 The Gauss-Markov model

The Gauss-Markov model is the work horse in statistical estimation. In its simplest form it could be written as $p(\boldsymbol{l}|\boldsymbol{x})$ only specifying the conditional probability for the observations given the parameters, thus being of type b) in Fig. 3. As soon as the observations $\boldsymbol{l}$ are available, we obtain the graphical model d), and using the Bayesian theorem we can derive the unknown parameters as $\mathrm{argmax}_x p(\boldsymbol{x}|\boldsymbol{l}) = \mathrm{argmax}_x p(\boldsymbol{l}|\boldsymbol{x})p(\boldsymbol{x})$ in case some prior information $p(\boldsymbol{x})$ about the parameters $\boldsymbol{x}$ is available. Self-calibrating bundle adjustment could be modelled a bit more expressivly, making the effect of the three parameter types, scene coordinates $\boldsymbol{k}$, orientation parameters $\boldsymbol{t}$, and additional parameters $\boldsymbol{p}$ onto the observed image coordinates $\boldsymbol{l}$ explicit, see Fig. 6.
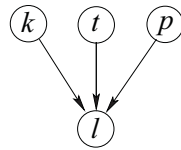
Fig. 6: Graphical model for bundle adjustment with coordinates $\boldsymbol{k}$, orientation parameters $\boldsymbol{t}$ and additional parameters $\boldsymbol{p}$ influencing the observations $\boldsymbol{l}_i$, e.g. the image coordinates. The joint probability is $p(\boldsymbol{l}, \boldsymbol{k}, \boldsymbol{t}, \boldsymbol{p}) = p(\boldsymbol{k})p(\boldsymbol{t})p(\boldsymbol{p}) \prod_i p(\boldsymbol{l}_i | \boldsymbol{k}, \boldsymbol{t}, \boldsymbol{p})$
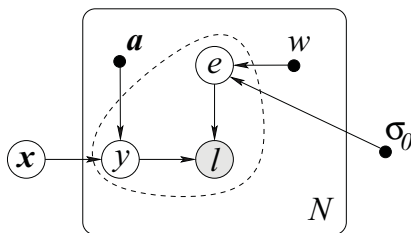
Fig. 7: Gauss-Markov model as Bayesian network. The black nodes represent given information, here the coefficients $\boldsymbol{a}$ of the (rows of the) design matrix, the weights w and the variance factor σ_0. The observations l result from a contamination of the unknown predicted observations, denoted by y, and the unknown measurement deviations e. The rounded box indicates the probabilistic model within the box is repeated N-times, indicating the independence of the N observed values l. In addition to the N fitted observations y and the measurement deviations e the parameter vector $\boldsymbol{x}$, which couples the fitted observations via the given coefficients $\boldsymbol{a}$, is unknown. Observe, this model contains the additive filtering model as core.

In case we want to model more details of the Gauss-Markov model, see Fig. 7, we can specify the dependency on the rows $\boldsymbol{a}_n$ of the design matrix, make the error free observations $y_n = \boldsymbol{a}_n^\mathsf{T} \boldsymbol{x}$ and the measurement deviations e_n explicit, and specify the individual weights w_n and the common factor σ_0 to obtain

$$\underline{l}_n = \boldsymbol{a}_n^\mathsf{T} \boldsymbol{x} + \underline{e}_n \qquad \sigma_n^2 = \sigma_0^2 / w_n \qquad (10)$$

together with the prior information

$$\underline{\boldsymbol{x}} \sim \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{\Sigma}_{x_0 x_0}) \qquad (11)$$

e.g. representing the coordinates of the control points together with their uncertainty, assuming very large covariances for the new unknown scene points. The joint probability is given in

the factorized form

$$p(\boldsymbol{l}, \boldsymbol{x}) = p(\boldsymbol{x}) \prod_n p(l_n | \boldsymbol{x}) . \qquad (12)$$

Observe, the Gauss-Markov model contains the filtering model from Fig. 4 as central part within the area with dashed boundary in Fig. 7, here in the form $l = y + e$.

We will not discuss general Bayesian nets here. Finding the optimal parameters in Bayesian nets with a tree structure, such as all examples, is linear in the number of nodes, if the number of variables per node is fixed. This high efficiency is exploited e.g. in Kalman filtering.

2.4 Markov Random Fields

Markov random fields are graphical models with undirected edges. Undirected graphs appear in image processing, image analysis, point cloud processing, but also in geodetic and photogrammetric networks.

In image processing the graph is induced by the regular structure of the pixels, yielding a regular pattern, see Fig. 8, lower left. In im-
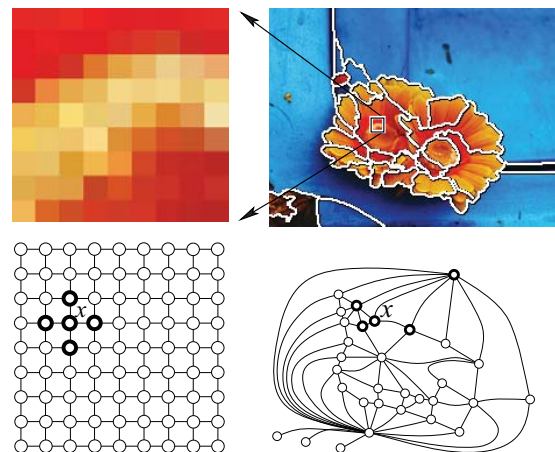
Fig. 8: Graphs of photogrammetric models as basis for Markov random fields. *Left:* pixel grid, neighbouring or adjacent pixels are assumed to have the same colour or the same class label. *Right:* region adjacency graph, neighbouring regions are assumed to have the same class label (image and segmentation from SCHINDLER & FÖRSTNER (2013)).

age analysis one often starts with partitioning the image into regions, the region adjacency

graph then may be the building block for image interpretation, which is equivalent to assigning a class name to each region, see Fig. 8, lower right. In point cloud processing the graph may result from a triangulation or a segmentation. Bundle adjustment, see Fig. 9 links the observed image coordinates $\boldsymbol{x}_{ij}$, not to be confused with the unknown parameters, with scene coordinates $\boldsymbol{k}_i$ and orientation parameters $\boldsymbol{t}_j$ via the collinearity constraints $0 = \mathrm{coll}_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{k}_i, \boldsymbol{t}_j)$.
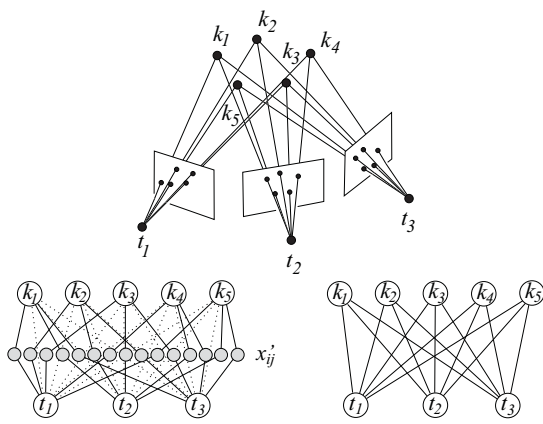
Fig. 9: Graphs of photogrammetric models as basis for Markov random fields. Bundle adjustment. *Lower left*: with 3-cliques (cliques with three nodes) linking the observations x_{ij} with the coordinates and the transformation parameters (some of the connections are shown dotted to keep the figure readable). *Lower right:* with 2-cliques (cliques with two nodes) linking coordinates and transformation parameters, observations as given values are eliminated, see DEL- LAERT & KAESS (2006).

A Markov random field having nodes x_i is characterized by the following property: The probability for a node $p(x_i|\mathcal{X} \setminus x_i)$ given the values of all other nodes is identical to the conditional probability $p(x_i|\mathcal{N}_i)$ given its neighbours:

$$p(x_i|\mathcal{X} \setminus x_i) = p(x_i|\mathcal{N}_i) . \qquad (13)$$

This is a generalization of the Markov property of time series where the current state x_t only depends on the previous thus neighbouring one x_{t-1}, explaining the name for the type of graphical model.

The joint probability $p(x_1, ..., x_N)$ of all variables x_n in a Markov random field can be written as the product of factors $\psi_i(\mathcal{X}_i)$ depend-

ing on the cliques in the corresponding graph, the clique $\mathcal{C}_i$ being a fully connected subgraph, i.e. a subgraph, where all nodes are connected, inducing the set $\mathcal{C}_i = \{x_{i_1}, ..., x_{i_{N_i}}\}$:

$$p(x_1, ..., x_N) = \frac{1}{Z} \prod_{i \in C} \psi_i(\mathcal{C}_i) . \qquad (14)$$

The functions $\psi_i(\mathcal{C}_i)$ are called potential functions and are assumed to be positive for all values the variables in the clique may have. They conceptually are *no probabilities*.

A probability of the structure (14) is called a Gibbs-distribution. The structure of this distribution is equivalent to the Markov property in (13). This follows from a theorem by HAMMERSLEY & CLIFFORD (1971), see also BUSCH (1992). In Fig. 8 the lower left network for the image grid shows only two-cliques, the lower right network of the region adjacency graph shows two- and three-cliques and the network for the adjustment in Fig. 9 shows three-cliques or two-cliques, depending on whether the observational nodes are included or excluded in the model.

2.4.1 Types of Markov random fields

There are different classes of Markov random fields (MRF), depending on the way observations are handled. For simplicity we refer to the grid type graph with maximal two-cliques, see Fig. 10.
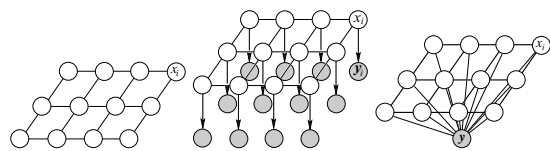
Fig. 10: Types of MRF's. *Left:* Markov random field as prior. *Middle:* hidden Markov random field, containing directed edges to the observations y_i. *Right:* Conditional random field.

A Markov random field may be used for modelling the prior distribution of the image:

$$p(\boldsymbol{x}) = \frac{1}{Z} \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) \qquad (15)$$

where only the potential functions $\psi_{ij}(x_i, x_j)$ need to be specified such that likely configurations of pairs (x_i, x_j) obtain large values $\psi_{ij}(x_i, x_j) > 0$. The key publication by GEMAN & GEMAN (1984) of the school of Grenander (GRENANDER 1976), addressed the problem of image restoration as a global optimization problem solved by stochastic relaxation in a Markov Chain Monte Carlo scheme.

A hidden Markov random field makes explicit the fact, that the class labels for the pixels are unknown. The naming disregards the fact that the edges to the observed nodes are directed. The observations y_i depend on the corresponding class labels x_i. Thus, the total probability $p(\boldsymbol{x}, \boldsymbol{y})$ of the network is

$$p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x}) = \qquad (16)$$
$$\frac{1}{Z} \prod_i p(y_i|x_i) \prod_{(ij)\in\mathcal{E}} \psi_{ij}(x_i, x_j)$$

where the prior $p(\boldsymbol{x})$ is the simple Markov random field (15). The likelihood $p(\boldsymbol{y}|\boldsymbol{x})$ can be factorized into a product of factors $\psi_i(x_i, y_i) := p_i(y_i|x_i)$ for given observations only depending on individual nodes.

Often one wants to use observations y_i which are taken from a region around the i-th pixel. Then they are not any more independent. This is the reason why one uses

Conditional random fields proposed by LAFFERTY et al. (2001) in the area of language processing and introduced to image analysis problems by KUMAR & HEBERT (2006) under the label *discriminative random fields*, though already e.g. GIMEL'FARB (1996) used models of this structure for texture analysis. The main idea is the following: one directly models the posterior probability by conditioning the probabilistic relations in the cliques, here between neighbouring pixels, on the given observations

$$p(\boldsymbol{x}|\boldsymbol{y}) = \qquad (17)$$
$$\frac{1}{Z} \prod_{i\in C_1} \psi_i(x_i, \boldsymbol{y}) \prod_{(ij)\in C_2} \psi(x_i, x_j, \boldsymbol{y})$$

where the one-cliques C_1 contain all nodes, and the two-cliques C_2 all edges.

In all cases it is useful to replace the product of the probabilities and potentials by a sum of the negative logarithms. The negative logarithm of a probability can be interpreted as the (self-) information $I(x) = -\log p(x)$, being the relative surprise when observing the sample (SHANNON & WEAVER 1949). Correspondingly the negative logarithm $E(x) = -\log \psi(x)$ of the potentials is called the energy, as the concept has been developed in physics. Thus, the complete models can be written as sums.

When neglecting the given observed, fixed values $\boldsymbol{y}$, for the conditional random field model (17) we obtain

$$I(\boldsymbol{x}|\boldsymbol{y}) \quad = \quad \log Z + \sum_{i\in C_1} E_i(x_i) \quad (18)$$
$$+ \sum_{(ij)\in C_2} E_{ij}(x_i, x_j).$$

2.4.2 Tasks

Having fixed the structure of the model several tasks need to be addressed.

Inference. The first task is inference: In case the model is fully specified, one is interested in an estimate $\widehat{\boldsymbol{x}}$ for the optimal set of values for the unknowns $\{x_i\}$, e.g. the one which maximizes the probability $p(\boldsymbol{x}|\boldsymbol{y})$, leading to the maximum a posteriori (MAP) estimate, or, equivalently, the one which minimizes $I(\boldsymbol{x}|\boldsymbol{y})$. If the x_i represent class labels in an image, this will lead to an optimal labelling of all pixels or regions of the image. If the x_i represent the continuous values of the intensities of the pixels, one might obtain an optimal restoration of the image.

The problem in general is intractable, i.e. the computing time increases exponentially with the number of nodes, which is plausible when regarding the large search space, see the discussion of the number of images above. Only in very special cases finding the optimum is tractable: (a) in case the variables are continuous and follow a Gaussian distribution, the optimal solution can be achieved by solving a large linear equation system, (b) in case the variables are binary and the potential functions obey certain conditions, the problem can be

mapped to a network algorithm, namely finding the maximum flow or the minimal cut in a network (BOYKOV & KOLMOGOROV 2004). In all other cases only suboptimal algorithms exist (BISHOP 2006).

None of the algorithms yields the probability $p(\boldsymbol{x}|\boldsymbol{y})$ as this would require the determination of the normalization constant Z, which is $Z = \sum_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y})$, the sum over all possible images.

One simple algorithm uses the Markov property and iteratively determines the locally best estimate given all neighbours:

$$\widehat{x}_i^{(\nu+1)} = \mathrm{argmax}_{x_i} p\left(x_i \mid \mathcal{N}_i^{(\nu)}\right) \qquad (19)$$

where (ν) indicates the iteration number. The algorithm is known as *iterative conditional modes* (ICM). It often yields good results, but may get stuck in a local minimum of $I(\boldsymbol{x}|\boldsymbol{y})$. We will relate this algorithm to one used in adjustment theory.

As graphical models factor the joint probability into small factors, see (1) and its specialization for Bayesian nets and Markov random fields, see (16), they can be represented as so-called factor graphs, which allow for a homogeneous algorithmic handling of Bayesian nets and Markov random fields (KSCHISCHANG et al. 2001).

Learning. The second task is learning. It results from the difficulty in specifying the potentials in real world problems completely. Take an example: As we are free to choose the potential functions we could use posterior probabilities $p(x_i|\boldsymbol{y}_i)$ and $p(x_i, x_j|\boldsymbol{y}_{ij})$ of classifiers for labels x_i or label pairs (x_i, x_j). Both depend on the corresponding observed features $\boldsymbol{y}_i$ and $\boldsymbol{y}_{ij}$, and possibly global parameters $\boldsymbol{y}_g$, collected in the vector $\boldsymbol{y}$. Usually they are parametrized, e.g. in case one uses a maximum likelihood classifier based on a Gaussian distribution. Then the parameter vector $\boldsymbol{u}$ would contain the mean vectors and covariance matrices of the classifier to be learnt. Thus, we would obtain

$$\psi_i(x_i|\boldsymbol{u}) \qquad (20)$$
$$= p(x_i|\boldsymbol{y}_i, \boldsymbol{u}_1)\psi_{ij}(x_i, x_j|\boldsymbol{y}_{ij}, \boldsymbol{u}_2)$$

where the parameters $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2)$ need to be learned from the data. The complete model, now written using negative logarithms then is

$$I(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{u}) = \log Z(\boldsymbol{u}) + \qquad (21)$$
$$+ \sum_{i \in C_1} E_i(x_i, \boldsymbol{u}_1) + \sum_{(ij) \in \mathcal{C}_2} E_{ij}(x_i, x_j, \boldsymbol{u}_2)$$

where the normalization factor $1/Z(\boldsymbol{u})$ also depends on the unknown parameters. Given a large enough set of training data $(x_i, \boldsymbol{y}_i)$ and $(x_i, x_j, \boldsymbol{y}_{ij})$ the task is to learn, i.e. to estimate optimal parameters for $\boldsymbol{u}$. Again, the problem of estimating optimal parameters is intractable in general and only suboptimal solutions are known.

The following example wants to demonstrate the possibility to map classical geodetic networks to Markov random fields

2.5 *A Four-Node Network*

In order to demonstrate the flexibility of graphical models, especially of conditional random fields, we use a four-node graph with one edge missing as an example. We do not show the observational nodes, which would be linked to all four nodes, see Fig. 11.

2.5.1 A levelling network

The graph with four nodes representing the 1-cliques $\mathcal{C}_1 = \{1, 2, 3, 4\}$ and five edges representing the 2-cliques $\mathcal{C}_2 = \{(1, 2), (2, 3), (3, 4), (1, 4), (2, 4)\}$, has two 3-cliques $\mathcal{C}_3 = \{(1, 2, 4), (2, 3, 4)\}$ as there are only two fully connected triangles. Interpreting the graph as a levelling network restricts the modelling to 1- and 2-cliques, as we only have measured heights $l_i, i = 1, 3$ of the two control points and the five height differences $l_{ij}, (i, j) \in \mathcal{C}_2$.

We now model the joint probability with *maximal cliques*, thus by 2-cliques only. This is achieved by taking the potentials for the observed control points into on the potential of one of the corresponding the 2-cliques. The joint probability therefore is

$$\begin{aligned} p(\boldsymbol{x}) &= p(x_1, x_2, x_3, x_4) \\ &\propto \prod_{(i,j) \in \mathcal{C}_2} \psi_{ij}(x_i, x_j) \\ &\propto \psi_{12}(x_1, x_2) \bullet \dots \bullet \psi_{34}(x_3, x_4) . \end{aligned}$$
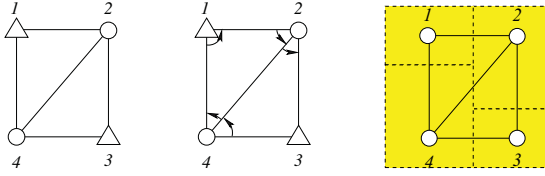
Fig. 11: *Left:* Leveling network as a conditional random field. The triangles indicate reference points with given heights h_1 and h_3, in principle being uncertain. The two unknown heights h_2 and h_4 are indicated by white circles. The edges in the graph indicate probabilistic relations between the two nodes in concern, which practically express the uncertainty of the five observed height differences. No observations relate three heights simultaneously, i.e. only two-cliques in the graph give rise to potentials. *Middle:* Angular network as a conditional random field. The triangles now represent control points in 2D with their uncertain Cartesian coordinates $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$. The coordinates $\boldsymbol{x}_2$ and $\boldsymbol{x}_4$ are unknown. The five edges in the graph again represent probabilistic relations between the corresponding nodes. Each angle measurement is indicated by arrows and depends on three points. Only two three-cliques in the graph give rise to potentials. *Right:* Classification of four parcels as a conditional random field. The nodes represent the unknown classes of the four fields. There may be prior probabilities on the classes, known from the type of geographic region. The edges again probabilistically constrain the corresponding classes, stating neighbouring fields to more likely belong to the same class than to different classes, which may be made dependent on observations, e.g. of the appearance of the field boundaries.

In order to choose adequate potential functions we assume the heights and the height differences to be normally distributed

$$
\begin{aligned}
\underline{l}_i &\sim \mathcal{N}(\mu_{x_i}, w_{x_i}^{-1}), i \in \mathcal{C}_1 \quad (22) \\
\underline{l}_{ij} &\sim \mathcal{N}(\mu_{x_j} - \mu_{x_i}, w_{ij}^{-1}), \quad (i,j) \in \mathcal{C}_2
\end{aligned}
$$

where we used the weights $w = 1/\sigma^2$. Therefore, we choose the following potential functions, being able to neglect the constant factors of the normal distribution

$$
\begin{aligned}
\psi_{12} &= \mathrm{e}^{-\frac{1}{2}(x_2 - x_1 - l_{12})^2 w_{12}} \mathrm{e}^{-\frac{1}{2}(x_1 - l_1)^2 w_1} \\
\psi_{14} &= \mathrm{e}^{-\frac{1}{2}(x_4 - x_1 - l_{14})^2 w_{14}} \quad (23) \\
\psi_{23} &= \mathrm{e}^{-\frac{1}{2}(x_3 - x_2 - l_{23})^2 w_{23}} \\
\psi_{24} &= \mathrm{e}^{-\frac{1}{2}(x_4 - x_2 - l_{24})^2 w_{24}} \\
\psi_{34} &= \mathrm{e}^{-\frac{1}{2}(x_4 - x_3 - l_{34})^2 w_{34}} \mathrm{e}^{-\frac{1}{2}(x_3 - l_3)^2 w_1} .
\end{aligned}
$$

Observe, we only used 2-cliques and integrated the prior information about the control points into one of the neighbouring 2-cliques. The potential functions essentially depend on the weighted squares of the differences v_i and v_{ij} between the unknown parameters and the measurements. Taking negative logarithms we therefore obtain the – not really surprising – result: We need to find the minimum of

$$
I(\boldsymbol{x}|\boldsymbol{l}) = \sum_{(i,j) \in \mathcal{C}_2} w_{ij} v_{ij}^2(x_i, x_j) \quad (24)
$$

here with $v^2 := -2\log\psi$, thus v_{12}^2 and v_{34}^2 containing the squared residuals at the control points. *The leveling network is a special Markov random field, namely a Gaussian random field.*

We now apply the iterative conditional mode algorithm to this problem. The normal equation system, resulting from setting $\partial I/\partial \boldsymbol{x} = \boldsymbol{0}$ and assuming weights $w_i = w_{ij} = 1$ is

$$
\begin{bmatrix} \mathbf{3} & -1 & \mathbf{0} & -1 \\ -1 & 3 & -1 & -1 \\ \mathbf{0} & -1 & \mathbf{3} & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} =
$$
$$
= \begin{bmatrix} -l_{12} - l_{14} + l_1 \\ l_{12} - l_{23} - l_{24} \\ l_{23} - l_{34} + l_3 \\ l_{34} + l_{14} + l_{24} \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} . \quad (25)
$$

The best estimate for the individual unknowns, given the others, therefore are

$$
\begin{aligned}
x_1^{(\nu+1)} &= \frac{1}{3}(h_1 + x_2^{(\nu)} + x_4^{(\nu)}) \\
x_2^{(\nu+1)} &= \frac{1}{3}(h_2 + x_1^{(\nu+1)} + x_3^{(\nu)} + x_4^{(\nu)}) \\
x_3^{(\nu+1)} &= \frac{1}{3}(h_3 + x_2^{(\nu+1)} + x_4^{(\nu)}) \\
x_4^{(\nu+1)} &= \frac{1}{3}(h_4 + x_1^{(\nu+1)} + x_3^{(\nu+1)} + x_2^{(\nu+1)})
\end{aligned}
$$

again with the iteration index (ν). This method of solving the normal equations iteratively is known as Gauss-Seidel method. It is guaranteed to converge in this case, as the normal equation matrix is symmetric and positive definite. This criterion of course is difficult to generalize for non-Gaussian situations. This procedure is equivalent to the so-called method of

iterative conditional modes for iteratively finding the optimal solution in a general Markov-random field using (13) in the form (BUSCH 1992)

$$x_i^{(\nu+1)} := \operatorname{argmax}_{x_i} p(x_i \mid \mathcal{N}_i^{(\nu)}(x_i)) . \quad (26)$$

A final remark refers to the sparsity of the model. In spite of the small size of the network the graph is not fully connected, as edge $(1, 3)$ is missing. This is reflected in the zero-entry $N_{13} = 0$ of the normal equation matrix. The zero at the $(1, 3)$ position indicates that given the other variables x_2 and x_4 the variables x_1 and x_3 are independent: formally $p(x_1, x_3|x_2, x_4) = p(x_1|x_2, x_4)p(x_3|x_2, x_4)$, as then the resulting normal equation system for x_1 and x_3 is diagonal. This is indicated by the boldtype numbers in the normal equation matrix in (25). The conditional independence results in a sparse graphical model. In adjustment theory it regularly is exploited for increasing the efficiency of computation of large networks. The sparseness of the normal equation matrix however is an indication for the sparsity of the complete model resulting from the type of measuring design, which therefore in a natural way leads to a sparse graphical model.

2.5.2 An angular network

We now use the same graph to represent a 2D network with measured angles. The nodes then represent random 2-vectors $\boldsymbol{x}_i$. The observations l_{ijk} are the angles $l_{ijk} := \alpha_{ijk} = \phi_{jk} - \phi_{ji}$ and depend on the coordinates of the points. Thus, we will need the two 3-cliques to represent the probability of the complete network:

$$p(\boldsymbol{x}|\boldsymbol{l}) \quad (27)$$
$$= \frac{1}{Z} \psi_{124}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4) \psi_{234}(\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) .$$

We assume the control point coordinates to be Gaussian distributed

$$\underline{\boldsymbol{l}}_i \sim \mathcal{N}(\boldsymbol{\mu}_{x_i}, w_i^{-1} l_2) , i = 1, 3 . \quad (28)$$

Again we assume the angular measurements to be normally distributed

$$\underline{l}_{ijk} \sim \mathcal{N}(\alpha_{ijk}(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k), w_{ijk}^{-1}) \quad (29)$$

with $(i, j, k) \in \mathcal{A} = \{(4,1,2), (1,2,4), (2,4,1), (3,4,2), (4,2,3)\}$. This yields the potentials

$$\psi_{124}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4) = \quad (30)$$
$$\mathrm{e}^{-\frac{1}{2}\left(|\alpha_{421}-l_{421}|^2 w_{421}+|\alpha_{142}-l_{142}|^2 w_{142}\right)}$$
$$\bullet\mathrm{e}^{-\frac{1}{2}\left(|\alpha_{214}-d_{214}|^2 w_{214}+|x_1-l_1|^2 w_1\right)}$$

and

$$\psi_{234}(\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) = \quad (31)$$
$$\mathrm{e}^{-\frac{1}{2}\left(|\alpha_{243}-l_{243}|^2 w_{243}+|\alpha_{324}-l_{324}|^2 w_{324}\right)}$$
$$\bullet\mathrm{e}^{-\frac{1}{2}\left(|x_3-l_3|^2 w_3\right)} .$$

2.5.3 A classification network

We finally use the graph for modelling a classification task. Let four agricultural fields be arranged as shown in the Fig. 11 right. The graph then represents the region adjacency.

One usually takes some spectral or texture features $\boldsymbol{y}_i$ in each region from the underlying image (not shown) as observations, and performs a Bayesian classification, i.e. labelling of the regions. This is achieved by maximizing the posterior probability $p(x_i|\boldsymbol{y}_i) \propto p(\boldsymbol{y}_i|x_i)p(x_i)$ using the likelihood $L(x_i) := p(\boldsymbol{y}_i|x_i)$, which is to be learnt from training data and some prior $p(x_i)$ on the occurrence of the different classes $x_i \in \{1, 2, ...\}$. It corresponds to four independent Bayesian nets of the type Fig. 3 d. The joint prior of all four nodes just is $p(\boldsymbol{x}) = \prod_{i=1}^{4} p(x_i)$, as the class membership or the labelling of the regions is assumed to be independent.

In case one can assume neighbouring regions often belong to the same class, a simple model for the prior would therefore be to assume neighbouring regions are likely belonging to the same class, and unlikely belonging to different classes. Assuming the probability for the complete network for both cases should differ by a factor 10, this can be expressed by the potential function for each edge

$$\psi(x_i, x_j) = \left\{ \begin{array}{l} 1, \text{ if } x_i = x_j \\ 0.1, \text{ if } x_i \neq x_j \end{array} \right. . \quad (32)$$

The prior for all nodes then is

$$p(\boldsymbol{x}) = \frac{1}{Z} \prod_{(ij) \in \mathcal{C}_2} \psi(x_i, x_j) \quad (33)$$

which leads to the complete probability

$$p(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{Z} \prod_{i \in \mathcal{C}_1} p(x_i | \boldsymbol{y}) \prod_{(ij) \in \mathcal{C}_2} \psi(x_i, x_j) \,.$$

(34)

This is the so-called Potts model for the prior in classification using a Markov random field. As discussed above, finding an optimal labelling is intractable. Examples with a small network are given by KORČ (2012).

3   Outlook and Conclusion

Graphical models are a powerful tool for communicating between users of a statistical model and experts in statistics. The paper showed various examples for models used in geodesy and photogrammetry where graphical models give insight into the internal structure of the models, especially making fixed, observed and unknown parameters explicit and showing the independence assumptions made.

For a graphical model, where the relationships are linear and the distributions are Gaussian, reasoning, especially optimization, leads to a linear equation systems guaranteeing a unique solution. For statistical optimization problems resulting from a graphical model in general no unique method exists. This does not only hold for parameter estimation problems, e.g. in the presence of outliers, but – with only few exceptions – for all classification problems due to the discrete nature of the search of the parameter space.

The paper shows various relations between modelling and estimation in geodetic and photogrammetric networks. The parsimony of the used models results from the special structure of most networks. Interpreting such networks as Markov random fields allows one to see the more general structure of the problems at hand. Examples are the modelling of cycle slips in GPS observations, where discrete and continuous variables occur simultaneously, or the change of the type of distribution, e.g. from Gaussian to Laplace, or the type of prior knowledge. These abilities not only build a bridge between models in geometry and in classification, but also between photogrammetry and remote sensing.

References

ACKERMANN, F., 1965: Fehlertheoretische Untersuchungen über die Genauigkeit photogrammetrischer Streifentriangulationen. – Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften **C 87**, München.

BILMES, J., 2000: Introduction to Graphical Models. http://www.ee.washington.edu/class/596/patrec/ (1.3.2013).

BISHOP, C., 2006: Pattern Recognition and Machine Learning. – Springer-Verlag, NY, USA.

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

BUSCH, A., 1992: Bayes-Statistik und Markoff-Felder für die Restaurierung digitaler Bilder. – Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften **C 396**, München.

DELLAERT, F. & KAESS, M., 2006: Square Root SAM: Simultaneous localization and mapping via square root information smoothing. – International Journal of Robotics Reasearch **25**: 2006.

GEMAN, S. & GEMAN, D., 1984: Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images. – IEEE T-PAMI **6** (6): 721–741.

GIMEL'FARB, G.L., 1996: Texture Modeling by Multiple Pairwise Pixel Interactions. – IEEE T-PAMI **18** (11): 1110–1114.

GRENANDER, U., 1976: Lectures in Pattern Theory. – Volumes I-III of Applied Mathematical Sciences, Springer Verlag, New York, USA.

HAMMERSLEY, J.M. & CLIFFORD, P., 1971: Markov fields on finite graphs and lattices. – unpublished, Berkeley, CA, USA. http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf (6.6.2013).

KORČ, F., 2012: Tractable Learning for a Class of Global Discriminative Models for Context Sensitive Image Interpretation. – PhD thesis, Department of Photogrammetry, University of Bonn.

KRAUS, K., 1972: Interpolation nach kleinsten Quadraten in der Photogrammetrie. – BuL **1**: 7–12.

KSCHISCHANG, F., FREY, B. & LOELIGER, H.-A., 2001: Factor graphs and the sum-product algorithm. – Information Theory, IEEE Transactions on **47** (2): 498–519.

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

LAFFERTY, J.D., MCCALLUM, A. & PEREIRA, F.C.N., 2001: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. – **Eighteenth** International Conference on Machine Learning, ICML '01: 282–289, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

LI, S.Z., 2000: Markov random field modeling in computer vision. – Springer-Verlag, London, UK.

MORITZ, H., 1978: Least squares collocation. – Revision of Geophysics and Space Physics **16** (3): 421–430.

MURPHY, K., 1998: A Brief Introduction to Graphical Models and Bayesian Networks. http://www.cs.ubc.ca/~murphyk/Bayes/bnintro. html (1.3.2013).

PEARL, J., 1988a: Embracing causality in default reasoning. – Artificial Intelligence **35** (2): 259–271.

PEARL, J., 1988b: Probabilistic Reasoning in Intelligent Systems. – Morgan Kaufmann Publishers.

PELZER, H., 1974: Zur Behandlung singulärer Ausgleichungsaufgaben. – Zeitschrift für Vermessungswesen **5**: 181–194, 479–488.

RASMUSSEN, C.E. & WILLIAMS, C.K.I., 2005: Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). – The MIT Press, Cambridge, MA, USA.

SCHILCHER, M., 1980: Empirisch-statistische Untersuchungen zur Genauigkeitsstruktur des photogrammetrischen Luftbildes. – Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften **C 262**, München.

SCHINDLER, F. & FÖRSTNER, W., 2013: Marching Front Graph Partitioning in Geometry and Image Processing. – PFG – Photogrammetrie, Fernerkundung, Geoinformation.

SCHROTH, R., 1986: Ein erweitertes mathematisches Modell der Aerotriangulation zur hochgenauen Punktbestimmung. – Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften **C 316**, München.

SHANNON, C.E. & WEAVER, W., 1949: The Mathematical Theory of Communication. – The University of Illinois Press. Urbana, IL, USA.

WINKLER, G., 2006: Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction (Stochastic Modelling and Applied Probability). – Springer-Verlag New York, Secaucus, NJ, USA.

Address of the Authors:

Prof. Dr.-Ing. Dr. h. c. mult. WOLFGANG FÖRSTNER, Josef-Schell-Str. 34, 53121 Bonn, e-mail: wf@ipb.uni-bonn.de

Article

A Trainable Markov Random Field for Low-Level Image Feature Matching with Spatial Relationships

TIMO DICKSCHEID & WOLFGANG FÖRSTNER, Bonn

Summary: Many vision applications rely on local features for image analysis, notably in the areas of object recognition, image registration and camera calibration. One important example in photogrammetry are fully automatic algorithms for relative image orientation. Such applications rely on a matching algorithm to extract a sufficient number of correct feature correspondences at acceptable outlier rates, which is most often based on the similarity of feature descriptions. When the number of detected features is low, it is advisable to use multiple feature detectors with complementary properties. When feature similarity is not sufficient for matching, spatial feature relationships provide valuable information. In this work, a highly generic matching algorithm is proposed which is based on a trainable Markov random field (MRF). It is able to incorporate almost arbitrary combinations of features, similarity measures and pairwise spatial relationships, and has a clear statistical interpretation. A major novelty is its ability to compensate for weaknesses in one information cue by implicitly exploiting the strengths of others.

Zusammenfassung: *Ein trainierbares Markoff-Zufallsfeld für die Zuordnung lokaler Bildmerkmale unter Berücksichtigung ihrer räumlichen Beziehungen.* Viele Anwendungen im Bereich des maschinellen Sehens nutzen lokale Merkmale für die Bildanalyse, insbesondere in den Bereichen Objekterkennung, Bildregistrierung und Kamerakalibrierung. Ein wichtiges Beispiel in der Photogrammetrie sind vollautomatische Algorithmen für die relative Kameraorientierung. Dazu muss aus den Bildmerkmalen verschiedener Bilder anhand eines Matchingalgorithmus eine ausreichende Anzahl von Zuordnungen mit vertretbarem Ausreißeranteil gewonnen werden. Die Suche nach Zuordnungen basiert dabei meist auf der Ähnlichkeit von Merkmalsbeschreibungen. Wenn die Anzahl der extrahierten Merkmale gering ist, macht es Sinn, mehrere möglichst komplementäre Merkmalsdetektoren gleichzeitig einzusetzen. Ist die Ähnlichkeit von Bildmerkmalen kein ausreichendes Kriterium für die Zuordnung, liefern räumliche Beziehungen von Merkmalen zusätzlich wertvolle Information. In dieser Arbeit stellen wir ein allgemeines Matchingverfahren vor, das auf einem trainierbaren Markoff-Zufallsfeld basiert. Es ermöglicht die gleichzeitige Berücksichtigung nahezu beliebiger Arten von Bildmerkmalen, Ähnlichkeitsmaßen und paarweisen räumlichen Beziehungen, und lässt sich statistisch klar interpretieren. Eine Besonderheit dieses Verfahrens ist seine Eigenschaft, Schwachpunkte einer Informationsquelle durch die Stärken einer anderen implizit auszugleichen.

1 Introduction

Many vision applications use local image features as a sparse representation of image content. In photogrammetry, local image features have been used successfully to build automatic algorithms for relative image orientation (POLLEFEYS et al. 2000, MAYER 2005, LÄBE & FÖRSTNER 2006, SNAVELY et al. 2006). In general, bundle adjustment benefits from feature matching for automatic extraction of tie points, if outliers are handled in a reliable way.
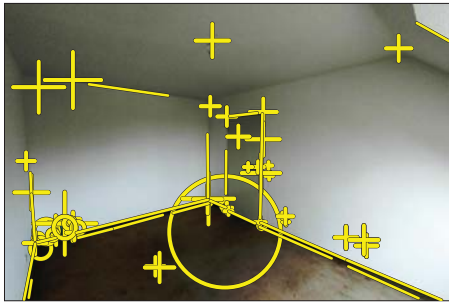
Fig. 1: Set of complementary image features covering an indoor scene. The plot includes junctions (FÖRSTNER et al. 2009) represented by crosses, blobs (LOWE 2004) represented by circles and straight line segments (FÖRSTNER 1994).

Fig. 2: Feature similarity and spatial appearance: p_1' in image $\mathcal{I}'$ has high similarity to p_1 in image $\mathcal{I}$, although it is not the right candidate. Inconsistency of the spatial relationship "is left of" with (p_2, p_2') indicates this.

Typical image features used as tie points include corners, junctions, dark and bright blobs or line segments Fig. 1. Selecting an appropriate detector is crucial, because the importance of properties like repeatability, localization accuracy, coverage or computational complexity depends on the task at hand. If the amount of detected features is too small, it becomes necessary to combine multiple detectors. In such a case, complementarity of features is an important aspect (DICKSCHEID et al. 2010).

An image feature defines a local image patch with a particular location, orientation, shape and size. In many applications, feature correspondences can be reliably determined by analysing the similarity of those patches (*appearance-based matching*, see section 2). Feature detection and matching have to reflect the expected range of image poses. We assume the general case of arbitrary mutual rotation and possibly large scale differences, referring to all close range applications or cases where no prior knowledge on the exterior orientation is available.

Feature appearance alone is not sufficient for reliable matching if many features have similar appearance, or if the descriptors have poor distinctiveness. In such cases the use of spatial feature relationships is known to improve the matching results. For example, consider the illustration in Fig. 2. Here, the blob feature patch p_1 in the left image has very similar appearance to p_1' in the right image if one allows for arbitrary rotations, but this is obviously not the right candidate. However, as p_1 is left of p_2 in $\mathcal{I}$, but

p_1' is right of p_2' in $\mathcal{I}'$, the spatial alignment gives us evidence about a possible misassignment.

A number of sophisticated methods for including spatial relationships into the matching process have been proposed, but most of them are tailored to a specific type of spatial relationship, feature type, or image data. They are therefore difficult to adapt to new matching problems.

In this work, a highly generic matching algorithm is proposed which is based on a trainable Markov random field (MRF). It is able to incorporate almost arbitrary combinations of features, feature descriptions, similarity measures and pairwise spatial relationships. The solution has a clear interpretation as the maximum a posteriori estimate of a binary classification problem, which consists in choosing a good subset from a coarse initial preselection of putative matches. A major novelty of this algorithm is its ability to compensate for weaknesses in one information cue by implicitly exploiting the strength of others.

Notation

We assume that we are given two images $\mathcal{I}$ and $\mathcal{I}'$. Feature detection and description gives us two sets $\mathcal{P}$ and $\mathcal{P}'$ of features. Each feature $p_i \in \mathcal{P}$ is itself a set $\{(x_i, y_i), \sigma_i, \alpha_i, d_i, \lambda_i\}$, where (x_i, y_i) is its location in the image given in pixels, usually referring to the centre of the local patch, σ_i is the scale given in pixels, which we assume to be proportional to the (mean) diameter of the local patch, α_i is its dominant orientation in radians, d_i is the descriptor for the local patch, and λ_i denotes the type of detector used. The features also carry uncertainty information for their elements. We will discuss

this aspect in section 3.2. A feature correspondence is a pair $v_n = (\boldsymbol{p}_i, \boldsymbol{p}'_j)$ where $\boldsymbol{p}_i \in \mathcal{P}$ and $\boldsymbol{p}'_j \in \mathcal{P}'$. Dissimilarity is one property of a correspondence, expressed by a distance measure $s_n = d(\boldsymbol{d}_i, \boldsymbol{d}'_j)$ on the descriptors.

2 Appearance-Based Feature Matching

In appearance-based feature matching, one usually assumes that two features are likely to correspond if (1) the similarity of their patches is high and (2) the similarity to other patches is significantly smaller, which directly corresponds to the criteria "similarity" and "exclusion" that ULLMAN (1979) already identified for a good visual mapping. To measure similarity, robust and distinctive *descriptors* are computed from the local patches. Corresponding image features can then be found by nearest neighbour search in the space of these descriptors. A de-facto standard, often denoted as *best matching*, works as follows:

1. Determine for each descriptor in one view its two nearest neighbours belonging to the other image.
2. Select the nearest neighbour as a correspondence only if the ratio between the two distances is significantly different from 1.

A more general formulation of this algorithm leads to the BESTMATCH-k algorithm: Instead of evaluating the ratio of distances to the best and second best neighbour (BESTMATCH-2), one may consider the distances to the $(k-1)$-th and k-th best neighbours. In the special case of $k = 1$, the nearest neighbour is always selected.

3 Spatial Relationships for Feature Matching

3.1 *Related Work*

In their early work, BARNARD & THOMPSON (1980) use an iterative relaxation labelling technique to select matches with locally similar image disparities. FÖRSTNER (1986) accepts only matches that are consistent under a global affine transformation, assuming that the scene can be reasonably approximated by a tilted plane. In the *relational matching* approach of (SHAPIRO & HARALICK 1987), an optimal

matching is found by minimizing the number of arbitrary spatial relationships that are not preserved by a final assignment. AGUILAR et al. (2009) proposed an iterative algorithm that constructs a consistent set of matches in terms of spatial nearest neighbourhood relationships. BAY et al. (2005) match straight line segments by first selecting the three most similar candidates per feature, and then iteratively removing matches that cause the highest number of sidedness violations (section 3.3). A remarkable feature of their algorithm is the boosting step, where previously discarded candidates are explicitly reintroduced in a post processing manner in case they become spatially consistent after the initial filtering. Most of these methods treat spatial consistency as a hard constraint, and tend to eliminate a significant amount of inliers.

To exploit appearance and spatial layout simultaneously, DELPONTE et al. (2006) exploit the properties of a *singular value decomposition (SVD)* to amplify the values of favourable matches in a *proximity matrix* $\boldsymbol{G} \in \mathbb{R}^{|\mathcal{P}| \times |\mathcal{P}'|}$, which captures the proximity and similarity of all pairs of features. TELL & CARLSSON (2002) proposed an interesting feature descriptor that itself captures aspects of spatial layout. Some recent methods cast spatial inconsistency and feature dissimilarity into a combined energy function to find the best matching as the one with minimum energy (SCHELLEWALD & SCHNÖRR 2005, CHOI & KWEON 2009, TORRESANI et al. 2008).

These algorithms provide no straightforward way to bring larger sets of relationships and feature dissimilarity measures with possibly significantly different strengths into such a joint formulation. Furthermore, the relative weighting of appearance and spatial consistency is not intuitive in most of these works.

3.2 *Homogeneous Point and Line Representations*

To derive spatial relationships for different types of features, we assume that we can always construct the normalized 2D homogeneous point $\mathbf{x}_i = [x_i, y_i, 1]^\top$ with 3×3 covariance matrix Σ_{xx} representing the position of an image feature $\boldsymbol{p}_i$. For line segments, we will use the midpoint for construct-
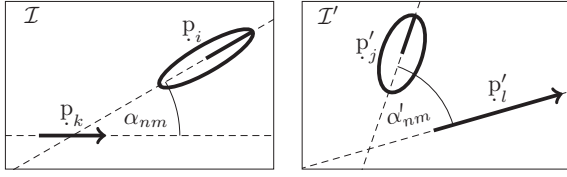
Fig. 3: Computation of the pairwise orientation difference $t_{nm}^\alpha = \min(|\alpha_{nm} - \alpha_{nm}'|, 2\pi - |\alpha_{nm} - \alpha_{nm}'|)$ for two matches $v_n = (\boldsymbol{p}_i, \boldsymbol{p}_j')$ and $v_m = (\boldsymbol{p}_k, \boldsymbol{p}_l')$.

ing $\mathbf{x}_i$, which usually has a strong localization error along the line, and a small error perpendicular to it. In a similar manner, we assume that the uncertain 2D homogeneous line $\mathbf{l}_i = \pm[\cos\alpha_i, \sin\alpha_i, -d]^\mathsf{T}$ with covariance matrix Σ_{ll} can be constructed from each feature $\boldsymbol{p}_i$. For point-like features, we use the centroid representation of straight line segments (MEIDOW et al. 2009, 3.1.2), where the centroid is the image location of the original point feature, and the direction is identified with the dominant gradient orientation within the local patch, as stored in the SIFT descriptor. As pointed out by LOWE (2004, section 5), we must expect this direction to have a standard deviation of about three degrees. Using these conversions, we define explicit operators $\mathbf{x}(\boldsymbol{p}_i)$ and $\mathbf{l}(\boldsymbol{p}_i)$, which return the uncertain homogeneous 2D point or line representation for a feature $\boldsymbol{p}_i$.

3.3 *Some Uncertain Spatial Relationships*

Here we give a short description of three different kinds of relationships, leading to statistically motivated soft constraints for the matching.

Consistency of angles

Consider the angle between two oriented features Fig. 3. We assume that the difference between these angles is rather small for valid pairs of correspondences, so that large differences indicate outliers. For two putative matches $v_n = (\boldsymbol{p}_i, \boldsymbol{p}_j')$ and $v_m = (\boldsymbol{p}_k, \boldsymbol{p}_l')$, we compute the enclosing angles

$$\alpha_{nm} = \alpha(\boldsymbol{p}_k) - \alpha(\boldsymbol{p}_i) \mod 2\pi \qquad (1)$$
$$\alpha_{nm}' = \alpha(\boldsymbol{p}_l') - \alpha(\boldsymbol{p}_j') \mod 2\pi . \qquad (2)$$

The difference $t_{nm}^\alpha \in (0, \pi)$ of the angles spanned in the two images is then given by

$$t_{nm}^\alpha = \min(|\alpha_{nm} - \alpha_{nm}'|, 2\pi - |\alpha_{nm} - \alpha_{nm}'|) . \qquad (3)$$

In case that the uncertainty of feature orientations varies significantly, the distances should additionally be normalized based on their standard deviations.

Consistency of distance

If two features are located close to each other in one view, we also expect their correspondences in another view to be close. This simple reasoning based on proximity was already suggested by ULLMAN (1979). We measure the distance between two feature locations, normalize it by the length of the image diagonal, and compare it to the same normalized distance of the two corresponding features in the second image, leading to the inconsistency measure

$$t_{nm}^d = t^d(v_n, v_m) = t^d(\boldsymbol{p}_i, \boldsymbol{p}_j', \boldsymbol{p}_k, \boldsymbol{p}_l') \qquad (4)$$
$$= \frac{|\mathbf{x}(\boldsymbol{p}_i) - \mathbf{x}(\boldsymbol{p}_k)|}{\sqrt{(N_x)^2 + (N_y)^2}} - \frac{|\mathbf{x}(\boldsymbol{p}_j') - \mathbf{x}(\boldsymbol{p}_l')|}{\sqrt{(N_x')^2 + (N_y')^2}} ,$$

using the vertical and horizontal dimensions N_x, N_y of an image $\mathcal{I}$ in pixels, with $t^d \in (-1, 1)$. Again, an additional benefit can be gained by replacing the Euclidean distances with the proper test statistic, i.e. by normalizing the distances using their standard deviations.

Consistency of pairwise sidedness

Consider again the example in Fig. 2. As $\boldsymbol{p}_1$ is left of $\boldsymbol{p}_2$, while $\boldsymbol{p}_1'$ is right of $\boldsymbol{p}_2'$, the spatial relationship "is left of" is inconsistent. This relationship is often denoted as *sidedness* or *ordering constraint*. In most existing algorithms, it is modelled as a hard constraint, based on a tolerance of a few pixels. We propose to implement a statistical test instead, which takes the orientation accuracy into account, leading to a third, binary-valued inconsistency measure

$$t_{nm}^s \in \{0, 1\} . \qquad (5)$$

For example, in Fig. 4, we would obtain $t_{12}^s = 0$ indicating that the sidedness relations of v_1 and v_2 are identical in both images. Accordingly,

the inconsistency of v_1 and v_3 would produce $t_{13}^s = 1$. How to determine the sidedness is explained in more detail in DICKSCHEID (2010, section 4.3.5).

4 A Trainable Markov Random Field for Feature Matching

4.1 Feature Matching as a Labelling Problem

Our goal is to find the most probable matching under a number of reasonable assumptions. Our method is based on the assumption that it is easy to obtain an initial set $\mathcal{V} = \{v_1, \cdots, v_N\}$ of putative feature correspondences that contains most of the true positives. This could be the set

$$\mathcal{V}^0 = \{(\boldsymbol{p}_i, \boldsymbol{p}'_j) \mid \boldsymbol{p}_i \in \mathcal{P}, \boldsymbol{p}'_j \in \mathcal{P}', \lambda_i = \lambda'_j\} \quad (6)$$

of all correspondences between features of the same type. In practice however, a significantly smaller set $\mathcal{V} \subset \mathcal{V}^0$ can be used which still contains the majority of true correspondences. It has been shown empirically in DICKSCHEID (2010, section 2.4) that an effective way to do so is to use the BESTMATCH-K algorithm, with the matching rank k varying for each type of detector and descriptor.

Feature matching can then be interpreted as the selection of a good subset of $\mathcal{V}$ by assigning a label l from the set $\mathcal{L} = \{0, 1\}$ to each element in $\mathcal{V}$. Then each element v_n of $\mathcal{V}$ takes the role of a binary random variable defined over the set $\mathcal{L}$. If $l_n = 1$, we say that "match n is selected", otherwise "match n is discarded". For simplicity we use the notation v_n for denoting the particular event $v_n = l_n$. A labelling

$$\boldsymbol{l} = \boldsymbol{f}(\mathcal{V}) = \{v_1, \cdots, v_N\} \quad (7)$$

of all variables is a *configuration*. The principle of interpreting feature matching as a labelling problem is illustrated by an artificial example in Fig. 4.

4.2 Statistical Model with Pairwise Spatial Relationships

Appearance-based matching with descriptors, as described in section 2, computes for each putative match $v_n \in \mathcal{V}$ the dissimilarity
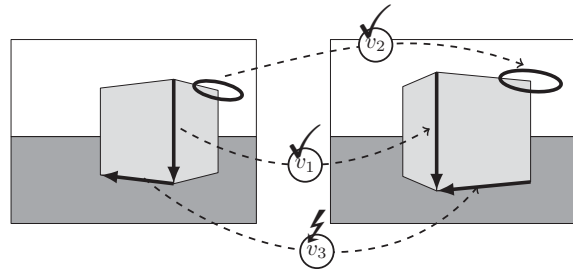
Fig. 4: Two artificial images of a scene with a cube, overlaid by three features represented by black ellipses and arrows. The features may lead to a set of three putative matches $\mathcal{V} = \{v_1, v_2, v_3\}$. The task is then to select an optimal subset of $\mathcal{V}$ by labelling each putative match. In this example, the labelling $\boldsymbol{f}(\mathcal{V}) = \{v_1 = 1, v_2 = 1, v_3 = 0\}$ is the desired solution, eliminating the spatially inconsistent match v_3.

$s_n \in \mathbb{R}$, which we collect in the vector $\boldsymbol{s} = [s_1, \cdots, s_N]$. It then treats the decision about a match v_n without taking any spatial relations into account. In the following, we use the index set $\mathcal{N} = \{1, \cdots, N\}$ over $\mathcal{V}$ and consider pairs of correspondences defined by index pairs $\mathcal{C}_2 \subseteq \mathcal{N} \times \mathcal{N}$.

At this point, and without loss of generality, we will ignore that features and descriptors might be of different type. We will come back to this aspect again in section 5.

Considering the set $\mathcal{T}^2 = \{\boldsymbol{t}_{nm} | (n, m) \in \mathcal{C}^2\}$ of pairwise spatial inconsistencies reveals statistical dependencies between the variables. Note that $|\boldsymbol{t}_{nm}| = G$ is the number of different spatial relationships used. For example, the inconsistency t_{13}^s referring to "sidedness" between v_1 and v_3 is high for the two putative correspondences v_1 and v_3 in Fig. 4. We must therefore expect that one of them is an outlier, even though the descriptor dissimilarities s_1, s_2 and s_3 might be small. In other words, after observing the spatial inconsistency, it would be naive to make independent decisions for each putative match.

In a statistical treatment, we would say that the random variable v_1 is now dependent on v_2. It also depends on its directly related observations, s_1 and t_{12}^s. This can be expressed by an undirected graphical model as shown by the *factor graph* in Fig. 5. Each white node of this graph represents one putative feature correspondence, while shaded nodes represent ob-
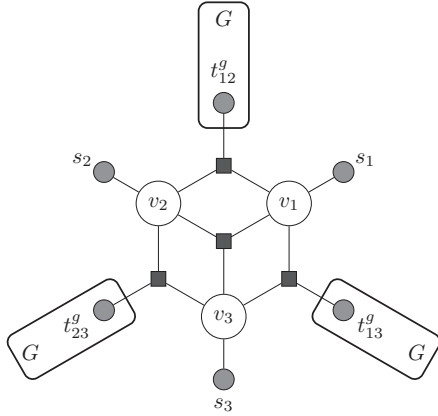
Fig. 5: Graph representing the information used for feature matching corresponding to the illustration in Fig. 4. Both descriptor dissimilarities s_n and inconsistencies t_{nm} of pairwise spatial relationships are taken into account. We follow the notation in BISHOP (2006): Observed values are represented by shaded nodes, maximum cliques (fully connected subgraphs) with more than two nodes are represented by a common black rectangle (*factor graphs*), and *plates* are used to illustrate multiple independent spatial inconsistency measures t_{nm}^g, with G being the number of different types of spatial relationships. The factor node in the centre expresses the joint prior probability for all v_n.

served dissimilarity measures. Here, values s_n refer to descriptor dissimilarities, which relate to one correspondence, while values t_{nm} refer to spatial inconsistency measures, defined over pairs of correspondences. Statistical dependencies between entities are represented by the edges of the graph. The black squares denote maximum cliques - fully connected subgraphs where all elements are conditionally dependent on each other.

In general, we make the following conditional independence assumptions:
1. All observations are mutually conditionally independent, given the correspondences.
2. The label of a putative match v_n does not depend on observations s_m or $t_{m,o}$ with $m, o \neq n$.

This model nicely supports our practical setup, as its statistical dependencies can be derived from training data.

We propose to solve the labelling problem by computing the maximum a posteriori estimate (MAP) of the variables in this model, given the observed data. Referring to the simple example

in Fig. 4, this can be done by maximizing the probability

$$p(v_1, v_2, v_3, s_1, s_2, s_3, \boldsymbol{t}_{12}, \boldsymbol{t}_{13}, \boldsymbol{t}_{23}) \qquad (8)$$

$$= P(v_1, v_2, v_3) \left[\prod_{n=1}^{3} p(s_n | v_n) \right]$$

$$\cdot p(\boldsymbol{t}_{23} | v_2, v_3) p(\boldsymbol{t}_{13} | v_1, v_3) p(\boldsymbol{t}_{12} | v_1, v_2)$$

$$= P(v_1, v_2, v_3) \left[\prod_{n=1}^{3} p(s_n | v_n) \right]$$

$$\cdot \prod_{g=1}^{G} p(t_{12}^g | v_1, v_2) p(t_{13}^g | v_1, v_3) p(t_{23}^g | v_2, v_3)$$

The first simplification is obtained by exploiting our conditional independence assumptions, using the fact that $p(a|b,c) = p(a|b)$ in case that a is conditionally independent of c. The expansion of $\boldsymbol{t}_{nm}$ in (8) also uses the assumption that all observations are mutually independent.

For a general problem with $|\mathcal{V}| = N$ putative matches, we will obtain a graph having N binary cliques (v_n, s_n) and $G|\mathcal{C}_2|$ ternary cliques (v_n, v_m, t_{nm}^g). The general joint probability of the variables reads

$$p(\boldsymbol{l}, \boldsymbol{s}, \mathcal{T}_2) \qquad (9)$$

$$= p(v_1, \dots, v_N, s_1, \dots, s_N, \boldsymbol{t}_{12}, \dots, \boldsymbol{t}_{(N-1)N})$$

$$= P(\boldsymbol{l}) \left[\prod_{n \in \mathcal{N}} p(s_n | v_n) \right] \prod_{\substack{(n,m) \\ \in \mathcal{C}_2}} p(\boldsymbol{t}_{nm} | v_n, v_m).$$

The factor $P(\boldsymbol{l})$ in (9) imposes a practical problem: It depends directly on the number of putative matches, which is unknown in advance. To make the formulation tractable for practical problems, we therefore make the following assumption for the joint probability:

$$P(\boldsymbol{l}) = P(v_1, \dots, v_N) \doteq \frac{1}{Z'} \prod_{\substack{(n,m) \\ \in \mathcal{C}_2}} P(v_n, v_m)$$

$$(10)$$

It leads to a significant simplification of the model, which now reads

$$p(\boldsymbol{l}, \boldsymbol{s}, \mathcal{T}_2) = \frac{1}{Z'} \left[\prod_{n \in \mathcal{N}} p(s_n | v_n) \right] \quad (11)$$

$$\cdot \prod_{\substack{(n,m) \\ \in \mathcal{C}_2}} p(\boldsymbol{t}_{nm} | v_n, v_m) P(v_n, v_m) \ .$$

As we will see later, we do not require specific knowledge about the partition function Z', because it does not affect the final solution.

By going from (9) to (11), we make an explicit model assumption. This leads to a restricted stochastic model which still corresponds to the original graphical model. The model assumption effectively drops the higher order cliques between putative matches v_n (i.e. the clique (v_1, v_2, v_3) in Fig. 5) in favour of a change of the pairwise potential functions (last factor of (11)).

4.3 *Global Minimization Problem*

By maximizing the density function (11) we realize a MAP estimate of the involved variables. This is equivalent to minimizing the energy function

$$E(\boldsymbol{l}, \boldsymbol{s}, \mathcal{T}_2) = - \sum_{n \in \mathcal{N}} \log p(s_n | v_n) \quad (12)$$

$$- \sum_{\substack{(n,m) \\ \in \mathcal{C}_2}} \left[\log P(v_n, v_m) + \log p(\boldsymbol{t}_{nm} | v_n, v_m) \right]$$

where we omit the term $1/Z'$ of the partition function, as it does not affect the solution. It is essentially a sum over functions of unary and binary cliques over $\mathcal{V}$, given the graphical model, and can therefore be directly interpreted as a Markov Random Field. Defining unary potentials $\boldsymbol{\theta}^1_{n;v_n}$ and binary potentials $\boldsymbol{\theta}^2_{nm;v_n v_m}$ as

$$\boldsymbol{\theta}^1_{n;v_n} = - \log p(s_n | v_n) \quad (13)$$

$$\boldsymbol{\theta}^2_{nm;v_n v_m} = - \log p(v_n, v_m) \quad (14)$$
$$- \log p(\boldsymbol{t}_{nm} | v_n, v_m) \ ,$$
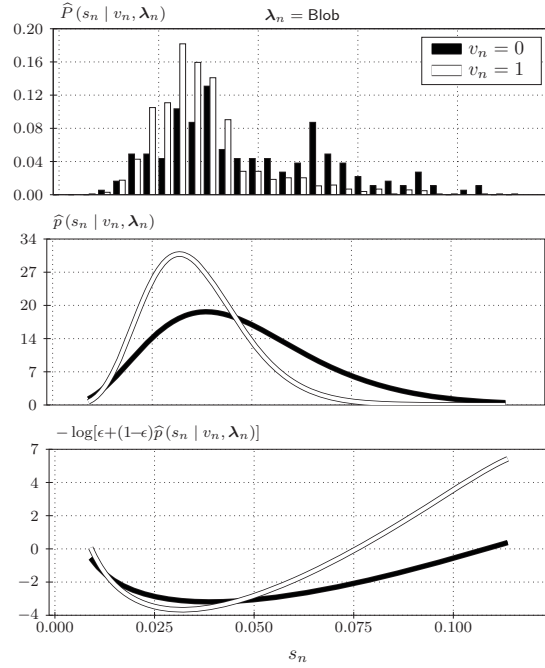
Fig. 6: *Top:* Normalized histograms of dissimilarities s_n for good ($v_n = 1$) and bad ($v_n = 0$) blob feature correspondences, computed by normalized Euclidean distances of SIFT descriptors. *Middle:* Approximation of the histograms using a Beta distribution, which is used as a parametric approximation $\widehat{p}(s_n | v_n, \boldsymbol{\lambda}_n)$ of the likelihood function. *Bottom:* Bounded negative log likelihood derived from $\widehat{p}(s_n | v_n, \boldsymbol{\lambda}_n)$, which we use for the energy potentials. The observations refer to the training dataset (section 6.1). Note that the theoretical range of the observations is $(0, 1)$, and that the Beta distribution is defined over the range $[0, 1]$. Here we only plot the relevant range; the densities are practically zero above $s_n \simeq 0.13$. The bound ϵ within the logarithm theoretically prevents the log likelihoods from reaching a limiting value for very rare values $s_n > 0.13$.

we can use the notation

$$E(f(\mathcal{V}), \boldsymbol{s}, \mathcal{T}_2; \boldsymbol{\theta}) \quad (15)$$
$$= \sum_{n \in \mathcal{N}} \boldsymbol{\theta}^1_{n;v_n} + \sum_{\substack{(n,m) \\ \in \mathcal{C}_2}} \boldsymbol{\theta}^2_{nm;v_n v_m} \ ,$$

which is very common in MRF theory. Note that we used $\boldsymbol{l} = f(\mathcal{V})$ according to (7). To find a good solution for the matching problem, given an initial set $\mathcal{V}$ of putative matches and observations $\mathcal{D} = \{\boldsymbol{s}, \mathcal{T}_2\}$, we finally search for a configuration with minimum energy (12),

so we look for an optimal solution

$$f^*(\mathcal{V}) = \underset{f(\mathcal{V})}{\arg\min}\ E(f(\mathcal{V}), \boldsymbol{s}, \mathcal{T}_2; \boldsymbol{\theta})\ . \quad (16)$$

We actually apply the LP-S linear programming relaxation going back to SCHLESINGER (1976) to solve (16), which gives a very good approximation of the global optimum (KUMAR et al. 2009). We use the commercial Mosek package (http://www.mosek.com) for solving the relaxed minimization problem, which provides an efficient implementation of the interior point algorithm for linear programs with up to thousands of variables. The practical complexity is polynomial with very good convergence properties. For sets $\mathcal{V}$ of putative correspondences with $N = |\mathcal{V}| < 500$, we usually solve the matching problem in a few seconds on a 2.4 GHz CPU. For sparsely textured scenes, N is typically smaller than 200, leading to negligible computation times for obtaining the optimal solution.

5 Learning the Potential Functions

Ground truth labellings for the data are obtained based on homographies (MIKOLAJCZYK et al. 2005), manually, or using the surface-based automatic annotation setup described in DICKSCHEID (2010, section 5). The latter one uses 3D point clouds from Laser scans as ground truth for the surfaces, which requires registration of the point clouds to the camera coordinate systems. The methods are indicated in the first row of Tab. 1.

For minimizing (15) one basically has to collect the potentials (13) and (14) for each node of the corresponding graph and feed them into the software. They are composed of the likelihoods of observed values, given the unknown labels and the prior probability $P(v_n, v_m)$. We will now derive some trainable parametric models for these potentials using the setup of detectors and descriptors described in section 6.1 and the spatial relationships introduced in section 3.3. We used 24 image pairs from indoor and outdoor architectural scenes for the training (section 6.2), where ground truth labellings of feature correspondences are computed using the setup described in DICKSCHEID (2010, section 5). For model fitting from ground truth data
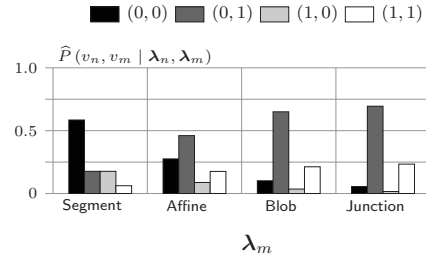
Fig. 7: Empirical fraction of pairs of putative matches, where the first match refers to straight line segments ($\boldsymbol{\lambda}_n = $ Segment) for different labellings l_n, l_m as observed on the training data. We obtain four groups, referring to the events $(v_n = 0, v_m = 0)$, $(v_n = 1, v_m = 0)$, $(v_n = 0, v_m = 1)$ and $(v_n = 1, v_m = 1)$. For example, if we select two matched line segments and two matched junction features from the set of putative matches, chances are around 70 percent that the line segment match is an outlier (label 0) according to the group of bars shown on the right.

we used the Statistics module of the Boost C++ Library's Math Toolkit (http://www.boost.org). We also tried independent datasets of the same image category and arrived at the same models with slightly different parameters. We did not investigate the dependence of the models on different image categories, so the potential functions that we present here should be considered as one particular realization of the algorithm with a focus on man-made scenes. We will evaluate this realization in section 6.

Dependency on the feature type

Our main motivation is to combine complementary feature types, descriptor dissimilarity measures with significantly different properties, and a whole range of spatial relationships simultaneously. Then the likelihoods depend formally on the type of feature λ^F, the type of descriptor λ^D, and the type of dissimilarity measure λ^M. We collect these for each putative match v_n in a tuple $\boldsymbol{\lambda}_n = (\lambda_n^F, \lambda_n^D, \lambda_n^M)$. The likelihoods defining the unary and binary potentials therefore actually read $p_n(s_n|v_n, \boldsymbol{\lambda}_n)$ and $p(t_{nm}|v_n, v_m, \boldsymbol{\lambda}_n, \boldsymbol{\lambda}_m)$. The prior becomes $P(v_n, v_m, \boldsymbol{\lambda}_n, \boldsymbol{\lambda}_m)$ accordingly.

5.1 *Unary Potentials*

Take a look at the normalized histogram on top of Fig. 6. It shows the dissimilarities of good

Tab. 1: Properties of the datasets used for our experiments.

Dataset	CLASS	BOAT	GRAFFITI	BLANK-12	BLANK-22	DRAGON
Ground Truth	manual	homography	homography	manual	manual	surface-based
Texture	sparse	strong	strong	very sparse	very sparse	sparse
3D structure	multiplanar	quasiplanar	planar	multiplanar	multiplanar	complex
Distortion	affine	rotation+scale	strong affine	affine	affine	affine
Overlap	$\sim 60\%$	$\sim 100\%$	$\sim 100\%$	$\sim 90\%$	$\sim 90\%$	$\sim 100\%$
# Images	8	6	6	12	22	6
Resolution	752×500	213×170	213×170	1203×800	752×500	800×600
Example						

$(v_n = 1)$ and bad $(v_n = 0)$ blob feature correspondences, referring to normalized Euclidean distances of SIFT descriptors.

Due to the normalization, the histogram shapes can be reasonably approximated by a Beta distribution $\mathrm{Beta}(s_n|a, b)$. We estimate its two parameters from training data (section 6.2) separately for the inlier and outlier distributions to obtain estimates for the class conditional likelihood functions $\widehat{p}(s_n|v_n = 0, \boldsymbol{\lambda}_n)$ and $\widehat{p}(s_n|v_n = 1, \boldsymbol{\lambda}_n)$, as shown in the middle of Fig. 6 for blob features.

The negative log likelihood $-\log \widehat{p}(s_n|v_n, \boldsymbol{\lambda}_n)$ that we actually use in the energy function (12) is shown in the bottom plot of Fig. 6. Note that we introduce a bound on the log likelihood by using $-\log[\epsilon + (1 - \epsilon)\widehat{p}(s_n|v_n, \boldsymbol{\lambda}_n)]$ with a small threshold $\epsilon = 0.001$. In practice, the bound only affects values s_n very close to the limits of the domain $[0, 1]$, which occur very rarely in practice.

We also model the dissimilarity likelihoods for other features by Beta distributions, as described in (DICKSCHEID 2010, section 4.3.3).

5.2 Binary Potentials

Priors

For each feature in one image, we preselect the k most similar features in the other image as its putative matches, where the parameter k differs between feature types $\boldsymbol{\lambda}$. For example, we select more putative matches per feature for straight line segments than for blob features, following the empirical investigations in

DICKSCHEID (2010, section 2.4). We must therefore expect different prior probabilities $P(v_n, v_m, \boldsymbol{\lambda}_n, \boldsymbol{\lambda}_m)$. The relative frequencies within the training data for pairs of matches, where the first match is a line segment, are shown in Fig. 7. Indeed we see the strong influence of different preselection criteria per feature type on the prior: For pairs containing one line segment match and one match of another type, it is most likely that the line segment match is an outlier. This reflects the fact that k is largest for the line segments. As $P(v_n, v_m, \boldsymbol{\lambda}_n, \boldsymbol{\lambda}_m)$ is a discrete probability, we can model it as a binomial distribution and use those relative frequencies within the binary potentials $\boldsymbol{\theta}^2_{nm;v_n v_m}$.

Likelihoods for pairs of correspondences

As an example for the likelihood $p(t_{nm}|v_n, v_m, \lambda_n, \lambda_m)$, we discuss the inconsistency t^α of angles spanned by two features (section 3.3). It is obvious that we can neither expect angles between pairs of correct matches to be always equal, nor angles between outliers to be always largely different. When investigating the empirical distribution of the consistency measures t^α on our training dataset, we see that they carry valuable information for our problem, though. The distribution for pairs of blob and junction feature matches is shown in the top row of Fig. 8. As in case of the unary potentials, we modelled the likelihoods using a Beta distribution, which corresponds strongly to the empirical distributions.

The distribution indicates that for small inconsistencies t^α between feature correspon-
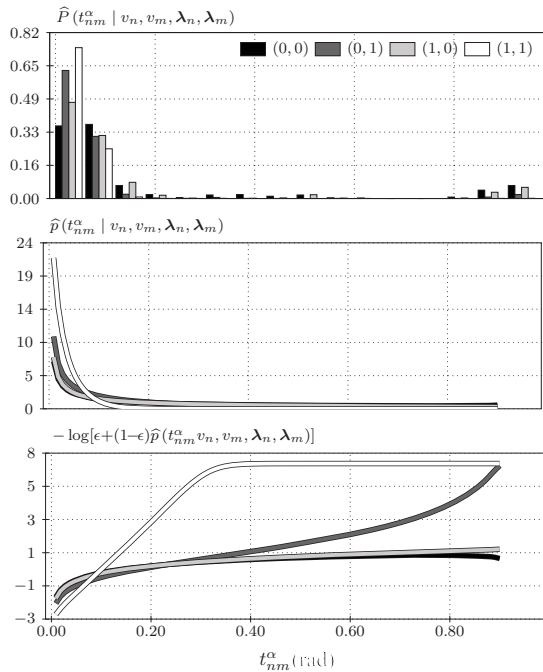
Fig. 8: *Top:* Normalized histograms of observations t_{nm}^{α} between blob and junction feature correspondences, denoting inconsistency of angles between pairs of oriented features. We obtain four distributions, referring to the events $(v_n = 0, v_m = 0)$, $(v_n = 1, v_m = 0)$, $(v_n = 0, v_m = 1)$ and $(v_n = 1, v_m = 1)$. *Middle:* Approximation of the histograms by Beta distributions, used as an estimate for the likelihood $p(t_{nm}^{\alpha} \mid v_n, v_m, \boldsymbol{\lambda}_n, \boldsymbol{\lambda}_m)$. *Bottom:* Bounded negative log likelihood derived from $\widehat{p}(t_{nm}^{\alpha} \mid v_n, v_m, \boldsymbol{\lambda}_n, \boldsymbol{\lambda}_m)$, which we use for the energy potentials. The observations refer to the training dataset. Note that the theoretical range of the observations is $(0, 1)$, and that the Beta distribution is defined over the range $[0, 1]$. Here we only plot the range of values that we observed on the training dataset.

Fig. 9: Visual matching results for an image pair of the CLASS dataset for the three methods described in section 6.1. Features depicted in white are correctly matched, features in grey are outliers. We see that the simple BESTMATCH-2 approach gives quite many outliers, especially among the line segments which have the weakest descriptors. Using a topological filter and boost stage (TOPOMATCH) removes a significant number of those. The results for our approach (MAPMATCH) contain more inliers, and at the same time the lowest outlier rate. Detailed results from more image pairs of the dataset are listed in Fig. 10.

dences of this type, it is most likely that both matches are inliers, referring to this observation only. Hence, stronger feature types will implicitly motivate the selection of weaker ones when the angular consistency is high. With increasing inconsistency, it becomes more probable that the blob correspondence is an outlier. This corresponds strongly to our initial assumptions. For very high inconsistencies, the labelling $(0, 0)$ is motivated, which means that both correspondences are likely to be outliers.

Very similar observations can be made for other combinations of feature types. The Beta distribution is a good continuous model for all real-valued dissimilarity and inconsistency measures that we investigated, however, other measures might require a different model.

6 Experiments

We will show that the feature matching algorithm with the parametric models derived in section 5 allows for significantly better matching results on sparsely textured scenes than the standard best-matching approach (BESTMATCH-2), which only takes descriptor dissimilarities into account. We also want to make sure that our results are comparable to the results obtained with the method of (BAY et al. 2005), which is specifically designed for sparsely textured scenes. We search for a matching algorithm that maximizes the number

of correct correspondences while not exceeding a critical outlier rate. Focussing on typical image orientation algorithms with a RANSAC scheme, the critical rate is at about 50%. So, if one matching algorithm returns 20 correct correspondences with no outliers, and another one returns 50 correct correspondences with 15 outliers, we consider the latter one to be better.

6.1 *Experimental Setup*

Intentionally we select a set of popular feature detection algorithms with high complementarity, and use descriptors with different distinctiveness and invariance properties in order to demonstrate the potential of our method. By keeping the set of detectors and descriptors together with their parameter settings fixed, all methods shown here have to cope with the same strengths and shortcomings of the features. Nevertheless, our experiments must not be understood as a comparison of detectors, but as a comparison of wide baseline matching methods.

Detectors

The classical LOWE blob detector (LOWE 2004) is based on the Laplacian and known to have very good scale and rotation invariance. We use the original implementation kindly provided by the author, starting with the original instead of the double image resolution for building the scale space pyramid. The FOP0 detector extracts junction points using the framework of FÖRSTNER (1994). These features are not scale invariant, and also sensitive to affine distortions. We use the original implementation of the author, with a manually determined but fixed estimate of 0.015% for the standard deviation of the image noise, related to an intensity range of 1. The MSER detector of MATAS et al. (2004) detects segmentation regions with complex shape. We use the widely used implementation provided by MIKOLAJCZYK et al. (2005), however approximate the local patches by a circular shape covering an equally sized area around the same centroid. As the affine invariance that MSER itself is able to produce gets lost hereby, we denote the features as MSER° instead. The EDGE detector from the framework of FÖRSTNER (1994) is included as a typical straight line segment detector.

Descriptors

For all but the EDGE features, we use SIFT descriptors computed using the original software provided by LOWE (2004). The feature orientations are taken from the dominant gradient orientation that is assigned to the descriptor. Descriptors for the FOP0 points are computed with a fixed window size of $3s = 12$(pixel). The straight line segments are coupled with our own implementation of the colour-histogram based descriptors of BAY et al. (2005). *These descriptors are significantly less distinctive than the SIFT descriptors.* Our implementation has been carefully compared to the implementation of the authors and leads to very similar results. Following BAY et al. (2005), the orientation of the line segments is defined by choosing the side with brighter image intensities to be left of the segment.

Matching Strategies

We show results for three different wide baseline stereo matching algorithms. The simplest and most common one is a classical descriptor-based best matching approach (BESTMATCH-2) with a 70% threshold, as described in section 2. Furthermore, we use a reimplementation of the method proposed by BAY et al. (2005), which will be denoted as TOPOMATCH in the following. It includes both the three-point- and the point-line topological filtering stages described in the paper, as well as the boosting step. *Although we reimplemented the method carefully, we cannot claim that the results apply directly to the original implementation of the authors.* Our own method is denoted as MAP-MATCH in the following.

6.2 *Image Data*

For training the likelihood functions and priors, we use observations measured from 24 pairs of images showing indoor and outdoor architectural scenes, which remain constant across all experiments. The training images are not part of the evaluation, i.e. we do not perform image specific training.

We show results based on five different datasets. The properties of the datasets are summarized in Tab. 1. The CLASS, BLANK-12
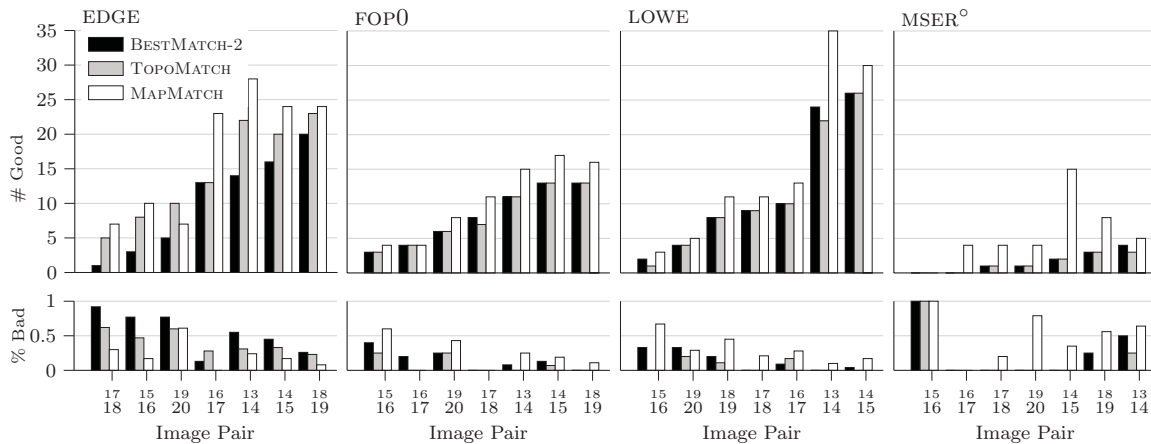
Fig. 10: Matching results for all neighbouring image pairs of the CLASS dataset, computed with the three wide baseline stereo matching algorithms described in section 6.1. Shown are the number of correct correspondences and the percentage of outliers for each feature type. The annotation has been done manually. We see that our approach (MAPMATCH) most often yields higher numbers of inliers than the others at slightly higher but acceptable outlier rates.

and BLANK-22 datasets used a fisheye lens, and have been corrected for radial distortion. As the complexity of the TOPOMATCH and MAPMATCH methods is too high for processing high resolution images with strong texture, we have downsampled the BOAT and GRAFFITI datasets (MIKOLAJCZYK et al. 2005) to reduce the amount of features in a mostly natural way.

6.3 *Results*

For investigating the success of a method referring directly to the extracted feature correspondences, we report the number of good correspondences (inliers) and the percentage of outliers for each matched image pair. Although we report the statistics separately for each feature type, the matching has been performed on all feature types simultaneously.

Sparsely textured datasets

Referring to the datasets with sparse texture, our approach MAPMATCH shows mostly superior matching results. The image pair of the CLASS dataset depicted in Fig. 9 provides a visual impression of the matching results for the different methods. The BESTMATCH-2 approach, relying only on descriptors, cannot compensate the weakness of the line segment descriptors, which results in many outliers among the line segment correspondences. Using the topological filter in the TOPOMATCH

method filters many of those outliers, but does not lead to a higher number of point feature correspondences. The MAPMATCH approach (bottom) achieves both effects quite well.

Fig. 10 shows detailed results for more image pairs of the CLASS dataset. Our approach yields a consistently higher number of inliers. The results for straight line segments are especially notable, as our algorithm also produces the overall smallest outlier rates. For other feature types however, it tends to have higher outlier rates than other methods.

For the BLANK-12 dataset Fig. 11, one obtains similar observations. The number of inliers is significantly higher for MAPMATCH over all considered image pairs and feature types, while the outlier rates are acceptable, sometimes even better than for the other two methods. In particular, MAPMATCH would allow to compute the epipolar geometry of the third pair $6/9$ quite robustly, with a total of 36 correct point matches (ignoring the line segments), while TOPOMATCH with 6 point matches is clearly at the borderline, and BESTMATCH-2 with 21 point matches significantly weaker. The TOPOMATCH implementation does not yield significantly more inliers than BESTMATCH-2, but has lower outlier rates. This is intuitive, considering that it removes matches with inconsistent spatial relationships.
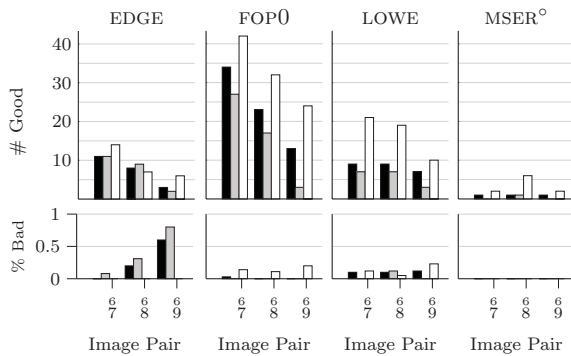
Fig. 11: Results for three image pairs with increasing baseline taken from the BLANK-12 dataset. The number of inliers is significantly higher for MAPMATCH, while the outlier rates are still good, sometimes also better than for the other two methods. In particular, MAPMATCH would allow to compute the epipolar geometry of the third pair $6/9$ quite robustly, with a total of 36 correct point matches (ignoring the line segments), while TOPOMATCH with 6 point matches is clearly at the borderline, and BESTMATCH-2 with 21 point matches significantly weaker.

Fig. 12: Matching results for all image pairs containing the first image of the BOAT dataset, computed with the three wide baseline stereo matching algorithms described in section 6.1. The annotation has been done based on plane homographies, which works only for point features. The image scale and rotation difference per image pairs increases significantly from left to right. Due to the strong scale differences, the number of FOP0 correspondences is so small for all methods that we don't show them here.

Standard datasets

The results for the BOAT dataset Fig. 12 show that our approach yields comparable results to the classical BESTMATCH-2. Note that here the image pairs are sorted by increasing scale and rotation difference between the images. For strong distortions, MAPMATCH yields more inliers than the BESTMATCH-2 approach, at the cost of a slightly higher outlier rate. Nevertheless it has a tendency to extract too many outliers at times, as can be seen in case of the affine region features for image pairs 1/4 in Fig. 12, and in case of the blobs for pair 1/6. The TOPOMATCH approach yields very similar results to BESTMATCH-2, with a tendency to extract even less matches. Note that although the line segments were used for matching in all of our experiments, they are not listed for the GRAFFITI and BOAT dataset, as the homography-based annotation cannot evaluate them automatically.

Results for straight line segments

The straight line features play a special role, as the matching of lines is in general more difficult due to the uncertainty of the location of the start-/endpoints, and in particular more difficult due to the weak descriptors used here.

On the investigated datasets, the MAPMATCH approach shows better results than both other methods referring to the line segments. At the same time, the TOPOMATCH method often shows better results for matching lines than BESTMATCH-2. We can therefore conclude that the spatial relationships seem to play indeed an important role for matching features with weak descriptors.

7   Conclusions and Future Work

It can be reasonable to use multiple complementary detectors in a vision system for increasing the amount of detected features. If the distinctiveness of feature descriptors is rather weak, as in case of sparsely textured scenes, or simple descriptors, spatial relationships provide important additional information for matching.

We have developed and described a generic method for modelling the matching problem with different detectors, descriptors and pairwise spatial relationships, which takes the structure of a binary classification problem and is consistent with MRF theory. Its potential functions have a clear statistical interpretation and can be trained from data using simple parametric models. The method therefore adapts
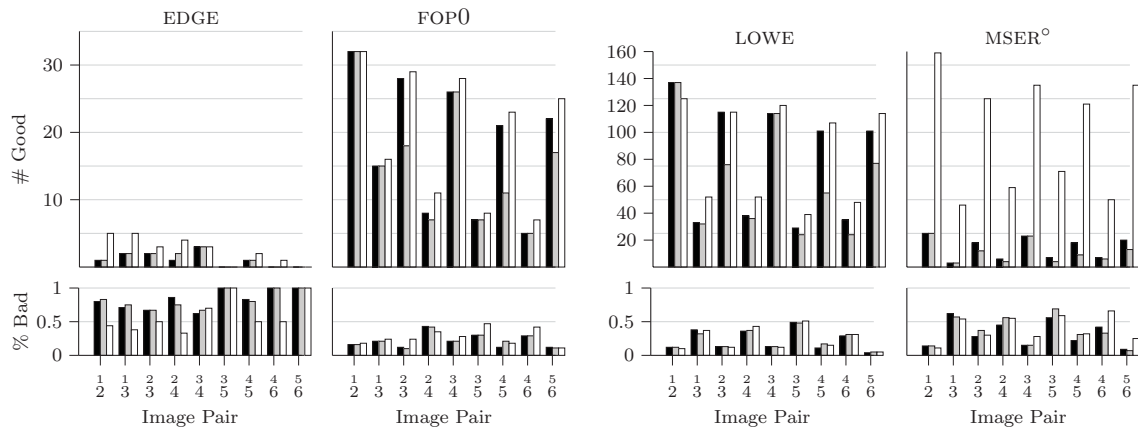
Fig. 13: Results for overlapping image pairs for the DRAGON dataset. The matching of EDGE features seems to be particularly difficult here for all three methods. The MAPMATCH approach solves it significantly better, though still not satisfyingly. For the other feature types, the MAPMATCH approach shows consistently better results in terms of higher number of inliers at comparable and satisfying outlier rates. Observe especially the affine blobs, where MAPMATCH extracts between 7 and 10 times more inliers, at a only slightly higher outlier rate.

very well to new matching problems and is straightforward to implement. We have implemented a particular instance of the algorithm which is able to produce better matching results on sparsely textured scenes compared to existing standard and specialized methods.

Simple descriptor-based matching is faster than our approach and still effective in case that many features are available. Therefore, we propose to fall back to this standard method when the amount of detected features is high.

Preselecting putative matches based on descriptor similarity is the most heuristic part of the proposed algorithm, so an investigation of more sophisticated criteria would be interesting. One may also achieve better results when choosing closer parametric approximations of the empirical likelihood distributions, perhaps by introducing mixture models. Finally, selecting more specific training images for particular matching problems can potentially give more accurate results. We have not yet investigated the effect of different training sets.

References

AGUILAR, W., FRAUEL, Y., ESCOLANO, F., MARTINEZ-PEREZ, M., ESPINOSA-ROMERO, A. & LOZANO, M., 2009: A Robust Graph Transformation Matching for Non-Rigid Registration. – Image and Vision Computing **27** (7): 897–910.

BARNARD, S.T. & THOMPSON, W.B., 1980: Disparity Analysis of Images. – IEEE Transactions on Pattern Analysis and Machine Intelligence.

BAY, H., FERRARI, V. & GOOL, L.V., 2005: Wide-Baseline Stereo Matching with Line Segments. – IEEE Conference on Computer Vision and Pattern Recognition **1**: 329–336, Washington DC, USA.

BISHOP, C.M., 2006: Pattern Recognition and Machine Learning. – Information Science and Statistics, Springer.

CHOI, O. & KWEON, I.S., 2009: Robust feature point matching by preserving local geometric consistency. – Computer Vision and Image Understanding **113** (6): 726–742.

DELPONTE, E., ISGRÒ, F., ODONE, F. & VERRI, A., 2006: SVD-matching using SIFT features. – Graphical models **68** (5-6): 415–431.

DICKSCHEID, T., 2010: Robust Wide-Baseline Stereo Matching for Sparsely Textured Scenes. – PhD thesis, Department of Photogrammetry, Institute of Geodesy and Geoinformation, University of Bonn.

DICKSCHEID, T., SCHINDLER, F. & FÖRSTNER, W., 2010: Coding Images with Local Features. – International Journal of Computer Vision 1–21: 10.1007/s11263-010-0340-z.

FÖRSTNER, W., 1986: A feature based correspondence algorithm for image matching. – International Archives of the Photogrammetry 26-3/3, ISP Commision III, Rovaniemi, Finland.

FÖRSTNER, W., 1994: A Framework for Low Level Feature Extraction. – European Conference on Computer Vision, Volume III: 383–394, Stockholm, Sweden.

FÖRSTNER, W., DICKSCHEID, T. & SCHINDLER, F., 2009: Detecting Interpretable and Accurate Scale-Invariant Keypoints. – IEEE International Conference on Computer Vision, Kyoto, Japan.

KUMAR, M.P., KOLMOGOROV, V. & TORR, P.H., 2009: An analysis of convex relaxations for map estimation of discrete mrfs. – Journal of Machine Learning Research **10**: 71–106.

LÄBE, T. & FÖRSTNER, W., 2006: Automatic Relative Orientation of Images. – 5th Turkish-German Joint Geodetic Days, Berlin.

LOWE, D.G., 2004: Distinctive Image Features from Scale-Invariant Keypoints. – International Journal of Computer Vision **60** (2): 91–110.

MATAS, J., CHUM, O., URBAN, M. & PAJDLA, T., 2004: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. – Image and Vision Computing **22**: 761–767.

MAYER, H., 2005: Robust Least-Squares Adjustment Based Orientation and Auto-Calibration of Wide-Baseline Image Sequences. – ISPRS Workshop BenCOS 2005: 11–17, Beijing, China.

MEIDOW, J., BEDER, C. & FÖRSTNER, W., 2009: Reasoning with uncertain points, straight lines, and straight line segments in 2d. – ISPRS Journal of Photogrammetry and Remote Sensing **64** (2): 125–139.

MIKOLAJCZYK, K., TUYTELAARS, T., SCHMID, C., ZISSERMAN, A., MATAS, J., SCHAFFALITZKY, F., KADIR, T. & GOOL, L.V., 2005: A Comparison of Affine Region Detectors. – International Journal of Computer Vision **65** (1/2): 43–72.

POLLEFEYS, M., KOCH, R., VERGAUWEN, M. & VAN GOOL, L., 2000: Automated Reconstruction of 3D Scenes from Sequences of Images. – ISPRS Journal Of Photogrammetry And Remote Sensing **55** (4): 251–267.

SCHELLEWALD, C. & SCHNÖRR, C., 2005: Probabilistic Subgraph Matching Based on Convex Relaxation. – International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'05) **3757**: 171–186, Springer.

SCHLESINGER, M., 1976: Sintaksicheskiy analiz dvumernykh zritelnikh signalov v usloviyakh pomekh (Syntactic Analysis of Two-Dimensional Visual Signals in Noisy Conditions). – Kibernetika **4**: 113–130.

SHAPIRO, L.G. & HARALICK, R.M., 1987: Relational matching. – Applied Optics **26** (10): 1845–1851.

SNAVELY, N., SEITZ, S.M. & SZELISKI, R., 2006: Photo tourism: exploring photo collections in 3d. – ACM SIGGRAPH **2006**: 835–846, New York, NY, USA.

TELL, D. & CARLSSON, S., 2002: Combining Appearance and Topology for Wide Baseline Matching. – European Conference on Computer Vision, Copenhagen: 68–81, Springer.

TORRESANI, L., KOLMOGOROV, V. & ROTHER, C., 2008: Feature correspondence via graph matching: Models and global optimization. – European Conference on Computer Vision: 596–609, Springer.

ULLMAN, S., 1979: The interpretation of visual motion. – MIT Press. Cambridge, MA.

Address of the Authors:

Dr.-Ing. TIMO DICKSCHEID, Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, D-56068 Koblenz, Tel.: +49-261-1306-0, Fax: +49-261-1306-5302, e-mail: dickscheid@bafg.de

Prof. Dr.-Ing. Dr. h.c. mult. WOLFGANG FÖRSTNER, Rheinische Friedrich-Wilhelms-Universität zu Bonn, Institut für Geodäsie und Geoinformation, Photogrammetrie, Nußallee 15, D-53115 Bonn, Tel.: +49-228-73-2713, Fax: +49-228-73-2712, e-mail: wf@ipb.uni-bonn.de

Article

DijkstraFPS: Graph Partitioning in Geometry and Image Processing

Falko Schindler & Wolfgang Förstner, Bonn

Summary: Data partitioning is a common problem in the field of point cloud and image processing applicable to segmentation and clustering. The general principle is to have high similarity of two data points, e.g. pixels or 3D points, within one region and low similarity among regions. This pair-wise similarity between data points can be represented in an attributed graph. In this article we propose a novel graph partitioning algorithm. It integrates a sampling strategy known as farthest point sampling with Dijkstra's algorithm for deriving a distance transform on a general graph, which does not need to be embedded in some space. According to the pair-wise attributes a Voronoi diagram on the graph is generated yielding the desired segmentation. We demonstrate our approach on various applications such as surface triangulation, surface segmentation, clustering and image segmentation.

Zusammenfassung: *DijkstraFPS: Graphpartitionierung in Geometrie und Bildverarbeitung*. Datenpartitionierung ist eine elementare Aufgabe im Bereich Punktwolken- und Bildverarbeitung, vor allem zur Segmentierung und zum Clustern. Das generelle Prinzip ist es, hohe Ähnlichkeit zwischen zwei Datenpunkten derselben Region und geringe Ähnlichkeit zwischen verschiedenen Regionen zu erreichen. Diese paarweise Ähnlichkeit kann als attributierter Graph auf den gegebenen Daten repräsentiert werden. In diesem Artikel stellen wir einen neuen Graphpartitionierungsalgorithmus vor. Er integriert eine Samplingstrategie namens Farthest Point Sampling mit dem Verfahren von Dijkstra zur Ableitung einer Distanztransformation auf einem allgemeinen Graphen, der nicht in einen Raum eingebettet sein muss. Gemäß der paarweisen Attribute wird ein Voronoi-Diagramm auf dem Graphen generiert, das die gewünschte Segmentierung liefert. Wir demonstrieren unseren Ansatz für verschiedene Anwendungen, wie die Oberflächentriangulierung, die Oberflächensegmentierung, das Clustering und die Bildsegmentierung.

1 Introduction

In the fields of point cloud and image processing many applications require the partitioning of the underlying data as a pre-processing step. For both, surface and image segmentation, a 2D manifold is to be partitioned into non-overlapping regions. But also line partitioning, reconstructing, i.e. triangulating, surfaces as well as clustering data points in high dimensional feature space involve partitioning the geometric or the feature space.

The number of possible partitionings of a dataset is extremely large. Already for a bi-nary partitioning of an image with N pixels into foreground and background one has 2^N possible partitionings. Therefore, no generally optimal technique for partitioning exists.

Methods for partitioning fall into two types. *Split and/or merge techniques* start from a dissimilarity and/or similarity measure within and/or between regions which are *iteratively* found by splitting the complete dataset and/or by merging the individual elements. A large number of partitioning techniques exists, e.g. based on quad- or octrees, normalized cuts (Shi & Malik 2000) – which are only optimal for one partitioning – or graph-cut based meth-

ods (BOYKOV & FUNKA-LEA 2006) – which are only optimal for binary partitioning and special similarity functions. Split and merge techniques are only describable by the local properties of the individual iteration steps and do not possess a global property.

Partitioning can be interpreted as *clustering* in a feature space, the features depending on the original data (FORSYTH & PONCE 2002). This immediately suggests to seek for modes of the density function induced by the features and finding the valley lines (COMANICIU & MEER 2002). The principle of *watershed algorithms* is the inverse (SZELISKI 2010): Regions are catchment areas bounded by the watershed lines of a gradient image, the gradient measuring the dissimilarity between neighbouring elements (VINCENT & SOILLE 1991, MEINE & KÖTHE 2005). The quality of the mean shift and watershed partitioning depends on the ability to define problem adequate features, why these methods often lead to oversegmentation, i.e. a too large number of regions requiring a subsequent merging step. However, both methods can be described by the global properties of their solution.

A segmentation is either used to reduce the complexity of subsequent algorithms as the number of basic elements usually is several orders of magnitudes larger than the number of regions, recently consequently named superpixels (VEKSLER et al. 2010, MESTER et al. 2011, ACHANTA et al. 2012). Alternatively, segmentation is understood as a first step towards a symbolic image description, where the regions are basic units for a subsequent image interpretation. In the last years this lead to the concept of semantic image segmentation, where the segmentation is understood and realized as supervised classification (ROTHER et al. 2004, ROSCHER 2012, ARBELAEZ et al. 2012).

Most techniques can be interpreted as graph partitioning. The graph is either given by the structure of the data as for digital images, or derived from the data by some similarity measure based on geometry alone as in point cloud processing, on feature similarity as in clustering, or on semantic closeness as in semantic segmentation.

We propose a new efficient split and merge type graph partitioning algorithm, which itera-

tively determines a set of Voronoi cells based on an application dependent metric. The strength of the algorithm lies in its ability to overwrite the partitioning of the previous step within the sequence of split and merge steps. Due to a careful choice of the similarity metric for the graph's edge attributes we are able to control the alignment of Voronoi edges according to our objective. We will describe how to choose the edge attributes for different applications like curve and surface reconstruction, curve and surface segmentation, clustering and image segmentation.

In section 2 we will recapitulate previous work on marching front based sampling and partitioning methods before we formulate a generalized graph partitioning algorithm in section 3. We demonstrate our graph partitioning method on various applications in section 4 and conclude in section 5.

2 Background and related Work

Our graph partitioning method is based on computing Voronoi diagrams on an edge attributed graph, the attributes containing some application dependent distance between two nodes. In discrete geometry there are two common algorithms for deriving a *distance map*: Dijktra's algorithm (DIJKSTRA 1959) and fast marching method (FMM, SETHIAN 1996).

Given one or more seed points, Dijkstra's algorithm computes the shortest path along *existing graph edges* from each vertex to its closest seed point yielding a distance map. Implicitly this yields path lengths, also called intrinsic or geodesic distances. Note that the *distance* is not necessarily the spatial distance, but the sum of all edge attributes along the path.

FMM, especially its formulation for meshed manifolds (KIMMEL & SETHIAN 1998), computes *surface intrinsic distances* on meshed surfaces. In contrast to Dijkstra's algorithm, however, paths can pass *through* triangles, thus are not restricted to triangular edges. Technically, both Dijkstra's algorithm and FMM solve the so-called Eikonal equation $|\nabla d(\mathbf{x})| = F(\mathbf{x})$ with the boundary condition $d(\mathbf{x}_0) = 0$. In terms of distances on meshed manifolds its interpretation is as follows: Given a seed point $\mathbf{x}_0$ and a function F defining the friction at

(a) Fast marching   **(b)** Dijkstra

(c) Adaptive friction   **(d)** Second front

(e) 20 iterations   **(f)** Final segmentation
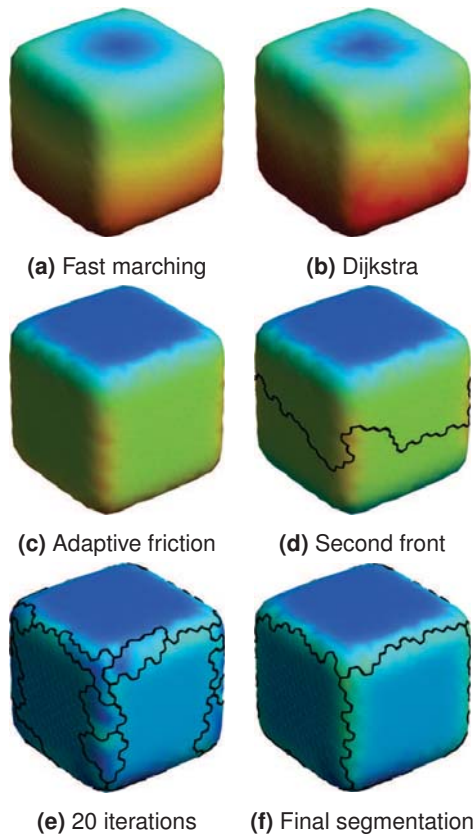
Fig. 1: (a): Demonstration of surface segmentation as proposed in SCHINDLER & FÖRSTNER (2011). Starting from a random seed point the geodesic distance to each other surface point is computed via fast marching (KIMMEL & SETHIAN 1998). (b): The distance is colour-coded from near (blue) to far (red). This distance map can be approximated with Dijkstra's algorithm. (c): Using a curvature adaptive friction function $F(\mathbf{x})$ we obtain small increases along planar regions and large gradients at sharp edges. (d): Adding the farthest point as new seed point and repeating Dijkstra's algorithm we obtain a segmentation into two segments. (e): The second wave front stops somewhere in the middle when newly computed distances exceed distances from the first iteration. After 20 iterations we clearly observe an oversegmentation. (f): To eliminate small segments along the edges we proposed a decremental segmentation step yielding the desired result with 6 segments only. The combination of incremental and decremental segmentation turned out to increase both robustness to data noise and independence of the random location of the first seed point (SCHINDLER & FÖRSTNER 2011).

each point $\mathbf{x}$ on the manifold, FMM yields the distance map $d(\mathbf{x})$ such that the gradient magnitude $|\nabla d(\mathbf{x})|$ is identical to the local friction $F(\mathbf{x})$. Then, the distance $d(\mathbf{x})$ is proportional to the arrival time of a propagating wave front starting at point $\mathbf{x}_0$ in case the local friction is inversely proportional to the local speed of the wave front. Both algorithms run in $O(N \log N)$ on sparse graphs, Dijkstra's algorithm of course only achieving an approximate solution.

Exploiting FMM for calculating exact distance maps on surfaces, MOENNING & DODGSON (2003a,b) propose a strategy for surface segmentation called fast marching farthest point sampling (FastFPS). It tries to find a set of *seed points* such that the corresponding Voronoi cells correspond to the desired partitioning of a surface. Their method starts from a random seed point $\mathbf{x}_0$. Every next seed point will be the point $\mathbf{x}^*$ with largest distance after updating the distance map d via FMM. After N iterations they obtain N seed points $\mathbf{x}_n$, $n = 1 \ldots N$, a distance map representing the distance to the closest of all N seed points and implicitly a Voronoi segmentation. PEYRÉ & COHEN (2004, 2006) apply a similar farthest-point strategy on meshed surfaces for segmentation, re-meshing and surface flattening using geodesic distances.

In SCHINDLER & FÖRSTNER (2011) we proposed FPS with Dijkstra's algorithm to robustly segment meshed surfaces (Fig. 1), taking into account the sub-optimality of Dijkstra's algorithm to generate a distance map. We also proposed an application specific stopping criterion to automatically determine a suitable number of seed points N. Such a stopping criterion can be found for other applications as well, but is out of scope of this paper. Further, we introduced a decremental segmentation strategy: We iteratively remove small segments by setting their vertex distances to infinity and re-run Dijkstra's algorithm starting from neighbouring segments.

Although FMM yields more accurate results, it is limited to manifolds. Dijkstra's algorithm approximates the geodesic distance, but for densely sampled manifolds the differences are small (Figs. 1a and 1b). More decisive, Dijkstra's algorithm does not require the un-

derlying graph to be embedded in some space. Thus, in case the graph is embedded, i.e. its nodes possess coordinates on a line, a surface or in a volume, one may partition the line, the surface or the volume according to the metric. Otherwise one may just partition the nodes of the graph according to distances encoded in the edges. This increases the flexibility of the proposed approach.

Here we transfer the idea of surface segmentation using FPS to a graph partitioning procedure. Therefore, we replace FMM by Dijkstra's algorithm to be able to handle more general graphs which are not embedded in some space. In the following, we present the novel algorithm, called DijkstraFPS, and demonstrate it with various applications in geometry and image processing.

3 Graph Partitioning

DijkstraFPS can be seen as a general graph partitioning algorithm, detached from the underlying semantics. We assume that the semantics of the partitioning problem can be coded in the distances between the nodes of a general graph. It does not need to be embedded in some space.

The procedure is given in Algorithm 1. It assumes, all pair-wise distances are pre-computed and encoded as *edge costs* between adjacent pairs of *nodes*.

The decremental segmentation step is almost identical. Only the FPS (lines 3–5 in Algorithm 1) are modified: Instead of initializing the front $\mathcal{Q}$ with the farthest point $s \leftarrow \operatorname{argmax}_n d_n$ with distance $d_s \leftarrow 0$ and label $l_s \leftarrow l^+$, we initialize the front $\mathcal{Q}$ with the neighbouring vertices n of the smallest segment l^- with distance $d_n \leftarrow \infty$ and label $l_n \leftarrow$ undefined. The *Dijkstra step* (lines 6–13) will propagate the front into the unlabelled region l^- and update all corresponding vertices.

Observe, only the distances of those vertices are updated which are closer to the current seed node (line 11), whose number diminishes with each additional seed node.

Fig. 2 illustrates the partitioning of a synthetic example graph with 7 nodes connected by 12 weighted edges. After three incremental Dijkstra steps (Figs. 2b–2d) one decremental step (Fig. 2e) is performed. Note that with

In: neighbours $\mathcal{N}$, costs F, #segments L_{inc}
Out: node distances **d**, node labels **l**
1   distances $\mathbf{d} \leftarrow \infty$, labels $\mathbf{l} \leftarrow$ undefined;
2   **for** $l^+ \leftarrow 1$ **to** L_{inc} **do**
3   pick new seed $s \leftarrow \operatorname{argmax}_n d_n$;
4   distance $d_s \leftarrow 0$ and label $l_s \leftarrow l^+$;
5   initialize new front $\mathcal{Q} \leftarrow \{s\}$;
6   **while** *front is not empty:* $\mathcal{Q} \neq \{\}$ **do**
7   select node $u \leftarrow \operatorname{argmin}_{q \in \mathcal{Q}} d_q$;
8   remove u from front $\mathcal{Q} \leftarrow \mathcal{Q} \setminus u$;
9   **foreach** *neighbour* $v \in \mathcal{N}_u$ **do**
10   new distance $d'_v \leftarrow d_u + F_u^v$;
11   **if** *new distance* $d'_v < d_v$ **then**
12   update $d_v \leftarrow d'_v, l_v \leftarrow l_u$;
13   add to front $\mathcal{Q} \leftarrow \mathcal{Q} \cup v$;

Algorithm 1: DijkstraFPS graph partitioning. New seed nodes *s* are added iteratively by choosing the farthest node w. r. t. node distances **d** (FPS, lines 3–5) that are constantly updated using Dijkstra's algorithm (lines 6–13) with edge costs F_u^v, yielding a labelling **l**.

DijkstraFPS graph partitioning all boundaries are possibly subject to change in a following Dijkstra step.

Within this paper we will restrict to undirected graphs only. Directed graphs, however, work as well, i.e. the propagation is cheaper in one direction than the other. Moreover, edge costs do not have to fulfill the triangle inequality of a metric space nor has the graph to be Euclideanly embeddable. Only negative costs are disallowed to avoid infinite loops during Dijkstra's algorithm.

Many applications suggest an embedding of the graph into a surface or a volume, leading to a graph based on a triangulation or tetrahedralization. Fig. 3 depicts two possible graph structures for a set of given 2D points. The triangle-based graph structure introduces one node per triangle (Fig. 3a). When labeling triangles via DijkstraFPS we obtain a segmentation boundary along triangular edges. The vertex-based graph structure introduces one node per vertex, i.e. 2D point, (Fig. 3b) yielding a segmentation boundary along Voronoi edges. In the following section 4 we will demonstrate both and point out when to use which structure.

(a) Synthetic graph with edge costs

(b) 1st Dijkstra step from random seed node

(c) 2nd Dijkstra step from farthest node from (b)

(d) 3rd Dijkstra step from farthest node from (c)

(e) Decremental step removing green segment
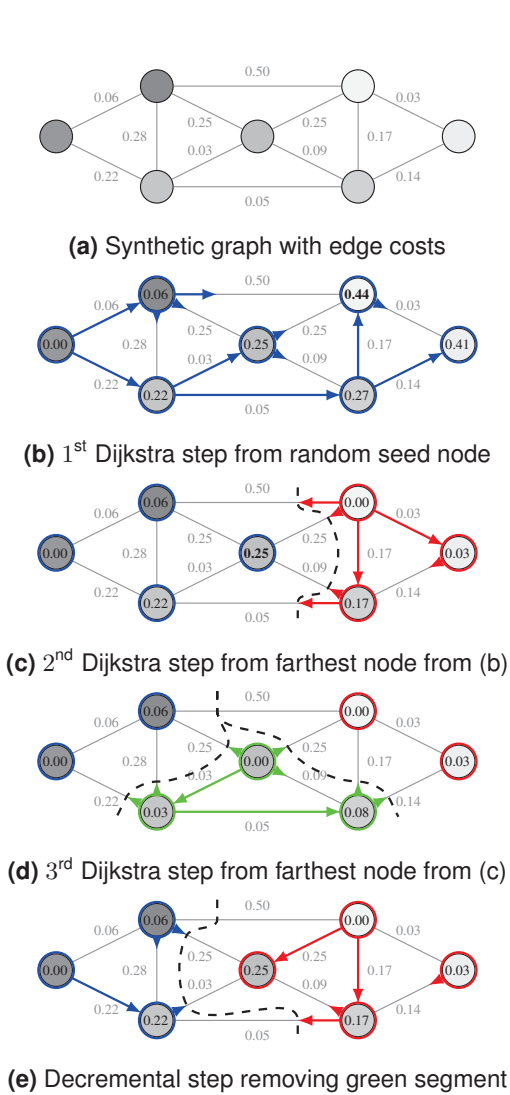
Fig. 2: Graph partitioning via DijkstraFPS. (a): Synthetic example graph with edge costs. The costs are computed as the absolute value of gray value differences. (b)–(d): Result of a Dijkstra step traversing from the very left node and incrementally summing up edge weights. The shortest path to each node is indicated by the blue arrows and leads to a distance printed within the circular nodes. The *farthest point* is the one with distance $d = 0.44$. A second Dijkstra step starts at this farthest node and updates node distances until an update would increase the distance value, e.g. to the left $0.00 + 0.50 > 0.06$. A third Dijkstra step starts at the centre node and again updates node distances creating a third segment shown in green. (e): Decremental step removing the green segment. All green distances are initialized to ∞ and Dijkstra starts propagating a front from neighbouring segments. The boundary slightly changes compared to (c).

(a) Triangle-based   **(b)** Vertex-based
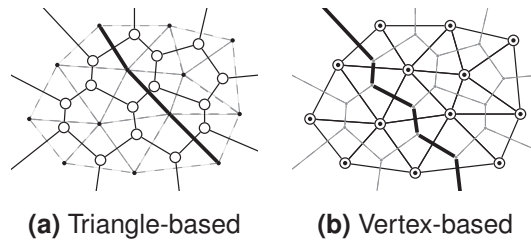
Fig. 3: Two possible graph structures for embedded graphs, here for 2D points (black dots). Graph nodes (circles) are either (a) Delaunay triangles or (b) Voronoi vertices. A possible segmentation boundary is shown as bold polyline along the Delaunay triangulation (a) or the Voronoi diagram (b).

4 Applications

The DijkstraFPS graph partitioning algorithm is applicable to many problems that can be expressed as attributed graph. In this section we will demonstrate four examples: *surface triangulation* by partitioning the 3D space and *curve reconstruction* by partitioning the 2D space, *surface segmentation* by partitioning a triangulated 2D manifold and *curve segmentation* by partitioning a 1D polyline, *clustering* by partitioning points in an nD – possibly non-metric – space, and *image segmentation* by partitioning pixels or superpixels of the 2D image plane.

In this article we ignore the possibility to formulate an application-specific stopping criterion (SCHINDLER & FÖRSTNER 2011). Instead, we will define the number of segments of the incremental (L_{inc}) and decremental (L_{dec}) segmentation manually.

4.1 Surface Triangulation

The problem of triangulating a surface given an unordered set of 3D points is also known as meshing or surface reconstruction. The points are to be connected by triangles such that the triangulation approximates the original surface well. The problem can be easily transferred to 2D space where the boundary is a curve.

The underlying assumption is that the points sample the curve densely enough. A common measure for characterizing the sampling density is the ϵ-sampling: It depends on the *local feature size* of a point $\mathbf{x}$ on a curve γ. Given the *medial axis* as the set of points with more than
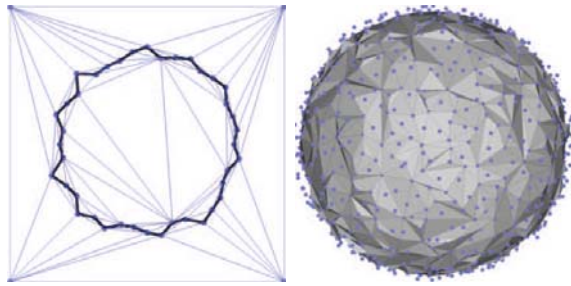
(a) 40 points on a circle with 2.5 % noise and 4 auxiliary corner points ($L_{inc} = 20$, $L_{dec} = 2$)

(b) 1000 points on a sphere with 1 % noise and 8 auxiliary points ($L_{inc} = 30$, $L_{dec} = 2$)

Fig. 4: Reconstruction results with the triangle-based graph structure (Fig. 3a). The original point cloud is shown with blue dots, the Delaunay triangulation with thin lines (omitted in 3D) and the boundary as bold polyline in 2D and surface in 3D.

(a) Sampled curve with increasing ϵ-sampling

(b) Delaunay triangulation and curve reconstruction

Fig. 5: (a): Dependence on sampling density. The original 2D cloud of 200 points samples a curve with increasing ϵ-sampling. (b): The curve reconstruction using the triangle-based partitioning approach (Fig. 3a, $L_{inc} = 100$, $L_{dec} = 2$) yields correct results until it breaks at a point with $\epsilon = 1.02$, as the sampling density on the right-hand side of the curve is not high enough anymore referring to its increasing curvature.

one closest point on the curve γ, the local feature size $\mathrm{lfs}(\mathbf{x})$ is the Euclidean distance of $\mathbf{x}$ to the medial axis. Then a set of points $\mathcal{X}$ is an ϵ-sample of a curve γ if every point $\mathbf{x} \in \gamma$ on the curve γ is within distance $\epsilon \cdot \mathrm{lfs}(\mathbf{x})$ of some point in $\mathcal{X}$ (EDELSBRUNNER 1998). We will use this concept to empirically characterize the success of a segmentation procedure by the maximum value ϵ may have, as larger values of ϵ correspond to lower sampling density.

The idea of determining a 2D polyline or 3D surface triangulation via partitioning is to divide the space into simplices and to partition these simplices into "inside" and "outside". In 3D space we reconstruct a surface by partitioning a Delaunay tetrahedralization or the corresponding 3D Voronoi complex. In 2D this corresponds to the reconstruction of a curve by partitioning a Delaunay triangulation or the corresponding 2D Voronoi complex (Fig. 3).

We choose to work with the Delaunay and Voronoi diagrams, since they are commonly used for this application. They were first exploited for surface reconstruction by BOIS-SONNAT (1984) and led to the concept of α-shapes (EDELSBRUNNER & MÜCKE 1994), *r-regular shapes* (ATTALI 1997) and the *crust* (AMENTA et al. 1998) with various derivatives like *conservative crust* (DEY et al. 1999), *power crust* (AMENTA et al. 2001) and *eigen-crust* (KOLLURI et al. 2004).

Since the partitioning divides all data points into two segments, the resulting boundary surface will be a closed surface within the Delaunay triangulated sampling volume.

We want to support the wave front to travel from an *inside* triangle to another *inside* or from an *outside* triangle to another *outside*, but hinder crossing the boundary. Under the above-mentioned sampling assumption a triangle edge that belongs to the boundary will be relatively short (Fig. 4a). Thus we construct our graph with edge costs simply being the inverse length of the common triangular edge. The inverse edge length can be raised to an exponent ≥ 2 in order to put more weight on very short edges, i.e. approximating an L_∞-norm.

The Delaunay triangulation is always bounded by the convex hull. If the desired segmentation boundary is part of the convex hull, it is not surrounded by triangles and thus can not be the boundary between differently labelled triangles. Therefore, we add auxiliary points to the original point cloud to extend the convex hull and avoid the boundary problem (Fig. 4a).

Fig. 4 shows results obtained with the described graph structure and edge costs on a synthetic 2D as well as on a 3D dataset. The point clouds are a circle with 40 points and a sphere with 1000 points disturbed with Gaussian noise. The circle is correctly reconstructed even when further reducing the number of points. The
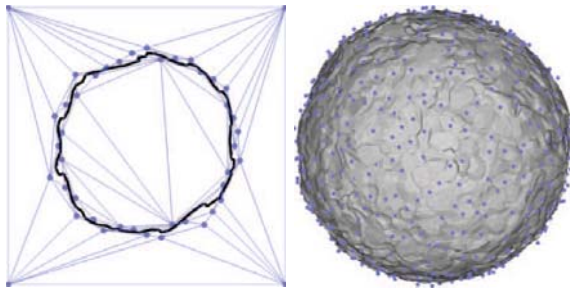
(a) 40 points on a circle with 2.5 % noise and 4 auxiliary corner points ($L_{inc} = 20$, $L_{dec} = 2$)

(b) 1000 points on a sphere with 1 % noise and 8 auxiliary points ($L_{inc} = 30$, $L_{dec} = 2$)

Fig. 6: Reconstruction results with the vertex-based graph structure (Fig. 3b). The original point cloud is shown with blue dots, the Delaunay triangulation with thin lines (omitted in 3D) and the boundary as bold polyline in 2D and surface in 3D. The boundary polyline/surface is drawn by linking edge centres of differently labelled vertices with the triangular/tetrahedral centroid.

topology of points on the surface of a sphere is a bit more complicated leading to some points not included in the triangulation. Their large deviation from the underlying surface contradicts the above-mentioned sampling assumption. The overall shape, however, is preserved.

In Fig. 5 we investigate the dependence on a dense sampling. The point cloud is generated as $(x, y) = (-t^2/2, -\sin(10t))$ with 200 uniformly sampled values $t \in [0, 2\pi]$. Since the x-intervals decrease from left to right, the local ϵ-sampling increases. The reconstruction via partitioning the Delaunay complex is shown as a bold line. It yields visually correct results until it breaks at a point with $\epsilon = 1.02$. In comparison: The popular *Crust* (AMENTA et al. 1998) and *Power Crust* (AMENTA et al. 2001) algorithms guarantee correct results only for $\epsilon < 0.252$ – TCHERNIAVSKI & STELLDINGER (2008) even claim $\epsilon < 0.1$ – while *Cocone* (AMENTA et al. 2000) and *Tight Cocone* (DEY & GOSWAMI 2003) require $\epsilon < 0.06$. Of course our test is much weaker than a theoretical proof, but still indicates robustness to rather low sampling density.

The alternative vertex-based structure would be the dual graph: Instead of representing triangles by nodes we use the vertices and link them by graph edges equivalent to the Delau-

nay edges. This representation might be better suited if the boundary sample points are disturbed by random noise such that an approximation is desired. By partitioning the vertices into *inside* and *outside* we obtain a boundary curve lying *in between* data points.

Graph edges crossing the boundary curve are usually rather short (Fig. 6a). Thus we define the edge costs as the inverse length of the crossed triangle edge.

Fig. 6 shows results obtained with this alternative graph structure on the very same synthetic datasets as in Fig. 4. The boundary is correctly reconstructed in between the given data points.

4.2 *Surface Segmentation*

The problem of segmenting a surface requires an algorithm to partition a surface into usually compact segments with similar elements – w. r. t. pre-defined properties. Here, we assume the surface to be represented by a triangular mesh, at each vertex being isomorphic to a disk.

To our knowledge there is only one other work that applies a front propagation method to surface segmentation: PAGE et al. (2003) use FMM as a final region growing step called "marching watersheds", referring to the popular watershed segmentation (BEUCHER & LANTUEJOUL 1979) that has been implemented for meshed surfaces by MANGAN & WHITAKER (1999).

For segmenting a surface in 3D space we will work with the very same graph structure as for reconstructing a boundary curve in 2D since we again want to partition a 2D manifold. Now, however, the manifold is non-planar and we exploit other geometric properties for defining edge costs.

For many applications in surface reconstruction and object modelling one wants to partition a triangular mesh into piece-wise planar regions. Thus we introduce high costs for graph edges linking non-planar nodes. For the triangle-based graph structure it is natural to compute triangle normals and derive costs from the angle between two of them. For the vertex-based graph structure vertex normals are needed that can be computed via principal component analysis from neighbouring points (HOPPE et

al. 1992), possibly robustified by computing the coordinate-wise median (ROUSSEEUW & LEROY 1987, section 3.2.1) of multiple neighbouring normals (SCHINDLER & FÖRSTNER 2011).

Analogous to partitioning adjacent triangles in 3D space we can partition straight line segments in 2D space as it is commonly done using the Ramer-Douglas-Peucker algorithm (RAMER 1972, DOUGLAS & PEUCKER 1973). Again the edge costs are derived from the angle between two line segments or – in the vertex-based graph structure – between two points.

Fig. 7 shows resulting segmentations in both 2D and 3D as well as for both graph structures. In the 2D examples the line and vertex normals are shown as red and blue arrows, respectively. The final segmentation is indicated with coloured triangles, lines and vertices. Even though both the cube and the square have rounded corners and the points are disturbed by 3 % (square) and 1 % (cube) Gaussian noise, the DijkstraFPS segmentation yields visually pleasing results.

Fig. 8 shows the segmentation result on a real dataset: It was reconstructed from 159 images using Bundler (SNAVELY et al. 2006), PMVS2 (FURUKAWA & PONCE 2010) and Poisson surface reconstruction (KAZHDAN et al. 2006).

4.3 Clustering

Within this section we focus on clustering data points via graph partitioning. We choose to create undirected links between each graph node and its $k = 25$ nearest neighbours (k-NN). In contrast to, e.g. Delaunay triangulation the complexity for k-NN is independent of the dimensionality. Edge costs are the squared Euclidean distance, possibly raised to an exponent ≥ 2.

We compare DijkstraFPS to common clustering algorithms on two synthetic datasets (Tab. 1):

The *k-means* algorithm (MACQUEEN 1965 iteratively computes the mean coordinates per class and updates class labels according to the closest mean. In case of different scatter the second step yields incorrect labels (dataset A).

The *Expectation–maximization* (EM) algorithm (DEMPSTER et al. 1977) not only estimates the means but also the class variances,

(a) $L_{inc} = 10$, $L_{dec} = 4$ **(b)** $L_{inc} = 10$, $L_{dec} = 4$

(c) $L_{inc} = 50$, $L_{dec} = 6$ **(d)** $L_{inc} = 50$, $L_{dec} = 6$
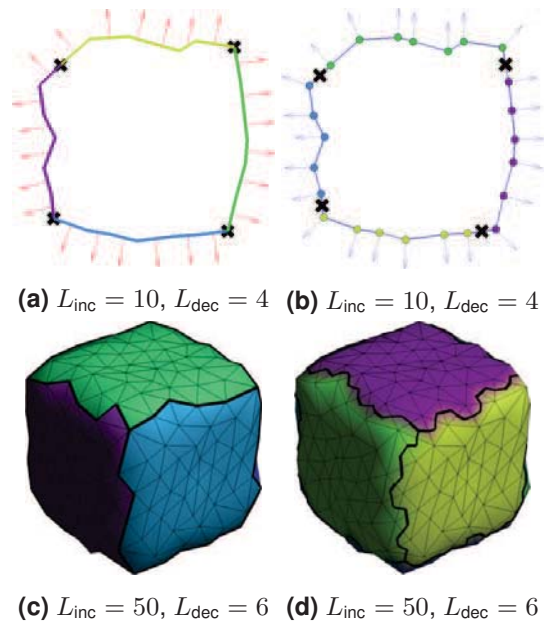
Fig. 7: Segmentation of a polyline in 2D space (top row: 24 points on a square, 3 % noise) and a surface in 3D space (bottom row: 240 points on a cube, 1 % noise) using the triangle-based (left) and vertex-based (right) graph structure. Edge costs are defined as the angle of the two normals. The normals are either computed for triangles, for lines or for vertices, depending on the dimensionality and the graph structure. In the 2D case they are shown as red and blue arrows. The resulting segmentation is indicated as coloured triangles, lines and points, respectively, as well as a bold, black boundary.

thus yields correct results (up to one falsely classified point) for dataset A. If it, however, assumes Gaussian distributions, it does not succeed with dataset B.

The *mean shift* clustering (COMANICIU & MEER 2002) iteratively replaces points by the centre of neighbouring points within a certain radius. It yields correct results for dataset A, but possibly too many clusters in the second example depending on the chosen radius. Here, a larger radius yields a similar result like *k-means*.

DijkstraFPS returns perfect point labels on both datasets shown in Tab. 1, certainly benefiting from the long edges between points of different clusters. On datasets with touching or overlapping distributions DijkstraFPS might fail.

Fig. 8: Surface segmentation on a visually reconstructed, meshed point cloud.

4.4 *Image Segmentation*

FPS has been used for progressive image sampling by ELDAR et al. (1997). As "farthest point" they choose a vertex of the Euclidean Voronoi diagram that maximizes some weighted distance function. Combining FPS with a marching scheme like FMM or Dijkstra's algorithm we can approximate the non-Euclidean Voronoi diagram and thus derive not only a suitable image sampling but an image segmentation as well.

The graph structure is inherently given by the pixel grid. One can, however, choose between a 4-neighbourhood (pixels horizontally and vertically connected), an 8-neighbourhood (pixels additionally diagonally connected) or a triangulation (diagonal connections in one direction only).

In contrast to other image segmentation algorithms we can use arbitrary colour similarity measures as edge costs and thus are not restricted to gray images or Euclidean colour spaces. We compute the Euclidean distance of two pixels in the hue-saturation-value (HSV) colour cone.

Fig. 9 shows three segmentation results. While FLOWER 1 and FLOWER 2 are successfully segmented, the approach fails for FLOWER 3. Instead of segmenting "flower" and "not flower" the *unsupervised* segmentation draws the boundary along the dark bold edges in the background.

Since the DijkstraFPS graph partitioning is not restricted to a regular pixel grid, we can work with irregular image regions such as superpixels. Further we can generate superpixels with DijkstraFPS by partitioning the image into more than two segments, e.g. 30 segments as
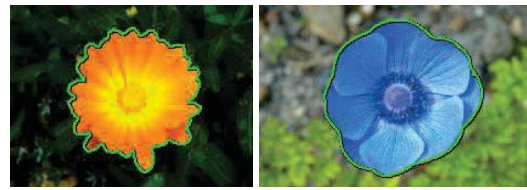
(a) FLOWER 1
($L_{inc} = 100$, $L_{dec} = 2$)

(b) FLOWER 2
($L_{inc} = 100$, $L_{dec} = 2$)

(c) FLOWER 3
($L_{inc} = 200$, $L_{dec} = 2$)

Fig. 9: DijkstraFPS image segmentation. Three test images are segmented using our proposed graph partitioning scheme. Edge costs are computed as squared Euclidean distances in hue-saturation-value (HSV) colour space. While FLOWER 1 (a) and FLOWER 2 (b) are perfectly segmented, in FLOWER 3 (c) the dark background edges attract the segmentation boundary more.

indicated by the white boundaries in Fig. 10a. Then we build a second graph with only 30 nodes and edges according to the superpixel adjacencies in the image. Edge costs are derived, e.g. from the average colours of the superpixels. As can be seen from the green boundary in Fig. 10a the final segmentation is very accurate due to the flexibility of 30 initial segments but also robust due to the larger number of pixels that contribute to the superpixel colour. Results for two images from the Berkeley Segmentation Dataset and Benchmarks 500 (BSDS500 AR-BELAEZ et al. 2011) are given in Figs. 10c and 10d.

For large images DijkstraFPS is rather slow, since it is very general and not optimized for the regular image grid. Of course other superpixel generating algorithms (LEVINSHTEIN et al. 2009, ACHANTA et al. 2012) can be used and might save significant computing time. The second segmentation step would be performed using DijkstraFPS. An evaluation of different superpixel schemes combined with our DijkstraFPS graph partitioning might be reasonable but is out of scope of this article.
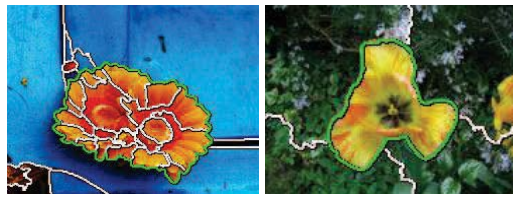
(a) FLOWER 3
($L_{inc} = 200$, $L_{dec} = 30$)

(b) FLOWER 4
($L_{inc} = 100$, $L_{dec} = 5$)
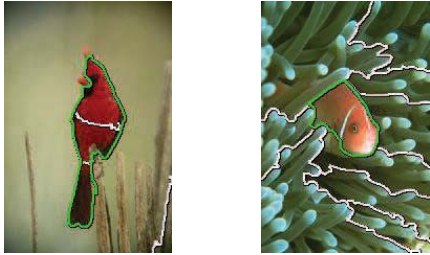
(c) 196027 (ARBE-
LAEZ et al., 2011)
($L_{inc} = 200$, $L_{dec} = 5$)

(d) 210088 (ARBE-
LAEZ et al., 2011)
($L_{inc} = 100$, $L_{dec} = 10$)

Fig. 10: Progressive image segmentation. (a): FLOWER 3 image from Fig. 9c segmented into 30 segments, shown with white boundaries. Then a graph with only 30 nodes is built from the obtained segmentation and again partitioned using DijkstraFPS, yielding the green boundary. (b), (c) and (d): More examples.

5 Conclusion

We proposed a new graph partitioning scheme based on Dijkstra's distance transform and farthest point sampling (DijkstraFPS) and showed how to construct graphs for solving various problems in geometry and image processing. The partitioning is guided by pre-defined edge costs that do not have to follow a Euclidean metric. The method is not restricted to a binary partition but can yield multiple segments, possibly in terms of an oversegmentation or – in terms of image segmentation – superpixels. The latter themselves can be nodes of a subsequent graph partitioning.

As shown in SCHINDLER & FÖRSTNER (2011) the algorithm can be augmented with an automatic, application-specific stopping criterion. Establishing such criteria for each above-mentioned application remains future work.

There is much space for improving the proposed edge costs. We focused on demonstrating the diversity of DijkstraFPS graph partitioning

Tab. 1: Clustering results of three common clustering algorithms and our DijkstraFPS graph partitioning. DijkstraFPS ($L_{inc} = 10$, $L_{dec} = 2$) yields 100 % correct results in both examples. The ellipses are 1-σ confidence intervals.
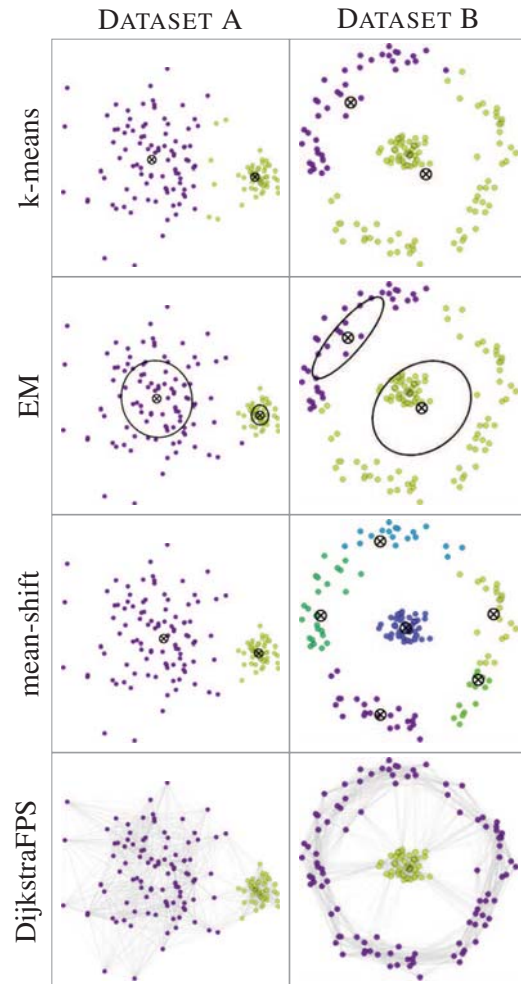
rather than on fine-tuning for specific applications.

Other possible applications of the method include triangulation and segmentation of full wave-form lidar point clouds, segmentation of radar images, hierarchical image segmentation as indicated in section 4.4, segmentation of image sequences and image segmentation based on texture similarity, which of course have to be empirically compared to standard algorithms used in that application.

References

ACHANTA, R., SHAJI, A., SMITH, K., LUCCHI, A., FUA, P. & SÜSSTRUNK, S., 2012: Slic superpixels compared to state-of-the-art superpixel

methods. – IEEE Transactions on Pattern Analysis and Machine Intelligence **34** (11): 2274–2282.

AMENTA, N., BERN, M. & EPPSTEIN, D., 1998: The crust and the β-skeleton: Combinatorial curve reconstruction. – Graphical Models and Image Processing **60**: 125–135.

AMENTA, N., CHOI, S., DEY, T.K. & LEEKHA, N., 2000: A simple algorithm for homeomorphic surface reconstruction. – Sixteenth Annual Symposium on Computational Geometry.

AMENTA, N., CHOI, S. & KOLLURI, R., 2001: The power crust. – Sixth ACM Symposium on Solid Modeling and Applications.

ARBELAEZ, P., MAIRE, M., FOWLKES, C. & MALIK, J., 2011: Contour detection and hierarchical image segmentation. – IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (5): 898–916.

ARBELAEZ, P., HARIHARAN, B., GU, C., GUPTA, S., BOURDEV, L.D. & MALIK, J., 2012: Semantic segmentation using regions and parts. – Computer Vision and Pattern Recognition.

ATTALI, D., 1997: r-regular shape reconstruction from unorganized points. – Thirteenth Annual Symposium on Computational Geometry: 248–253.

BEUCHER, S. & LANTUEJOUL, C., 1979: Use of watersheds in contour detection. – International Workshop on Image Processing, Real-time Edge and Motion Detection.

BOISSONNAT, J.-D., 1984: Geometric structures for three-dimensional shape representation. – ACM Transactions on Graphics **3**: 266–286.

BOYKOV, Y. & FUNKA-LEA, G., 2006: Graph cuts and efficient n-d image segmentation. – International Journal of Computer Vision **70** (2): 109–131.

COMANICIU, D. & MEER, P., 2002: Mean shift: a robust approach toward feature space analysis. – IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (5): 603 –619.

DEMPSTER, A.P., LAIRD, N.M. & RUBIN, D.B., 1977: Maximum likelihood from incomplete data via the em algorithm. – Journal of the Royal Statistical Society **39** (1): 1–38.

DEY, T.K., MEHLHORN, K. & RAMOS, E.A., 1999: Curve reconstruction: Connecting dots with good reason. – Fifteenth Annual Symposium on Computational Geometry.

DEY, T.K. & GOSWAMI, S., 2003: Tight cocone: A water-tight surface reconstructor. – Eighth ACM Symposium on Solid Modeling and Applications.

DIJKSTRA, E.W., 1959: A note on two problems in connexion with graphs. – Numerische Mathematik **1** (1): 269–271.

DOUGLAS, D.H. & PEUCKER, T.K., 1973: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. – Cartographica: The International Journal for Geographic Information and Geovisualization **10** (2): 112–122.

EDELSBRUNNER, H. & MÜCKE, E.P., 1994: Three-dimensional alpha shapes. – ACM Transactions on Graphics **13**: 43–72.

EDELSBRUNNER, H., 1998: Shape reconstruction with delaunay complex. – Lecture Notes in Computer Science **1380**: 119–132.

ELDAR, Y., LINDENBAUM, M., PORAT, M. & ZEEVI, Y.Y., 1997: The farthest point strategy for progressive image sampling. – IEEE Transactions on Image Processing **6** (9): 1305–1315.

FORSYTH, D.A. & PONCE, J., 2002: Computer vision: a modern approach. – Prentice Hall Professional Technical Reference.

FURUKAWA, Y. & PONCE, J., 2010: Accurate, dense, and robust multi-view stereopsis. – IEEE Transactions on Pattern Analysis and Machine Intelligence **32** (8): 1362–1376.

HOPPE, H., DEROSE, T., DUCHAMP, T., MCDONALD, J. & STUETZLE, W., 1992: Surface reconstruction from unorganized points. – Nineteenth Annual Conference on Computer Graphics and Interactive Techniques: 71–78.

KAZHDAN, M., BOLITHO, M. & HOPPE, H., 2006: Poisson surface reconstruction. – Fourth Eurographics symposium on Geometry processing, SGP'06: 61–70, Eurographics Association, Aire-la-Ville, Switzerland.

KIMMEL, R. & SETHIAN, J.A., 1998: Computing geodesic paths on manifolds. – National Academy of Sciences **95**: 8431–8435.

KOLLURI, R., SHEWCHUK, J.R. & O'BRIEN, J.F., 2004: Spectral surface reconstruction from noisy point clouds. – Eurographics Symposium on Geometry Processing.

LEVINSHTEIN, A., STERE, A., KUTULAKOS, K.N., FLEET, D.J. & DICKINSON, S.J., 2009: Turbopixels: Fast superpixels using geometric flows. – IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (12): 2290–2297.

MACQUEEN, J., 1965: Some methods for classification and analysis of multivariate observations. – Fifth Berkeley Symposium on Mathematical Statistics and Probability.

MANGAN, A.P. & WHITAKER, R.T., 1999: Partitioning 3d surface meshes using watershed segmentation. – IEEE Transactions on Visualization and Computer Graphics **5** (4): 308–321.

MEINE, M. & KÖTHE, U., 2005: Image segmentation with the exact watershed transform. – Visualization, Imaging and Image Processing.

MESTER, R., CONRAD, C. & GUEVARA, A., 2011: Multichannel segmentation using contour relax-

ation: Fast super-pixels and temporal propagation. – Scandinavian Conference on Image Analysis.

MOENNING, C. & DODGSON, N.A., 2003a: Fast marching farthest point sampling. – Eurographics 2003, Granada, Spain.

MOENNING, C. & DODGSON, N.A., 2003b: Fast marching farthest point sampling for point clouds and implicit surfaces. – Technical Report 565. University of Cambridge, Computer Laboratory.

PAGE, D.L., KOSCHAN, A.F. & ABIDI, M.A., 2003: Perception-based 3d triangle mesh segmentation using fast marching watersheds. – Computer Vision and Pattern Recognition: 27–32.

PEYRÉ, G. & COHEN, L., 2004: Surface segmentation using geodesic centroidal tesselation. – Second International Symposium on 3D Data Processing, Visualization and Transmission: 995–1002.

PEYRÉ, G. & COHEN, L., 2006: Geodesic remeshing using front propagation. – International Journal of Computer Vision **69**: 145–156.

RAMER, U., 1972: An iterative procedure for the polygonal approximation of plane curves. – Computer Graphics and Image Processing **1** (3): 244–256.

ROSCHER, R., 2012: Sequential Learning using Incremental Import Vector Machines for Semantic Segmentation, PhD thesis. Univerity of Bonn.

ROTHER, C., KOLMOGOROV, V. & BLAKE, A., 2004: Grabcut: Interactive foreground extraction using iterated graph cuts. – ACM Transactions on Graphics **23**: 309–314.

ROUSSEEUW, P.J. & LEROY, A.M., 1987: Robust regression and outlier detection. – John Wiley & Sons, Inc.

SCHINDLER, F. & FÖRSTNER, W., 2011: Fast marching for robust surface segmentation. – Photogrammetric Image Analysis, Lecture Notes in Computer Science **6952**: 147–158.

SETHIAN, J.A., 1996: A fast marching level set method for monotonically advancing fronts. – National Academy of Sciences of the USA **93**: 1591–1595.

SHI, J. & MALIK, J., 2000: Normalized cuts and image segmentation. – IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (8): 888–905.

SNAVELY, N., SEITZ, S.M. & SZELISKI, R., 2006: Photo tourism: Exploring image collections in 3d. – ACM **25**: 835–846.

SZELISKI, R., 2010: Computer Vision: Algorithms and Applications. – Springer.

TCHERNIAVSKI, L. & STELLDINGER, P., 2008: A thinning algorithm for topologically correct 3d surface reconstruction. – International Conference on Visualization, Imaging and Image Processing: 119–124.

VEKSLER, O., BOYKOV, Y. & MEHRANI, P., 2010: Superpixels and supervoxels in an energy optimization framework. – European Conference on Computer Vision ECCV **10**: 211–224.

VINCENT, L. & SOILLE, P., 1991: Watersheds in digital spaces: An efficient algorithm based on immersion simulations. – IEEE Transactions on Pattern Analysis and Machine Intelligence **13**: 583–598.

Address of the Authors:

Dipl.-Ing. FALKO SCHINDLER, Rheinische-Friedrich-Wilhelms-Universität Bonn, Institut für Geodäsie und Geoinformation, Photogrammetrie, Nußallee 15, D-53115 Bonn, Tel.: +49-228-73-2908, e-mail: falko.schindler@uni-bonn.de.

Prof. Dr.-Ing. Dr. h. c. mult. WOLFGANG FÖRSTNER, Josef-Schell-Str. 34, D-53121 Bonn, e-mail: wf@ipb.uni-bonn.de

Finding Poly-Curves of Straight Line and Ellipse Segments in Images

Susanne Wenzel & Wolfgang Förstner, Bonn

Summary: Simplification of given polygons has attracted many researchers. Especially, finding circular and elliptical structures in images is relevant in many applications. Given pixel chains from edge detection, this paper proposes a method to segment them into straight line and ellipse segments. We propose an adaption of Douglas-Peucker's polygon simplification algorithm using circle segments instead of straight line segments and partition the sequence of points instead the sequence of edges. It is robust and decreases the complexity of given polygons better than the original algorithm. In a second step, we further simplify the poly-curve by merging neighbouring segments to straight line and ellipse segments. Merging is based on the evaluation of variation of entropy for proposed geometric models, which turns out as a combination of hypothesis testing and model selection. We demonstrate the results of `circlePeucker` as well as merging on several images of scenes with significant circular structures and compare them with the method of Patraucean et al. (2012).

Zusammenfassung: *Segmentierung von Pixelketten in Geraden- und Ellipsenelemente*. Die Detektion runder und elliptischer Strukturen ist relevant für viele Anwendungen. Die Reduktion der Komplexität gegebener Polygone ist für sich ein interessantes Forschungsthema. Diese Arbeit stellt ein Verfahren zur Segmentierung von Pixelketten einer Kantendetektion in Geraden- und Ellipsensegmente vor. Der erste Schritt besteht in einer Adaption des Douglas-Peucker Algorithmus, in der Kreise anstelle von Geraden zur Partitionierung verwendet werden und die Punkt- statt der Kantensequenz partitioniert wird. Das Verfahren ist robust und reduziert die Komplexität der gegebenen Polygone stärker als der originale Algorithmus. In einem zweiten Schritt vereinfachen wir diese Vorsegmentierung durch das Verschmelzen benachbarter Segmente zu Geraden- und Ellipsensegmenten und stützen uns dabei auf die Entropieänderung. Wir zeigen die Ergebnisse der Vorsegmentierung als auch der folgenden Vereinfachung anhand verschiedener Bilder von Szenen, die signifikante kreisförmige Strukturen aufweisen und vergleichen sie mit dem Algorithmus von Patraucean et al. (2012).

1 Introduction

Polygon simplification is interesting from several points of view. First, in terms of compact description of spatial data, e.g. in the context of image description. Second, in terms of generalisation, e.g. in the context of cartography or resolution dependent visualization of polygons.

On the other hand finding circular and elliptical structures in images is relevant in terms of compact image description and further image interpretation. Most image interpretation systems which use bottom up image features, thus not just pure pixel information, are based on key point or edge detection. Directly identifying circular and elliptical structure gives rise to much more informative image features from bottom up (Chia et al. 2012, Jurie & Schmid 2004).

In this paper we propose a two-step polygon simplification algorithm that approximates a given set of ordered points in 2D by a sequence of straight line and ellipse segments. The poly-curves are intended to be at least C^0, thus positional continuous. Although the algorithm is applicable to any kind of ordered 2D points we assume pixel chains within images, see Fig. 1. The first intuition behind our ap-

Fig. 1: Finding line and ellipse sections. Left: pixel chains. Middle: segmentation into circle chains using `circlePeucker`. Right: aggregation and classification into straight line and ellipse segments. Blue: lines. Orange: circular and elliptical arcs.

proach is that arbitrary smooth curves can be locally characterised by an osculating circle. We will use this in the first step of the algorithm where we simplify the given set of points by a sequence of circle segments.

But due to perspective distortions, in general there will be almost no circles in images. All circles in object space are projected to ellipses in image space, ellipses in object space are projected to ellipses, anyway. Only in rare cases, the image of 3D circles or 3D ellipses are mapped to hyperbola, namely in case they partially are behind the camera. The situation is different if the circles are sitting in a set of parallel planes and the viewing direction intentionally has been taken orthogonal to these planes or the image has been rectified to mimic this situation. Then almost no ellipses will occur in the images, and the proposed method can directly be transferred by replacing ellipses by circles.

Therefore, eventually the pixel chain is represented by a sequence of straight line and ellipse segments. This way we are more flexible representing curved lines. Please note, that circles are part of the representation, as they are just special ellipses.

The proposed method consists of two steps, see Fig. 1. Given the pixel chains within the image, we first iteratively segment the region boundary into circular segments. This yields an over-segmentation due to the non existence of real circle segments. Second, we merge neighbouring segments to straight line and ellipse segments based on statistical reasoning, namely hypothesis testing and model selection. This step optimally estimates lines and ellipses in a least squares sequence.

One might argue, why not directly segment a pixel chain into ellipses, but first look for circle segments, and then group them to ellipses.

There are two main reasons for the two-step procedure:

1.) The slope, curvature or curvature change functions of the ellipse are no simple functions, which allow to identify elliptical segments, as this is the case for straight lines (constant slope) and circles (constant curvature), 2.) there is no simple local measure telling whether a local segment belongs to an ellipse or not: Analysing the curvature, distinguishes circles and straight lines. One would need the second derivatives of the curvature to capture the properties of a local ellipse element, as an ellipse has two more degrees of freedom, than a circle. But determining fourth derivatives is very unstable.

There are two main contributions of this paper. First, for region boundary segmentation we propose an adaption to Douglas-Peucker's algorithm (DOUGLAS & PEUCKER 1973) which is based on circles as basic geometric elements and partitions the sequence of points instead of the sequence of edges. Second, we adapt the idea of variation of entropy by BEDER (2005) and statistically optimal merge neighboured segments while optimally fitting lines and ellipses. The whole process depends on two parameters, namely the precision of the edge extraction and the expected accuracy of the straight line and ellipse segments. The first one is an internal precision which guides the edge extraction, the second one is what the user defines to be and might guide the degree of generalization. Both can be estimated from training data.

Therefore, setting these parameters once is sufficient: The process works stable for all of our experiments using the same parameter set.

Related work

To our knowledge there is no work about an adaption of Douglas-Peucker's algorithm to the use of circles instead of lines as basic elements. However, proposals exists to simplify polygons by sets of circular arcs for the efficient storage of polylines. GÜNTHER & WONG (1990) proposed the so called Arc Tree which represents arbitrary curved shapes in a hierarchical data structure with small curved segments at the leaves of a balanced binary tree. MOORE et al. (2003) proposed a method for polygon simplification using circles. They aim on closed poly-

gons given by a set of 2D points. Based on medial axis from Voronoi polygons they propose a population of circles which they afterwards filter to get a set of circles which best approximate the given polygon. The final representation of the polygon consists of circles represented by centre and radius and tangents which link neighbouring circles. No work on using ellipses for improving storage requirements are known to us.

Finding ellipses in images has attracted many researchers. Some of them use Hough-transform methods which tend to be slow. Most techniques start from pixel chains. Early works focussed on ellipse fitting, e.g. PAVLIDIS (1983) and PORRILL (1990), later focussed on unbiased estimates, e.g. WU (2008) and LIBUDA et al. (2006). We are interested in the more general problem of describing the pixel chains by sequences of line and ellipse segments, a problem already addressed in AL-BANO (1974), however, neither enforcing ellipses, nor looking for a best estimate for ellipses. WEST & ROSIN (1992) and ROSIN & WEST (1995) performed a segmentation of sequences into lines and ellipses in a multistage process. They first segment a 2D-curve into straight lines. Afterwards sequences of line segments are segmented into arcs restricted to their endpoints. One might interpret this step as merging sequences of lines to elliptical arcs. Model selection is done implicitly by evaluating a significance measure to each proposed segment, which is based on its geometry, purely. However, their criteria are non-statistical, thus cannot easily be adapted to varying noise situations. JI & HARALICK (1999) criticised this and proposed a statistically valid criterium. Starting from Rosin's output of arc segmentation they merge pairs of arcs belonging to the same ellipse. Moreover they also group non-adjacent arcs and exploit the sign of the arcs for grouping. Proposals for merging are validated via hypothesis testing. They showed only few results on comparably easy images. NGUYEN & KERAUTRET (2011) also addressed the segmentation of pixel chains into lines and ellipses. It is based on a discrete representation of tangents, circles, and an algebraic fitting through neighbouring arcs only using some key points (boundary and mid-

point, instead of the complete pixel chain. Recently PATRAUCEAN et al. (2012) proposed a parameterless line segment and elliptical arc detector. They use an ellipse fitting algorithm which uses both, the algebraic distance of the conic equation and deviation from the gradient direction. Their model selection aims at avoiding false negatives, by controlling the number of false positives. Realizing the principle of "non-accidentalness" their method adapts to noise, which explains their visually appealing results. Their validation and model selection criteria, however, are based on fixed tolerance bands. Also they do not enforce any continuity between neighbouring segments.

Our scope is to segment pixel chains into straight line and ellipse segments, exploiting the knowledge about their statistical properties both w.r.t. detectability as well as w.r.t. accuracy.

Notation

Geometric elements are named with calligraphic letters, e.g. χ is the name of a point, whereas x is its Cartesian representation. Homogeneous vectors and matrices are denoted with upright bold letters, e.g. $\mathbf{x}$ and $\mathbf{C}$.

The rest of the paper is organized as follows. First, we describe the segmentation of region boundaries into circle elements based on the idea of Douglas-Peucker's algorithm. The merging procedure to obtain line and ellipse segments is explained in section 3. This section also gives details about model selection by variation of entropy and by the principle of minimum description length. Finally, section 4 presents results on synthetic and real data and compares them with the method of PATRAUCEAN et al. (2012).

2   Region Boundary Segmentation

Given a set of ordered points in 2D we aim at a partitioning into groups joining a common geometric element, specifically circular segments. We use the feature extraction procedure as described in FÖRSTNER (1994) and FUCHS & FÖRSTNER (1995). It includes an automatic noise estimation and an edge preserving filter as described in FÖRSTNER (2000). In contrast to many

other procedures it delivers region boundaries as well as thin lines in the form of chains of points with sub-pixel coordinates. For finding fine details, we use 0.7 pixel for the differentiation and 1.0 pixel for the integration scale. No blow up of the images is performed, as proposed by KÖTHE (2003) for fully exploiting the resolution.

2.1 *Algorithm*

Our concept of region boundary segmentation is based on the well known Douglas-Peucker algorithm (DOUGLAS & PEUCKER 1973). This algorithm is designed to simplify polygons. Therefore, it recursively splits the sequence of polygon edges into larger edges, until the distance of an eliminated point to the corresponding edge is below a threshold t. Thus neighbouring edges share a common point. In contrast, we want to recursively split the sequence of points $X = \{x_i\}$ until each subsequence can be approximated by a circular arc well enough. Thus neighbouring sequences are meant not to share a common point.

We realize this by first determining the mid points $x'_i = \frac{1}{2}(x_i + x_{i+1})$, $i = 1, ..., I - 1$, leading to a sequence $X' = \{x'_i\}$, which is a factor 2 smoother in variance than the original. Each edge in the sequence X' corresponds to a point x_i in the original sequence, except for the start and the end point. We now recursively partition the sequence of edges of X' into segments, which approximate the points x'_i by a circular arc up to a pre-specified tolerance t. A segment is split at that point x'_i where the distance to the circular arc is maximum. In order to enforce continuity, we fix the start and end point of the segments and determine the best fitting arc, see below.

The algorithm for approximating a polyline by a sequence of circles, called `circlePeucker`, is given in Alg. 1. It uses (1) function `fitArc(X)` for fitting a circular arc segment S to a given set of points X constraining it to the start and end point, and (2) a function `distXS(X, S)` for determining the index i_b and the distance d_{max} of the point with the largest distance of the points X to an arc segment S. The algorithms recursively splits the chain until the largest distance of a point to the corresponding arc is below a

In: Ordered set of points $X = \{x_1 \dots x_I\}$, tolerance t
Out: List of segments O

1   **if** $I = 2$ **then** $O = \{1, I\}$, **return**;
2   $S = \texttt{fitArc}(X)$;
3   $(d_{max}, i_b) = \texttt{distXS}(X, S)$;
4   **if** $d_{max} > t$ **then**
5   partition at i_b:
      $X_1 = \{x_1 \dots x_{ib}\}$,
      $X_2 = \{x_{ib} \dots x_{end}\}$;
6      $O_1 = \texttt{circlePeucker}(X_1, t)$;
7      $O_2 = \texttt{circlePeucker}(X_2, t)$;
8      $O = O_1 \cup O_2$;
9   **end**
10   **return**

Algorithm 1: function `circlePeucker`

pre-set threshold t. As result we get a list O of N circle segments, each segment represented as a list of indices $\{i'_n\}$, $n = 1, ..., N$. Thus, we call $O' = \texttt{circlePeucker}(X', t)$. The edges $(i', i' + 1)$ of the segments in O' correspond to the sought points x_i, except for the start and the end point, which are added to the first and the last segment. This yields the required partitioning O of the original point sequence.

2.2 *Fitting Circle Segments*

The algorithm `fitArc(X)`, needed in Alg. 1 line 2, constrains the circle to the starting and the endpoint of the current polygon segment. Additionally, we determine the distances $d = [d_i]$ of the involved points x_i to the arc segment between x_1 and x_I of S, not to the whole circle. Thus, the distance of a point to a segment is the minimum of the distance to the footpoint on the segment or the distance to the start or endpoint.

A circle usually has three degrees of freedom, but by restricting the arc to two points there is just one degree of freedom left. We parametrize the arc segment by its height h and solve the following optimization problem

$$\hat{h} = \operatorname{argmin}_h \left(\| d(X, S(x_s, x_e, h)) \|_L \right). \quad (1)$$

For a robust estimate we choose the L_1-norm ($L = 1$), thus we optimize h such that the sum

of absolute distances of all points to the arc segment is minimized.

3   Merging

Given the pre-segmentation O of section 2 which is assumed to be over-segmented, we aim at a simplification by merging neighbouring segments which share the same model.

The pre-segmentation is based on circles, but in general there are almost no circles in natural images as they suffer from perspective distortions. Thus, our final segmentation is meant to consist of segments of straight lines and ellipses. From the pre-segmentation we just take the information about which points belong to one segment and ignore the parameters of the fitted circles.

The final representation is achieved by fitting straight lines and ellipses through neighboured segments and single segments using all points belonging to them. This is different from ROSIN & WEST (1995) who only use the endpoints from the pre-segmentation.

3.1   *Fitting Ellipses*

We perform maximum-likelihood estimations for fitting lines and ellipses, respectively, to the data. For line fitting we refer to standard literature, e.g. MCGLONE (2004).

Fitting ellipses is not trivial. We have to make sure an arc segment to be an ellipse and not a parabola or hyperbola.

We represent conics with the symmetric 3×3-matrix $\mathsf{C} = \begin{bmatrix} \boldsymbol{C}_{hh} & c_{h0} \\ c_{0h}^T & c_{00} \end{bmatrix}$ using homogeneous coordinates $\mathbf{x}$ for the points on the conic $\mathbf{x}^\top \mathsf{C} \mathbf{x} = 0$. To ensure the conic to be an ellipse the homogeneous part of the conic must fulfil $|\mathbf{C}_{hh}| > 0$. Therefore, we use Fitzgibbon's constraint (FITZGIBBON et al. 1999) which is equivalent to

$$|\mathbf{C}_{hh}| = 1 \,. \qquad (2)$$

This is a valid choice, as the conic representation is homogeneous. We end up with a maximum likelihood estimation following a Gauss-Helmert model with the constraint (2). Parameters are initialized using the direct method of Fitzgibbon (FITZGIBBON et al. 1999).

As a result we not only obtain the ellipse parameters but also the estimated variance σ^2 of the data and covariance matrix $\boldsymbol{\Sigma}$ of the parameters, which we use for the subsequent tests.

3.2   *Merging Segments Based on Variation of Entropy*

Deciding whether two neighbouring segments belong to the same model may be based on a statistical hypothesis test. As hypothesis tests aim at rejecting the null hypothesis, they can be used as sieve for keeping false hypothesis: Therefore, we use hypothesis testing for reliably identifying breakpoints between segments *not* belonging to the same model, by testing the null-hypothesis that they belong to the same segment.

Deciding which model fits the data best, i.e. whether a curved line is best approximated by a line or an ellipse, is a typical model selection problem and may be solved by the principle of minimal description length (MDL). This may be directly applied to isolated segments.

Merging segments based on hypothesis testing lacks on the risk of accepting large changes in geometry, in case the parameters of the proposed model are very uncertain. Therefore, we follow the idea of variation of entropy by BEDER (2005). He derives an information theoretical measure for the increase of uncertainty of a model due to adding new observations. This is equivalent to the change of entropy of the probability density function of the model's parameters. Following BEDER (2005), the change of entropy can be split into two parts. One depends on the increase of randomness due to new observations and is related to hypothesis testing. The other depends on the change of geometric uncertainty due to new data, respectively.

The differential entropy of a probability density function $p(x)$ is given by $h(p) = -\int p(x) \log p(x) \mathrm{d}x$. It reflects the randomness of a stochastic variable x. In case of a D-dimensional normally distributed random variable $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the entropy is given by (COVER & THOMAS 1991)

$$h(p) = 0.5 \log \left[(2\pi e)^D |\boldsymbol{\Sigma}| \right] . \qquad (3)$$

Now, assume a segmentation O of points

$\mathcal{X} = \{\mathcal{X}_1 \cup \ldots \cup \mathcal{X}_N\}$ into N segments. Further assume, we already found a model $\mathcal{M}_n$ fitting the points $\mathcal{X}_n$ of segment n, e.g. a line $\mathcal{M}_n = l_n$. Now, we propose the points $\mathcal{X}_{n+1}$ of the neighbouring segment to belong to model $\mathcal{M}_n$, too. Without loss of generality we might argue on the neighbouring segments $n = 1$ and $n + 1 = 2$. Now, let the parameters of proposed model $\mathcal{M}_1$ be $\hat{\boldsymbol{\theta}}_1 \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1)$, with the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_1 = \hat{\sigma}_1^2 \boldsymbol{\Sigma}_1$ of the parameters, depending on the theoretical covariance matrix $\boldsymbol{\Sigma}_1$ and the estimated variance factor $\hat{\sigma}_1^2 = \Omega_1 / R_1$, derived from the weighted sum Ω_1 of the squared residuals and the redundancy R_1 of the estimation process.

When adding new observations $\mathcal{X}_{n+1}$ we estimate $\hat{\boldsymbol{\theta}}_2 \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_2)$ from $\mathcal{X}_{n,n+1} = \{\mathcal{X}_n \cup \mathcal{X}_{n+1}\}$ and obtain $\hat{\sigma}_2^2$ and the theoretical covariance matrix $\boldsymbol{\Sigma}_2$.

To validate the agreement of such two groups of observations concerning one of the two models $\mathcal{M}_1$ and $\mathcal{M}_2$ we analyse the change of entropy caused by adding new observations:

$$\Delta h_{\mathcal{M}} = h(\mathcal{N}(\hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_2)) - h(\mathcal{N}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1)) \quad (4)$$

The parameters of a model $\mathcal{M}$ typically are given by adjustment theory. Thus, we know the variance factor $\hat{\sigma}^2$ and the empirical covariance matrix $\hat{\boldsymbol{\Sigma}} = \hat{\sigma}^2 \boldsymbol{\Sigma}$. Using (3) we get

$$\Delta h_{\mathcal{M}} = \underbrace{0.5 \log\left(\hat{\sigma}_2^2 / \hat{\sigma}_1^2\right)}_{\Delta h_0} + \underbrace{0.5 \log(|\boldsymbol{\Sigma}_2| / |\boldsymbol{\Sigma}_1|)}_{-\Delta h_g}$$
$$(5)$$

The first term Δh_0 is closely related to the Fisher test statistic

$$\hat{\sigma}_2^2 / \hat{\sigma}_1^2 \sim \mathcal{F}(\Delta R, R_1) \quad (6)$$

with redundancy R_1 and $\Delta R = R_2 - R_1$, which is used to test whether the second set of observations fits the model estimated by the first set. It reflects the increase of randomness due to including new observations. The term Δh_g reflects the increase in randomness due to the geometric change of the model.

Therefore, we argue in the sense of hypothesis testing. Given a threshold $T_S = \mathcal{F}^{-1}(S, \Delta R, R_1)$ with significance level S by the inverse of Fisher distribution, there is no statistical reason to reject the hypothesis that both

sets of observations fit the model if

$$\Delta h_0 < 0.5 \log T_S \quad (7)$$

which means that both sets of observations fit the model due to uncertainty in estimated parameters.

To bound the risk of large changes in geometry we further bound the increase of entropy by Δh_g. BEDER (2005) found this bound to be at the same order of magnitude as the increase of Δh_0. We use $T_g = T_0 + \frac{1}{2} \log T_S$ with a model dependent additional constant T_0 which we empirically found to be equal to the number of parameters of the current model, e.g. $T_0 = 2$ in case of lines or $T_0 = 5$ in case of ellipses. This compensates for a decrease in condition number of the covariance matrix caused by merging, therefore increasing with the number of parameters.

In case $R_1 = 0$ we cannot use an estimated variance factor $\hat{\sigma}_1^2$, but use the theoretical value σ_1^2 instead. Thus, (5) degenerates to

$$\Delta h_{\mathcal{M}} = 0.5 \log\left(\hat{\sigma}_2^2 / \sigma_1^2\right) + 0.5 \log(|\boldsymbol{\Sigma}_2| / |\boldsymbol{\Sigma}_1|).$$
$$(8)$$

Now, the ratio $\hat{\sigma}_2^2 / \sigma_1^2 \sim \chi_{R_2}^2$ and we derive the threshold T_S from the inverse of χ^2 distribution.

Please note, that the proposed approach is asymmetric in evaluation of $\mathcal{X}_{n,n+1}$ and $\mathcal{X}_{n+1,n}$. The asymmetry is compensated by always checking whether the smaller of two neighbouring segments can be merged with the larger one, thus the larger segment is taken to be $\mathcal{M}_1^*$.

3.3 *Model Selection*

We have seen how to use the variation of entropy to merge neighbouring segments to lines or ellipses, respectively. But the entropy criterion may not favour one of these two models. Then we select the one with smallest description length. This happens in case of long segments having very small curvature. Here the segments may be approximated either by a long line or by an ellipse segment having small curvature.

We evaluate the description length for merged models from their residuals. We use the modified Akaike criterion (AKAIKE 1974) $\text{MDL}_{\text{AIC}} = -2 \log p(\boldsymbol{l}|\hat{\boldsymbol{\theta}}) + 2U$ using
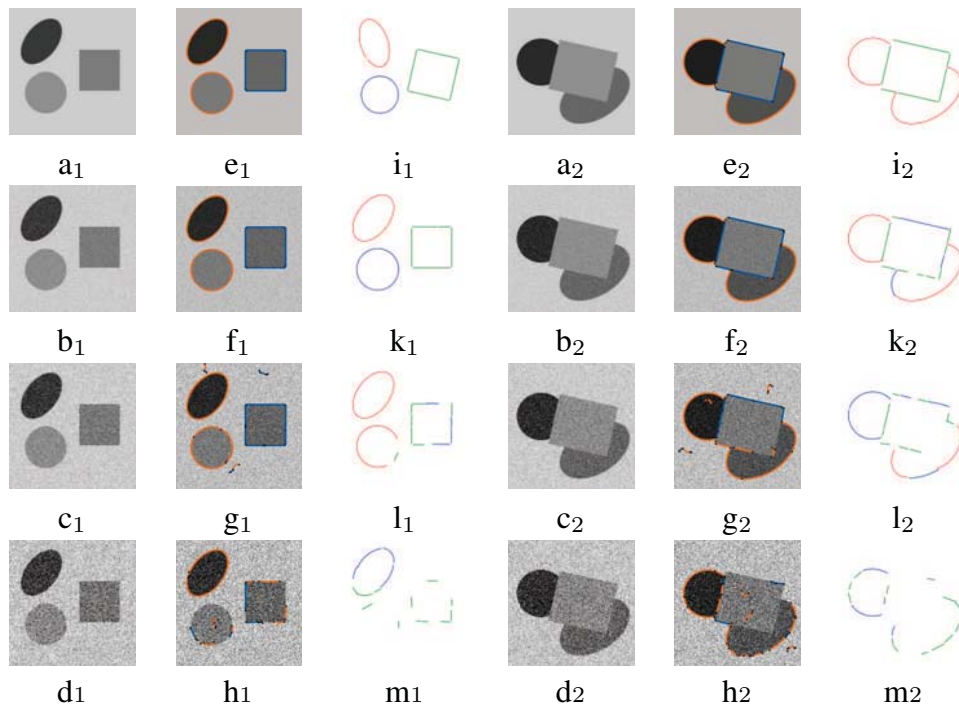
Fig. 2: Comparison to ELSD under different noise conditions (Best viewed in colour). a_*: given image. b_* - d_*: $\sigma_n = 10, 20, 40$ grayvalues. e_* - h_*: our final results. Colours, see Fig. 1. i_* - m_*: results of ELSD. Red - elliptical arcs. Blue - circular arcs. Green: lines.

the log-likelihood function of data l and estimated parameters $\widehat{\theta}$ and the number of parameters U. In case of normally distributed observations the log-likelihood function is equal to the sum of weighted squared residuals and we get

$$\text{MDL}_{\text{AIC}} = \Omega^2 + 2U. \qquad (9)$$

Now, after having collected all criteria, we start the simplification of the given pre-segmentation. This is done in a greedy manner where we try to simplify the polygon while considering given pixel chains and while keeping the change of geometry slow.

3.4 Algorithm

For each segment $o_n \in O$ we initialize lines l_n and ellipses C_n, if possible, i.e. we estimate model parameters $\{\hat{\theta}_1\}_n$, covariances $\{\Sigma_1\}_n$ and residuals $\{v\}_n$. Let us call them models $\mathcal{M}_n^l$ and $\mathcal{M}_n^C$, respectively. For all neighbouring elements we propose merging, i.e. estimate parameters $\{\hat{\theta}_2\}_{n,n+1}$, covariances $\{\Sigma_2\}_{n,n+1}$ and residuals $\{v\}_{n,n+1}$ of all

potentially merged models. Let us call them models $\mathcal{M}_{n,n+1}^l$ and $\mathcal{M}_{n,n+1}^C$, respectively. For these models $\mathcal{M}_n^*$ and $\mathcal{M}_{n,n+1}^*$ we evaluate $T_S = \mathcal{F}^{-1}(S, \Delta R, R_1)$, Δh_0 and Δh_g using (5) or (8). To simplify notation, we avoid the index $(n, n+1)$ in the following. If $\Delta h_0 < \frac{1}{2}\log T_S$ and $\Delta h_g < T_0 + \frac{1}{2}\log T_S$ we add the proposed model to the set of merging proposals $\mathcal{P}$.

We require the geometrical change to be as small as possible when merging two segments. Therefore, we may choose the model $\mathcal{M}$ from $\mathcal{P}$ with smallest Δh_g. But note that we can not compare changes in entropy between line and ellipses. These are different models of different complexity, thus we are not allowed to pick the model with smallest Δh_g from the whole set $\mathcal{P}$. In a greedy process we start with lines, i.e. first merging all lines, which fulfil the requirements and afterwards merging all ellipses. More precisely, we pick these proposed merged line l^* from segments o_n and o_{n+1} with smallest Δh_g. If merging these two segments to an ellipse is a valid choice, too, we choose the line model if $\text{MDL}(l^*) < \text{MDL}(C^*)$.

After merging two segments, we update Δh for all affected segments and again pick the best proposal concerning Δh_g. If there are no line proposals left we continue with ellipses the same way, except from evaluating MDL.

All segments left, e.g. those segments that could not merge, are tested whether their curvature is significantly different from 0. If so, they become an ellipse, if not they become a line. To be precise, we perform variance propagation on the curvature and perform hypothesis tests on the 95% significance level.

4 Results

This section presents some results. We give details about parameters, show the resulting segments and discuss the success of the merging step by means of some statistics. We compare our results to those from ELSD (elliptical line segment detector) (PATRAUCEAN et al. 2012) as this is state of the art and there exists code as well as an online demo to process own images using fixed parameters.

Parameter setting

There are just a few parameters to choose and these are well understandable and stable for all tested images.

From experiments we found the standard deviation of edge pixels $\sigma_e = 0.1$ [pixel]. For this we set the tolerance t for the pre-segmentation using `circlePeucker` to $t = 3 \cdot \sigma_e$. Due to compression artefacts and image distortions, lines in images often are not that smooth and we set the variance for grouping a factor three larger than σ_e. Thus, for fitting and merging lines and ellipses, the uncertainty of each pixel is assumed to be isotropic $\boldsymbol{\Sigma}_{pp} = (3 \cdot \sigma_e)^2 \boldsymbol{I}_2$.

The significance level for the Fischer-test-statistic in (7) is set to $S = 0.95$. The additive constant for evaluating the bound of Δh_g in (5) is set to the number of parameters of the current model, $T_0 = U$.

As our purpose is the segmentation of given pixel chains and not the interpretation of the image the identification of spurious scatter is out of scope. Our algorithm works stable even for very small chains. Nevertheless, to simplify the visualization we do not show short pixel chains, say shorter than 10 or 20 [pixels], depending on the structure of the image.

Synthetic data

First we investigate the noise sensitivity of the procedure using synthetic images, see Fig. 2. When changing the noise σ_n of an image from $\sigma_n(0)$ to $\sigma_n(k)$ the standard deviation of edge pixels by $\sigma_e(\sigma_n(k)) = \sqrt{1 + \frac{\sigma_n^2(k)}{\sigma_n^2(0)}} \cdot \sigma_e(0)$ is adapted, where we assumed the noise of the image to be $\sigma_n(0) = 2$ [gr]. The parameters of edge detection are not changed.

Please note that the proposed algorithm works quite stable up to a certain degree of noise. As long as the contrast is high, geometric elements are reliable and accurate detected.

Douglas-Peucker vs. `circlePeucker`

Next we show the effect of the circle-version of the Peucker-Algorithm. We compare the results of `circlePeucker` to the original Douglas-Peucker algorithm when used as pre-segmentation for the final merging step. The results are given in Fig. 3 and Tab. 1.

Fig. 3 shows the results for two natural images when using the classical Douglas-Peucker algorithm and the new `circlePeucker`, respectively, as pre-segmentation for the final merging step as described in section 3. We see that both algorithms perfectly approximate the given data. This is due to the tight threshold for the maximum distance to a fitted geometric element which is the same in both cases. But we realize, just by visual inspection, that our new segmentation reduces the number of segments significantly. For a quantitative evaluation of this reduction, we count the total number of segments for each processed image when using the original Douglas-Peucker and our new segmentation, respectively. Tab. 1 gives these numbers for each processed image together with the total number of evaluated pixel chains and the final number of segments after the merging step. We see that the new circle-based pre-segmentation reduces the number of segments by almost 50 % compared to the line-based Douglas-Peucker algorithm. The merging step further reduces the number of segments by about 25 %.

We show some of the advantages of pre-segmentation using `circlePeucker` by some details. E.g. the capital O of the STOP-sign actually consists of four arcs instead of

Fig. 3: Pre-segmentation `circlePeucker` vs. Douglas-Peucker (Best viewed in colour). a_*: Given images as used in PATRAUCEAN et al. (2012). b_*: Pre-Segmentation using Douglas-Peucker. c_*: Pre-Segmentation using `circlePeucker`. e_* and f_*: Final segmentation using b_* and c_*, respectively. d_* results by ELSD. Colours, see Figs. 1 and 2.

Tab. 1: Statistics of simplification. The number of objects in the first column refers to the number of evaluated pixel chains per image. The second column gives the number of line segments using the classical Douglas-Peucker (DP) algorithm. The third column gives the number of circle segments using `circlePeucker` (new). Fourth and fifth column give the number of segments for the final segmentation results.

	No. objects	No. segments			
		pre-segmentation		final	
		DP	new	DP	new
worm (Fig. 3)	187	1961	834	839	613
stop (Fig. 3)	159	1233	560	608	458
window (Fig. 5)	331	2226	868	941	726
icosahedron (Fig. 5)	1071	8991	6177	4517	3235
arcade (Fig. 4)	844	5908	2563	4675	2242
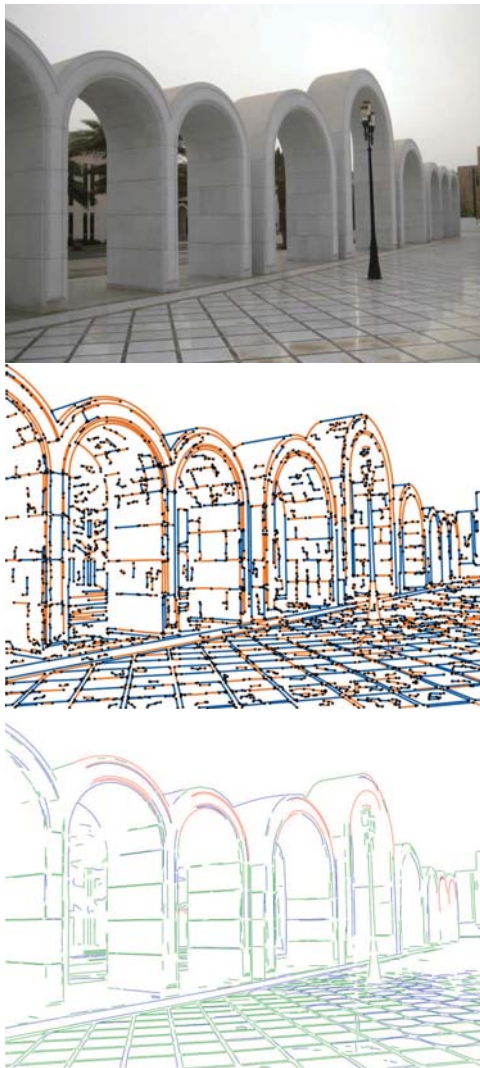
Fig. 4: More results on real images (Best viewed in colour). Arcade. From top to bottom: given image, our final result using `circlePeucker` for pre-segmentation, result of ELSD. Colours, see Figs. 1 and 2.

one ellipse. When using `circlePeucker` we get this result exactly. While using Douglas-Peucker tends to approximate arcs by lines, obviously. The main reason for this is the identification of break points candidates when evaluating the pre-segmentation. Obviously, `circlePeucker` identifies points of changing curvature more likely than Douglas-Peucker. The same effect can be observed for the boundary lines around the worm.

Comparison to ELSD

We give two more results on natural images in Figs. 4 and 5 and compare our results to those from from ELSD (PATRAUCEAN et al. 2012) in Figs. 2 to 5. Let us take the worm of the book cover shown in Fig. 3. ELSD resolves nearby edges, e.g. the black boundaries of the worm. The pre-processing of our method identifies these as (dark) lines, which are then simplified. The slightly curved boundaries of the letters W or B are straightened by ELSD, while better resolved by our method. ELSD simplifies too much, e.g. the ellipse of O in the STOP-sign. While ELSD detects the lines independently, our method segments the edge pixel chains, therefore at sharp corners occasionally an additional short segment is preserved, e.g. the rectangles within Fig. 2.

To summarize we see, the pre-segmentation using `circlePeucker` correctly identifies arc segments and especially their breakpoints. By itself these are promising results and improve the standard algorithm in terms of reducing the number of breakpoints of a given polygon while preserving the geometry.

The merging step identifies elliptical arcs correctly and further reduces the number of segments of most given pixel chains. Lines are identified in most cases, if not this might be due to distortions, especially for long lines.

5 Conclusion

We presented a line simplification approach which approximates given pixel chains by a sequence of lines and elliptical arcs. For this we proposed an adaption to Douglas-Peucker's algorithm for the use of circles instead of straight lines. Furthermore, we developed an approach for the simplification of such a segmentation by merging neighbouring segments due to their agreement to a joint geometric model in terms of bounded variation of entropy. The approach depends on just a few parameters which are clearly explained by a priori knowledge about edge detection accuracy. Depending on the assumed edge accuracy we showed very accurate results. We showed the effects of polyline segmentation and simplification on several images with comparable good results referring to an state of the art algorithm. We proved the success of merging in terms of the reduction rate of number of segments per object. We believe that the final segmentation gives rise to useful

Fig. 5: More results on real images (Best viewed in colour). Left: Gothic window. Right: cropped icosahedron. From top to bottom: given image, our final result using `circlePeucker` for pre-segmentation, result of ELSD. Colours, see Figs. 1 and 2.

high level image features as input for an image interpretation system.

Acknowledgement

References

AKAIKE, H., 1974: A new look at the statistical model identification. – IEEE Transaction on Automatic Control **AC-19**: 716–723: System identification and time-series analysis.

ALBANO, A., 1974: Representation of digitized contours in terms of conic arcs and straight-line segments. – Computer Graphics and Image Processing **3** (1): 23–33.

BEDER, C., 2005: Agglomerative grouping of observations by bounding entropy variation. – DAGM, LNCS **3663**: 101–108.

CHIA, A., RAJAN, D., LEUNG, M. & RAHARDJA, S., 2012: Object recognition by discriminative combinations of line segments, ellipses and appearance features. – PAMI **34** (9): 1758–1772.

COVER, T.M. & THOMAS, J., 1991: Elements of Information Theory. – Wiley.

DOUGLAS, D.H. & PEUCKER, T.K., 1973: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. – Cartographica **10** (2): 112–122.

FITZGIBBON, A.W., PILU, M. & FISHER, R.B., 1999: Direct least-squares fitting of ellipses. – PAMI **21** (5): 476–480.

FÖRSTNER, W., 1994: A framework for low-level feature extraction. – ECCV **801**/1994: 383–394.

FÖRSTNER, W., 2000: Image preprocessing for feature extraction in digital intensity, color and range images. – Geomatic Methods for the Analysis of Data in Earth Sciences **95**/2000: 165–189.

FUCHS, C. & FÖRSTNER, W., 1995: Polymorphic grouping for image segmentation. – ICCV.

GÜNTHER, O. & WONG, E., 1990: The arc tree: An approximation scheme to represent arbitrary curved shapes. – CVGIP **51** (3): 313–337.

JI, Q. & HARALICK, R.M., 1999: A Statistically Efficient Method for Ellipse Detection. – ICIP: 730–734.

JURIE, F. & SCHMID, C., 2004: Scale-invariant shape features for recognition of object categories. – CVPR **2**: II–90 – II–96.

KÖTHE, U., 2003: Edge and Junction Detection with an Improved Structure Tensor. – DAGM, LNCS **2781**: 25–32.

LIBUDA, L., GROTHUES, I. & KRAISS, K.-F., 2006: Ellipse detection in digital image data using geometric features. – VISAPP: 175–180.

McGLONE, J.C., 2004: Manual of Photogrammetry. – 5th edition edn. American Society for Photogrammetry and Remote Sensing.

MOORE, A., MASON, C., WHIGHAM, P.A. & THOMPSON-FAWCETT, M., 2003: The use of the circle tree for the efficient storage of polygons. – GeoComputation.

NGUYEN, T.P. & KERAUTRET, B., 2011: Ellipse detection through decomposition of circular arcs and line segments. – ICIAP: 554–564.

PATRAUCEAN, V., GURDJOS, P. & VON GIOI, R.G., 2012: A parameterless line segment and elliptical arc detector with enhanced ellipse fitting.. – ECCV **7573**: 572–585.

PAVLIDIS, T., 1983: Curve fitting with conic splines. – ACM Transaction Graphics **2** (1): 1–31.

PORRILL, J., 1990: Fitting ellipses and predicting confidence envelopes using a bias corrected kalman filter. – Image and Vision Computing **8** (1): 37–41.

ROSIN, P.L. & WEST, G.A.W., 1995: Nonparametric Segmentation of Curves into Various Representations. – PAMI **17** (12): 1140–1153.

WEST, G.A.W. & ROSIN, P.L., 1992: Multi-stage Combined Ellipse and Line Detection. – BMVC: 197–206.

WU, J., 2008: Robust Real-Time Ellipse Detection by Direct Least-Square-Fitting. – International Conference on Computer Science and Software Engineering **1**: 923–927.

Address of the Authors:

Dipl.-Ing. SUSANNE WENZEL, Rheinische Friedrich-Wilhelms-Universität zu Bonn, Institut für Geodäsie und Geoinformation, Photogrammetrie, Nußallee 15, D-53115 Bonn, Tel.: +49-228-73-2713, Fax: +49-228-73-2712 and Prof. Dr.-Ing. Dr. h.c. mult. WOLFGANG FÖRSTNER, Josef-Schell-Str. 34, 53121 Bonn, e-mail: {susanne}{wf}@ipb.uni-bonn.de

Article

Bundle Adjustment and System Calibration with Points at Infinity for Omnidirectional Camera Systems

JOHANNES SCHNEIDER & WOLFGANG FÖRSTNER, Bonn

Summary: We present a calibration method for multi-view cameras that provides a rigorous maximum likelihood estimation of the mutual orientation of the cameras within a rigid multi-camera system. No calibration targets are needed, just a movement of the multi-camera system taking synchronized images of a highly textured and static scene. Multi-camera systems with non-overlapping views have to be rotated within the scene so that corresponding points are visible in different cameras at different times of exposure. By using an extended version of the projective collinearity equation all estimates can be optimized in one bundle adjustment where we constrain the relative poses of the cameras to be fixed. For stabilizing camera orientations – especially rotations – one should generally use points at the horizon within the bundle adjustment, which classical bundle adjustment programs are not capable of. We use a minimal representation of homogeneous coordinates for image and scene points which allows us to use images of omnidirectional cameras with single viewpoint like fisheye cameras and scene points at a large distance from the camera or even at infinity.

We show results of our calibration method on (1) the omnidirectional multi-camera system Ladybug 3 from Point Grey, (2) a camera-rig with five cameras used for the acquisition of complex 3D structures and (3) a camera-rig mounted on a UAV consisting of four fisheye cameras which provide a large field of view and which is used for visual odometry and obstacle detection in the project MoD (DFG-Project FOR 1505 "Mapping on Demand").

Zusammenfassung: *Bündelausgleichung und Systemkalibrierung mit Punkten im Unendlichen für omnidirektionale Kamerasysteme.* In diesem Artikel stellen wir eine Kalibrierungsmethode für Multikamerasysteme vor, welche eine strenge Maximum-Likelihood-Schätzung der gegenseitigen Orientierungen der Kameras innerhalb eines starren Multikamerasystems ermöglicht. Zielmarken werden nicht benötigt. Das synchronisiert Bilder aufnehmende Kamerasystem muss lediglich in einer stark texturierten statischen Szene bewegt werden. Multikamerasysteme, deren Bilder sich nicht überlappen, werden innerhalb der Szene rotiert, so dass korrespondierende Punkte in jeder Kamera zu unterschiedlichen Aufnahmezeitpunkten sichtbar sind. Unter Verwendung einer erweiterten projektiven Kollinearitätsgleichung können alle zu schätzenden Größen in einer Bündelausgleichung optimiert werden. Zur Stabilisierung der Kameraorientierungen – besonders der Rotationen – sollten Punkte am Horizont in der Bündelausgleichung verwendet werden, wozu klassische Bündelausgleichungsprogramme nicht in der Lage sind. Wir benutzen eine minimale Repräsentation für homogene Koordinaten für Bild- und Objektpunkte, welche es uns ermöglicht, mit Bildern omnidirektionaler Kameras wie Fisheye-Kameras und mit Objektpunkten, welche weit entfernt oder im Unendlichen liegen, umzugehen.

Wir zeigen Ergebnisse unserer Kalibrierungsmethode für (1) das omnidirektionale Multikamerasystem Ladybug 3 von Point Grey, (2) ein Kamerasystem mit fünf Kameras zur Aufnahme komplexer 3D-Strukturen und (3) ein auf eine Drohne montiertes Kamerasystem mit vier Fisheye-Kameras, welches ein großes Sichtfeld besitzt und zur visuellen Odometrie und zur Hinderniserkennung im Projekt MoD (DFG-Projekt FOR 1505 „Mapping on Demand") verwendet wird.

1   Introduction

1.1   *Motivation*

The paper presents a rigorous bundle adjustment for the estimation of the mutual camera orientations in a rigid multi-camera system. It is based on an extended version of the projective collinearity equation which constrains the relative poses of the cameras to be fixed, whereby all estimates can be optimized in one bundle adjustment. Further it enables the use of image and scene points at infinity like the bundle adjustment "BACS" (Bundle Adjustment for Camera Systems) presented in SCHNEIDER et al. (2012).

Bundle adjustment is the work horse for orienting cameras and determining 3D points. It has a number of favourable properties: It is statistically optimal in case all statistical tools are exploited, highly efficient in case sparse matrix operations are used, useful for test field free self calibration, and can be parallelized to a high degree. In this paper we want to extend bundle adjustment with the estimation of the parameters of the mutual camera orientation in a multi-camera system.

Multi-camera systems are used to increase the resolution, to combine cameras with different spectral sensitivities (Z/I DMC, Vexcel Ultracam) or – like *omnidirectional cameras* – to augment the effective aperture angle (Blom Pictometry, Rollei Panoscan Mark III). Following SCARAMUZZA (2008), omnidirectional cameras have a viewing range of more than a half-sphere, such as multi-cameras systems, catadioptric cameras including mirrors, or also special fisheye lenses, such as the Lensagon BF2M15520. Additionally, multi-camera systems gain importance for the acquisition of complex 3D structures.

Far or even ideal points, i.e. points at infinity, e.g. points at the horizon or luminous stars are effective in stabilizing the orientation of cameras, especially their rotations.

In order to exploit the power of a bundle adjustment, it therefore needs to be extended to handle multi-camera systems and image and scene points at infinity, see Fig. 1.

1.2   *Notation*

Vectors and matrices are typed slanted boldface, e.g. $\boldsymbol{x}$ and $\boldsymbol{R}$. Homogeneous vectors and matrices are typed upright, e.g. $\mathbf{x}$ and $\mathbf{M}$. The
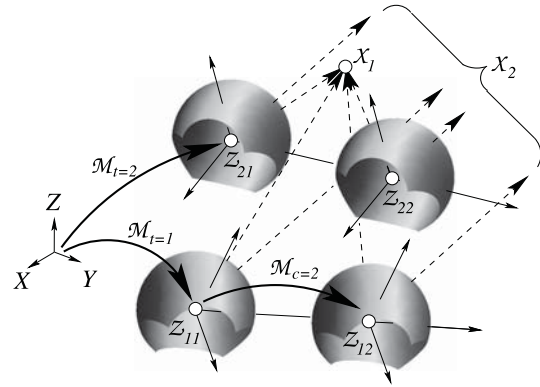
Fig. 1: A two-camera system with fisheye cameras $c = 1, 2$ with projection centers $\mathcal{Z}_{tc}$, rigid motion $\mathcal{M}_c$ and time-varying motion $\mathcal{M}_t$, having a field of view larger than $180°$ shown at two exposure times $t = 1, 2$ observing two points $\mathcal{X}_i, i = 1, 2$, one being close-by, the other at infinity. Already a block adjustment with a single camera moving over time will be stabilized by points at infinity.

skew 3×3-matrix $\boldsymbol{S}(\boldsymbol{a})$ induces the cross product, hence $\boldsymbol{S}(\boldsymbol{a})\boldsymbol{b} = \boldsymbol{a} \times \boldsymbol{b}$, vertical vector concatenation e.g. of $\boldsymbol{x}$ and $\boldsymbol{y}$ is written in Matlab syntax $[\boldsymbol{x}; \boldsymbol{y}] = [\boldsymbol{x}^\mathsf{T}, \boldsymbol{y}^\mathsf{T}]^\mathsf{T}$.

1.3   *The Idea*

The classical collinearity equations for image points $\mathcal{x}'_{it}([x'_{it}; y'_{it}])$ of scene point $\mathcal{X}_i([X_i; Y_i; Z_i])$ in camera t with rotation matrix $\mathcal{R}_t([r_{kk'}])$ with k and $k' = 1, ..., 3$ and projection center $\mathcal{Z}_t([X_{0t}; Y_{0t}; Z_{0t}])$ read as

$$x'_{it} = \frac{r_{11}(X_i - X_{0t}) + r_{21}(Y_i - Y_{0t}) + r_{31}(Z_i - Z_{0t})}{r_{13}(X_i - X_{0t}) + r_{23}(Y_i - Y_{0t}) + r_{33}(Z_i - Z_{0t})} \quad (1)$$

$$y'_{it} = \frac{r_{12}(X_i - X_{0t}) + r_{22}(Y_i - Y_{0t}) + r_{32}(Z_i - Z_{0t})}{r_{13}(X_i - X_{0t}) + r_{23}(Y_i - Y_{0t}) + r_{33}(Z_i - Z_{0t})} \quad (2)$$

Obviously, these equations are not useful for far points or ideal points, as small angles between rays lead to numerical instabilities or singularities. They are not useful for bundles of rays of omnidirectional cameras, because rays perpendicular to the viewing direction, as they may occur with fisheye cameras, cannot be transformed into image coordinates. This would require different versions of the collinearity equation depending on the type of sensor as one would need to integrate the camera model into the bundle adjustment.

We can avoid these disadvantages by using homogeneous coordinates $\mathbf{x}'_{it}$ and $\mathbf{X}_i$ for im-

age and scene points, a calibration matrix K_t and the motion matrix M_t, containing the pose parameters of the camera system, in: $\mathbf{x}'_{it} = \lambda_{it}[\mathsf{K}_t \mid \mathbf{0}]\mathsf{M}_t^{-1}\mathbf{X}_i = \lambda_{it}\mathsf{P}_t\mathbf{X}_i$. Obviously, (a) homogeneous image coordinates allow for ideal image points, even directions opposite to the viewing direction, (b) homogeneous scene coordinates allow for far and ideal scene points, and including an additional motion is simply an additional factor.

However, this leads to two problems. As the covariance matrices $\Sigma_{\mathbf{x}'_{it}\mathbf{x}'_{it}}$ of homogeneous vectors are singular, the optimization function of the maximum likelihood estimation $\sum_{it} |\mathbf{x}_{it} - \lambda_{it}\mathsf{P}_t\mathbf{X}_i|^2_{\Sigma_{x'_{it}x'_{it}}}$ formally cannot be used. A minor, but practical problem is the increase of the number of unknown parameters, namely the Lagrangian multipliers, which are necessary when fixing the length of the vectors $\mathbf{X}_i$. In large bundle adjustments with more than a million scene points this prohibitively increases the number of unknowns by a factor 5/3.

1.4 *Task and Challenges*

The task is to model the projection process of a camera system as the basis for a bundle adjustment for a multi-view camera system, which (a) consists of mutually fixed single-view cameras, (b) allows the single cameras to be omnidirectional, requiring to explicitly model the camera rays and (c) which allows for far or ideal scene points for stabilizing the configuration. The model formally reads as

$$\chi_{itc} = \mathcal{P}_c(\mathcal{M}_c^{-1}(\mathcal{M}_t^{-1}(\mathcal{X}_i))) \qquad (3)$$

with the I scene points $\mathcal{X}_i, i = 1, ..., I$, the T motions $\mathcal{M}_t, t = 1, ..., T$ of the camera system from the scene coordinate system, the C motions $\mathcal{M}_c, c = 1, ..., C$ of each single camera from the camera system, which makes the mutual orientation explicit, the projection $\mathcal{P}_c$ into the camera systems $c = 1, ..., C$, and the observed image points χ_{itc} of scene point i in camera c at time/pose t.

In order to realize this we need to be able to represent bundles of rays together with their uncertainty, using uncertain direction vectors, to represent scene points at infinity using homogeneous coordinates, and minimize the number of parameters to be estimated. The main challenge lies in the inclusion of the statistics into an adequate minimal representation.

2 Related Work

Multi-camera systems are proposed by many authors. E. g. MOSTAFA & SCHWARZ (2001) present an approach to integrate a multi-camera system with GPS and INS. NISTÉR et al. (2004) discuss the advantage to use a stereo video rig in order to avoid the difficulty with the scale transfer. SAVOPOL et al. (2000) report on a multi-camera system for an aerial platform to increase the resolution. In all cases, the multi-view geometry is only used locally.

Orientation of a stereo rig is discussed in HARTLEY & ZISSERMAN (2000, p. 493). MOURAGNON et al. (2009) propose a bundle solution for stereo rigs working in terms of direction vectors, but they minimize the angular error without considering the covariance matrix of the observed rays. FRAHM et al. (2004) present an approach for orienting a multi-camera system, however not applying a statistically rigorous approach. MUHLE et al. (2011) discuss the ability to calibrate a multi-camera system in case the views of the individual cameras are not overlapping. IKEDA et al. (2003) describe a geometric and photometric camera calibration for omnidirectional multi-camera systems using a calibration board and a total station. CARRERA et al. (2011) calibrate a general multi-camera system by mapping each camera individually and applying a global bundle adjustment afterwards. ZOMET et al. (2001) discuss the problem of re-calibrating a rig of cameras due to changes of the internal parameters. Bundle adjustment of camera systems are extensively discussed in the thesis of KIM (2010).

Uncertain geometric reasoning using projective entities is extensively presented in KANATANI (1996), but only using Euclideanly normalized geometric entities and allowing the estimation for some single geometric entities only. HEUEL (2004), eliminating these deficiencies, proposes an estimation procedure which does not eliminate the redundancy of the representation and also cannot easily include elementary constraints between observations, see MEIDOW et al. (2009). The following developments are based on the minimal representation schemes proposed in FÖRSTNER (2012) which reviews previous work and generalizes e.g. BARTOLI (2002).

3   Concept

3.1   *Model for a moving single-view Camera*

3.1.1   Image coordinates as observations

Using homogeneous coordinates

$$\mathbf{x}'_{it} = \lambda_{it} \mathsf{P}_t \mathbf{X}_i = \lambda_{it} \mathsf{K}_t R_t^\mathsf{T} [I_3 \mid -\boldsymbol{Z}_t] \mathbf{X}_i \quad (4)$$

with a projection matrix

$$\mathsf{P}_t = [\mathsf{K}_t \mid \mathbf{0}_{3\times1}] \mathsf{M}_t^{-1}, \quad \mathsf{M}_t = \begin{bmatrix} R_t & \boldsymbol{Z}_t \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$$

makes the motion of the camera explicit. It contains for each pose t: the projection center $\boldsymbol{Z}_t$ and the rotation matrix R_t, describing the translation and rotation between the scene coordinate system and the camera system, and the calibration matrix K_t, containing parameters for the principal point, the principal distance, the affinity, and possibly lens distortion, see MC-GLONE et al. (2004, 3.149 ff.) and (10). In case of an ideal camera with principal distance c, thus $\mathsf{K}_t = \mathrm{Diag}([c, c, 1])$, and Euclidean normalization of the homogeneous image coordinates with the k-th row $\mathbf{A}_{t,k}^\mathsf{T}$ of the projection matrix P_t

$$\mathbf{x}'^e_{it} = \frac{\mathsf{P}_t \mathbf{X}_i}{\mathbf{A}_{t,3}^\mathsf{T} \mathbf{X}_i} = \begin{bmatrix} \mathbf{A}_{t,1}^\mathsf{T} \mathbf{X}_i / \mathbf{A}_{t,3}^\mathsf{T} \mathbf{X}_i \\ \mathbf{A}_{t,2}^\mathsf{T} \mathbf{X}_i / \mathbf{A}_{t,3}^\mathsf{T} \mathbf{X}_i \\ 1 \end{bmatrix} \quad (5)$$

we obtain (1) and (2), e.g. $x'_{it} = \mathbf{A}_{t,1}^\mathsf{T} \mathbf{X}_i / \mathbf{A}_{t,3}^\mathsf{T} \mathbf{X}_i$.

Observe the transposition of the rotation matrix in (4), which differs from HARTLEY & ZISSERMAN (2000, (6.7)), but makes the motion of the camera from the scene coordinate system into the current camera system explicit, see KRAUS (1997).

3.1.2   Ray directions as observations

Using the directions from the cameras to the scene points we obtain the collinearity equations

$$^k\mathbf{x}'_{it} = \lambda_{it}{}^k\mathsf{P}_t \mathbf{X}_i = \lambda_{it} R_t^\mathsf{T} (\boldsymbol{X}_i - \boldsymbol{Z}_t)$$
$$= \lambda_{it}[I_3 \mid \mathbf{0}]\mathsf{M}_t^{-1}\mathbf{X}_i . \quad (6)$$

Instead of Euclidean normalization, we now perform spherical normalization $\mathbf{x}^s = \mathsf{N}(\mathbf{x}) =$ $\mathbf{x}/|\mathbf{x}|$, where $|\mathbf{x}|$ is the length of $\mathbf{x}$, yielding the collinearity equations for camera bundles

$$^k\mathbf{x}'^s_{it} = \mathsf{N}(^k\mathsf{P}_t \mathbf{X}_i) . \quad (7)$$

We thus assume the camera bundles to be given as T sets $\{^k\mathbf{x}_{it}, i \in \mathcal{I}_t\}$ of normalized directions for each time t of exposure. The unknown parameters are the six parameters of the motion in $^k\mathsf{P}_t$ and the three parameters of each scene point. Care has to be taken with the sign: We assume the negative Z-coordinate of the camera system to be the viewing direction. The scene points then need to have non-negative homogeneous coordinate $X_{i,4}$, which in case they are derived from Euclidean coordinates via $\mathbf{X}_i = [\boldsymbol{X}_i; 1]$ always is fulfilled. In case of ideal points, we therefore need to distinguish between the scene points $[\boldsymbol{X}_i; 0]$ and $[-\boldsymbol{X}_i; 0]$ which are points at infinity in opposite directions.

As a first result we observe: The difference between the classical collinearity equations and the collinearity equations for camera bundles is twofold. 1.) The unknown scale factor is eliminated differently: Euclidean normalization leads to the classical form in (5), spherical normalization leads to the bundle form in (7). 2.) The calibration is handled differently: In the classical form it is made explicit, here we assume the image data to be transformed into camera rays taking the calibration into account. This will make a difference in modelling the individual cameras during self-calibration, a topic we will not discuss in this paper.

3.1.3   Handling far and ideal scene points

Handling far and ideal scene points can easily be realized by also using *spherically* normalized coordinates $\mathbf{X}_i^s$ for the scene points leading to

$$^k\mathbf{x}'^s_{it} = \mathsf{N}(^k\mathsf{P}_t \mathbf{X}_i^s) . \quad (8)$$

Again care has to be taken with points at infinity.

The confidence ellipsoid of 3D points can be used to visualize the achieved precision, in case the points are not too far. For a simultaneous visualization of confidence ellipsoids of 3D points which are close and far w.r.t. the origin one could perform a stereographic projection of the 3D-space into a unit sphere, i.e.

$\boldsymbol{X} \mapsto \boldsymbol{X}/(1+|\boldsymbol{X}|)$ together with the transformation of the confidence ellipsoids. The relative poses of points close to the origin then will be preserved, far points will sit close to the boundary of the sphere. Their uncertainty in distance to the origin then can be inferred using their distance to the boundary of the sphere.

3.2 Model for Sets of Camera Systems

With an additional motion $\mathsf{M}_c(R_c, \boldsymbol{Z}_c)$ for each camera of the camera system we obtain the general model for camera bundles

$$^k\mathbf{x}_{itc}^{\prime s} = \mathsf{N}\left([I_3 \mid \mathbf{0}_{3\times1}]\,\mathsf{M}_c^{-1}\,\mathsf{M}_t^{-1}\,\mathbf{X}_i^s\right) \quad (9)$$

which makes all elements explicit: The observed directions $\boldsymbol{x}_{itc}^{\prime}(^k\mathbf{x}_{itc})$ represented by normalized 3-vectors, having two degrees of freedom, unknown or known scene point coordinates $\mathcal{X}_i(\mathbf{X}_i^s)$, represented by spherically normalized homogeneous 4-vectors, having 3 degrees of freedom, unknown pose $\mathcal{M}_t$ of camera system, having 6 parameters for each time a set of images was taken and unknown calibration M_c containing the relative pose of the cameras which are assumed to be rigid over time, having 6 parameters per camera. We refer relative poses to the first camera as reference camera with $R = I_3$ and $\boldsymbol{Z} = \mathbf{0}$.

3.3 Generating Camera Directions from observed Image Coordinates

In most cases the observations are made using a digital camera whose sensor is approximately planar. The transition to the directions of the camera rays needs to be performed before starting the bundle adjustment. As mentioned before, this requires the internal camera geometry to be known. Moreover, in order to arrive at a statistically optimal solution, one needs to transfer the uncertainty of the observed image coordinates to the uncertainty of the camera rays. As an example we discuss two cases.

3.3.1 Perspective cameras

In case of perspective cameras with small image distortions, we can use the camera-specific and
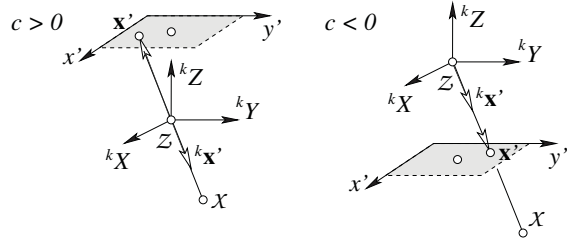
Fig. 2: The direction of the homogeneous image coordinate vector and the direction of the ray is different depending on the sign of the principal distance c.

maybe temporally varying calibration matrix

$$\mathsf{K}(\boldsymbol{x}', \boldsymbol{q}) = \begin{bmatrix} c & cs & x_H' + \Delta x(\boldsymbol{x}', \boldsymbol{q}) \\ 0 & c(1+m) & y_H' + \Delta y(\boldsymbol{x}', \boldsymbol{q}) \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

for the forward transformation

$$^g\mathbf{x}' = \mathsf{K}(\boldsymbol{x}', \boldsymbol{q})\,^k\mathbf{x}'^s \quad (11)$$

leading to the observable image coordinates $^g\mathbf{x}'$. The ^{g} indicates that the mapping can handle general distortions via additional parameters $\boldsymbol{q}$. Besides the basic parameters, namely the principal distance c with the image plane $^kZ = c$, the shear s, the scale difference m, and principal point $\boldsymbol{x}_H'$, the calibration matrix contains additive corrections for modelling lens distortion or other deviations, which depend on the additional parameters $\boldsymbol{q}$ and on $\boldsymbol{x}$, the position of the image point. In case of small deviations (11) can easily be inverted. However, one must take into account the different signs of the coordinate vector and the direction from the camera to the scene point (Fig. 2),

$$^k\mathbf{x}'^s \approx s\,\mathsf{N}\left(\mathsf{K}^{-1}(^g\boldsymbol{x}', \boldsymbol{q})\,^g\mathbf{x}'\right) \quad (12)$$

with $s \in \{-1, +1\}$ such that $^k x_3'^s < 0$. This relation is independent of the sign of the third element of the calibration matrix. Given the covariance matrix $\boldsymbol{\Sigma}_{^g x'^g x'}$ of the image coordinates, the covariance matrix of $^k\mathbf{x}'$ can be determined by variance propagation, omitting the dependency of the calibration matrix on the point coordinates $\boldsymbol{x}'$. Note that a point $^g\mathbf{x}'$ at infinity corresponds to the direction $^k\mathbf{x}'$ perpendicular to the viewing direction.
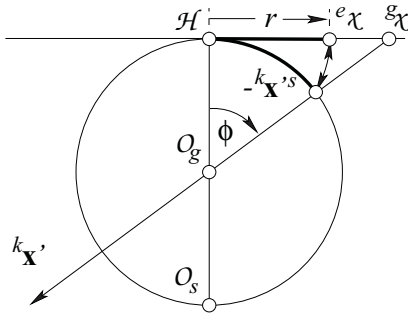
Fig. 3: Relation between sensor point, viewing direction and viewing ray.

3.3.2 Omnidirectional single-viewpoint cameras

As an example for an omnidirectional single-view camera we take a camera with a fisheye-lens. We model the fisheye objective with the equidistant-model described in ABRAHAM & FÖRSTNER (2005). The interior orientation of a camera is determined separately by camera calibration according to ABRAHAM & HAU (1997) using Chebyshev polynomials. Using the equidistant-projection and applying all corrections, we obtain image points $^e\chi$ which lie closer to the principal point $\mathcal{H}$ than the gnomonic projections $^g\chi$ of the scene points (Fig. 3). The ray direction $^k\mathbf{x}'^s$ can be derived from $^e\chi$ by using the normalized radial distance $r' = |^e\boldsymbol{x}|$ growing with the angle ϕ between the viewing direction and the camera ray.

Again, the uncertainty of the image coordinates can be transformed to the uncertainty of the direction $^k\mathbf{x}'^s$ of the camera ray via variance propagation. In all cases the covariance matrix of the camera ray is singular, as the normalized 3-vector only depends on two observed image coordinates.

3.4 *The Estimation Procedure*

The collinearity equations in (9) contain three equations per observed camera ray and four parameters for each homogeneous scene point, though, both being unit vectors. Therefore, the corresponding covariance matrices are singular and more than the necessary parameters are contained in the equations. We therefore want to reduce the number of parameters to the necessary minimum. We do this after linearization.

3.4.1 Linearization and update for pose and relative pose parameters

Linearization of the non-linear model leads to a linear substitute model which yields correction parameters that allow to derive corrected approximate values. We start with approximate values for the poses of the reference camera given for every time of exposure $t = 1, ..., T$ by R_t^a for the rotation matrix and $\boldsymbol{Z}_t^a$ for the projection center, the relative poses from the reference camera to each other camera $c = 2, ..., C$ given by R_c^a and $\boldsymbol{Z}_c^a$, $\mathbf{X}_i^{sa}$ for the $i = 1, ..., I$ spherically normalized scene points, and $\mathbf{x}_{itc}^a = \mathsf{N}\left([I_3 \mid \boldsymbol{0}_{3\times1}] \, \mathsf{M}_c^{-1} \, \mathsf{M}_t^{-1} \, \mathbf{X}_i^s\right)$ for the normalized directions.

The Euclidean coordinates will be simply corrected by $\boldsymbol{Z} = \boldsymbol{Z}^a + \Delta\boldsymbol{Z}$, the three parameters $\Delta\boldsymbol{Z}$ are to be estimated. The rotation matrix will be corrected by pre-multiplication with a small rotation, thus by $R = R(\Delta\boldsymbol{R}) R^a \approx (I_3 + S(\Delta\boldsymbol{R}))R^a$, where the small rotation $R(\Delta\boldsymbol{R})$ depends on a small rotation vector $\Delta\boldsymbol{R}$ that is to be estimated.

3.4.2 Reduced coordinates and update of coordinates

The correction of the unit vectors is performed using reduced coordinates (FÖRSTNER 2012). These are coordinates, say the 2-vector $\boldsymbol{x}_r$ of the direction $\mathbf{x}^s$, in the two-dimensional tangent space $\mathrm{null}\left(\mathbf{x}^{sa\mathsf{T}}\right) = [\boldsymbol{r}, \boldsymbol{s}]$ of the unit sphere S^2 evaluated at the approximate values $\mathbf{x}^{sa}$

$$\boldsymbol{x}_r = \mathrm{null}^\mathsf{T}\left(\mathbf{x}^{sa\mathsf{T}}\right)\mathbf{x}^s = \begin{bmatrix} \boldsymbol{r}^\mathsf{T}\mathbf{x}^s \\ \boldsymbol{s}^\mathsf{T}\mathbf{x}^s \end{bmatrix}. \quad (13)$$

The corrections $\Delta\boldsymbol{x}_r$ of these reduced coordinates are estimated. This leads to the following update rule

$$\mathbf{x}^s = \mathsf{N}\left(\mathbf{x}^{sa} + \mathrm{null}\left(\mathbf{x}^{sa\mathsf{T}}\right)\Delta\boldsymbol{x}_r\right). \quad (14)$$

Obviously, the approximate vector $\mathbf{x}^{sa}$ is corrected by

$$\Delta\mathbf{x} = \mathrm{null}\left(\mathbf{x}^{sa\mathsf{T}}\right)\Delta\boldsymbol{x}_r \quad (15)$$

and then spherically normalized to achieve the updated values $\mathbf{x}^s$.

Using (13) we now are able to reduce the number of equations per direction from three to two, making the two degrees of freedom of the

observed direction explicit. This results in pre-multiplication of all observation equations on (9) with $\text{null}^\mathsf{T}\left({}^k\mathbf{x}_{itc}^{sa\mathsf{T}}\right)$. Following (15) we use the substitution $\Delta\mathbf{X}_i^s = \text{null}\left(\mathbf{X}_i^{sa\mathsf{T}}\right)\Delta\boldsymbol{X}_{r,i}$ when linearizing the scene coordinates. Then we obtain the linearized model

$$
{}^k\boldsymbol{x}_{r,itc} + \widehat{\boldsymbol{v}}_{x_{r,itc}} \tag{16}
$$
$$
= J^\mathsf{T} R_c^{a\mathsf{T}} R_t^{a\mathsf{T}} S(\boldsymbol{X}_{i0}^a)\widehat{\Delta\boldsymbol{R}}_t
$$
$$
- X_{ih}^a J^\mathsf{T} R_c^{a\mathsf{T}} R_t^{a\mathsf{T}} \widehat{\Delta\boldsymbol{Z}}_t
$$
$$
+ J^\mathsf{T} R_c^{a\mathsf{T}} S\left(R_t^{a\mathsf{T}}(\boldsymbol{X}_{i0}^a - X_{ih}^a \boldsymbol{Z}_t^a)\right)\widehat{\Delta\boldsymbol{R}}_c
$$
$$
- X_{ih}^a J^\mathsf{T} R_c^{a\mathsf{T}} \widehat{\Delta\boldsymbol{Z}}_t
$$
$$
+ J^\mathsf{T}[I_3 \mid \mathbf{0}_3](\mathsf{M}_c^a)^{-1}(\mathsf{M}_t^a)^{-1}\text{null}\left(\mathbf{X}_i^{a\mathsf{T}}\right)\widehat{\Delta\boldsymbol{X}}_{ri}
$$

with

$$
J = \frac{1}{|\mathbf{x}|}\left(I_3 - \frac{\mathbf{x}\mathbf{x}^\mathsf{T}}{\mathbf{x}^\mathsf{T}\mathbf{x}}\right)\text{null}(\mathbf{x}^\mathsf{T})\bigg|_{\mathbf{x} = {}^k\mathbf{x}_{itc}^a} \tag{17}
$$

and the partitioned homogeneous vector $\mathbf{X}^s = [\boldsymbol{X}_0; X_h]$ depending on $\widehat{\Delta\boldsymbol{R}}_t$, $\widehat{\Delta\boldsymbol{Z}}_t$, $\widehat{\Delta\boldsymbol{R}}_c$, $\widehat{\Delta\boldsymbol{Z}}_c$ and $\widehat{\Delta\boldsymbol{X}}_{r,i}$.

We now arrive at a well-defined optimization problem: find $\widehat{\Delta\boldsymbol{X}}_{r,i}$, $\widehat{\Delta\boldsymbol{R}}_t$, $\widehat{\Delta\boldsymbol{Z}}_t$, $\widehat{\Delta\boldsymbol{R}}_c$, $\widehat{\Delta\boldsymbol{Z}}_c$ minimizing

$$
\Omega\left(\widehat{\Delta\boldsymbol{X}}_{r,i}, \widehat{\Delta\boldsymbol{R}}_t, \widehat{\Delta\boldsymbol{Z}}_t, \widehat{\Delta\boldsymbol{R}}_c, \widehat{\Delta\boldsymbol{Z}}_c\right) \tag{18}
$$
$$
= \sum_{itc} \widehat{\boldsymbol{v}}_{r,itc}^\mathsf{T} \Sigma_{x_{r,itc}x_{r,itc}}^{-1} \widehat{\boldsymbol{v}}_{r,itc}
$$

with the regular 2×2-covariance matrices

$$
\Sigma_{x_{r,itc}x_{r,itc}} \tag{19}
$$
$$
= {}^k J_s^\mathsf{T}({}^k\mathbf{x}_{itc}^a)\begin{bmatrix} \Sigma_{x_{itc}x_{itc}} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{bmatrix} {}^k J_s({}^k\mathbf{x}_{itc}^a).
$$

4 Experiments

4.1 *Implementation Details*

We have implemented the bundle adjustment as a Gauss-Markov model in Matlab. Observations are bundles of rays, the interior orientations of all cameras of the camera system are assumed to be known. To overcome the rank deficiency we define the gauge by introducing seven centroid constraints on the approximate

values of the scene points. This results in a free bundle adjustment, where the trace of the co-variance matrix of the estimated scene points is minimal. We can robustify the cost function by down-weighting measurements whose residual errors are too large by minimizing the robust Huber cost function HUBER (1981).

For the initialization sufficiently accurate approximate values for the scene point coordinates $\mathbf{X}_i^a$ and for the translation and the rotation of the relative poses M_c^a of each camera in the camera system of the reference camera as well as for the poses M_t^a of the reference camera in the scene coordinate system at the times of synchronized exposure are needed. Firstly, we determine the pose of each camera without considering the cameras as a rigid multi-camera rig using the bundle adjustment program Au-relo provided by LÄBE & FÖRSTNER (2006). With the first camera as the reference camera we then determine approximate values for all $c = 2, ..., C$ relative poses M_c^a robustly using the median quaternion and median translation over all $t = 1, ..., T$ unconstrained estimated relative poses. Scene points are triangulated by using all corresponding image points that are consistent with the approximated relative poses and the poses of the reference camera in the scene coordinate system. Ray directions with large residuals are discarded.

4.2 *Test on Correctness and Advantage*

We first check the correctness of the implemented model and then show the advantage of including far points or points with glancing intersections within the bundle adjustment based on a simulated scenario.

4.2.1 Simulated scenario

We simulated a multi-camera system moving on a radiused square, observing 50 close scene points and 10 scene points far away at the horizon, i.e. at infinity (Fig. 4). The multi-camera system contains three single-view cameras. Every scene point is observed by a camera ray from all 20 positions of the camera system. The simulated set-up provides a high redundancy of observations. Assuming the standard deviation of an image coordinate to be 0.3 pixel and
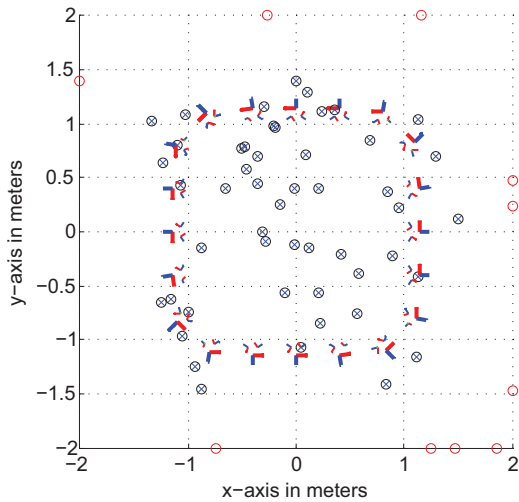
Fig. 4: Simulation of a moving multi-camera system (poses of reference camera shown as bold tripods) with loop closing. Scene points nearby (crossed dots) and at the horizon (empty dots) being numerically at infinity are observed.

Fig. 5: Multi-camera system consisting of five overlapping perspective camera views: Infrared camera on top, RGB camera in the middle and three monochromatic cameras. The distances from the RGB camera to the others are about 10 cm.

a principal distance of 500 pixel, we add normally distributed noise with $\sigma_l = 0.3/500$ radiant on the spherically normalized camera rays to simulate the observation process (Fig. 4). As initial values for the bundle adjustment we randomly disturb both the generated spherical normalized homogeneous scene points $\mathbf{X}_i^s$, which are directions, by $6°$, the generated motion parameters of the reference camera R_t and $\mathbf{Z}_t$ of M_t by $3°$ and 2 cm, and the relative pose parameters of the remaining cameras R_c and $\mathbf{Z}_c$ of M_c by $3°$ and 10 % of the relative distances between the projection centres.

The iterative estimation procedure stops after eight iterations, when the maximum normalized observation update is less than 10^{-6}. The residuals of the observed image rays in the tangent space of the adjusted camera rays, which are approximate angles between the rays in radiants, do not show any deviation from the normal distribution. The estimated a posteriori variance factor $\widehat{\sigma}_0^2 = 0.9925^2$ approves the a priori stochastic model with the variance factor $\sigma_0^2 = 1$. In order to test if the estimated orientation parameters and scene point coordinates represent the maximum likelihood estimates for normally distributed noise of the observations, we have generated the same simulation 2000 times with different random noise. The mean of the estimated variance factors is not significantly different from one, indicating an unbiased estimator with minimum variance. These results confirm the correctness of the approach and implementation.

4.2.2 Decrease of rotational precision excluding far points

Bundle adjustment programs, such as Aurelo, cannot handle scene points with glancing intersections, e.g. with maximal intersection angles lower than $\gamma = 1$ gon, which therefore are excluded in the estimation process to avoid numerical difficulties. Far scene points, however, can be observed over long periods of time and therefore should improve the quality of the rotation estimation significantly. We investigate the decrease of precision of the estimated rotation parameters of $\widehat{R}_t$ and $\widehat{R}_c$ when excluding scene points with glancing intersection angles. In detail, we will determine the average empirical standard deviation $\sigma_{\alpha_t} = \widehat{\sigma}_0 \sqrt{\operatorname{tr} \Sigma_{\widehat{R}_t \widehat{R}_t}/3}$ and $\sigma_{\alpha_c} = \widehat{\sigma}_0 \sqrt{\operatorname{tr} \Sigma_{\widehat{R}_c \widehat{R}_c}/3}$ for all estimated rotation parameters and report the average decrease of precision by excluding far points. They are determined by the geometric mean, namely $\exp\left[\sum_t^T \log(\sigma'_{\alpha_t}/\sigma_{\alpha_t})/T\right]$ and $\exp\left[\sum_c^{C-1} \log(\sigma'_{\alpha_c}/\sigma_{\alpha_c})/(C-1)\right]$, where σ'_{α_t} and σ'_{α_c} represent the resulting average empirical standard deviations when scene points whose maximal intersection angle are lower than a threshold γ are excluded.

We determine the decrease of precision for the estimated rotation parameters by excluding a varying number of scene points at infinity on the basis of the introduced simulation
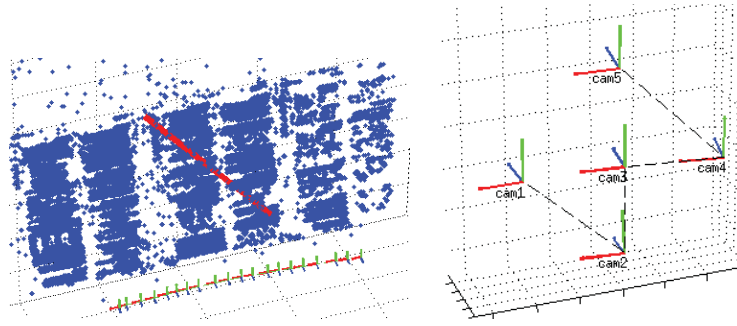
Fig. 6: Left: Illustration of the estimated scene points and poses of the reference camera. The red line denotes the known length on a poster for scale definition. Right: The estimated relative poses.

of a moving multi-camera system. Again we generate 50 scene points close to the multi-camera positions and vary the number of scene points at infinity to be 5, 10, 20, 50 and 100. The resulting average decrease in precision of the estimated rotations in $\widehat{\mathsf{M}}_c$ is 6.21 %, 8.98 %, 19.90 %, 42.29 % and 75.60 % and in $\widehat{\mathsf{M}}_t$ 7.15 %, 11.77 %, 27.67 %, 54.56 % and 91.28 %, respectively. This strongly proves the points at infinity to have a highly relevant positive influence on the rotational precision.

4.3 Calibration of Multi-Camera Systems

4.3.1 Calibration with overlapping views

We now describe the calibration of the camera system shown in Fig. 5 with highly overlapping views, which is used for 3D reconstruction of vines. In order to determine the relative poses of the multi-camera system we apply the bundle adjustment to 100 images of a wall draped with highly textured posters. The images were taken at 20 stations in a synchronized way. We use Aurelo without considering the known relative orientation between the cameras to obtain an initial solution for each camera and the scene points. The dataset contains 593,412 image points and 63,140 observed scene points.

Starting from an a priori standard deviation of the image coordinates of $\sigma_l = 1$ pixel, the a posteriori variance factor is estimated with $\widehat{\sigma}_0^2 = 0.11^2$ indicating the automatically extracted Lowe points to have an average precision of approximately 0.1 pixel. This high precision of the point detection results mainly from the good images and the calibration quality of the camera used. Fig. 6 illustrates the estimated

scene points and poses as well as the estimated relative poses.

The estimated uncertainty of the estimated rotations of the cameras with regard to the reference camera is 0.1 mgrad – 0.2 mgrad around the viewing direction axis and 0.4 mgrad – 0.8 mgrad orthogonal to it. We scale the photogrammetric model by using a measured distance of 1.105 m with an error of about 0.1 %. The uncertainty of the estimated relative translations is 0.02 mm – 0.04 mm in viewing direction and 0.1 mm – 0.2 mm orthogonal to it.

4.3.2 Multi-camera system Ladybug 3

The omnidirectional multi-camera system Ladybug 3 consists of six cameras, five of which are mounted in a circular manner, one showing upwards, together covering 80 % of the full viewing sphere. Neighbouring images only have a very small overlap, which is too weak for system calibration without additional information. We have mounted the omnidirectional multi-camera system Ladybug 3 on a robot (Fig. 7a), which executes a circular movement with a radius of 50 cm in a highly textured room while the Ladybug is taking synchronized images. This ensures overlapping images of different cameras at different times of the exposure. Approximate values for this image sequence consisting of 150 images taken by the five horizontal cameras at 30 exposure times are obtained with Aurelo that provides 135,012 image points of 24,078 observed scene points. The resulting 150 camera poses are shown in Fig. 7b.

After applying our bundle adjustment the estimated a posteriori variance factor amounts to $\widehat{\sigma}_0^2 = 0.25^2$ using a priori stochastic model with $\sigma_l = 1$ pixel for the image points, indicating the

Fig. 7: (a) Ladybug 3 on robot, (b) Estimated camera poses.

Fig. 8: Comparison of relative poses: estimated (solid) and manufacturer given (dashed).
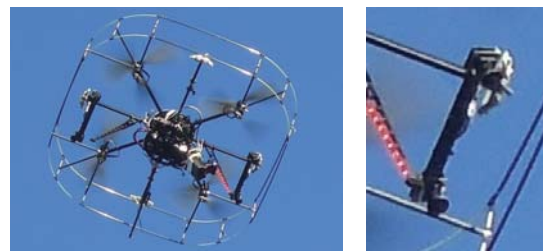
Fig. 9: Illustration of the UAV and the stereo cameras used in the MoD-project. One stereo pair is looking forward and one backwards giving a wide field of view.

automatically extracted Lowe points to have a quite good precision. Two parallel walls with a known distance of 7.01 m can be estimated out of the estimated scene points, which is used to define the scale between the estimated relative camera poses. The estimated rotation parameters show a very high precision. The maximal deviation to the manufacturer's calibration parameters is $0.6°$. The estimated uncertainties of the rotations and translations between the cameras are in the order of 1.5 mgrad – 2.5 mgrad and 0.1 mm – 0.2 mm, respectively.

To compare the estimated poses with the ones provided by the manufacturer we apply a rigid transformation which minimizes the distances between the estimated and given projection centers. The resulting estimated relative poses in Fig. 8 show significant translational deviations in the order of 1 mm – 4 mm compared to the manufacturer's calibration parameters.

The interior angles differ from a regular pentagon where each interior angle is $108°$ by up to $13°$. Possible reasons for the deviations are to few observed scene points near to the camera system and that we have used a different interior orientation for each camera from our own calibration, which is different from that of the manufacturer.

4.3.3 Multi-camera system with fisheye lenses

We make the same investigation on an image sequence consisting of 96 images taken by four synchronized cameras with Lensagon BF2M15520 fisheye lenses having a field angle up to $185°$. The cameras are mounted on an UAV (unmanned aerial vehicle) to generate two stereo pairs, one looking ahead and one looking backwards, providing a large field of view (Fig. 9). The UAV moves along a circle at a height of 5 m above a parking lot while rotating around its own axis, providing four overlapping images at each time of exposure.

Fisheye objectives cause severe distortions at the image boundary. Thus, in order to find corresponding points using the SIFT-operator we need to use a transformation between overlapping images which is very similar to a conformal projection, i.e. one that preserves angles because the SIFT operator is translation,

Fig. 10: Sample images of the Ladybug 3 dataset.

scale and rotation invariant. For this reason we transform the original images using the stereographic fisheye model. This ensures a conformal mapping between two different images when observing a scene at infinity as they themselves are conformal mappings of the spherical image of the scene. We obtain low deviations from a similarity transformation for locally planar points not too close to the cameras, fulfilling the preconditions for rotation and scale invariant SIFT-matching.

Aurelo provides approximate values for the 96 camera poses and 81,821 image points of 15,344 observed scene points which are transformed into image directions using (16). After the bundle adjustment the estimated variance factor is $\widehat{\sigma}_0^2 = 1.47^2$ using an a priori stochastic model with $\sigma_l = 1$ pixel for the image points, indicating a quite poor precision of the point detection. The cause for this low precision, which still needs to be analyzed, may be a lower image quality caused by both, the fisheye projection, or vibrations. The uncertainty of the estimated rotations and translations between the cameras within a stereo pair is 2 mgrad – 6 mgrad and 0.5 mm – 1.5 mm, respectively, and the uncertainty of the estimated rotations and translations between the forward and backward looking stereo camera systems is 5 mgrad – 9 mgrad and 1.5 mm – 2.5 mm.

4.4 Decrease of rotational Precision excluding far Points

In order to examine the decrease of the rotational precision of the estimated camera system poses we apply the bundle adjustment to an image sequence consisting of 360 images taken by four of the six cameras of the multicamera system Ladybug 3 (Fig. 10) excluding and including far points. The Ladybug 3 is mounted on a hand-guided platform and is triggered for one shot per meter with the help of an odometer. Approximate values are obtained with Aurelo by combining the individual cameras into a single virtual camera by adding distance-dependent corrections to the camera rays (SCHMEING et al. 2011).

The dataset contains 10,891 of 26,890 scene points observed with maximal intersection angles per point significantly lower than $\gamma = 1$ gon (histogram in Fig. 11a). The average standard deviation of each estimated rotation parameter is shown in Fig. 11b showing the individual gain in precision, sorted according to ascending rotational standard deviation. For some images the gain obviously is very large. The gain is mainly obtained due to a higher number of observed scene points at the individual poses, which can be seen in the scatter plot in Fig. 11c. Some of the estimated rotations show very large differences in the precision, demonstrating the relevance of the far scene points in the Ladybug 3 dataset. The use of far points results in an almost constant precision of the rotation parameters over all camera stations, compared to the results of the bundle adjustment if far points are excluded. The estimated a posteriori variance factor is $\widehat{\sigma}_0^2 = 1.05^2$ using an a priori stochastic model 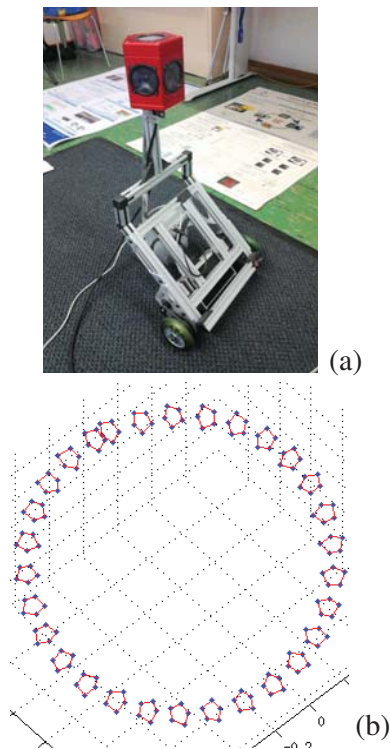with $\sigma_l = 1$ pixel for the image points, indicating a quite poor precision of the point detection which mainly results from the low image quality.

5 Conclusions and Future Work

We proposed a rigorous bundle adjustment for omnidirectional and multi-view cameras which enables an efficient maximum likelihood estimation using image and scene points at infinity and which can be used to calibrate a general

(a) Number of scene points with small intersection angles.

(b) Average empirical standard deviation of estimated rotations.

(c) Scatter plot of σ_{α_t} against the number of observed scene points at t.
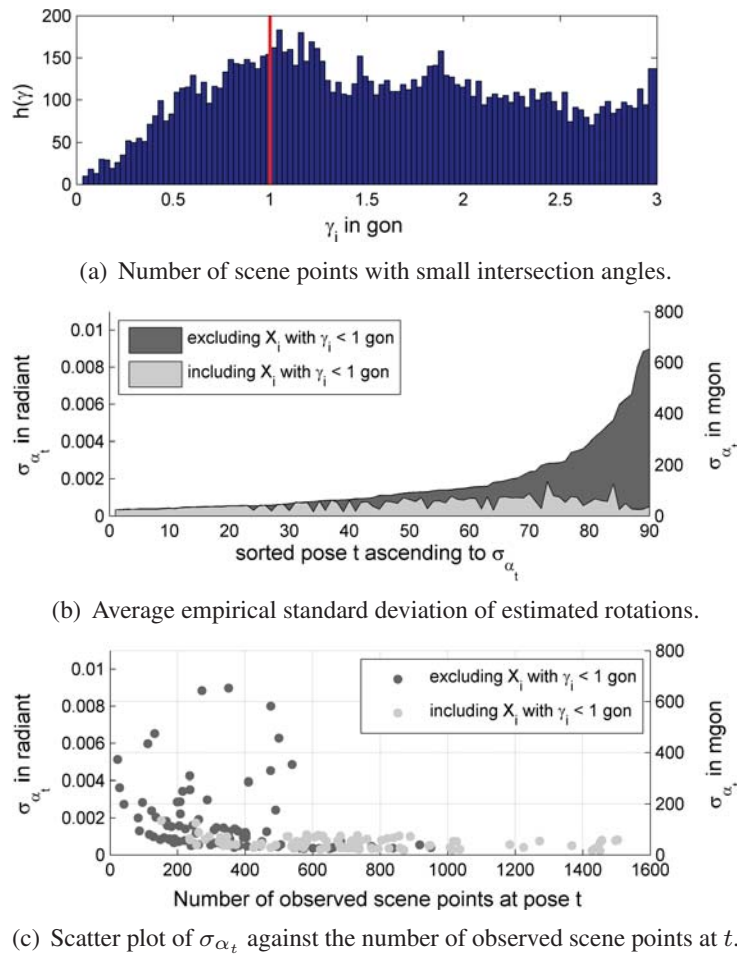
Fig. 11: The histogram in (a) shows the number of scene points in the multi-camera dataset with small intersection angles. The average precision σ_{α_t} determined by excluding and including scene points with $\gamma < 1$ gon for all poses $t = 1, ..., T$ is compared to each other in (b) and against the number of observed scene points in (c).

multi-camera system. Our experiments on simulated data show that scene points at the horizon can stabilize the orientation of the camera rotations significantly. Future work will focus on improving the precision of the relative poses by testing different image acquisition strategies. Furthermore, we are developing a fast C-implementation and eventually will extend our software by a self calibration part.

Software

The Matlab code of the proposed bundle adjustment will be made available at: `www.ipb.uni-bonn.de/bacs`.

6 Acknowledgements

References

ABRAHAM, S. & FÖRSTNER, W., 2005: Fish-Eye-Stereo Calibration and Epipolar Rectification. – ISPRS Journal of Photogrammetry & Remote Sensing **59** (5): 278–288.

ABRAHAM, S. & HAU, T., 1997: Towards Autonomous High-Precision Calibration of Digital Cameras. – Videometrics V, SPIE Annual Meeting **3174**: 82–93.

BARTOLI, A., 2002: On the Non-linear Optimization of Projective Motion Using Minimal Parameters. – 7th European Conference on Computer Vision - Part II: 340–354.

CARRERA, G., ANGELI, A. & DAVISON, A.J., 2011: SLAM-Based Automatic Extrinsic Calibration of a Multi-Camera Rig. – IEEE International Conference on Robotics and Automation: 2652–2659.

FÖRSTNER, W., 2012: Minimal Representations for Testing and Estimation in Projective Spaces. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2012** (3): 209–220.

FRAHM, J.-M., KÖSER, K. & KOCH, R., 2004: Pose estimation for Multi-camera Systems. – Pattern Recognition, 26th DAGM-Symposium: 286–293.

HARTLEY, R.I. & ZISSERMAN, A., 2000: Multiple View Geometry in Computer Vision. – Cambridge University Press.

HEUEL, S., 2004: Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction. – LNCS **3008**, Springer.

HUBER, P.J., 1981: Robust Statistics. – John Wiley, New York, NY, USA.

IKEDA, S., SATO, T. & YOKOYA, N., 2003: Calibration Method for an Omnidirectional Multi-camera System. – SPIE **5006**: 499–507.

KANATANI, K., 1996: Statistical Optimization for Geometric Computation: Theory and Practice. – Elsevier Science.

KIM, J.-H., 2010: Camera Motion Estimation for Multi-Camera Systems. – PhD thesis, School of Engineering, ANU College of Engineering and Computer Science, The Australian National University.

KRAUS, K., 1997: Photogrammetry. – Dümmler Verlag Bonn: Vol.1: Fundamentals and Standard Processes. Vol.2: Advanced Methods and Applications.

LÄBE, T. & FÖRSTNER, W., 2006: Automatic Relative Orientation of Images. – 5th Turkish-German Joint Geodetic Days.

McGLONE, C.J., MIKHAIL, E.M. & BETHEL, J.S., 2004: Manual of Photogrammetry. – American Society of Photogrammetry and Remote Sensing.

MEIDOW, J., BEDER, C. & FÖRSTNER, W., 2009: Reasoning with Uncertain Points, Straight Lines, and Straight Line Segments in 2D. – International Journal of Photogrammetry and Remote Sensing **64**: 125–139.

MOSTAFA, M.M. & SCHWARZ, K.-P., 2001: Digital Image Georeferencing from a Multiple Camera System by GPS/INS. – ISPRS Journal of Photogrammetry and Remote Sensing **56** (1): 1–12.

MOURAGNON, E., LHUILLIER, M., DHOME, M. & DEKEYSER, F., 2009: Generic and Real-time Structure from Motion Using Local Bundle Adjustment. – Image and Vision Computing **27** (8): 1178–1193.

MUHLE, D., ABRAHAM, S., HEIPKE, C. & WIGGENHAGEN, M., 2011: Estimating the Mutual Orientation in a Multi-Camera System with a Non Overlapping Field of View. – 2011 ISPRS Conference on Photogrammetric Image Analysis.

NISTÉR, D., NARODITSKY, O. & BERGEN, J., 2004: Visual Odometry. – IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 652–659.

SAVOPOL, F., CHAPMAN, M. & BOULIANNE, M., 2000: A Digital Multi CCD Camera System for Near Real-Time Mapping. – International Archives of Photogrammetry and Remote Sensing, Volume XXXIII. ISPRS Congress, Amsterdam, The Netherlands.

SCARAMUZZA, D., 2008: Omnidirectional Vision: from Calibration to Robot Motion Estimation. – PhD thesis, ETH Zurich.

SCHMEING, B., LÄBE, T. & FÖRSTNER, W., 2011: Trajectory Reconstruction Using Long Sequences of Digital Images From an Omnidirectional Camera. – 31th DGPF Conference.

SCHNEIDER, J., SCHINDLER, F., LÄBE, T. & FÖRSTNER, W., 2012: Bundle Adjustment for Multi-camera Systems with Points at Infinity. – ISPRS Annals of Photogrammetry, Remote Sensing and the Spatial Information Sciences, 22nd Congress of the ISPRS: 75–80, Daniel Verlag, Bonn.

ZOMET, A., WOLF, L. & SHASHUA, A., 2001: Omni-rig: Linear Self-recalibration of a Rig with Varying Internal and External Parameters. – Eighth IEEE International Conference on Computer Vision: 135–141.

Address of the Authors:

M.Sc. JOHANNES SCHNEIDER, Rheinische-Friedrich-Wilhelms-Universität Bonn, Institut für Geodäsie und Geoinformation, Photogrammetrie, Nußallee 15, D-53115 Bonn, Tel.: +49-228-73-2901, e-mail: johannes.schneider@uni-bonn.de.
Prof. Dr.-Ing. Dr. h. c. mult. WOLFGANG FÖRSTNER, Josef-Schell-Str. 34, D-53121 Bonn, e-mail: wf@ipb.uni-bonn.de

Article

A Robust Iterative Kalman Filter Based On Implicit Measurement Equations

RICHARD STEFFEN, Mülheim

Summary: In the field of photogrammetry, computer vision and robotics recursive estimation of time dependent processes is an important task. Usually Kalman filter based techniques are used which rely on explicit model functions that directly and explicitly describe the effect of the parameters on the observations. However, some problems naturally result in implicit constraints between the observations and the parameters, for instance all those resulting in homogeneous equation systems. By implicit we mean that the constraints are given by equations that are not easily solvable for the observation vector. We propose an iterative extended Kalman filter based on implicit measurement equations. The derived filter is useful for various applications, where the possibility to use implicit constraints simplifies the modeling. As an extension, we introduce a robustification technique similar to TING et al. (2007) and HUBER (1981) which down-weights the influence of potential outliers. The feasibility of the proposed framework is demonstrated at a number of typical computer vision applications.

Zusammenfassung: *Robuster iterativer Kalman-Filter mit implizierten Beobachtungsgleichungen*. In der Photogrammtrie, der Computer Vision und der Robotik finden rekursive Schätzungen ein weites Anwendungsspektrum. Üblicherweise werden in diesem Zusammenhang Kalman-Filter-basierte Techniken angewendet, welche auf expliziten Beobachtungsmodellen basieren, die den Effekt der Beobachtungen auf die Parameter direkt und explizit beschreiben. Einige Probleme sind jedoch aufgrund ihrer Natur als implizite Bedingungen zwischen den Parametern und den Beobachtungen formuliert, wie zum Beispiel Bedingungen unter Verwendung homogener Koordinaten. Unter impliziten Bedingungen verstehen wir Gleichungen, welche nicht trivial nach einem Beobachtungsvektor aufgelöst werden können. Diese Arbeit präsentiert einen iterativen erweiterten Kalman-Filter, welcher die Verwendung impliziter Beobachtungsgleichungen ermöglicht. Als eine Erweiterung führen wir ein Schema zur Robustifizierung nach TING et al. (2007) und HUBER (1981) ein, welches den Einfluss potenzieller Ausreißer reduziert. Die Nützlichkeit dieses Werkzeuges wird an einigen typischen Beispielen aus dem Bereich der Bildverarbeitung demonstriert.

1 Introduction

Recursive estimation and Kalman filtering is a classical technique (KALMAN 1960) and has been widely used in robotics and computer vision (WELCH & BISHOP 1995) to tackle problems such as positioning, object reconstruction, object tracking or calibration tasks. So far, the applications of Kalman filter techniques was limited to problems where the observations are represented by an explicit function in the unknown parameters.

However, many problems encountered in computer vision naturally result in implicit constraints between the observations and the parameters (HARTLEY & ZISSERMAN 2000, HEUEL 2004, PERWASS et al. 2005, HEUEL 2001). An example is the iterative estimation of a 2d line from 2d point observations (section 3).

Although it is always possible to reduce the solution of an implicit problem to the solution of an explicit problem (KOCH 1999, p.231ff) by introducing pseudo parameters, it is often much easier and straightforward to specify the measurement equations as implicit constraints relating the state vector to the observation vector.

The linear Kalman filter using explicit functions has been extended to the case of non-linear functions known as the extended Kalman filter (EKF) to deal with non-linear models and an iterative scheme (IEKF) to take into account the change of the linearization point. The only publication the author knows considering implicit constraints within Kalman filtering is SOATTO et al. (1994). However, the method only allows for a single iteration, equivalent to the classical EKF without taking into account the advantage of the IEKF.

The focus of this paper is on providing the foundations for integrating implicit constraints into a Kalman filter using the iterative scheme of the IEKF in order to enrich the application domain of the popular Kalman filter approach. As an enhancement to STEFFEN & BEDER (2007), we show that this novel framework is a generalization of the classical iterative Kalman filter and we demonstrate the benefit of the introduced approach for a variety of applications.

The traditional Kalman filter consists of two steps, a time update and a measurement update. Since the challenges for implicit constraints are in applying the measurement update, we mainly investigate the integration of implicit constraints to the update step.

Recently, the unscented Kalman filter (JULIER & UHLMANN 1997) has obtained a lot of attention which aims at improving the stochastic properties of the filter. Instead, our work aims at simplifying the specification of measurement equations instead.

This work is structured as follows: first we derive the prerequisites for the recursive estimation algorithm based on implicit measurement functions in section 2.1. Then we show, how outliers can be detected and the algorithms robustness can be improved. The final algorithm is given in section 2.3. Finally we present examples of computer vision tasks benefitting from the proposed algorithm in section 3. This comprises different tasks such as pose estimation, point cloud fitting and structure from motion.

2 Kalman Filter Update using Implicit Constraints

In this section we first derive the classical Kalman filter update using explicit measurement equations. Based on this concept, we show how to incorporate implicit constraints and adress the treatment of outliers. Section 2.3 provides a summary of the final algorithm.

2.1 *Update Estimation*

The Kalman filter consists of a dynamic model (prediction step) and a correction step (update step). The dynamic model is given by a non-linear function h which provides a predicted state vector $\bar{p}$ by $\bar{p} = h(p) + \epsilon$, with p as the previous state vector and ϵ as additional Gaussian random noise that influences the state vector. In addition, the state vectors uncertainty is given by its covariance matrix C_{pp} that has to be computed for the predicted state vector $\bar{C}_{pp}$ by error propagation from the previous time step. After the prediction, we usually obtain measurements to update the predicted state vector. Note that this process can be interpreted as a weighted mean (2), (3) of the predicted state itself and the state implied by the measurements. The measurement equation is given by

$$z + v = f(p), \qquad (1)$$

with z as the observation vector, v as the estimated residuals and $f(p)$ as an explicit function of the state vector. In general, f is non-linear and we have to update the state vector iteratively by $p^\nu = p^{\nu-1} + \Delta p^\nu$ until the convergence point has been reached ($\Delta p^\nu = 0$). The index ν indicates the iteration counter. The update vector Δp^ν can be obtained by an iterative maximum likelihood estimate using

$$\Delta p^\nu = F v^\nu \qquad (2)$$
$$= F\left(z - f(p^{\nu-1}) - A(\bar{p} - p^{\nu-1}) \right)$$

with A as the Jacobian of f w.r.t. the updated state vector p^ν and F as the influence or gain matrix by

$$F = \bar{C}_{pp} A^\mathsf{T} (C_{zz} + A\bar{C}_{pp} A^\mathsf{T})^{-1}, \qquad (3)$$

where C_{zz} denotes the covariance matrix of the observations. The last term $A(\bar{p} - p^{\nu-1})$ introduces the correction in the iterative extended Kalman filter (WELCH & BISHOP 1995). The covariance matrix of the estimated state vector can be obtained by the error propagation of (2) via

$$C_{pp} = (I - FA)\bar{C}_{pp}. \tag{4}$$

See WELCH & BISHOP (1995) or THRUN et al. (2001) for further details.

In the following, we derive a recursive estimation scheme for the case of non-linear implicit measurement constraints using the weighted mean idea. In complete analogy to the classical explicit Kalman filter we start with a parameter vector $\bar{p}_1$ (the state vector) and its covariance matrix $\bar{C}_1$ resulting from some prediction step. This state shall be updated according to a newly acquired measurement vector z that implicitly constrains the parameter vector. By implicit we mean that the measurement model is given by a non-linear implicit function

$$g(p, z) = 0 \tag{5}$$

relating the unknown parameter vector p to the observation vector z. Such an implicit observation model equation is often much easier to obtain than the explicit function $z = f(p)$ required by the classical Kalman filter. Note that every explicit function is easily made implicit by subtraction of the measurement vector $0 = g(p, z) = f(p) - z$.

We start by analyzing how a new parameter vector can be estimated from those observations alone by looking at the Taylor expansion of the observation model equation in (5)

$$0 \approx g(p^\nu, z^\nu) + A(p - p^\nu) + B^\mathsf{T}(\hat{z} - z^\nu)$$
$$= g(p^\nu, z^\nu) + A\Delta p + B^\mathsf{T}(\hat{z} - z + z - z^\nu) \tag{6}$$

containing the Jacobians

$$A = \left.\frac{\partial g(p, z)}{\partial p}\right|_{z^\nu, p^\nu} \quad B = \left.\frac{\partial g(p, z)^\mathsf{T}}{\partial z}\right|_{z^\nu, p^\nu}. \tag{7}$$

By rearranging this equation, we obtain the linear (left) and non-linear (right) contradiction part

$$A\Delta p + B^\mathsf{T}(\hat{z} - z) = -g(p^\nu, z^\nu) - B^\mathsf{T}(z - z^\nu), \tag{8}$$

which are equal at the convergence point with the final estimated observation vector $\hat{z}$. Given enough such observations, the maximum likelihood estimate of the parameter vector p is obtained by iteratively updating (FÖRSTNER & WROBEL 2004) similarly to the classical Kalman filter by

$$p^\nu = p^{\nu-1} + \Delta p^\nu, \tag{9}$$

with the non-linear contradiction c_g in

$$\Delta p^\nu = CA^\mathsf{T}(B^\mathsf{T} C_{zz} B)^{-1} c_g^\nu \tag{10}$$

using the covariance matrix

$$C = (A^\mathsf{T}(B^\mathsf{T} C_{zz} B)^{-1} A)^{-1} \tag{11}$$

and the non-linear contradiction vector

$$c_g^\nu = -g(p^{\nu-1}, z^{\nu-1}) - B^\mathsf{T}(z - z^{\nu-1}). \tag{12}$$

We can also compute the residuals of the observations (MIKHAIL & ACKERMANN 1976)

$$v^\nu = \hat{z} - z$$
$$= C_{zz} B(B^\mathsf{T} C_{zz} B)^{-1}(c_g^\nu - A\Delta p^\nu) \tag{13}$$

yielding the linearization point for the next iteration $z^{\nu+1} = z + v^\nu$ and p^ν from (9).

This estimation scheme for the computation of a parameter vector from a given observation set using implicit constraints is also known as the Gauss-Helmert model. Now we combine this estimation scheme with the state vector from the prediction step to achieve an iterative recursive update. To do so, we interpret the predicted state vector $\bar{p}$ as a direct observation z_1 of the new state vector, which fits into the above estimation scheme using the model equation

$$0 = g_1(p, z_1) = p - z_1 \tag{14}$$

and the observations $z_1 = \bar{p}$ having the covariance matrix $C_{zz_1} = C_{pp_1} = \bar{C}_{pp}$. Because this constraint is linear, the Jacobians are in this case simply $A_1 = I$ and $B_1 = -I$ and indepen-

dent of the linearization point. Considering the prediction of the state vector alone, we would obtain $\boldsymbol{p} = \bar{\boldsymbol{p}}$ so that $\boldsymbol{c}_{g_1} = \boldsymbol{0}$. As the measurement update is supposed to influence the state vector and thereby the joint linearization point, we have to cope with the change of the contradiction incurred by this change, which we call $\Delta \boldsymbol{c}_{g_1}$ in the following to reflect this important property. Plugging the direct measurement (14) and its Jacobians into (12) yields the contradiction

$$\Delta \boldsymbol{c}_{g_1} = \bar{\boldsymbol{p}} - \boldsymbol{p}^\nu \qquad (15)$$

because $\boldsymbol{g}_1(\boldsymbol{p}^\nu, \boldsymbol{z}_1^\nu) = \boldsymbol{0}$ is immediately fulfilled for the direct observation. Therefore, also the equation

$$C_{pp_1}^{-1} \Delta \boldsymbol{p}_1 = C_{pp_1}^{-1} \Delta \boldsymbol{c}_{g_1} \qquad (16)$$

holds, which will become useful in the following.

We are now ready to formulate the recursive estimation as a weighted mean process of two variables being the predicted state $\boldsymbol{p}_1$ on the one hand and the estimated state from the novel observations $\boldsymbol{p}_2$ on the other hand. Hence, the state update is given by

$$\Delta \boldsymbol{p} = (C_{pp_1}^{-1} + C_{pp_2}^{-1})^{-1} (C_{pp_1}^{-1} \Delta \boldsymbol{p}_1 + C_{pp_2}^{-1} \Delta \boldsymbol{p}_2). \qquad (17)$$

Substituting (16) and (10) into this weighted mean update, we obtain

$$\Delta \boldsymbol{p} = \overbrace{(C_{pp_1}^{-1} + C_{pp_2}^{-1})^{-1}}^{C_{pp}}$$
$$(C_{pp_1}^{-1} \Delta \boldsymbol{c}_{g_1} + A_2^\mathsf{T} (B_2^\mathsf{T} C_{zz} B_2)^{-1} \boldsymbol{c}_{g_2}). \qquad (18)$$

Using the well known matrix inversion (WOODBURY 1950) identity

$$(K + LN^{-1}M)^{-1} =$$
$$K^{-1} - K^{-1}L(N + MK^{-1}L)^{-1}MK^{-1},$$

we can reformulate (18) and finally get

$$\Delta \boldsymbol{p}^\nu = F \boldsymbol{c}_{g_2} + (I - FA_2)\Delta \boldsymbol{c}_{g1}, \qquad (19)$$

with the substitution

$$F = C_{pp_1} A_2^\mathsf{T} (B_2^\mathsf{T} C_{zz} B_2 + A_2 C_{pp_1} A_2^\mathsf{T})^{-1}, \qquad (20)$$

yielding the iterative update $\boldsymbol{p}^{\nu+1} = \boldsymbol{p}^\nu + \Delta \boldsymbol{p}^\nu$. The residuals are computed using (13)

$$\boldsymbol{v}_1 = -\Delta \boldsymbol{c}_{g_1} + \Delta \boldsymbol{p}^\nu \qquad (21)$$
$$\boldsymbol{v}_2 = C_{zz} B_2 (B_2^\mathsf{T} C_{zz} B_2)^{-1} (\boldsymbol{c}_{g_2} - A_2 \Delta \boldsymbol{p}^\nu) \qquad (22)$$

allowing to compute the contradiction for the next iteration

$$\boldsymbol{c}_{g_2} = -\boldsymbol{g}_2(\boldsymbol{p}^\nu, \boldsymbol{z}_2 + \boldsymbol{v}_2) + B_2^\mathsf{T} \boldsymbol{v}_2. \qquad (23)$$

Finally, note that the new covariance matrix of the state vector is given by

$$C_{pp} = (I - FA_2)\bar{C}_{pp}. \qquad (24)$$

As already mentioned, the explicit model of the classical Kalman filter can be transformed to an implicit model into the form $\boldsymbol{0} = \boldsymbol{f}(\boldsymbol{p}) - \boldsymbol{z}$. Applying this to the novel framework, the Jacobian w.r.t. the observation becomes B_2 becomes the negative identity matrix $-I$. Therefore, the gain matrix in (20) boils down to the classical gain matrix in (3). The contradiction in (23) becomes

$$\boldsymbol{c}_{g_2} = -(\boldsymbol{f}(\boldsymbol{p}) - \boldsymbol{z} + \boldsymbol{v}) + \boldsymbol{v}$$
$$\boldsymbol{c}_{g_2} = \boldsymbol{z} - \boldsymbol{f}(\boldsymbol{p}) \qquad (25)$$

Substitute $\boldsymbol{c}_{g_2}$ and (15) into the update equation (19) we obtain

$$\Delta \boldsymbol{p}^\nu = F(\boldsymbol{z} - \boldsymbol{f}(\boldsymbol{p}^{\nu-1}))$$
$$+ (I - FA_2)(\bar{\boldsymbol{p}} - \boldsymbol{p}^{\nu-1})$$
$$= (\bar{\boldsymbol{p}} - \boldsymbol{p}^\nu)$$
$$+ F\left(\boldsymbol{z} - \boldsymbol{f}(\boldsymbol{p}^{\nu-1}) - A_2(\bar{\boldsymbol{p}} - \boldsymbol{p}^{\nu-1})\right)$$

as the iterative update to the approximate values $\boldsymbol{p}^\nu = \boldsymbol{p}^{\nu-1} + \Delta \boldsymbol{p}^\nu$. To be equal to the classical iterative Kalman filter, this can be rewritten into (2) using the predicted state at $\nu = 0$ as the reference state. As a conclusion, the classical iterative Kalman filter is just a specialization of the novel framework derived here. Therefore, there is no need to compare the novel Kalman filter with the classical Kalman filter as from theory both are equal. A full proof can be found in (STEFFEN 2009).

The presented algorithm is based on a least squares optimization, which is known to be very

sensitive to outliers. In the following section, we show how the robustness of the presented method can be increased by re-weighting the observations.

2.2 Robustification by Re-Weighting

The classical Kalman filter as well as the estimation scheme presented so far minimizes the squared residuals of the observations, which is known to be sensitive to outliers. We now show how outliers may be detected by considering the plausibility of the computed residuals with respect to the expected uncertainty. By reducing the influence of such observations on the estimation, the robustness can be increased.

The weighted mean process is mainly influenced by two error effects. First, an erroneous dynamic model results in an erroneous prediction. Second, noisy observations yield a correction effect to the estimated state.

In TING et al. (2007) a robust outlier detection is presented for the classical Kalman filter. We will adapt this technique using an alternative re-weighting method proposed in HUBER (1981). Assuming an error free prediction, the update of the observations in $\boldsymbol{v}_2$ is normal distributed with zero mean. In this case, we are able to detect outliers by simply normalizing v_2 with the inverse observation covariance $\boldsymbol{C}_{zz}$ and reweight the observations accordingly.

However, in realistic applications the prediction model does not always hold true. Its effect on the improvement of the observations in $\boldsymbol{v}_2$ can not be modeled in general and depends on the system noise of the dynamic model. For instance, in the structure-from-motion problem an error in the camera position orthogonal to the viewing direction results in a consistent translation fraction in the image coordinates.

One common way to solve this problem is to approximate the complex deformation of the estimated observation $\widehat{\boldsymbol{z}} = \boldsymbol{z} + \boldsymbol{v}$. This can be done by choosing an approximation function depending on the expected deformations. In the case of image observations, a homography could be a good choice. The robust estimation of this function and the detection of the outliers can then be done by a RANSAC based approach or by a robustified least square solution. However, such a procedure for outlier detection is often quite expensive.
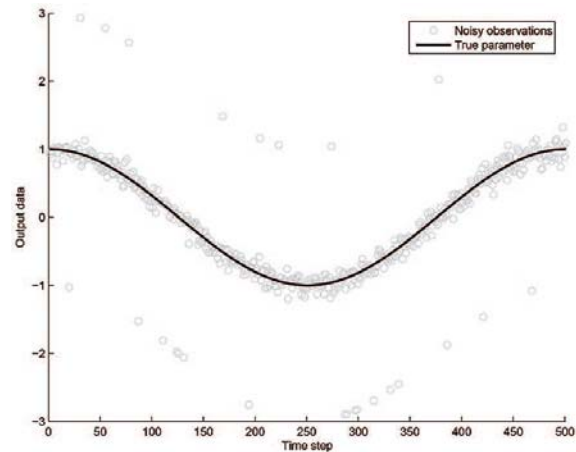
Fig. 1: Full cosine wavelength 2π, sampled with 500 samples, noise is 0.05, system noise 0.01, 5 percent outliers with strength of 2, iteration to convergence.
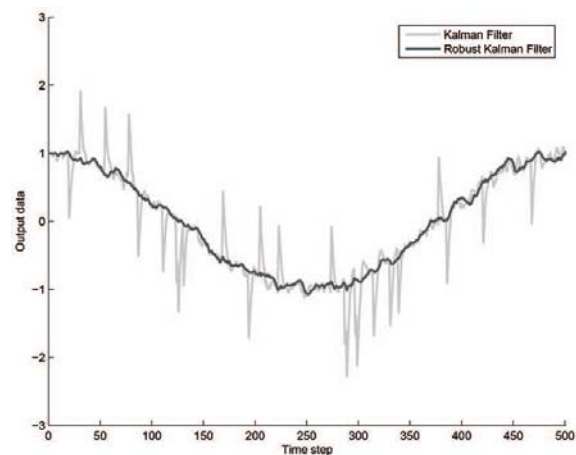
Fig. 2: Recursive estimation using the non-robustified and the robustified version of the Kalman filter.

From another point of view, the influence of the erroneous prediction is small if the system noise is large enough to compensate for the prediction error, which should be the case for a well approximating dynamic model. Thus, we are able to robustify the update by reweighting the observations in the following sense.

We first normalize the residuals $\boldsymbol{v}_{2_j}$ with the uncorrelated observation standard deviation to get standard normal distributed test values

$$c_j = \frac{v_{2_j}}{\sigma_j}. \qquad (26)$$

One can argue that the normalization should be

done using σ_v from C_{vv}. However, in the case that all observations can be assumed to have the same influence (equal redundancy number) to the parameter vector, it is suitable to use σ_z instead. The absolute test values c_j allow to decide for each single observation, whether there is a reason to consider it as an outlier. We then compute a variance factor w_j for each observation according to HUBER (1981)

$$w_j = \begin{cases} 1 & \text{if } \|c_j\| \leq k \\ \frac{\|c_j\|}{k} & \text{if } \|c_j\| > k \end{cases}, \quad (27)$$

which does not alter the observations within the range of k times the expected standard deviation and reduces the effect of observations outside this range on the estimation. To perform the desired re-weighting, we use the observation covariance matrix

$$C_{zz}^{(\nu)} = diag(\boldsymbol{w}) C_{zz}^{(0)} \quad (28)$$

instead of the initially given covariance matrix $C_{zz}^{(0)}$ in each iteration.

Following the experimental validation of TING et al. (2007), we also demonstrate the robustification on the one dimensional estimation of a cosinus curve containing some outliers. In Fig. 1 the noisy observations with 5% of outliers are shown. Fig. 2 shows the non robust and the robust version of the estimated curve parameters. The robustification yields a much smoother estimate not being perturbed by the outliers.

2.3 *The final Algorithm*

We now summarize the recursive estimation algorithm. From a previous estimation or prediction step of the filter, a current state vector $\boldsymbol{p}_1$ together with its covariance C_{11} is known. We gather additional observations $\boldsymbol{z}_2$ together with their covariance matrix C_{22} in a subsequent measurement step. The following algorithm may then be applied to update the state vector accordingly

1. set $\Delta \boldsymbol{p} = \boldsymbol{0}$
2. set $\boldsymbol{p} = \bar{\boldsymbol{p}}$
3. set $\boldsymbol{v}_1 = \boldsymbol{0}$
4. set $\boldsymbol{v}_2 = \boldsymbol{0}$, hence $\boldsymbol{z}^0 = \boldsymbol{z}$
5. Iterate until $\Delta \boldsymbol{p}^\nu$ is sufficiently small
   (a) compute Jacobians A_2 and B_2 at $\boldsymbol{p}^\nu$ and $\boldsymbol{z}^\nu$
   (b) compute the gain matrix F according to (20)
   (c) compute $\boldsymbol{c}_{g_2}$ according to (23)
   (d) compute $\Delta \boldsymbol{c}_{g_1}$ according to (15)
   (e) compute $\Delta \boldsymbol{p}^\nu$ according to (19)
   (f) update $\boldsymbol{p}^\nu$ with $\Delta \boldsymbol{p}^\nu$
   (g) compute $\boldsymbol{v}_1$ according to (21)
   (h) compute $\boldsymbol{v}_2$ according to (22)
   (i) update $\boldsymbol{z}^\nu = \boldsymbol{z} + \boldsymbol{v}_2$
   (j) compute normalized test values according to (26)
   (k) compute variance factor for all observations with (27)
   (l) compute reweighted observation covariance matrix for the next iteration
6. compute C_{pp} according to (24) .

After the algorithm is converged, we finally obtain the updated state vector $\boldsymbol{p}$ together with its covariance matrix C_{pp}. The only problem specific part is the computation of the Jacobians in step 5a, which has to be adapted. This completes the measurement update using the implicit constraint and a subsequent time update may be performed. Also note that for implicit measurement equations obtained directly from explicit equations by subtraction, the presented algorithm yields the same results as the classical iterated extended Kalman filter.

3 Exemplary Applications

In this section we present some examples for the usefulness of the proposed algorithm. Given that theoretically both our proposed framework and the traditional Kalman filter are equivalent, we do not provide a numerical example. More precisely, the results of the proposed algorithm is equivalent to the results of the traditional Kalman filters when using the explicit constraints as well when using the pseudo parameterization. Our motivation is to demonstrate the straightforwardness using implicit measurement equations for Kalman filter based estimation. Note, the observation in a Kalman filter must be minimal represented. For instance, using homogeneous observations lead to a singular covariance matrix of the observations and the Kalman filter update will fail.

3.1 *Line and Plane Estimation*

The recursive estimation of basic primitives from an uncertain point cloud is useful in a lot

of applications, e.g. in the task of urban scene reconstruction (POLLEFEYS et al. 2008). In this section we demonstrate solutions for a recursive 2d line and a recursive 3d plane estimation. A 2d line can be represented given a set of two parameters $l = \{a, b\}$. For every Euclidean point $\boldsymbol{x}_i = \{x_i, y_i\}$ the incidence with the line l is given by

$$y_i = ax_i + b. \tag{29}$$

Note, we are not able to reformulate this equation for both observations x_i and y_i explitely, but we can easily rewrite (29) in the implicit form

$$0 = g(\boldsymbol{p}, \boldsymbol{z}) = ax_i + b - y_i. \tag{30}$$

Using our proposed method to estimate the line parameter recursively with (30) as measurement update equation, the state vector contains $\boldsymbol{p} = [a, b]^\mathsf{T}$ and the observation vector the observed 2d points $\boldsymbol{z} = [x_i, y_i]^\mathsf{T}$. To handle vertical lines, (30) should be replaced by $0 = \mathbf{l}^\mathsf{T}\mathbf{x}$ with $\mathbf{l}$ as the homogeneous line with the unknown parameter d (distance from origin) and ϕ (angle of the line) and $\mathbf{x}$ as the homogeneous 2d point.

In a similar way, we are able to recursively estimate a plane which approximates a given point cloud. The incidence of an homogeneous 3d point $\mathbf{X}_i = [X_i, Y_i, Z_i, 1]^\mathsf{T}$ and a plane $\mathbf{A} = [\boldsymbol{n}, -d]^\mathsf{T}$ can be expressed as a simple bilinear constraint (Fig. 3).

$$0 = \mathbf{A}^\mathsf{T}\mathbf{X}_i. \tag{31}$$

Again, this constraint can not be easily reformulated to an explicit observation measurement function.

3.2 *Pose Estimation with parameterizable Shapes*

Pose estimation from a single camera observation is another interesting application. In this section, we give an example for the pose estimation of a simple sphere. In the subsequent section, we extend this approach to a more general formulation.

Assume that we are able to observe the silhouette of a projected sphere with a known radius (Fig. 4). For every point $\mathbf{x}_i$ on the silhou-
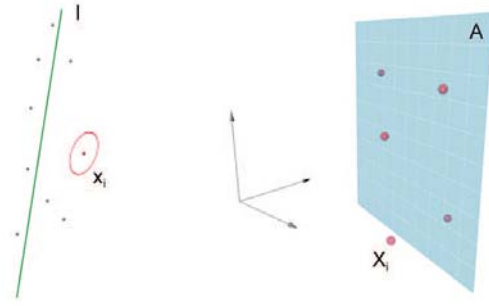
Fig. 3: Left: Recursive 2d line estimation from uncertain 2d points. Right: Recursive 3d Plane estimation from an uncertain 3d point cloud.

ette we can formulate an implicit measurement equation in the following way:

The projection ray $\mathbf{L}_i = [\mathbf{L}^h, \mathbf{L}^0]^\mathsf{T}$ in Pluecker representation can be obtained using a known camera orientation by

$$\mathbf{L}_i = \bar{\mathsf{P}}_L^\mathsf{T}\mathbf{x}_i = \begin{bmatrix} \mathbf{L}^h \\ \mathbf{L}^0 \end{bmatrix} = \begin{bmatrix} \mathsf{C} \\ \mathsf{D} \end{bmatrix}\mathbf{x}_i \tag{32}$$

The inverse projection matrix $\bar{\mathsf{P}}_L$ can be computed as a function of the camera orientation and its calibration (HEUEL 2004). The distance between the projection ray of the silhouette points $\mathbf{X}_i$ and the center of the sphere $\mathbf{X}_c$ has to be the known radius R. We can express this constraint implicitly using $\mathsf{S}(\bullet)$ as the skew matrix of a vector by

$$\begin{aligned} 0 &= \frac{\|\mathsf{S}(\mathbf{L}_i^h)\mathbf{X}_c + \mathbf{L}_i^0\|}{\|\mathbf{L}_i^h\|} - R \\ &= \frac{\mathsf{S}(\mathsf{C}\mathbf{x}_i)\mathbf{X}_c + \mathsf{D}\mathbf{x}_i}{\|\mathsf{C}\mathbf{x}_i\|} - R. \end{aligned} \tag{33}$$

Estimating the center of a moving sphere observing the projected silhouette and assuming a linear motion model, the state vector contains the spheres center and its velocity. The observation vector contains the observed Euclidean silhouette image coordinates. The parameters C, D and R are assumed to be known and constant.

3.3 *Pose Estimation with meshed Shapes*

Pose estimation from a single camera often deals with known shapes given as 3d meshes,
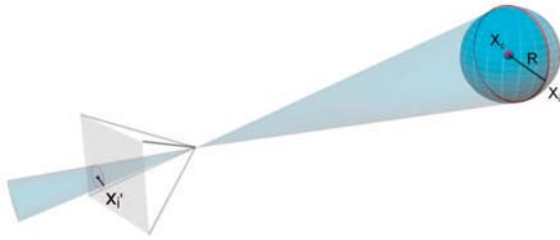
Fig. 4: Recursive 3d pose estimation of a sphere with known radius using observed silhouette points in an image.

(ROSENHAHN et al. 2004). Extracting the silhouette in the projected image of the observed object, we can use an edge (from the mesh) to point (from the silhouette) incidence as the measurement equation, similar to the well known ICP algorithm.
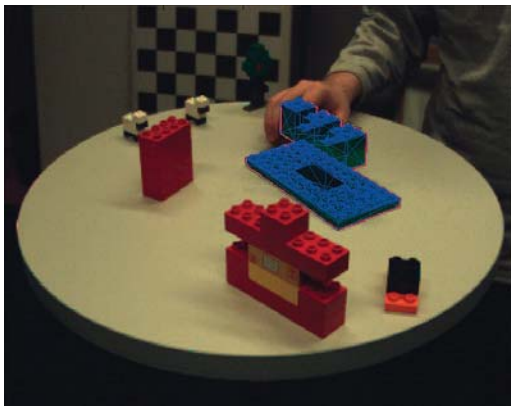
Fig. 5: Recursive 3d pose estimation with a single camera of a known object. The object is represented as a meshed shape with observed silhouette points in an image (Image and silhouette data provided by B. ROSENHAHN, MPI).

The pose of an arbitrary 3d object can be represented by its Euclidean position r and its orientation $\mathbf{q}$ represented as quaternion. Position and orientation are combined in the motion matrix

$$\mathsf{M} = \mathsf{M}(\boldsymbol{r}, \mathbf{q}) = \begin{bmatrix} R(\mathbf{q}) & \boldsymbol{r} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}. \qquad (34)$$

Applying the motion to the homogeneous 3d mesh point coordinates $\mathbf{X}_i$ given in a local object system and projecting them into the image space, we get the image coordinates of the pro-

jected mesh points with

$$\mathbf{x}_i = \mathsf{PM}\mathbf{X}_i. \qquad (35)$$

Mesh edges are defined by two connected points $\{a, b\}$. The congruent line in the image space can be obtained by

$$\mathbf{l}_{ab} = \mathsf{S}(\mathbf{x}_a)\mathbf{x}_b. \qquad (36)$$

The incidence of the observed silhouette point $\mathbf{x}'$ and this projected edge is given by a bilinear contraint

$$0 = \mathbf{l}_{ab}^\mathsf{T}\mathbf{x}'. \qquad (37)$$

Substitute (35) and (36) into (37) we get the final contraint

$$0 = (\mathsf{S}(\mathsf{PM}\mathbf{X}_a)\mathsf{PM}\mathbf{X}_b)^\mathsf{T}\mathbf{x}'. \qquad (38)$$

Using our recursive update model, the state vector contains the position and orientation of the 3d object and its velocities assuming a linear prediction model given by

$$\boldsymbol{p} = \begin{bmatrix} \boldsymbol{r} & \mathbf{q} & \dot{\boldsymbol{r}} & \dot{\mathbf{q}} \end{bmatrix}^\mathsf{T}. \qquad (39)$$

The measurement vector contains the observed Euclidean 2d silhouette points.

3.4 *Artificial Horizon Estimation*

In the field of robotics and navigation, an artificial horizon identification can be useful to stabilize the system over a long period of time as shown in NETO et al. (2011). In case of real measurements, the horizon line measurement is uncertain. This leads to undesirable effects in subsequent processes. In this case, the first choice is a Kalman filter based smoothing.

Using an implicit measurement update equation leads to an astonishling simple solution. The reference horizon in a normalized camera coordinate system can be expressed as a line in homogeneous representation by

$$\mathbf{l}^o = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

We denote the desired pitch by ω and roll by κ. In homogeneous representation, we use the simple measurement equation by the incidence
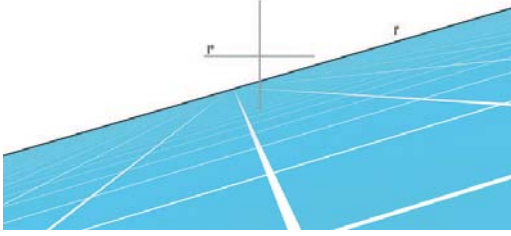
Fig. 6: Artificial horizon estimation from a single horizon line.

formulation

$$\mathbf{0} = S(\mathbf{l})R_\kappa R_\omega \mathbf{l}^o. \qquad (40)$$

Again, using our recursive update model the state vector contains the roll and pitch angles and their velocities by $\boldsymbol{p} = \begin{bmatrix} \kappa & \omega & \dot{\kappa} & \dot{\omega} \end{bmatrix}^\mathsf{T}$. The measurement vector contains a line representation, for instance by angle and distance from origin (*hessian normal form*) or by *gradient and intercept* $\{a, b\}$, (29).

3.5 Structure from Motion with Points and Lines

Following the approach of DAVISON (2003), the motion of a single camera can be described by the following state vector

$$\boldsymbol{p} = \begin{bmatrix} \boldsymbol{r} & \mathbf{q} & \dot{\boldsymbol{r}} & \dot{\mathbf{q}} & \boldsymbol{X}_1 \dots \boldsymbol{X}_i & \mathbf{L}_1 \dots \mathbf{L}_j \end{bmatrix}^\mathsf{T} \qquad (41)$$

comprising the camera state followed by a set of feature parameters. The camera trajectory is represented by its actual position $\boldsymbol{r}$, its orientation quaternion $\mathbf{q}$, its velocity vector $\dot{\boldsymbol{r}}$ and its angular velocity vector $\dot{\mathbf{q}}$. The 3d point coordinates are represented by their Euclidean points $\boldsymbol{X}_i$. Additionally we introduce 3d lines represented by its Pluecker coordinates $\mathbf{L}_i$. The interior camera parameters are assumed to be known.

We assume a linear time update model. In the approach of DAVISON (2003) the measurement model for object points is based on the co-linearity equations, which can be written as homogeneous equations

$$\mathbf{x}_i = \mathsf{P}\boldsymbol{X}_i \quad \text{with} \quad \mathsf{P} = \mathsf{K}R(\mathbf{q}) \begin{bmatrix} I_{3\times 3} | -\boldsymbol{r} \end{bmatrix}. \qquad (42)$$

As our approach is able to cope with implicit functions, we formulate the co-linearity constraint using the cross-product such that the co-linearity equations can be stated as an implicit equation

$$S(\mathbf{x}_i)\mathsf{P}\mathbf{X}_i = -S(\mathsf{P}\mathbf{X}_i)\mathbf{x}_i = \mathbf{0}. \qquad (43)$$

Obviously, those implicit constraints are equivalent to the explicit constraints used in DAVISON (2003). Also observe that they are non-linear in the camera pose parameters. Using only the point observations in our filter, the results are equal to the results achieved using a classical iterative extended Kalman filter.

Now let us incorporate the lines in the same way. The co-linearity equations for lines can be represented in homogeneous coordinates by

$$\mathbf{l}_j = \mathsf{P}_L \mathbf{L}_j \quad \text{with} \quad \mathsf{P}_L\limits_{3\times 6} = \mathsf{P}_L(\mathsf{P}), \qquad (44)$$

where P_L can be computed as a function of the camera orientation and its calibration (HEUEL 2004). This constraint can not be easily expressed as non homogeneous observations in an explicit formulation. Using implicit constraints we can reformulate (44) in the same easy way to

$$S(\mathbf{l}_j)\mathsf{P}_L\mathbf{L}_j = -S(\mathsf{P}_L\mathbf{L}_j)\mathbf{l}_j = \mathbf{0}. \qquad (45)$$

Using two constrains for points in (43) and two constrains for lines in (45) for every image point and line as an observation, we can solve the Kalman filter based update combining points and lines in one update step.

4 Conclusion

We presented a novel derivation of a recursive estimation framework in a Kalman filter approach which allows us to use implicit measurement constraint equations rather than being restricted to explicit ones. By using implicit constraints, the task of modeling recursive estimation schemes is eased significantly. Furthermore, we presented an improvement to the framework in order to deal with outliers in the observations. Instead of the elimination of observations, we used a re-weighting method.

We demonstrated the usefulness of this new algorithm on exemplary classical computer vision tasks.

The presented method is applicable to a broad range of time driven estimation problems, especially including all those resulting in homogeneous equation systems.

References

DAVISON, A.J., 2003: Real-time simultaneous localisation and mapping with a single camera. – 9th International Conference on Computer Vision: 674–679.

FÖRSTNER, W. & WROBEL, B., 2004: Mathematical concepts in photogrammetry. – MCGLONE, J.C., MIKHAIL, E.M. & BETHEL, J. (eds.): Manual of Photogrammetry, ASPRS: 15–180.

HARTLEY, R. & ZISSERMAN, A., 2000: Multiple View Geometry in Computer Vision. – Cambridge University Press.

HEUEL, S., 2001: Points, lines and planes and their optimal estimation. – Pattern Recognition, 23rd DAGM Symposium, LNCS **2191**: 92–99, Springer.

HEUEL, S., 2004: Uncertain Projective Geometry - Statistical Reasoning for Polyhedral Object Reconstruction. – Springer.

HUBER, P., 1981: Robust Statistics. – Wiley, New York, USA.

JULIER, S. & UHLMANN, J., 1997: A new extension of the Kalman filter to nonlinear systems. – International Symposium Aerospace/Defense Sensing, Simulations and Controls, Orlando, FL, USA.

KALMAN, R.E., 1960: A new approach to linear filtering and prediction problems. – Journal of Basic Engineering: 35–45.

KOCH, K.-R., 1999: Parameter Estimation and Hypothesis Testing in Linear Models. – 2nd edn., Springer.

MIKHAIL, E.M. & ACKERMANN, F., 1976: Observations and Least Squares. – University Press of America.

NETO, A.D.M., VICTORINO, A.C., FANTONI, I. & ZAMPIERI, D.E., 2011: Robust horizon finding algorithm for real-time autonomous navigation based on monocular vision. – 14th International IEEE Conference on Intelligent Transportation Systems, ITSC 2011.

PERWASS, C., GEBKEN, C. & SOMMER, G., 2005: Estimation of geometric entities and operators from uncertain data. – 27. Symposium für Mustererkennung, DAGM 2005, Wien, LNCS: 459–467 Springer-Verlag, Berlin, Heidelberg.

POLLEFEYS, M., NISTÉR, D., FRAHM, J.-M., AKBARZADEH, A., MORDOHAI, P., CLIPP, B., ENGELS, C., GALLUP, D., KIM, S.J., MERRELL, P., SALMI, C., SINHA, S.N., TALTON, B., WANG, L., YANG, Q., STEWÉNIUS, H., YANG, R., WELCH, G. & TOWLES, H., 2008: Detailed real-time urban 3d reconstruction from video. – International Journal of Computer Vision **78** (2-3): 143–167.

ROSENHAHN, B., PERWASS, C. & SOMMER, G., 2004: Pose estimation of 3d free-form contours: **62**: 267–289.

SOATTO, S., FREZZA, R. & PERONA, P., 1994: Motion estimation on the essential manifold. – ECCV '94: Third European Conference-Volume II on Computer Vision: 61–72, Springer-Verlag, London, UK.

STEFFEN, R., 2009: Visual SLAM from image sequences acquired by unmanned aerial vehicles. – PhD thesis, Institute of Photogrammetry, University of Bonn.

STEFFEN, R. & BEDER, C., 2007: Recursive estimation with implicit constraints. – HAMPRECHT, F., SCHNÖR, C. & JÄHNE, B. (eds.): DAGM 2007, LNCS **4713**: 194–203, Springer.

THRUN, S., BURGARD, W. & FOX, D., 2001: – Probabilistic Robotics (Intelligent Robotics and Autonomous Agents).

TING, J., THEODOROU, E. & SCHAAL, S., 2007: A kalman filter for robust outlier detection. – IEEE International Conference on Intelligent Robotics Systems.

WELCH, G. & BISHOP, G., 1995: An introduction to the Kalman filter. – Technical report. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

WOODBURY, M., 1950: Inverting modified matrices. – Memorandum 42, Statistics Research Group Princeton, New Jersey, USA.

Address of the Author:

Dr.-Ing. RICHARD STEFFEN, Löhberg 78, D-45468 Mülheim a.d. Ruhr, Tel: +49-208-4500025, e-mail: rsteffen@messbild.de

Article

Estimation and Mapping of Carbon Stocks in Riparian Forests by using a Machine Learning Approach with Multiple Geodata

Leonhard Suchenwirth, Michael Förster, Berlin, Friederike Lang, Freiburg & Birgit Kleinschmit, Berlin

Summary: Floodplain ecosystems offer valuable carbon sequestration potential. In comparison to other terrestrial ecosystems, riparian forests have a considerably higher storage capacity for organic carbon (C_{org}). However, a scientific foundation for the creation of large-scale maps that show the spatial distribution of C_{org} is still lacking. In this paper we explore a machine learning approach using remote sensing and additional geographic data for an area-wide high-resolution estimation of C_{org} stock distribution and evaluate the relevance of individual geofactors. The research area is the Danube Floodplain National Park in Austria, one of the very few pristine riparian habitats left in Central Europe. Two satellite images (Ikonos and RapidEye), historical and current topographic maps, a digital elevation model (DEM), and mean groundwater level (MGW) were included. We compared classifications of C_{org} stocks in vegetation, soils, and total biomass based on two, three, four, and five classes. The results showed that a spatial model of C_{org} in riparian forests can be generated by using a combination of object-based image analysis (OBIA) and classification and regression trees (CART) algorithm. The complexity of floodplains, where patterns of C_{org} distribution are inherently difficult to define, clearly exacerbated the challenge of achieving high classification accuracy. In assessing the relevance of individual geofactors, we found that remote sensing parameters are more important for the classification of C_{org} in vegetation, whereas parameters from auxiliary geodata, e.g. elevation or historical riverbeds, have more influence for the classification of soil C_{org} stocks. This was also confirmed by a comparative linear multiple regression analysis.

Zusammenfassung: *Schätzung und Kartierung von Kohlenstoffvorräten in Auwäldern mithilfe eines Ansatzes des maschinellen Lernens und verschiedenartigen Geodaten.* Auenökosysteme haben ein hohes Speicherpotenzial für organischen Kohlenstoff (C_{org}), auch im Vergleich zu anderen terrestrischen Ökosystemen. Allerdings fehlt eine wissenschaftliche Grundlage für die Schaffung von großmaßstäbigen Karten, die die räumliche Verteilung des C_{org} zeigen. In diesem Beitrag untersuchen wir einen Ansatz des maschinellen Lernens mittels Fernerkundungs- und zusätzlichen geografischen Daten für eine flächendeckende hochauflösende Abschätzung der C_{org}-Verteilung und bewerten die Relevanz der einzelnen Geofaktoren. Das Untersuchungsgebiet ist der Nationalpark Donau-Auen in Österreich, eines der wenigen unberührten Auenhabitate in Mitteleuropa. Zwei Satellitenbilder (Ikonos und RapidEye), historische und aktuelle topografische Karten, das digitale Geländemodell und Grundwasserdaten wurden einbezogen. Wir verglichen die Klassifizierung des C_{org}-Gehalts in Vegetation, Boden und Gesamtbiomasse in zwei, drei, vier und fünf Klassen. Die Ergebnisse zeigen ein räumliches Modell der C_{org}-Verteilung in Auwäldern mit der Kombination einer objektbasierten Bildanalyse (OBIA) und einem CART (Klassifikations- und Regressionsbaum) -Algorithmus. Die Komplexität der Auen, in denen Muster von C_{org}-Verteilung von Natur aus schwer zu definieren sind, erschwerte es, eine hohe Klassifizierungsgenauigkeit zu erzielen. Bei der Beurteilung der Relevanz einzelner Geofaktoren zeigte sich, dass die Fernerkundungsparameter wichtig für die Klassifizierung von C_{org} in der Vegetation sind, während die Höhe oder die Lage des historischen Flussbetts mehr Einfluss auf die Klassifizierung des C_{org}-Gehalts im Boden haben. Dies wurde auch durch eine vergleichende lineare multiple Regression bestätigt.

---

1   Introduction

Floodplain ecosystems offer valuable carbon sequestration potential. Riparian forests have a considerably higher storage capacity for organic carbon (C_{org}) than other terrestrial ecosystems (Cierjacks et al. 2010, Hoffmann et al. 2009, Mitra et al. 2005). Among the different floodplain compartments, it is essential to pay special attention to riparian forest vegetation, but also to soils, which often dominate C_{org} pools (Baritz et al. 2010, Harrison et al. 1995, Hofmann & Anders 1996, Kooch et al. 2012, Lal 2005).

Despite the importance of floodplains for carbon sequestration, a scientific foundation for creating large-scale maps showing the spatial distribution of C_{org} is still lacking. Carbon distribution can be mapped at a global or national level, but regional validation is usually not available (Gibbs et al. 2007, Groombridge & Jenkins 2002, UNEP-WCMC 2008). In particular, there are no maps showing the actual allocation of the C_{org} storage within riparian soils and vegetation at the local or regional level. Various studies have focussed on C_{org} stocks in ecosystems, such as in alder fens (Busse & Gunkel 2002), coastal plain floodplains (Giese et al. 2000), boreal lakes in Ontario (Hazlett et al. 2005) or timber plantations in Scandinavia (Backéus et al. 2005, Cao et al. 2010). In tropical and subtropical wetlands there has been research on mangroves and shrimp farms in Thailand (Matsui et al. 2009), seasonal sequestration in the Okavango delta (Mitsch et al. 2010) and Panama (Grimm et al. 2008). Cierjacks et al. (2011) provided statistical models on the spatial distribution of C_{org} stocks in Danubian floodplain vegetation and soils. Rheinhardt et al. (2012) used indicators based on the distance to river for biomass estimations in a river system in North Carolina. However, these studies rely on data collected by cost-intensive field surveys. For improving the estimation of C_{org}, including larger or less accessible wetland and riparian areas, combined methods of remote sensing, geographic information systems (GIS) and machine learning are promising techniques.

A wide range of remote sensing methods (Farid et al. 2008, Munyati 2000, Ozesmi & Bauer 2002) and in particular object-based image analysis (OBIA) (Kollár et al. 2011, Rokitnicki-Wojcik et al. 2011, Wagner 2009) have been used for mapping of wetland habitats. However, these studies related to the differentiation of vegetation classes and did not focus on the assessment of biomass or C_{org}.

In addition, various remote sensing analyses of C_{org} stocks have been done for non-floodplain habitats, but most of these studies have focused either on C_{org} stocks in soil (Behrens & Scholten 2006, McBratney et al. 2003) or in vegetation (Awaya et al. 2004, Hilker et al. 2008, Olofsson et al. 2008). So far, no studies on the estimation of total C_{org} stocks in riparian forests have been done. And despite advances in remote sensing and geodata analysis, these techniques have not yet been applied to the analysis and estimation of area-wide C_{org} stocks in floodplains.

Goetz et al. (2009) distinguished three approaches for using remote sensing data to map carbon stocks. In the simplest method, the stratify and multiply (SM) approach, e.g. as used by Mayaux et al. (2004) or Suchenwirth et al. (2012), a single value or a range of values is assigned to each class of land cover, vegetation type, or other site characteristic. This approach is limited due to the range of biomass within any given thematic class and the ambiguities concerning the identification of given types. The second approach, combine and assign (CA), extends the SM approach to a wider range of spatial data to improve classifications (Gibbs et al. 2007). It has the advantage of using finer spatial units of aggregation and weighted data layers, but is limited due to the moot representativeness of class values and difficulties in acquiring consistent information as the study area size increases. The third approach, direct remote sensing (DR), uses machine learning techniques and extends satellite measurements directly to maps, i.e., a classification algorithm is trained to develop an optimized set of rules through iterative repeated data analysis (Breiman 2001) for the estimation of biomass and carbon (Baccini et al. 2012). This approach results in continuous values for biomass based on easily understandable rules, such as those described for the Amazon basin (Saatchi et al. 2007), Russian forests (Houghton et al. 2007), or the African continent (Williams et al. 2007).

SUCHENWIRTH et al. (2012) used remote sensing data and a digital elevation model to map carbon densities in a floodplain. They used an OBIA approach to classify vegetation types. The total carbon storage of soils and vegetation was quantified using a Monte-Carlo simulation for all classified vegetation types, and spatial distribution was mapped.

We want to improve this method by including additional data and using a machine learning technique. Due to the complexity of the spatial distribution of C_{org} in the Danube floodplains (CIERJACKS et al. 2010, 2011, SUCHENWIRTH et al. 2012), and the amount, variety, and variable consistency of available data, our goal is to establish a machine learning approach for an area-wide modeling of C_{org} stocks. To include remote sensing data and several additional geodata, we chose a classification and regression tree (CART) approach (BREIMAN et al. 1984, LOH 2011).

The specific aims of this paper are as follows:

(1) to evaluate a machine learning algorithm (CART) for estimating and mapping C_{org} stocks in vegetation (C_{org_veg}), soil (C_{org_soil}) and total biomass (vegetation, soil and deadwood; C_{org_tot}) in riparian forests based on classification accuracies, and (2) to rank the parameters in terms of their ability to predict C_{org}.

2   Materials and Methods

2.1 *Research Area*

The research area has a size of 11.3 km^2 and is situated within the Danube Floodplain National Park (*Nationalpark Donau-Auen*) in Austria (16.66° E, 48.14° N). The national park is located between the Austrian capital Vienna and the Slovak capital Bratislava and stretches along the river Danube for about 36 km (Fig. 1). The river has an average width of about 350 m, and the banks are generally fixed by riprap. Only a few human impacts on the area happened apart from the construction of the *Hubertusdamm* dike in the 19th century to protect areas on the northern riverbank from inundation. In the 1960s, natural forest structures were altered by widespread planting of hybrid poplars (*Populus* x *canadensis*), especially on the southern riverbank. In 1996, the area was declared a national park, and thus commercial enterprises were banned within its precincts. Despite of the mentioned human interventions, the area remains one of the last large pristine riparian habitats in Central Europe and has been recognized by the International Union for Conservation of Nature (IUCN) as a Riverine Wetlands National Park, Category II.

The national park's environmental features include the secondary streams (the Danube river itself is an international waterway), side channels and oxbow lakes, gravel banks, riparian forests and meadows, reed beds and xeric habitats. Within the forests, we can dif-
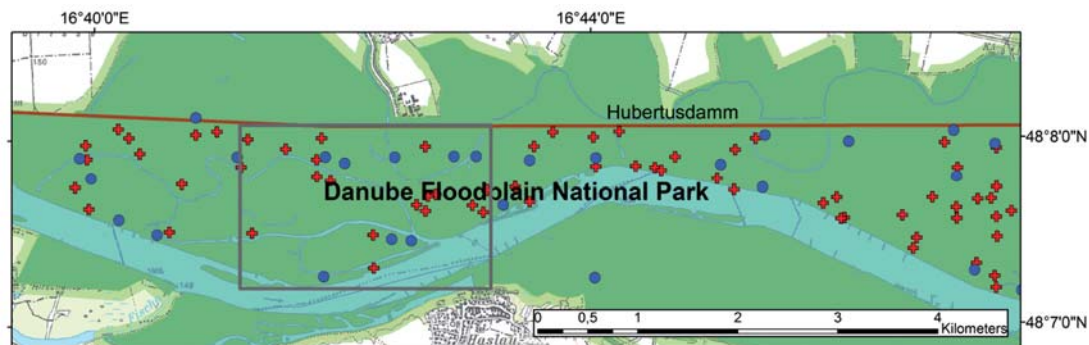
Fig. 1: Research Area, green: Danube Floodplain National Park, red cross: locations of the terrestrial sample points training data, blue dot: test data. The red line represents the Hubertusdamm dike. The grey box represents the outline of the subsets in Fig. 2.

ferentiate between hardwood forest (dominated by *quercus robur*, *fraxinus excelsior* and *acer campestre*), softwood forest (dominated by *salix alba* and *acer negundo*) and cottonwood forest (consisting of hybrid poplar plantations of the 1960ies) (Cierjacks et al. 2010). The main soil type is haplic fluvisol (calcaric). Calcaric gleysols are less important. The climate is continental with a mean annual temperature of 9.8 °C and a mean annual precipitation of 533 mm [Schwechat climate station, 48°07' N, 16°34' E, 184 m above sea level (ZAMG 2002)].

The mean carbon storage in the area was estimated as 359.1 Mg C ha^{-1} (472,186 Mg in an area of 13.1 km^2) by Cierjacks et al. (2010).

2.2 *Data*

The following available comprehensive data from the research area were included in the analysis: two very high spatial resolution (VHSR) satellite images from Ikonos and RapidEye sensor, historical and current topographic maps, a digital elevation model (DEM), and data on the mean groundwater level (MGW).

We purchased a preprocessed cloudfree Ikonos 2 image, recorded on April 22, 2009 with a spatial resolution of 1.0 m (panchromatic) and 4 m (multispectral), as well as a satellite image from RapidEye recorded on August 1, 2009 and processed at L3A with a spatial resolution of 5.0 m (multispectral), provided by the German Aerospace Centre. Both images were provided in the UTM WGS 1984 projected coordinate system and were reprojected into the Austrian MGI M34 projected coordinate system. We used this local system as the majority of local data was also projected in this way.

In addition to the spectral values, several ratios and texture parameters (Haralick et al. 1973) were calculated (Tab. 1). A digital elevation model derived from lidar data was used to compute height and slope. Increased slope values can suggest former riverbeds of the main stream or overgrown side channels, which can serve as an indicator of softwood (Suchenwirth et al. 2012), which cannot be detected directly through spectral values. Also

the height above ground has been included in the knowledge-base. The following vegetation types were determined by OBIA from the Ikonos image and the DEM: meadow, reed bed, cottonwood, softwood and hardwood forests (Suchenwirth et al. 2012).

Historical and current topographic maps were provided by the Austrian Federal Office for Metrology and Survey (Österreichisches Bundesamt für Eich- und Vermessungswesen, BEV). The historical maps are derived from three topographic land surveys, the First (1764–1806), the Second (1806–1869) and the Third Military Mapping Survey (1868–1880). We digitized the riverbeds and channels as well as oxbows and coded them, either if there was a historic water body or not. A groundwater model indicating median ground water depth was provided by the Vienna University of Technology.

During two terrestrial surveys in 2008 and 2010, a total of 104 samples from vegetation and soil were taken [69 samples in 2008 (Cierjacks et al. 2010) and 35 samples in 2010 (Rieger et al. 2013), Fig. 1]. All data were collected in a stratified randomized sampling design throughout the research area in 10 x 10 m plots. In each sample plot, forest stand structure was measured and soil samples were taken. A detailed description of the C_{org} calculation is given by Cierjacks et al. (2010) and Rieger et al. (2013). These data were randomly separated in training data (70 %) and test data (30 %) for the classification.

2.3 *Methods*

We developed a spatial model for the estimation and mapping of C_{org} stocks in soils and vegetation based on a machine learning algorithm. For this, we chose a classification and regression tree (CART) approach. CART creates classification rules in the shape of a decision tree. Decision trees show hierarchical rules according to which a dataset is classified. At the beginning of a decision tree is the basic population of the data. During the classification process, the dataset is divided according to binary rules (Breiman et al. 1984, Loh 2011, Quinlan 1986). The advantages of CART include the flexibility to handle a broad

Tab. 1: Available geodata, derived parameters and used abbrevations.

Available geodata	Derived parameters	Abbreviation
Ikonos image (April 22, 2009)	Blue channel	Ikonblu
	Green channel	Ikongrn
	Red channel	Ikonred
	Near infrared channel	Ikonnir
	NDVI (normalized difference vegetation index) (TUCKER 1979, ROUSE et al. 1973)	Ikonndvi
	Vegetation classification derived by OBIA (SUCHENWIRTH et al. 2012)	Classification
RapidEye image (August 1, 2009)	Blue channel	b1-REblue
	Green channel	b2-REgreen
	Red channel	b3-REred
	RedEdge channel	b4-REredEdge
	Near infrared channel	b5-REnir
	NDVI	RE_NDVI
	Transformed NDVI $[((b5+b3)+0.5)^{0.5}]$ (DEERING et al. 1975)	tNDVI
	modNDVI $[(b5-b4)/(b5+b4-2*b1)]$ (DATT 1999)	modNDVI
	b4NDVI $[(b5-b4)/(b5+b4)]$ (GITELSON & MERZLYAK 1994)	b4NDVI
	Solar Reflectance Index $[b5/b3]$ (ROUSE et al. 1973)	b4sri
	$[b2-b1]$	b2mb1
	$[b3-b1]$	b3mb1
	$[b3-b2]$	b3mb2
	$[b5-b4]$	b5mb4
	$[b3/b1]$	b3db1
	$[b4/b2]$	b4db2
	$[b5/b2]$	b5db2
	Texture parameters (HARALICK et al. 1973) Gray-level co-occurrence matrix (GLCM) homogeneity	GLCM_Homogeneity
	GLCM mean	GLCM_Mean
	GLCM correlation	GLCM_Correlation
	GLCM contrast	GLCM_Contrast
	Gray-level difference vector (GLDV) entropy	GLDV_Entropy
Digital elevation model	Elevation	DEM
	Slope	slope
Historical and current topographic maps	Existence of historic riverbed during: First Military Mapping Survey (1773 – 1781)	hist1
	Second Military Mapping Survey (1806 – 1869)	hist2
	Third Military Mapping Survey (1868 – 1880)	hist3
	Current distance to river based on current topographic map ÖK50	dist
Ground water model	Ground water level	MGW
C_{org} ground survey data from 2008 and 2010	Above ground carbon stocks	C_{org_veg}
	Below ground carbon stocks	C_{org_soil}
	Total carbon stocks	C_{org_tot}

range of response types, such as numeric and categorical data, the ease and robustness of construction, and the ease of interpretation (DE'ATH & FABRICIUS 2000).

For our work, we used the software package eCognition 8.7.1. It allowed us to combine CART and OBIA and thus make use of the vast amount of data including remote sensing and other spatially continuous geodata. OBIA has been successfully applied to classifications of diverse habitats from wetlands (KOLLÁR et al. 2011, ROKITNICKI-WOJCIK et al. 2011) and floodplains (WAGNER 2009) to forests (CHUBEY et al. 2006) and drylands (LALIBERTE et al. 2007). The CART approach in eCognition is based on the original algorithms described by BREIMAN et al. (1984) and has been implemented by the OPENCV-WIKI (2010) and eCognition (ECOGNITION 2012).

The ground survey dataset containing total carbon stocks was grouped into classes (Tab. 2) as were the separate stocks for vegetation and soil. We compared classifications of above ground biomass (C_{org_veg}), below ground biomass for soil depth up to 1 m (C_{org_soil}) and total carbon stocks (C_{org_tot}) using classifications based on two, three, four and five quantile classes. We used quantiles in order to have equal numbers of samples for each class. We applied this approach for different numbers of classes to define an optimum number of classes with acceptable classification accuracy.

The OBIA was performed on a multiresolution segmentation with a scale parameter of 200 and the homogeneity criterion including a shape of 0.1 and a compactness of 0.5. Each

spectral band of the RapidEye and Ikonos satellite imagery, as well as each additional geodata layer was weighted equally. However, calculated indices or ratios were not further weighted. Equal segmentation settings were used for all classifications in order to facilitate the comparability of area units among the classifications.

The internal CART algorithm was trained with the respective quantile classes and applied onto the parameters using the "classifier" tool in the software package eCognition 8.7.1, with a classifier depth of 10, a minimum sample count of 6 and 9 cross validation folds.

To evaluate the accuracy of the individual classifications, we calculated the overall accuracy (OA). We additionally decided to follow the suggestions of PONTIUS & MILLONES (2011) who recommend the use of allocation and quantity disagreement for accuracy assessment rather than the use of kappa. The two measures are described as follows:

a) Allocation disagreement (AD) is the number of pixels that have a less than optimal spatial allocation in the comparison map with respect to the reference map. Allocation disagreement is the distance above the quantity disagreement line.

b) Quantity disagreement (QD) is the absolute difference between the number of pixels of a certain class in the reference map and the number of pixels of the same class in the comparison map.

The lower the values of allocation and quantity disagreement, the better is the accu-

Tab. 2: C_{org} ranges (Mg C_{org} ha^{-1}) for C_{org_veg}, C_{org_soil}, and C_{org_tot} stocks for different numbers of classes.

class	Five quantile classes			Four quantile classes			Three quantile classes			Two quantile classes		
	C_{org_veg}	C_{org_soil}	C_{org_tot}	C_{org_veg}	C_{org_soil}	C_{org_tot}	C_{org_veg}	C_{org_soil}	C_{org_tot}	C_{org_veg}	C_{org_soil}	C_{org_tot}
1	< 55.0	<132.8	<231.0	< 75.0	<140.0	<255.5	< 86.5	<161.0	<281.0	<134.9	<186.4	<325.9
2	55.0 - 99.9	132.8 - 173.9	231.0 - 300.0	75.0 -135.0	140.0 - 186.5	255.5 - 326.9	86.5 - 180.0	161.0 - 203.2	281.0 - 373.0	>135.0	>186.5	>326.0
3	100.0 -134.0	174.0 - 197.3	300.1 - 360.9	135.1 - 200.0	186.5 - 227.0	327.0 - 407.0	>180.0	>203.2	>373.0			
4	134.1 - 193.0	197.4 - 240.0	361.0 - 445.0	>200.0	>227.0	>407.0						
5	>193.0	>240.0	>445.0									

racy. Both disagreement values are calculated as percentages.

Furthermore, we calculated for each classification the root-mean-square error (RMSE), frequently used to check the internal model quality with the advantage of being independent of the number of used classes (Kanevski et al. 2009, Richter et al. 2012). For our application, we used the arithmetic mean of each class (of the training plots) as the estimated value, and used the terrestrial value of each test plot as the measured value.

To calculate the relevance of the individual datasets, we summarized the use frequency of the individual parameters, normalized by the overall sum of all use frequencies. Additionally, we considered how many parameters derived from a specific dataset were applied,

normalized by the total number of the available parameters of a certain dataset. Erasmi et al. (2013) described the concept as "normalized importance".

3 Results

3.1 *Modelled C_{org} Distribution and Accuracies*

Fig. 2 shows the classification results in the form of maps for a part of the research area. The subset comprises all classes and all environmental features inside the research area. We can see that C_{org_veg} stocks are equally scattered across the area, while C_{org_soil} stocks increase as the distance to the river increases.

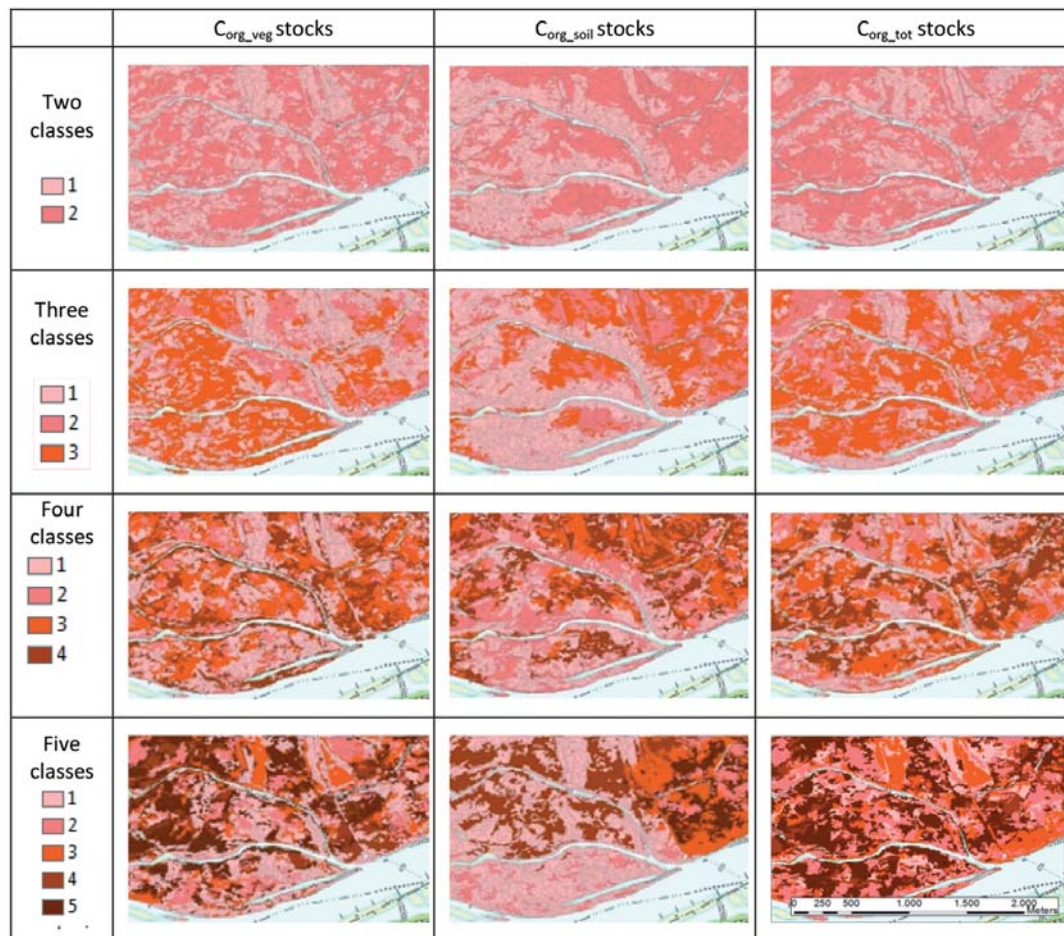

Fig. 2: Modelled distribution of C_{org_veg}, C_{org_soil}, and C_{org_tot} stocks for different numbers of classes. The increasing amount of stored C_{org} is represented by colour graduations increases from pink to red to brown.

The influence is less visible for C_{org_tot} but can still be seen for a classification with four classes.

We compared the derived accuracies (OA, AD, QD) for C_{org_veg}, C_{org_soil}, and C_{org_tot} stocks for all numbers of classes (Fig. 3), as well as RMSE. The comparison of classification ac-

curacies for C_{org_veg}, C_{org_soil}, and C_{org_tot} stocks revealed that the accuracy is highest for two classes and lowest for five classes (Fig. 3). Models with three or four classes range in between and represent a good compromise between complexity and acceptable accuracy.
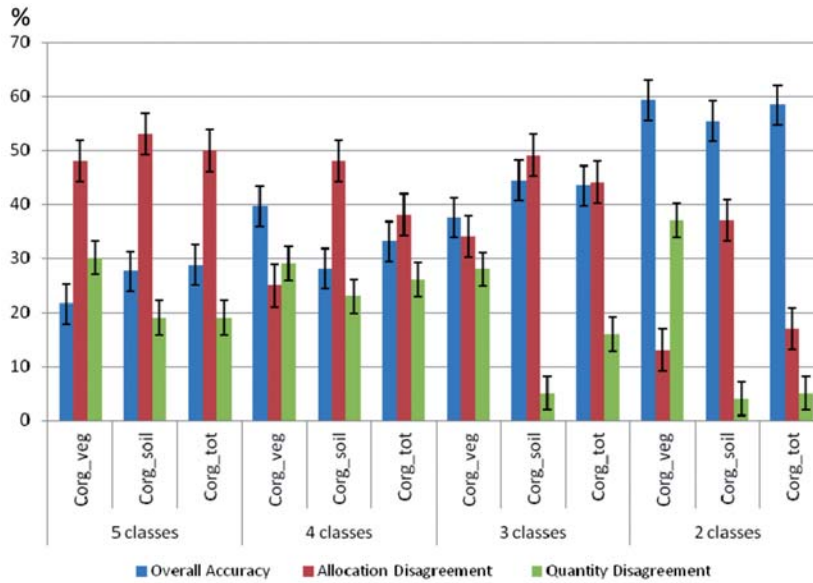

Fig. 3: Overall accuracy, allocation, and quantity disagreement in percent for classifications of C_{org_veg}, C_{org_soil}, C_{org_tot} based on five, four, three, and two classes.
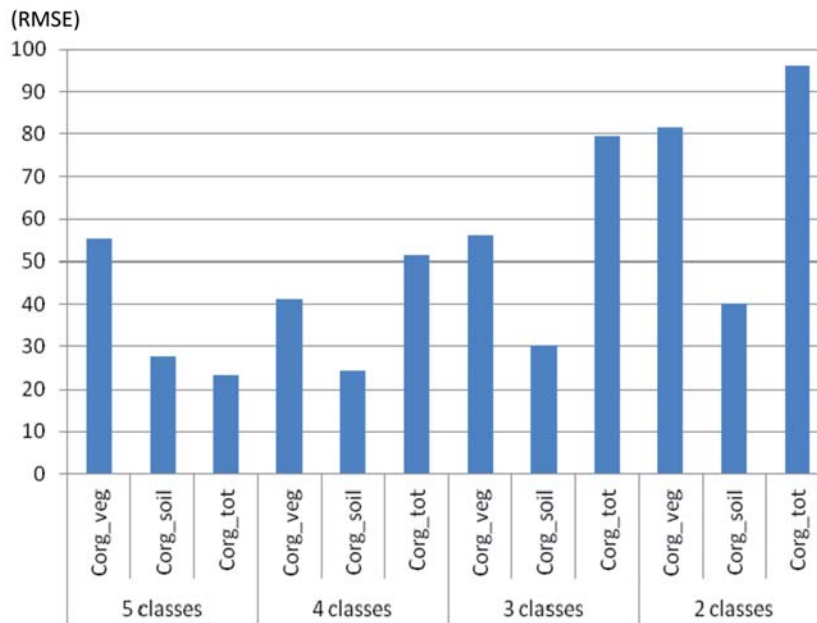

Fig. 4: Root-mean-square error for classifications of C_{org_veg}, C_{org_soil}, C_{org_tot} based on five, four, three, and two classes.

With regard to the model quality, we can examine Fig. 4. Classifications with fewer classes show higher RMSE values, e.g. more than 90 for C_{org_tot} two quantile classes, than classifications with more classes. The lowest RMSE values are below 25 for C_{org_soil} with four classes and C_{org_tot} with four classes.

3.2 Parameter Relevance

In the following we analyze the use frequency of the individual datasets and parameters. Tab. 3 shows the results for classifications with all quantile classes for C_{org_veg}, C_{org_soil} and C_{org_tot}.

For RapidEye parameters, the relevance ranged from 3.6 % (C_{org_soil} two classes) to 25.6 % (C_{org_tot} five classes). As the number of classes grows, the parameter relevance rises. For texture parameters, the relevance ranged from 4.6 % (C_{org_soil} 5 classes) to 29.5 % (C_{org_veg} four classes) with no clear indication of which number of classes provided the best results. The overall parameter relevance for Ikonos was lower. It ranged from 0 % (C_{org_soil} two or three classes) to 9.6 % (C_{org_veg} two classes) which could be explained by the acquisition date of April, when full leaf-out had not occurred yet.

For DEM parameters relevance ranged from 0 % (C_{org_veg} three classes; C_{org_soil} five classes) to the highest overall share of 52.1 % (C_{org_soil} two classes). The MGW reached the highest parameter relevance for all classification runs (32.7 % / 18.3 % / 26.2 %), with the relevance ranging from 0 % (C_{org_soil} two and four classes; C_{org_tot} five classes) to 43.2 % (C_{org_tot} two classes). For the "distance to river" parameter, the relevance ranged from 0 % (C_{org_soil} two and four classes) to 50.4 % (C_{org_soil} five classes), with this parameter achieving greater relevance when greater numbers of classes are used. For the parameters based on the existence of historical riverbeds, the relevance ranged from 0 % (C_{org_veg} two, three and four classes; C_{org_soil} five classes; C_{org_tot} two, four and five classes) to 36.0 % (C_{org_soil} two classes), and was important only when classifying C_{org_soil} classes.

To illustrate the importance of single parameters, Figs 5a–c give an exemplary insight of the parameter relevance of classifications with four classes for C_{org_veg}, C_{org_soil}, and C_{org_tot}. For C_{org_veg}, there are 16 parameters (RapidEye: 6; texture: 4; Ikonos: 2; DEM: 2; MGW and distance: 1 each), where the index *b4db2* (i.e. RapidEye's RedEdge divided by green channel) is the most important with more than 23 %. For C_{org_soil}, there are eleven

Tab. 3: Dataset relevance for classifications of C_{org_veg}, C_{org_soil}, and C_{org_tot} stocks.

		RapidEye	Texture	Ikonos	DEM	MGW	Distance to river	Historic maps
C_{org_veg}	5cl	14.5	22.5	5.5	6.3	16.5	31.7	3.0
	4cl	12.0	12.9	5.0	25.3	37.0	7.8	0.0
	3cl	21.8	20.6	3.0	0.0	34.3	20.2	0.0
	2cl	5.9	23.8	9.6	7.3	42.8	10.7	0.0
	Average	13.5	20.0	5.8	9.8	32.7	17.6	0.7
C_{org_soil}	5cl	4.1	4.6	1.7	0.0	39.2	50.4	0.0
	4cl	13.1	29.5	8.4	13.0	0.0	0.0	36.0
	3cl	5.2	18.4	0.0	6.6	33.9	16.6	19.3
	2cl	3.6	8.4	0.0	52.1	0.0	0.0	35.8
	Average	6.5	15.2	2.5	17.9	18.3	16.7	22.8
C_{org_tot}	5cl	25.6	9.8	5.0	11.6	0.0	48.0	0.0
	4cl	4.3	20.8	5.1	13.5	34.5	21.8	0.0
	3cl	9.8	7.6	8.2	8.4	27.0	35.9	3.0
	2cl	9.4	19.7	5.5	22.2	43.2	0.0	0.0
	Average	12.3	14.5	6.0	13.9	26.2	26.4	0.7

Fig. 5a: Parameter relevance for C_{org_vet} classifications based on 4 quantile classes (all abbreviations are explained in Tab. 1).
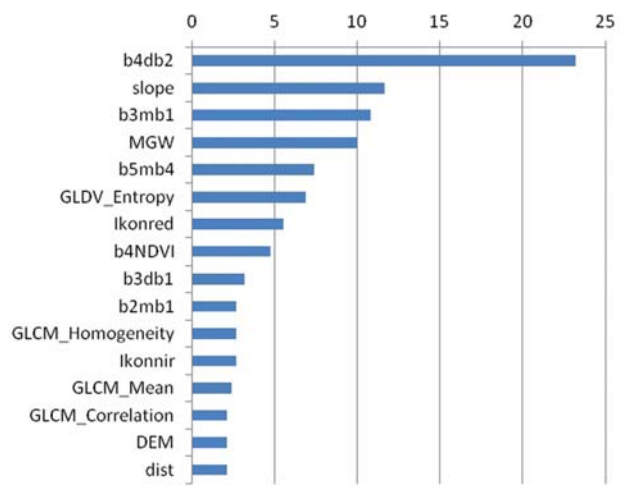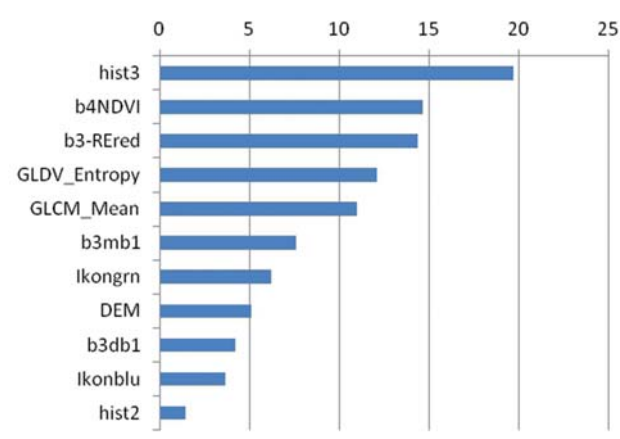

Fig. 5b: Parameter relevance for C_{org_soil} classifications based on 4 quantile classes (all abbreviations are explained in Tab. 1).
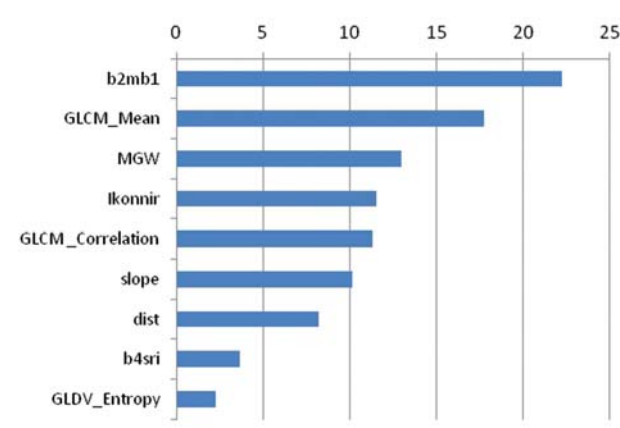

Fig. 5c: Parameter relevance for C_{org_tot} classifications based on 4 quantile classes (all abbreviations are explained in Tab. 1).

parameters (RapidEye: 4; texture: 2; Ikonos: 2; historical maps: 2; DEM: 1), of which *hist3* (existence of riverbed between 1868 to 1880) is the most relevant with almost 20 %. For C_{org_tot}, there are in total nine parameters (Rapid-Eye: 2; texture: 3; Ikonos: 1; MGW, DEM and distance: 1 each), of which *b2mb1* (RapidEye's green channel minus blue channel) is the most important one with more than 22 %.

4   Discussion

4.1   *Classification Results and Accuracies*

Our study provides a novel technique for the estimation and mapping of C_{org} stocks in floodplains based on remote sensing and additional geodata. It could be used to generate C_{org} inventories in other temperate wetlands, especially forested floodplains where ground assessment is difficult or impossible. The visualization of the individual classes shows complex distribution patterns of C_{org} stocks. Despite of the cluttered structure and the heterogeneous distribution within the different classes, the majority of classifications show that higher C_{org_soil} stocks have developed at a certain distance to the main riverbed of the Danube and its side arms. This is best illustrated by classifications with two but also four classes of C_{org_soil}. These lateral gradients were also described by CIERJACKS et al. (2010, 2011). In comparison, the patterns of C_{org_veg} and C_{org_tot} were less predictable. Classifications are very speckled for every model and a fully consistent classification is difficult due to the type of the terrain. This reflects the complexity of floodplain habitats in general, and the detailed intricacy of riparian C_{org} stocks in particular and also has been shown by SAMARITANI et al. (2011) and SUCHENWIRTH et al. (2012). For the particular case of the Danube floodplain, this may also be related to the widespread planting of hybrid poplars in the 1960s, which altered the natural vegetation structure of hardwood and softwood forests.

Surprisingly, the accuracy of the C_{org_soil} stock models was similar to the accuracy of the C_{org_veg} stock models. Predictive variables derived from remote sensing and other geodata serve as proxies for recent environmental conditions that control vegetation properties. Soil organic matter, in contrast, can accumulate over hundreds of years. Thus relations of C_{org_soil} stocks to recent environmental conditions might not be expected. It is likely that the variations in C_{org_soil} stocks found in our study are mainly due to variations in the C_{org} stocks of the upper soil horizons, which in turn are affected by recent environmental conditions. Furthermore, the position of historic riverbeds, a parameter with strong and long-lasting influence on soil organic matter content, was considered (Figs. 3 and 5b).

Predictably, an increase in the number of classes goes along with a more speckled appearance of the classification and overall accuracy decreases. Here, we have to keep in mind that a classification with fewer classes will automatically result in higher accuracy, and therefore the differences simply reflect the higher chance of misclassifications.

Similarly to the overall accuracy, allocation disagreement as well as quantity disagreement values decreased, i.e., the accuracy improved, with fewer classes. An exception is the very high quantity disagreement value for C_{org_veg} based on two classes.

The RMSEs (Fig. 4) provides a measure independent of the number of used classes. The RMSEs "mirror" the results of accuracy assessment, with lower RMSEs for classifications with higher class numbers. Especially for C_{org_soil} accuracies.

For assessing the performance of the CART approach we also compared our results with a linear multiple regression analysis for estimating C_{org_soil}, C_{org_veg}, and C_{org_tot}. Results showed that for C_{org_soil} regression (model intercept $p = 0.0069$; $F = 3.3789$) groundwater level was the most important parameter ($p = 0.0177$; $y = -11.275x + 1833.4$; $R^2 = 0.8657$).

For C_{org_tot} regression (model intercept $p = 2.3833^{-9}$; $F = 6.5114$), the green RapidEye channel ($p = 0.0145$; $y = -0.0756x + 584.28$; $R^2 = 0.5619$) and the red Ikonos channel ($p = 0.0188$; $y = -0.4198x + 426.33$; $R^2 = 0.5244$) were the most important parameters.

For C_{org_veg} regression (model intercept $p = 1.7728^{-6}$; $F = 7.7927$), the green RapidEye channel ($p = 0.0099$; $y = -0.0482x + 335.83$; $R^2 = 0.5301$) and red Ikonos channel ($p = 0.0081$;

$y = -0.3752x + 208.54$; $R^2 = 0.5562$) have the highest importance among the parameters.

The regression confirms our findings that remote sensing parameters are more important for the classification of C_{org_veg}, whereas parameters from auxiliary geodata have more influence on the classification of C_{org_soil} stocks.

It is worth discussing whether and which other additional parameters should be taken into consideration for the detection and modelling of C_{org} distributions in floodplains. Data on forest management practices or local sinks may be considered but were not available on a spatially inclusive and comprehensive level.

In general, Rocchini et al. (2013) argue that the classification of remotely sensed images for the derivation of ecosystem-related maps which also includes the estimation of C_{org} is commonly based on clustering of spatial entities within a spectral space with the implication that it is possible to divide the gradual variability of the Earth's surface into a finite number of discrete non-overlapping classes, which are exhaustively defined and mutually exclusive. Given the continuous nature of many ecosystem properties this approach is often inappropriate; especially as standard data processing and image classification methods involve the loss of information as continuous quantitative spectral information is being degraded into a set of discrete classes. For wetlands, Ozesmi & Bauer (2002) pointed out the limitations of remote sensing for classification and suggest the use of multi-temporal data for an improvement of classification accuracy. For remote sensing in wetlands, Adam et al. (2010) attribute the frequently observed limitations to the low spatial and spectral resolution in comparison to narrow vegetation units that characterize wetland ecosystems.

There may also be concerns about the reliability of terrestrial data. Error propagation may always be a source of uncertainty for the mapping of ecosystems (Rocchini et al. 2013). Our basic survey data have been collected very densely and thoroughly, but transferability to other terrains may become challenging.

Overall, we can conclude that the detection of floodplain characteristics is a challenging task. As for the appropriate number of classes, we consider three or four to be optimal. The accuracy is higher in comparison to a model with five classes, but the complexity is better represented than in a plain dichotomy of data and space created by merely two classes. Dillabaugh & King (2008) found an optimal number of three classes for their classifications of biomass in riparian marshes in Ontario.

Regarding our first research aim, a model approach with four classes seems to perform best. However, the concept of applying segregative classes remains to a certain extent debatable. Therefore, an approach with classes based on fuzzy logic (Zadeh 1989) should be considered in future works to improve the predictive capability of the C_{org} model.

A general point of criticism might apply to the question of why to classify a continuous variable with separate classes. Even though a continuous regression may seem more appropriate, we wanted to create statistically set classes and to follow the concept of different C_{org} concentrations in different compartments of the floodplains. For further planning applications, the regional managers would always apply an ordinal scale, e.g. high, medium, low. The provision of an estimate about the optimal class size for C_{org} might be valuable in terms of its practical application.

A further point of debate remains the sampling design. The random division of terrestrial survey data into 70 % training data and 30 % test data and repeated analysis would probably provide a better estimate about the uncertainties within the calibration and validation data. Repeated measurements could give an insight into the quality of the cal/val information and, in consequence, provide knowledge about the optimal sampling size and spatial distribution of these data. In further analysing steps a repeated calculation with varying samples is envisaged.

4.2 *Use of Parameters*

Regarding the application of parameters and their use frequency, classification of C_{org_veg} relied to a higher percentage on remotely sensed parameters like RapidEye, Texture, and Ikonos than did the classification of C_{org_soil} or C_{org_tot} stocks.

The fact that remotely sensed parameters, especially RapidEye parameters, are the most important factors for the classification of C_{org_veg} provides further evidence of the relevance of satellite imagery for the estimation of biomass, including C_{org} (GIBBS et al. 2007, NEEFF et al. 2005, RHEINHARDT et al. 2012). SCHUSTER et al. (2012) in particular proved the special relevance of the RedEdge channel for vegetation classification. It is nevertheless remarkable that MGW and the distance to the river played a more dominant role in the classification of C_{org_veg} and C_{org_tot} stocks than C_{org_soil} stocks, although one could assume that median groundwater would be a comparatively less decisive factor for vegetation than for soil biomass and resulting C_{org}. Still, similar findings for fine-root and above-ground biomass which also clearly reflected groundwater depths in the same study area support our results (RIEGER et al. 2013). For the case of distance to river, the differences within the parameter relevance (Fig. 5b) for C_{org_soil} is a specific characteristic and shows the variability of classification models. While remotely sensed parameters play the dominant role in all classifications, it is striking that the most important parameter for the C_{org_soil} classification are the historical riverbeds (Figs. 5a–c).

The case is different for the classification of C_{org_soil} stocks, where remote sensing based rules had in some cases less than 50 % influence towards the classification. In contrast, the application frequency of DEM and historical riverbeds – parameters not derived from remote sensing – was more common for the classifications of C_{org_soil} compared to C_{org_veg}. These parameters have already been used successfully in other studies (CIERJACKS et al. 2011, SAMARITANI et al. 2011) to determine C_{org} stocks. Concerning the use of historical maps, it should be kept in mind that our maps only provide information on roughly the last 250 years, whereas C_{org} stocks in soil are the consequence of geomorphologic and pedogenetic processes that have taken place over centuries and millennia.

In general, the assessment of the relevance of individual parameters for the C_{org} model showed that spectral information from remote sensing provides direct information about above ground biomass, while information on soil characteristics can only be explained indirectly through vegetation. This is due to the fact that C_{org_soil} reflects not only recent vegetation, but accumulations over centuries. This is reflected in the high relevance of historical maps for this factor (Fig. 5b) which emphasizes the potential of soils to serve as a memory of previous site conditions, such as historical inundations and changes in riverbeds that often occurred prior to present-day land management practices.

5   Conclusion and Outlook

Our study provides a machine learning approach to model C_{org} stock distributions in riparian forests. We aimed to evaluate a machine learning algorithm (CART) and determine the relevance of individual variables derived from the geodata for the estimation.

Overall, a spatial model of C_{org} in riparian forests could be generated using CART. With the use of geographic datasets, it was possible to show the spatial distribution in terms of a cartographic representation generated by classification. Yet, classification accuracy remains a challenge due to the high complexity of floodplains where patterns of C_{org} distribution are inherently difficult to define.

The evaluation of the relevance of the individual parameters derived from the geodata revealed that remote sensing parameters are more important for the classification of C_{org_veg}, than for the classification of C_{org_soil}. This is also the case for MGW and the distance to the river. In contrast, parameters derived from auxiliary geodata such as DEM and historical maps were more decisive for the classification of C_{org_soil} than C_{org_veg}. C_{org_tot} stocks fell in between in terms of application frequency of remote sensing and other parameters. Therefore, depending on the target (C_{org_soil} or C_{org_veg}), different parameters should be considered when analyzing the spatial distribution of carbon storage.

The application of data-mining approaches to remote sensing and other geodata is helping to automate and facilitate estimations of C_{org} in riparian forests. In addition, information on vegetation structure might improve the C_{org_soil} model. Each classification model

highlights the complex interrelations between C_{org} stocks and the external geofactors. In particular, vegetation cover and resulting C_{org_veg} seems to reflect recent site conditions while C_{org_soil} reflects both recent conditions and past processes. In this way, our model contributes to a better understanding of the importance and relationships of C_{org} cycling in floodplain ecosystems. Consequently, this work may serve as a local case study for a well and densely-surveyed area and contribute to improve methods of C_{org} estimation and monitoring in other floodplain areas with similar conditions in temperate climates. It might help to improve formal frameworks such as European biomass inventory (GALLAUN et al. 2010), REDD, and Kyoto protocols (BÖTTCHER et al. 2009, IPCC 2000, OBERSTEINER et al. 2009, PAOLI et al. 2010, UNEP-WCMC 2008).

Acknowledgements

References

ADAM, E., MUTANGA, O. & RUGEGE, D., 2010: Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. – Wetlands Ecology and Management **18** (3): 281–296.

AWAYA, Y., TSUYUKI, S., KODANI, E. & TAKAO, G., 2004: Potential of Woody Carbon Stock Estimation Using High Spatial Resolution Imagery: A Case Study of Spruce Stands. – SHIYOMI, M., KAWAHATA, H., KOIZUMI, H., TSUDA, A. & AWAYA, Y. (eds.): Global Environmental Change in the Ocean and on Land: 425–440, Terrapub, Tokyo, Japan.

BACCINI, A., GOETZ, S.J., WALKER, W.S., LAPORTE, N.T., SUN, M., SULLA-MENASHE, D., HACKLER, J., BECK, P.S.A., DUBAYAH, R., FRIEDL, M.A., SAMANTA, S. & HOUGHTON, R.A., 2012: Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. – Nature Climate Change **2** (3): 182–185.

BACKÉUS, S., WIKSTRÖM, P. & LÄMÅS, T., 2005: A model for regional analysis of carbon sequestration and timber production. – Forest Ecology and Management **216** (2005): 28–40.

BARITZ, R., SEUFERT, G., MONTANARELLA, L. & VAN RANST, E., 2010: Carbon concentrations and stocks in forest soils of Europe. – Forest Ecology and Management **260** (3): 262–277.

BEHRENS, T. & SCHOLTEN, T., 2006: Digital soil mapping in Germany – a review. – Journal of Plant Nutrition and Soil Science **169** (3): 434–443.

BÖTTCHER, H., EISBRENNER, K., FRITZ, S., KINDERMANN, G., KRAXNER, F., MCCALLUM, I. & OBERSTEINER, M., 2009: An assessment of monitoring requirements and costs of 'Reduced Emissions from Deforestation and Degradation'. – Carbon Balance and Management **4** (1): 7.

BREIMAN, L., FRIEDMAN, J.H., OLSHEN, R.A. & STONE, C.J., 1984: Classification and regression trees. – Wadsworth & Brooks / Cole Advanced Books & Software, Monterey, CA, USA.

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

BUSSE, L.B. & GUNKEL, G., 2002: Riparian alder fens – source or sink for nutrients and dissolved organic carbon? 2. Major sources and sinks. – Limnologica – Ecology and Management of Inland Waters **32** (1): 44–53.

CAO, T., VALSTA, L. & MÄKELÄ, A., 2010: A comparison of carbon assessment methods for optimizing timber production and carbon sequestration in Scots pine stands. – Forest Ecology and Management **260** (10): 1726–1734.

CHUBEY, M.S., FRANKLIN, S.E. & WULDER, M.A., 2006: Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. – Photogrammetric Engineering and Remote Sensing **72** (4): 383–394.

CIERJACKS, A., KLEINSCHMIT, B., BABINSKY, M., KLEINSCHROTH, F., MARKERT, A., MENZEL, M., ZIECHMANN, U., SCHILLER, T., GRAF, M. & LANG, F., 2010: Carbon stocks of soil and vegetation on Danubian floodplains. – Journal of Plant Nutrition and Soil Science **173** (5): 644–653.

CIERJACKS, A., KLEINSCHMIT, B., KOWARIK, I., GRAF, M. & LANG, F., 2011: Organic matter distribution in floodplains can be predicted using spatial and vegetation structure data. – River Research and Applications **27:** 1048–1057.

Datt, B., 1999: A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves. – Journal of Plant Physiology **154** (1): 30–36.

de'Ath, G. & Fabricius, K.E., 2000: Classification and regression trees: a powerful yet simple technique for ecological data analysis. – Ecology **81** (11): 3178–3192.

Deering, D.W., Rouse, J.W., Haas, R.H. & Schell, J.A., 1975: Measuring "forage production" of grazing units from Landsat MSS data. – 10th International Symposium Remote Sensing of Environment **II:** 1169–1178.

Dillabaugh, K.A. & King, D.J., 2008: Riparian marshland composition and biomass mapping using Ikonos imagery. – Canadian Journal of Remote Sensing **34** (2): 143–158.

eCognition, 2012: eCognition Developer Reference Book 8.8. – Trimble Germany GmbH, Munich.

Erasmi, S., Riembauer, G. & Westphal, C., 2013: Mapping habitat diversity from multi-temporal RapidEye and RADARSAT-2 data in Brandenburg, Germany. – Borg, E., Daedelow, H. & Johnson, R. (eds): 5th RESA Workshop, Neustrelitz, March **2013:** 75–89, GITO Berlin.

Farid, A., Goodrich, D.C., Bryant, R. & Sorooshian, S., 2008: Using airborne lidar to predict Leaf Area Index in cottonwood trees and refine riparian water-use estimates. – Journal of Arid Environments **72** (1): 1–15.

Gallaun, H., Zanchi, G., Nabuurs, G.-J., Hengeveld, G., Schardt, M. & Verkerk, P.J., 2010: EU-wide maps of growing stock and aboveground biomass in forests based on remote sensing and field measurements. – Forest Ecology and Management **260** (3): 252–261.

Gibbs, H.K., Brown, S., Niles, J.O. & Foley, J.A., 2007: Monitoring and estimating tropical forest carbon stocks: making REDD a reality. – Environmental Research Letters **2** (4), doi 10.1088/1748-9326/2/4/045023.

Giese, L.A., Aust, W.M., Trettin, C.C. & Kolka, R.K., 2000: Spatial and temporal patterns of carbon storage and species richness in three South Carolina coastal plain riparian forests. – Ecological Engineering **15** (Supplement 1): 157–170.

Gitelson, A. & Merzlyak, M.N., 1994: Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves – Spectral Features and Relation to Chlorophyll Estimation. – Journal of Plant Physiology **143** (3): 286–292.

Goetz, S., Baccini, A., Laporte, N., Johns, T., Walker, W., Kellndorfer, J., Houghton, R. & Sun, M., 2009: Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. – Carbon Balance and Management **4** (1): 2.

Grimm, R., Behrens, T., Märker, M. & Elsenbeer, H., 2008: Soil organic carbon concentrations and stocks on Barro Colorado Island – Digital soil mapping using Random Forests analysis. – Geoderma **146** (1–2): 102–113.

Groombridge, B. & Jenkins, M.D., 2002: World atlas of biodiversity: Earth's living Resources in the 21st century. – Prepared by UNEP World Monitoring Centre University of California Press, Berkeley, CA, USA.

Haralick, R.M., Shanmugam, K. & Dinstein, I.H., 1973: Textural Features for Image Classification. – IEEE Transactions on Systems, Man and Cybernetics **SMC-3** (6): 610–621.

Harrison, A.F., Howard, P.J.A., Howard, D.M., Howard, D.C. & Hornung, M., 1995: Carbon storage in forest soils. – Forestry **68** (4): 335–348.

Hazlett, P.W., Gordon, A.M., Sibley, P.K. & Buttle, J.M., 2005: Stand carbon stocks and soil carbon and nitrogen storage for riparian and upland forests of boreal lakes in north-eastern Ontario. – Forest Ecology and Management **219** (1): 56–68.

Hilker, T., Coops, N.C., Wulder, M.A., Black, T.A. & Guy, R.D., 2008: The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements. – Science of the Total Environment **404** (2–3): 411–423.

Hoffmann, T., Glatzel, S. & Dikau, R., 2009: A carbon storage perspective on alluvial sediment storage in the Rhine catchment. – Geomorphology **108** (1–2): 127–137.

Hofmann, G. & Anders, S., 1996: Waldökosysteme als Quellen und Senken für Kohlenstoff – Fallstudie ostdeutsche Länder. – Beiträge Forstwirtschaft und Landschaftsökologie **30** (1): 9–16.

Houghton, R.A., Butman, D., Bunn, A.G., Krankina, O.N., Schlesinger, P. & Stone, T.A., 2007: Mapping Russian forest biomass with data from satellites and forest inventories. – Environmental Research Letters **2** (4), doi 10.1088/1748-9326/2/4/045032.

IPCC, 2000: Special report on land use, land-use change and forestry. – Cambridge University Press, Cambridge, UK.

Kanevski, M., Timonin, V. & Pozdnukhov, A., 2009: Machine learning algorithms for spatial data analysis and modelling. – EFPL Press, Lausanne, Switzerland.

Kollár, S., Vekerdy, Z. & Márkus, B., 2011: Forest Habitat Change Dynamics in a Riparian Wetland. – Procedia Environmental Sciences **7** (0): 371–376.

Kooch, Y., Hosseini, S.M., Zaccone, C., Jalilvand, H. & Hojjati, S.M., 2012: Soil organic carbon sequestration as affected by afforestation: the Darab Kola forest (north of Iran) case study. – Journal of Environmental Monitoring **14** (9): 2438–2446.

Lal, R., 2005: Forest soils and carbon sequestration. – Forest Ecology and Management **220** (2005): 242–258.

Laliberte, A.S., Rango, A., Herrick, J.E., Fredrickson, E.L. & Burkett, L., 2007: An object-based image analysis approach for determining fractional cover of senescent and green vegetation with digital plot photography. – Journal of Arid Environments **69** (1): 1–14.

Loh, W.-Y., 2011: Classification and regression trees. – Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **1** (1): 14–23.

Matsui, N., Suekuni, J., Nogami, M., Havanond, S. & Salikul, P., 2009: Mangrove rehabilitation dynamics and soil organic carbon changes as a result of full hydraulic restoration and re-grading of a previously intensively managed shrimp pond. – Wetlands Ecology and Management **18** (2): 233–242.

Mayaux, P., Bartholomé, E., Fritz, S. & Belward, A., 2004: A new land-cover map of Africa for the year 2000. – Journal of Biogeography **31** (6): 861–877.

McBratney, A.B., Mendonça, Santos, M.L. & Minasny, B., 2003: On digital soil mapping. – Geoderma **117** (1-2): 3–52.

Mitra, S., Wassmann, R. & Vlek, P.L.G., 2005: An appraisal of global wetland area and its organic carbon stock. – Current Science **88**, Bangalore, India.

Mitsch, W., Nahlik, A., Wolski, P., Bernal, B., Zhang, L. & Ramberg, L., 2010: Tropical wetlands: seasonal hydrologic pulsing, carbon sequestration, and methane emissions. – Wetlands Ecology and Management **18** (5): 573–586.

Munyati, C., 2000: Wetland change detection on the Kafue Flats, Zambia, by classification of a multitemporal remote sensing image dataset. – International Journal of Remote Sensing **21**: 1787–1806.

Neeff, T., de Alencastro Graça, P.M., Dutra, L.V. & da Costa Freitas, C., 2005: Carbon budget estimation in Central Amazonia: Successional forest modeling from remote sensing data. – Remote Sensing of Environment **94** (4): 508–522.

Obersteiner, M., Huettner, M., Kraxner, F., McCallum, I., Aoki, K., Bottcher, H., Fritz, S., Gusti, M., Havlik, P., Kindermann, G., Rametsteiner, E. & Reyers, B., 2009: On fair, effective and efficient REDD mechanism design. – Carbon Balance and Management **4** (1): 11.

Olofsson, P., Lagergren, F., Lindroth, A., Lindström, J., Klemedtsson, L., Kutsch, W. & Eklundh, L., 2008: Towards operational remote sensing of forest carbon balance across Northern Europe. – Biogeosciences **5** (3): 817–832.

OpenCV-Wiki, 2010: Decision Trees. – http://opencv.willowgarage.com/documentation/cpp/ml_decision_trees.html (21.1.2013).

Ozesmi, S.L. & Bauer, M.E., 2002: Satellite remote sensing of wetlands. – Wetlands Ecology and Management **10** (5): 381–402.

Paoli, G., Wells, P., Meijaard, E., Struebig, M., Marshall, A., Obidzinski, K., Tan, A., Rafiastanto, A., Yaap, B., Ferry Slik, J., Morel, A., Perumal, B., Wielaard, N., Husson, S. & D'Arcy, L., 2010: Biodiversity Conservation in the REDD. – Carbon Balance and Management **5** (1): 7.

Pontius, R.G. & Millones, M., 2011: Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. – International Journal of Remote Sensing **32** (15): 4407–4429.

Quinlan, J.R., 1986: Induction of decision trees. – Machine Learning **1** (1): 81–106.

Rheinhardt, R., Brinson, M., Meyer, G. & Miller, K., 2012: Integrating forest biomass and distance from channel to develop an indicator of riparian condition. – Ecological Indicators **23** (0): 46–55.

Richter, K., Atzberger, C., Hank, T.B. & Mauser, W., 2012: Derivation of biophysical variables from Earth observation data: validation and statistical measures. – Journal of Applied Remote Sensing **6** (1): 063557–063551.

Rieger, I., Lang, F., Kleinschmit, B., Kowarik, I. & Cierjacks, A., 2013: Fine root and aboveground carbon stocks in riparian forests: the role of diking, environmental gradients and dominant tree species. – Plant and soil: 1–13, Springer, doi 10.1007/s11104-013-1638-8.

Rocchini, D., Foody, G.M., Nagendra, H., Ricotta, C., Anand, M., He, K.S., Amici, V., Kleinschmit, B., Förster, M., Schmidtlein, S., Feilhauer, H., Ghisla, A., Metz, M. & Neteler, M., 2013: Uncertainty in ecosystem mapping by remote sensing. – Computers & Geosciences **50**: 128–135, Elsevier.

Rokitnicki-Wojcik, D., Wei, A. & Chow-Fraser, P., 2011: Transferability of object-based rule sets for mapping coastal high marsh habitat among different regions in Georgian Bay, Canada. – Wetlands Ecology and Management: 1–14.

Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W., 1973: Monitoring vegetation systems in the Great Plains with ERTS. – Third ERTS Symposium **1973:** 309–317, Washington, DC, USA.

SAATCHI, S.S., HOUGHTON, R.A., DOS SANTOS ALVA-LÁ, R.C., SOARES, J.V. & YU, Y., 2007: Distribution of aboveground live biomass in the Amazon basin. – Global Change Biology **13** (4): 816–837.

SAMARITANI, E., SHRESTHA, J., FOURNIER, B., FROSSARD, E., GILLET, F., GUENAT, C., NIKLAUS, P.A., PASQUALE, N., TOCKNER, K., MITCHELL, E.A.D. & LUSTER, J., 2011: Heterogeneity of soil carbon pools and fluxes in a channelized and a restored floodplain section (Thur River, Switzerland). – Hydrology and Earth System Sciences **15** (6): 1757–1769.

SCHUSTER, C., FÖRSTER, M. & KLEINSCHMIT, B., 2012: Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. – International Journal of Remote Sensing **33** (17): 5583–5599.

SUCHENWIRTH, L., FÖRSTER, M., CIERJACKS, A., LANG, F. & KLEINSCHMIT, B., 2012: Knowledge-based classification of remote sensing data for the estimation of below- and above-ground organic carbon stocks in riparian forests. – Wetlands Ecology and Management **20** (2): 151–163.

TUCKER, C.J., 1979: Red and photographic infrared linear combinations for monitoring vegetation. – Remote Sensing of Environment **8** (2): 127–150.

UNEP-WCMC, 2008: Carbon and biodiversity: a demonstration atlas. – UNEP-WCMC, Cambridge, UK.

WAGNER, I., 2009: The Danube Floodplain Habitats – application of the Object-based Image Analysis approach. – CAR, A., GRIESEBNER, G. &

STROBL, J. (eds): Geospatial Crossroads @ GI_Forum '09: 218–227, Wichmann, Heidelberg.

WILLIAMS, C., HANAN, N., NEFF, J., SCHOLES, R., BERRY, J., DENNING, A.S. & BAKER, D., 2007: Africa and the global carbon cycle. – Carbon Balance and Management **2** (1): 3.

ZADEH, L.A., 1989: Knowledge Representation in Fuzzy Logic. – IEEE Transactions on Knowledge and Data Engineering **1** (1): 89–100.

ZAMG, 2002: Klimadaten von Österreich 1971–2000. – Zentralanstalt für Meterorologie und Geodynamik, Vienna, Austria.

Addresses of the Authors:

LEONHARD SUCHENWIRTH, MICHAEL FÖRSTER & BIRGIT KLEINSCHMIT, Technical University of Berlin, Geoinformation in Environmental Planning Lab, Straße des 17. Juni 145, 10623 Berlin, Germany, e-mail: {leonhard.suchenwirth}{michael.foerster}{birgit.kleinschmit}@tu-berlin.de

FRIEDERIKE LANG, University of Freiburg, Institute of Soil Science and Forest Nutrition, Bertoldstraße 17, 79098 Freiburg, Germany, e-mail: fritzi.lang@bodenkunde.uni-freiburg.de

Article

Analysis of Crop Reflectance for Estimating Biomass in Rice Canopies at Different Phenological Stages

Martin Leon Gnyp, Kang Yu, Helge Aasen, Cologne, Yinkun Yao, Shanyu Huang, Yuxin Miao, Beijing, China & Georg Bareth, Cologne

Summary: This paper contributes an assessment for estimating rice (*Oryza sativa L., irrigated lowland rice*) biomass by canopy reflectance in the Sanjiang Plain, China. Hyperspectral data were captured with field spectroradiometers in experimental field plots and farmers' fields and then accompanied by destructive aboveground biomass (AGB) sampling at different phenological growth stages. Best single bands, best two band-combinations, optimised simple ratio (SR), and optimised normalized ratio index (NRI), as well as multiple linear regression (MLR) were calculated from the reflectance for the non-destructive estimation of rice AGB. Experimental field data were used as the calibration dataset and farmers' field data as the validation dataset. Reflectance analyses display several sensitive bands correlated to rice AGB, e.g. 550, 670, 708, 936, 1125, and 1670 nm, which changed depending on the phenological growth stages. These bands were detected by correlograms for SR, NRI, and MLR with an offset of approximately ± 10 nm. The assessment of the three methods showed clear advantages of MLR over SR and NRI in estimating rice AGB at the tillering and stem elongation stages by fitting and evaluating the models. The optimal band number for MLR was set to four to avoid overfitting. The best validated MLR model ($R^2 = 0.82$) at the tillering stage was using four bands at 672, 696, 814 and 707 nm. Overall, the optimized SR, NRI, and MLR have a great potential in non-destructive estimation of rice AGB at different phenological stages. The performance against the validation dataset showed R^2 of 0.69 for SR and R^2 of 0.70 for NRI, respectively.

Zusammenfassung: *Reflexionsanalyse zur Abschätzung der Biomasse von Reis in unterschiedlichen phänologischen Stadien.* Dieser Beitrag versucht eine Bewertung zur Biomassenabschätzung von Kulturreis (Oryza sativa L., bewässerter Flachland-Kulturreis) mit Hilfe von Bestandsreflexion in der Sanjiang Ebene, China. Hyperspektrale Daten wurden in Freilandexperimenten und in Feldern von Landwirten mit Feldspektroradiometern gemessen. Nach den Spektralmessungen wurde die oberirdische Biomasse destruktiv in unterschiedlichen phänologischen Wachstumsstadien gemessen. Beste einzelne Bänder, beste Zweibandkombinationen, optimierter Simple Ratio (SR) und Normalisierter Ratio Index (NRI), sowie Multiple Regressionsanalyse (MLR) wurden anhand von Reflexionsdaten der Freilandexperimente berechnet. Die Daten von den Feldern der Landwirte wurden als Validierungsdatensatz verwendet. Die Reflexionsanalysen zeigen mehrere zur Biomasse korrelierend sensitive Bänder, z. B. 550 nm, 670 nm, 708 nm, 936 nm, 1125 nm und 1670 nm, welche sich in Abhängigkeit von phänologischen Wachstumsstadien änderten. Diese Bänder wurden mittels Korrelogramme für SR, NRI und MLR mit einem Versatz von ca. ± 10 nm detektiert. Die Bewertung der drei Methoden zeigte deutliche Vorteile von MLR gegenüber SR und NRI in der Biomassenabschätzung für Reis im Bestockungs- und Ährenschwellenstadium. Die optimale Bandanzahl für MLR wurde auf vier festgesetzt, um eine Überanpassung zu vermeiden. Das beste MLR-Modell ($R^2 = 0.82$) zum Bestockungsstadium basiert auf vier Bändern (672 nm, 696 nm, 814 nm und 707 nm). Die Analyse von hyperspektralen Reflexionsdaten zur Optimierung von VIs oder MLR hat ein großes Potential in der Biomassenabschätzung für Reis in unterschiedlichen phänologischen Stadien. Dies wird durch die gute Übertragbarkeit ($R^2 = 0.69$ für SR und $R^2 = 0.70$ für NRI) der optimierten Modelle in die landwirtschaftliche Praxis unterstrichen.

---

1 Introduction

In the field of crop science, the aboveground dry biomass (AGB) and nutrient use efficiency are considered to be the major factors for determining the final yield (RAUN & JOHNSON 1999). AGB influences at each phenological stage the amount of grain production, since the yield is defined as the amount of grain, straw, and AGB. Furthermore, knowledge of crop development characteristics and its spatial and temporal variation in the field are useful for determining crop requirements such as N-fertilisation as closely as possible and for achieving acceptable yields, e.g. for rice (FAGIERA 2007).

Traditional methods to estimate AGB involve direct destructive measurements in the field, which are time-consuming, expensive, and require intensive field work. In the context of precision agriculture, proximal sensing is a promising and well investigated tool to avoid the destructive approach (GEBBERS & ADAMCHUK 2010). Field canopy reflectance can be measured with portable handheld or mobile spectroradiometer, e.g. Yara N-Sensor (AGRICON 2013) and can be used to support farmer's decisions on crop management such as fertilisation, pest management, or irrigation.

Hyperspectral measurements in the field can also be used as groundtruth or for model development in analysing satellite imagery. The disadvantage of hyperspectral and multispectral satellite images is the high dependence on a clear sky at the image aquisition time, while spectroradiometers can be used in the field with some cloud cover for approximately 3–4 hours around solar noon. Spectroradiometers are fast and the most important non-destructive devices. They have a continuous acquisition of all reflectance values in a given spectral range of 350 nm – 2500 nm with a high spectral resolution of < 2 nm – 5 nm (MILTON et al. 2009, ORTENBERG 2011).

In many studies, the in-field reflectance measurements are acquired and required for calibrating satellite-borne hyperspectral data using, e.g. EO-1 Hyperion imagery (PSOMAS et al. 2011, KOPPE et al. 2012) or of airborne-based data, e.g. HyMap imagery (CHO & SKIDMORE 2009). The goal of these studies is to estimate AGB on a large scale using airborne or satellite-borne remote sensing data by applying evaluated models. The hyperspectral satellite EnMap, which is scheduled for 2016, will provide data with two separate sensors for the aquisition of VNIR and SWIR in the spectral domain of 420 nm – 2450 nm with 30 m ground resolution similar to the Hyperion data (SCHWIND et al. 2012). By simulating the spectral properties of EnMap with field spectroradiometer data, models can be developed and evaluated for EnMap-based estimation of AGB (KAUFMANN et al. 2010).

The relationship between reflectance and agricultural crop characteristics has been investigated in many studies in the last decades (THENKABAIL et al. 2000). Most of those studies focused on nitrogen, leaf area index (LAI), or yield estimation, but rarely on AGB estimation, since there is a strong correlation between LAI and AGB (SHIBAYAMA & AKIYAMA 1989, FILLELA & PENUELAS 1994). Studies that use hyperspectral ground data to estimate AGB have been carried out for grass, wheat, and for rice (RICHARDSON et al. 1983, SHIBAYAMA & AKIYAMA 1986, ANDERSON & HANSON 1992, SERRANO et al. 2000, HANSEN & SCHJOERRING 2003, LI et al. 2010, WANG et al. 2008, BAJWA et al. 2010).

The common method to estimate AGB from reflectance is based on the application, improvement, or development of spectral indices (SIs). Optimised or improved standard SIs such as SR and NRI represent best band selections based on correlograms (THENKABAIL et al. 2000, STROPPIANA et al. 2009, KOPPE et al. 2010). In addition, MLR has been applied in several studies, since this method provides flexibility in the choice of bands (SHIBAYAMA & AKIYAMA 1989, TAKAHASHI et al. 2000, YU et al. 2013). Futhermore, MLR is more reliable than SIs that is saturated at high LAI levels (YANG & CHEN 2004, HABOUDANE et al. 2004). It is simpler and more flexible for the adoption by growers and crop consultants than partial least square (PLS) (BAJWA et al. 2010).

Many AGB studies were conducted in greenhouses under controlled conditions or the spectral reflectance was measured in the laboratory (SONG et al. 2011), but not in the field. Furthermore, most results, proposed SIs, or developed models were not validated us-

ing an independent dataset to test the transferability of the models. Previous studies have often focused only on calibrating wavebands to a crop parameter, and have not adequately evaluated the performance of their results in an independent dataset (LU 2006, CHO & SKIDMORE 2009).

The main two objectives of this study are (i) to investigate the potential for rice AGB estimation from the canopy reflectance and (ii) to develop and evaluate the proposed AGB estimation models. The first step contains an analysis of the AGB variation, the analysis of the relationship between canopy reflectance and N-application and AGB. In a second step, three methods to estimate AGB by the reflectance are tested with a calibration dataset using experimental field data: Single bands, best two band-combinations (SR and NRI), and MLR. Finally, the three methods are transferred to an independent dataset using farmers' field data under conventional management.

2 Material and Methods

2.1 Study Area and Experimental Design

The research was carried out at the Qixing farm (47.2 °N, 132.8 °E) in Jiansanjiang, Heilongjiang Province, Northeast China. The farm is located in the Sanjiang Plain (Fig. 1), which is an alluvial plain from three rivers (Heilongjiang, Songhua, Wusuli), and covers about 109,000 km^2. The area is characterised by a sub-humid continental monsoon climate. The mean annual temperature is about 2 °C and the mean precipitation sums up to 550 mm per year. The frost-free period is about 130 days long. The rice fields belong to the northernmost cropping rice system in China and to the northernmost ones worldwide. The rice is sown in mid-April (in heated greenhouses), is transplanted after the frost period to the field from the middle to the end of May, and is harvested around end of September. The fields are controlled flooded and manured with N-fertiliser before transplanting the seedlings. Four to five seedlings (120–150 seedlings/m^2) are planted at one position forming a so-called hill. The irrigation is usually stopped 30 days before harvest.

Two field experiments were carried out at two sites (Keyansuo and Qixing research station) in Jiansanjiang and were used as the calibration dataset. They represented a wide range of growth conditions by N-fertiliser input conducted in a split-plot design: 0, 60, 75, 90, 105, 120, and 150 kg N ha^{-1} in 2007 (146 plots), and 0, 35, 70, 105, and 140 kg N ha^{-1} in 2008 (88 plots) and 2009 (95 plots). The plot size was approximately 20 m^2 each. The widely used rice variety *Kongyu131* (28 hills/m^2) was cultivated in all experiments. In addition to the experimental fields, 9 farmers' fields were selected in 2007–2009 and used as the validation dataset. They were managed by the farmers according to their usual practices. The size of these fields varied from 12 to 27 ha, where each field contained several plots with a mean

Fig. 1: Study area in the north-east corner of China.

size of 1,400 m². In most cases, the cultivar *Kongyu131* was planted.

2.2 Hyperspectral Data Collection

Hyperspectral and agronomic data were collected in 2007–2009. Before taking the spectral reflectance, the average number of tillers in each hill was determined per plot in order to measure the reflectance of representative plants (Fig. 2). Canopy spectral reflectance was measured with two non-imaging passive sensors by ASD (Analytical Spectral Devices, Inc., Boulder, CO, USA): QualitySpec® Pro in the wavelength domain of 350 nm – 1800 nm in 2007 and 2009, and FieldSpec3® Pro in the wavelength domain of 350 nm – 2500 nm in 2008. Both devices have a sampling interval of 1.4 nm in the VNIR and 2 nm in the SWIR domain. The measurements were taken from 9 a.m. to 1 p.m. LMT, mostly under cloudfree conditions in the field. Every 10 to 15 minutes, calibration measurements were taken with a white reference panel ($BaSO_4$) and were repeated depending on illumination changes. A default viewing angle (*α*) of 25° and a measuring height (*h*) of 1 m above the canopy created a field of view (*A*) of 0.15 m² with a radius (*r*) of 22 cm (1) and (2).

$$r(m) = h \times tan(\alpha \mid 2) \tag{1}$$

$$A(m^2) = \pi \times r^2 \tag{2}$$

Fig. 2: Hyperspectral data collection in paddy rice (distance to canopy: 0.3 m).

Ten sample counts in the spectrum averaging (settings in the ASD software) were repeated at 6 positions per plot. They were averaged per plot in order to reduce the atmospheric influence, e.g. clouds and wind, and field conditions, e.g. planting in rows. Overall, approximately 14,000 spectra (unaveraged) were collected from 2007 to 2009.

2.3 Aboveground Biomass (AGB) Collection

The AGB was measured destructively by clipping three (booting to heading stage) to five (tillering to stem elongation stage) hills of the measured rice plants. All plant samples were rinsed with water, the roots were clipped, and then the samples were divided in their plant organs leaf, stem and head. They were oven dried at 105 °C for 30 minutes, and dried at 70 °C until constant weight. The AGB was weighed later. In this study, the combined total dry AGB was used and not the individual AGBs from the different organs (leaf, stem, head). Altogether, 1,685 AGB samples were collected from the tillering to heading stages.

2.4 Spectral Indices (SIs) and Stepwise Multiple Linear Regression (MLR)

Single bands or combinations of up to four different bands were tested for their explanatory value. In addition, spectral indices representing two bands were analysed: simple ratio (SR) and normalised difference vegetation index (NDVI). They are widely used for the prediction of biophysical quantities of crops and were developed by Jordan (1969) and Rouse et al. (1974). In this study, the focus is on using the best band-combinations to optimise the SR and NDVI for AGB estimation due to the saturation of the NDVI (Haboudane et al. 2004). The optimised NDVI is also known as normalised ratio index (NRI) and was suggested to determine the best band-combinations (Thenkabail et al. 2000, Sims & Gamon 2002). All possible combinations were computed from the wavelengths in the domain of 350 nm – 1800 nm. The two band-combi-

nations were calculated with a self-developed Java program, analysed and plotted as a contour diagram using MATLAB 7.0 software (MATLAB 2013). Due to the noises caused by water absorption in the SWIR domain, the bands from 1330 nm to 1480 nm, and 1770 nm to 1800 nm were excluded from the analyses. Only the best two band-combinations are presented in the results. The SR and NRI equations are defined as (3) and (4):

$$SR = \frac{\rho_1}{\rho_2} \text{; where } \rho_1 > \rho_2 \qquad (3)$$

$$NRI = \frac{\rho_1 - \rho_2}{\rho_1 + \rho_2} \text{; where } \rho_1 > \rho_2 \qquad (4)$$

where
ρ reflectance value

For the analysis of AGB in relation to one feature, irrespectively if a single band or index was used, correlation analysis was applied. The method attempts to model the relationship between two or more variables by fitting a linear regression equation to observed data. Single bands, but also combination of two, three, and four different bands were tested using a stepwise multiple linear regression (MLR). This allows selecting predictors of dependent variable based on statistical criteria. The observed data is the dependent variable of the model. In this study, the AGB is the dependent variable and the single bands are the independent variables. In total, 1,250 bands were analysed in SPSS 20.0 (SPSS 2013) and the best MLR models are presented in the results. The MLR equation is defined as (5):

$$y = a + b_1 \times \rho_{b_1} + b_2 \times \rho_{b_2} + \ldots + b_i \times \rho_{b_i} \qquad (5)$$

where
y	multiple linear regression (MLR)
a	mathematical constant
$b_1, b_2, \ldots, b_i$	coefficients
$\rho_{b1}, \rho_{b2}, \ldots, \rho_{bi}$	reflectances

2.5 *Data Analysis and Statistics*

Original spectral data were used to average the six spectra per plot. The spectra were not smoothed, but significant outliers were excluded from the analysis. In addition, the stepwise MLR provided a method of feature reduction and a statement about the optimal band number to estimate AGB. Basic analyses were conducted like descriptive statistics of AGB, analyses of canopy spectra under different N-rates, and growth stages before using the data as a calibration dataset. The calibrated models were validated using an independent dataset to test the transferability of the models. The following statistic parameters, root-mean-square error (RMSE) and relative error (RE) against the observed mean, were used to calculate the fitness between the observed and estimated data. All statistical analyses were conducted in SPSS 20.0 and Statistica 6.0 (STATISTICA 2013).

3 Results

3.1 *Temporal AGB Variation*

AGB production and development are dependent on crop growth conditions such as weather, soil and nutrition. Tab. 1 illustrates the temporal AGB variation in diverse years and growth stages for the experimental sites. Generally, AGB production tends to increase from the tillering to the heading stage. Overall, it ranges from 0.1 t/ha to 14.1 t/ha across all stages and years. The rice crop had a high variation in AGB (CV > 30 %), especially during the early growth stages. During the later stages booting and heading, the variation was lower (CV < 30 %). Temporal variation between the three years is significant. In 2009, the temperature was lower than in the previous years, so the AGB production was lower with a mean AGB value of 0.8 t/ha – 7.0 t/ha. Highest values were observed in 2008.

Tab. 1: Descriptive statistics of AGB on the experimental fields.

	Stage	n	Min (t/ha)	Max (t/ha)	Mean (t/ha)	SD (t/ha)	CV (%)
2007	Tillering	146	0.1	2.0	0.9	0.46	46.5
	Stem Elongation	74	1.6	5.7	3.4	0.88	26.1
	Booting	49	2.9	7.5	5.6	1.05	18.8
	Heading	114	3.3	12.4	7.6	1.96	26.1
2008	Tillering	40	0.1	1.8	0.9	0.46	50.5
	Stem Elongation	40	0.9	2.9	1.6	0.49	31.3
	Booting	88	2.9	8.8	5.3	1.36	25.6
	Heading	88	4.4	14.1	9.0	1.83	20.4
2009	Tillering	91	0.2	1.6	0.8	0.34	41.3
	Stem Elongation	95	0.3	2.2	1.2	0.52	42.0
	Booting	95	1.4	6.6	3.5	1.26	36.1
	Heading	95	4.6	9.7	7.0	1.15	16.5
All	Tillering	277	0.1	2.0	0.9	0.43	46.3
	Stem Elongation	209	0.3	5.7	2.1	1.19	58.0
	Booting	232	1.4	8.8	4.7	1.57	34.0
	Heading	297	3.3	14.1	7.8	1.87	24.1

n = Number of samples, SD= Standard deviation, CV= Coefficient of variation

3.2 Canopy Reflectance Spectra under Different N-rates and at Different Growth Stages

The reflectance spectra of the rice canopies clearly indicated differences in AGB or LAI that resulted from different N-rates and at various phenological growth stages. The spectra of the experiment plots with five different N-rates at the booting stage in 2008 were taken as an example to display the response pattern of canopy reflectance (Fig. 3, left). Generally, the reflectance spectra tended to increase with rising LAI in the NIR (700 nm – 1100 nm) and SWIR (1100 nm – 1800 nm) regions, whereas the opposite occurred in the VIS (500 nm – 700 nm) region. The canopy LAI responded to N-application. Higher reflectance response occurred with lower N-application. Especially, in the green (500 nm – 600 nm) and red (600 nm – 700 nm) regions, obvious visible differences were detected. There was a high increase in the five spectra in the ReddEdge region (670 nm – 740 nm), which were

clustered here as one spectrum. Mostly, the differences between the first (0 kg·N/ha) and the second N-rate (35 kg·N/ha) were not significant in the reflectance. Similarities were also observed for the fourth (105 kg·N/ha) and fifth (140 kg·N/ha) N-rate.

Rice canopies showed a diverse reflectance at different growth stages. As an example, a dataset of four stages in 2008 was used to display the response of a rice crop from the tillering to the heading stage (Fig. 3, right). At an early phenological stage, the reflectance was mainly influenced by the soil and water of the paddy field, where the AGB production was low due to low LAI.

Generally, the reflectance increased from the tillering to the booting stage and decreased from the heading stage due to starting senescence of the plant. Maximum reflectance was observed at the booting stage. The canopy LAI and the biochemical components of the plant changed at different growth stages, which evidently influenced the reflectance.
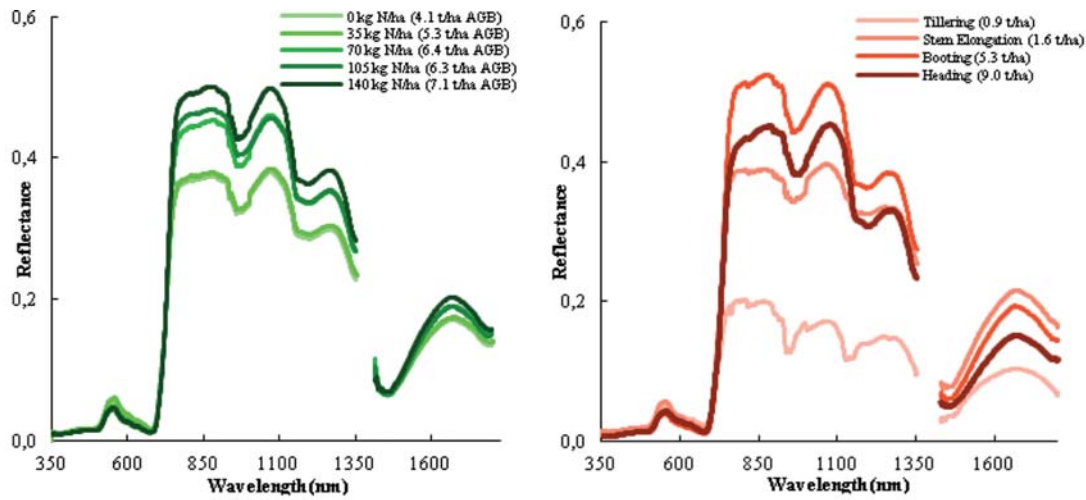
Fig. 3: Left: changes of rice canopy reflectance with varied N-rates at the booting stage in 2008, right: changes in reflectance at different growth stages in 2008.
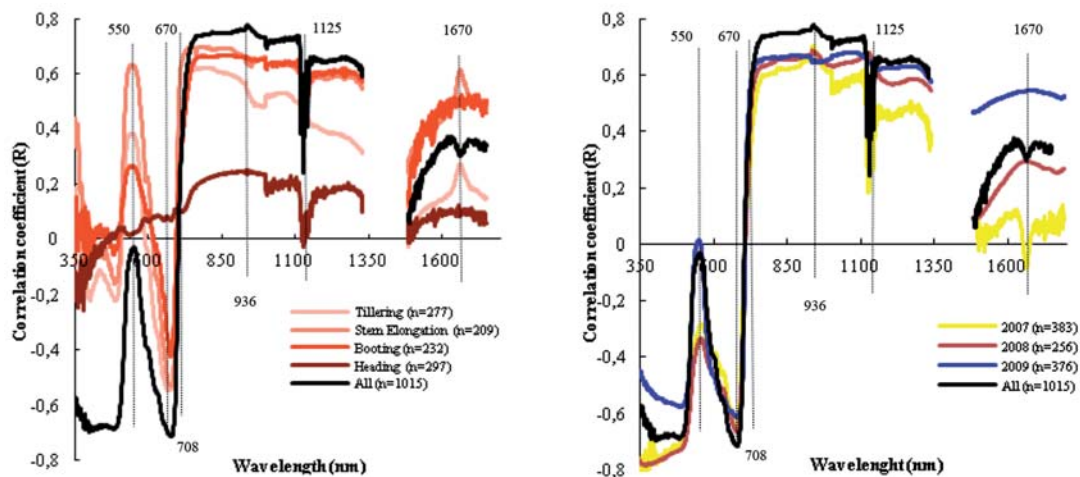

Fig. 4: Left: the correlation coefficients (R) between AGB and canopy reflectance at different phenological stages, right: in different years.

3.3 *Relationship between AGB and Spectral Reflectance*

The correlation coefficients between AGB and canopy reflectance at different phenological stages and in different years are presented in Fig. 4. First of all, the correlation between AGB and reflectance at different stages is described (Fig. 4, left). The pattern of the R (correlation coefficient) curves was similar at the different stages and across all stages except for the heading stage. Lowest absolute R values were observed at the tillering stage as a

result of low crop development at this growth stage. Highest R values were recorded at the stem elongation stage and across all stages. Maximum negative R values were observed at tillering and across all stages at wavelengths of 670 nm, which corresponds to high solar radiation absorption by chlorophyll pigments. In the RedEdge region, a high increase of R values was detected, which is coincident with reflectance increase of vegetation in this domain.

Greatest positive R values were observed in the NIR shoulder at the stem elongation

stage and at wavelengths of 936 nm across all stages. Two peaks are noticeable in the SWIR domain. The first one is located as a local minimum in the reflectance at wavelengths of 1125 nm at the booting and heading stages, which is not detected in the early stages of tillering and stem elongation. Additionally, the plotted R curves show some noises in the SWIR domain for the booting and heading stage. The noises were only observed in the 2007 data due to partly cloudy sky during the measurements. As a second peak, a local maximum in the R curves is observed at wavelengths of 1670 nm at tillering and stem elongation.

The plotted curves for the correlation coefficients (R) between AGB and canopy reflectance show strong differences from year to year (Fig. 4, right). In general, the relationship seems to be diverse in all three years. Across all years, the highest R values are observed (R > 0.75), and in 2007 the lowest (R > 0.6). In summary, sensitive bands are located at around 550, 670, 708, 936, 1125 and 1670 nm.

3.4 *Model Calibration by single Bands, SR and NRI*

The coefficient of determination between AGB and single bands, best SR, and best NRI was calculated. The best single bands and two band-combinations are shown in Tab. 2 at each growth stage and across all three years. Generally, the best SR and NRI always produce higher R^2 values as one single band. SR and NRI show a very similar performance in
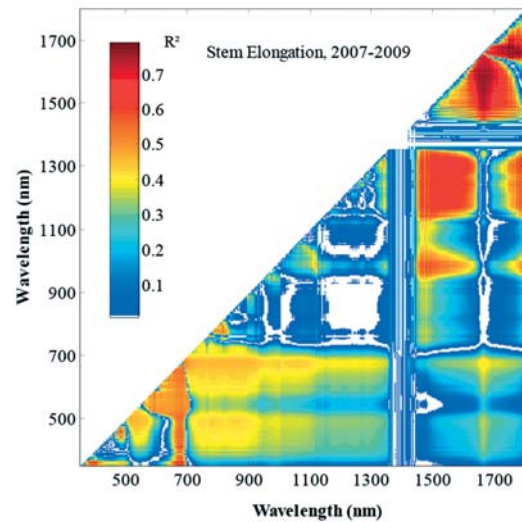

Fig. 5: Best two band-combinations for NRI at the stem elongation stage.

R^2 values except at the stem elongation stage. Moreover, they are similar in the band-combinations ± 20 nm). At the tillering stage, the SR (ρ_{822}, ρ_{716}) has the best performance (R^2 = 0.58). At the stem elongation stage, the NRI (ρ_{1678}, ρ_{1575}) displayed the best results (R^2 = 0.75) versus other stages (Tab. 2, Fig. 5). At the booting stage, SR (ρ_{695}, ρ_{513}) and NRI (ρ_{695}, ρ_{515}) performed similarly (R^2 = 0.54) using almost equal bands. Due to the changes in the canopy reflectance by biochemical components of the plant, all AGB predictors result in lower R^2 values (< 0.3) at the heading stage.

However, across the whole monitored season, high R^2 values were observed (R^2 > 0.6) for a high sample number (n = 1015). SR (ρ_{713},

Tab. 2: Single bands, SR and NRI model calibration at different growth stages (2007–2009 pooled data).

Stages (2007–2009)	n	Single Band ρ_1		SR ρ_1/ρ_2		NRI $(\rho_1-\rho_2)/(\rho_1+\rho_2)$	
		ρ_1	R^2	ρ_1, ρ_2	R^2	ρ_1, ρ_2	R^2
Tillering	277	672	0.344	822,716	**0.582**	799,711	0.559
Stem Elongation	209	780	0.487	1760,1325	0.528	1678,1575	**0.758**
Booting	232	854	0.443	695,513	**0.541**	695,515	**0.541**
Heading	297	380	0.066	800,789	0.218	800,789	**0.293**
All	1015	936	0.629	713,550	**0.757**	713,533	0.743

ρ_{550}) showed its best performance here ($R^2 = 0.75$). The best single band was 936 nm ($R^2 = 0.62$), which is also important for the MLR models.

3.5 MLR Model Calibration

MLR analyses were conducted in two directions: i) assessment of the optimal band number, ii) MLR-models with 1–4 single bands. Generally, with respect to the first direction, the MLR models explain 50–93 % of the variation in AGB (Fig. 6). The highest performance was observed at the stem elongation stage ($R^2 = 0.93$ with 18 bands), the lowest at the heading stage ($R^2 = 0.50$ with 17 bands). At the tillering stage, the R^2 reached a value of 0.82 with 19 bands and at the booting stage, a value of 0.60 with 7 bands. The accuracy of the MLR models was quite good. Across all stages, 35 bands explained 88 % of the bio-
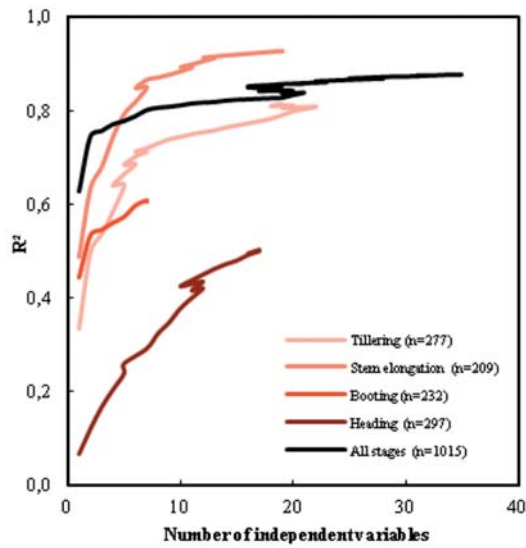

Fig. 6: Relationship between band number and performance (R^2) of the MLR models.

Tab. 3: Stepwise MLR models at diverse growth stages (MLR = multiple linear regression, adj. = adjusted, SE = Standard error of the estimator).

Stage	Model	Bands	Regression equation	R^2	$R^2_{adj.}$	SE
Tillering	MLR-1	1	AGB = 1.638 - 19.715 ρ_{672}	0.334	0.332	0.348
n = 277	MLR-2	2	AGB = 1.051 - 16.182 ρ_{672} + 2.369 ρ_{1052}	0.498	0.494	0.303
	MLR-3	3	AGB = 1.189 + 12.021 ρ_{672} + 4.321 ρ_{1052} - 28.093 ρ_{696}	0.536	0.530	0.292
	MLR-4	4	AGB = 0.824 - 31.879 ρ_{672} + 70.423 ρ_{696} + 12.658 ρ_{814} - 60.408 ρ_{707}	0.641	0.636	0.257
Stem	MLR-1	1	AGB = -0.190 + 9.658 ρ_{780}	0.487	0.485	0.858
Elongation	MLR-2	2	AGB = 0.220 + 91.228 ρ_{780} - 86.560 ρ_{763}	0.633	0.629	0.728
n = 209	MLR-3	3	AGB = 0.518 + 72.709 ρ_{780} - 63.048 ρ_{763} -22.077 ρ_{1489}	0.679	0.674	0.682
	MLR-4	4	AGB = 1.277 + 26.058 ρ_{780} - 24.207 ρ_{763} - 66.429 ρ_{1489} + 30.298 ρ_{1662}	0.744	0.739	0.611
Booting	MLR-1	1	AGB = 1.374 + 8.697 ρ_{854}	0.443	0.441	1.180
n = 232	MLR-2	2	AGB = 1.772 + 17.572 ρ_{854} - 19.088 ρ_{729}	0.533	0.529	1.083
	MLR-3	3	AGB = 1.555 + 24.049 ρ_{854} - 14.718 ρ_{729} - 10.531 ρ_{1172}	0.545	0.539	1.071
	MLR-4	4	AGB = 1.147 + 28.131 ρ_{854} - 20.997 ρ_{729} - 12.945 ρ_{1172} + 69.257 ρ_{377}	0.560	0.552	1.056
Heading	MLR-1	1	AGB = 9.970 - 256.497 ρ_{380}	0.066	0.062	1.816
n = 297	MLR-2	2	AGB = 7.798 - 255.759 ρ_{380} + 5.190 ρ_{1083}	0.119	0.113	1.766
	MLR-3	3	AGB = 7.842 - 263.838 ρ_{380} + 20.485 ρ_{1083} - 16.632 ρ_{1003}	0.166	0.157	1.722
	MLR-4	4	AGB = 6.445 - 488.717 ρ_{380} + 26.101 ρ_{1083} - 23.102 ρ_{1003} + 349.618 ρ_{406}	0.203	0.192	1.686
All	MLR-1	1	AGB = -1.534 + 20.179 ρ_{936}	0.629	0.628	1.889
n = 1015	MLR-2	2	AGB = -0.111 + 29.231 ρ_{936} - 25.930 ρ_{1659}	0.745	0.745	1.565
	MLR-3	3	AGB = 0.137 + 46.188 ρ_{936} - 24.244 ρ_{1659} - 18.150 ρ_{762}	0.757	0.756	1.529
	MLR-4	4	AGB = 0.481 + 37.904 ρ_{936} - 30.875 v_{1659} - 27.087 ρ_{762} + 18.429 ρ_{1027}	0.771	0.770	1.485

mass variability. However, the results indicated that the best MLR model was dependent on the number of independent variables. A higher number of variables cause higher R^2 values, but the number of bands has a limit. Fig. 6 demonstrates the relationship between the band number and the performance of the MLR models at each stage. The relation is illustrated as a curve for each stage. The curves are characterised by four features: Linear rise, saturation, again linear rise and finally asymptotic trend. Despite the stepwise increasing of bands, the performance of the models showed an indication of overfitting (after 2–3 bands). iii) For this reason, the optimal number was set to four. The best MLR models are listed in Tab. 3 separately for each stage and across all stages.

Tab. 4: Calibration results versus validation results (NRI = normalized ratio index, SR = simple ratio, MLR = multiple linear regression).

Calibration dataset		(2007–2009)	Validation dataset			
Stage	R^2	**Model** $_{(\rho)}$	**RMSE**	**RE**	**R^2**	Stage
	0.56	NRI $_{(799,711)}$	0.58	54.8	0.70	
	0.58	SR $_{(822, 716)}$	0.57	53.9	0.69	
Tillering	0.30	MLR-1 $_{(672)}$	0.66	57.6	0.29	Tillering
n = 277	0.44	MLR-2 $_{(672, 1052)}$	0.61	53.0	0.55	n = 92
	0.47	MLR-3 $_{(672, 1052, 696)}$	0.56	53.0	0.62	
	0.56	MLR-4 $_{(672, 696, 814, 707)}$	0.47	44.5	0.82	
	0.76	NRI $_{(1678, 1575)}$	1.25	59.4	0.38	
	0.53	SR $_{(1760, 1325)}$	1.36	64.8	0.45	
Stem Elongation	0.49	MLR-1 $_{(780)}$	1.88	89.3	0.18	Stem Elongation
n = 209	0.63	MLR-2 $_{(780, 763)}$	1.28	60.8	0.31	n = 130
	0.68	MLR-3 $_{(780, 763, 1489)}$	1.15	54.8	0.51	
	0.74	MLR-4 $_{(780, 763, 1489, 1662)}$	1.16	55.1	0.42	
	0.54	NRI $_{(695, 515)}$	1.89	49.3	0.00	
	0.54	SR $_{(695, 513)}$	3.49	90.8	0.00	
Booting	0.44	MLR-1 $_{(854)}$	1.89	49.3	0.14	Booting
n = 232	0.53	MLR-2 $_{(854, 729)}$	2.07	53.8	0.06	n = 257
	0.54	MLR-3 $_{(854, 729, 1172)}$	1.98	49.7	0.12	
	0.56	MLR-4 $_{(854, 729, 1172, 377)}$	1.86	48.4	0.11	
	0.29	NRI $_{(800, 789)}$	2.73	51.3	0.19	
	0.22	SR $_{(800, 789)}$	2.76	51.9	0.19	
Heading	0.02	MLR-1 $_{(380)}$	3.12	58.6	0.01	Heading
n = 297	0.04	MLR-2 $_{(380, 1083)}$	3.12	58.6	0.01	n = 191
	0.05	MLR-3 $_{(380, 1083, 1003)}$	3.09	58.1	0.02	
	0.06	MLR-4 $_{(380, 1083, 1003, 406)}$	2.86	53.8	0.12	
	0.74	NRI $_{(713, 533)}$	2.86	76.0	0.42	
	0.76	SR $_{(713, 550)}$	2.76	77.9	0.70	
All	0.60	MLR-1 $_{(936)}$	2.57	72.7	0.48	All
n = 1015	0.72	MLR-2 $_{(936, 1659)}$	2.42	68.4	0.53	n = 670
	0.74	MLR-3 $_{(936, 1659, 762)}$	2.44	68.9	0.54	
	0.75	MLR-4 $_{(936, 1659, 762, 1027)}$	2.49	70.2	0.55	

The results indicate that regression equations and the significant bands vary between the phenological stages. The performance of the models was improved by adding stepwise an additional independent band. For example at stem elongation, the R^2 values increased from 0.48 (MLR-1) to 0.74 (MLR-4) and across all stages from 0.62 (MLR-1) to 0.77 (MLR-4). Due to the high sample number, all results are significant at $p < 0.0001$ except for the heading stage. It is striking that many of the bands are located in the NIR and SWIR domain, but only some in the VIS domain.

3.6 *Calibration against Validation*

The calibrated (SR-, NRI-, MLR-) models were validated by an independent dataset (Tab. 4). Generally, the models using 2–4 single bands are the most promising for estimating AGB. Only one band explained AGB variability the least. In the case of MLR, the models tended to overfit already when fitting with 2–3 bands. This caused a slight increase of the R^2 and the RMSE values, though more bands were used for modeling.

At the tillering stage, the top identified models all used RED (672 nm, 692 nm) and NIR bands (707 nm – 1052 nm), while at the stem elongation stage NIR bands (763 nm, 780 nm) and SWIR bands (1325 nm – 1760 nm) dominated, and across all stages GREEN bands (533 nm, 550 nm), NIR bands (713 nm – 1027 nm) and a SWIR band (1659 nm) were selected. In most cases, the validated MLR-models provided the best results with highest R^2 values ($R^2 = 0.82$ at the tillering stage, $R^2 = 0.51$ at the stem elongation stage). The performance of the models at the booting and heading stages was worse ($R^2 < 0.19$). In addition, the RMSE values increased till to 3.49 t/ha. These models are useless for regionalisation. Across all stages, the SR showed its best performance ($R^2 = 0.70$). The RMSE values are reasonable, but the RE shows relatively high values (RE > 40 %). This fact can be explained through the different management of the experimental fields (manual work by fieldworkers) and farmers' fields (mechanical work by tractors and airplanes) and the different plot size (small scale farm-

ing in experiments and large scale in farmers' fields). When the SR-, NRI-, and MLR-models were evaluated with data from farmers' fields, the R^2 values were significantly smaller (0.38–0.51 at the stem elongation stage and 0.42–0.70 across all stages). However, at the tillering stage, the R^2 values were significantly higher (0.62–0.82) than the calibration values.

4　Discussion and Conclusion

In comparison to reflectance-based estimation of AGB for wheat or other cereals, the estimation of rice AGB is linked with a lower relationship between reflectance and AGB resulting in a lower R^2 performance of the models. Single bands, optimised SR and NRI, as well as MLR-based methods were able to explain 80 % – 90 % of the biomass variability in wheat, e.g. Zhu et al. (2008), Bao et al. (2009), Koppe et al. (2010), but only 60 % – 80 % in rice, e.g. Patel et al. (1985), Shibayama & Akiyama (1989), Takahashi et al. (2000), Wang et al. (2008), and Bajwa et al. (2010).

In this study, the performance of the investigated indices SR and NRI ($R^2 = 0.75$) was in the range of published studies, while MLR performed better ($R^2 = 0.93$). The bands of NRI were similar as in the study of Stroppiana et al. (2009) for rice AGB across all stages. Their analysis indicated highly correlated AGB ($R^2 > 0.9$) in the RedEdge. In the case of MLR, the optimal bands in this investigation were comparable to those by Wang et al. (2008). They also set the optimal band number to four to estimate AGB of rice, and they detected several bands in the SWIR domain. There was a clear cluster of SWIR bands, especially at the stem elongation stage, when the rice AGB was highly correlated with LAI (Gnyp et al. 2012). These bands represent the maximum reflectance in the 1500 nm – 1800 nm domain and are sensitive to lignin, starch and protein (Kumar et al. 2003).

Furthermore, one key objective of our study was the transfer of the optimised SIs and newly developed MLR models, which were investigated from data collected in field experiments, to real practice: Farmers' rice fields under conventional management. In this study, the up-scaling from the experimental to the

farmers' field level was promising using 2–4 bands by SR, NRI and MLR at the tillering ($R^2 = 0.82$) and the stem elongation stage ($R^2 = 0.51$). Prosperous scaling approaches from plots to the regional level were demonstrated to estimate LAI (Lausch et al. 2012), plant diseases (Laudien & Bareth 2006), or AGB in wheat (Koppe et al. 2012). But also high resolution and multi-temporal TerraSAR-X images indicated a high potential ($R^2 = 0.80$) of estimating rice AGB, e.g. Koppe et al. (2013). However, these studies focused on SIs or radar backscatter to transfer the models.

Several problems occurred during upscaling in this study, e.g., RedEdge shift and model overfitting. Best performing bands changed at different growth stages due to an increase or decrease of canopy reflectance in diverse spectral regions. Probably, the shift of the RedEdge and changes in the chlorophyll concentration due to the different N-application or different management in experimental and farmers' fields caused problems in model transferability. Temporal changes in RedEdge and chlorophyll concentration are a known fact, e.g. Horler et al. (1983) and Fillela & Penuelas (1994). The overfitting occurred with 2–3 bands and reduced the model predictability by MLR. Therefore, more bands improved the statistical relation, but not the model transferability. Nevertheless, since MLR is widely used for AGB estimation, PLS could probably improve the results. Hansen & Schjoerring (2003) showed that PLS performed better in predicting AGB compared to SIs. In a study for AGB estimation of rice, Thakahashi et al. (2000) achieved a better fitting by MLR, but a worse predictability than by PLS. Their results are comparable to this study at the stem elongation stage. In addition, the AGB variability is also diverse at each stage. The highest variability in AGB was observed, especially at the stem elongation stage. Probably, the water of the paddy field, but also wet soil could partly have some impact in the models at the tillering and stem elongation stages, when the plant cover was low. Since the same rice cultivar was planted in the experimental as well as in the farmers' fields, it should have no influence on the models. The different management such as controlled conditions in experimental and uncontrolled ones in farmers' fields might

be a problem. Up-scaling from small experimental to larger farmers' fields often yielded lower model predictability, e.g. Li et al. (2010) and Psomas et al. (2011), like in this study at the stem elongation stage.

Several different calculations could have been carried out in this study to partially avoid overfitting with MLR. For example, prior studies have shown that PLS, support vector machine (SVM), principle component analysis (PCA) and neural network approaches can also partly help to avoid this problem. In addition, the comparison of estimation of fresh and dry AGB would be of great value, since several studies displayed a better predictability of fresh rice AGB, e.g. Yang & Chen (2004). Due to the changing of the reflectance characteristics in the different plant organs, the AGB estimation should be investigated in the different organs leaf, stem and head as well. Future studies should involve data from large fields for calibration and for validation, which should be independent of each other. Moreover, sensor fusion could improve the validation as well, as shown for the radar and hyperspectral data combination by Koppe et al. (2012). Better development or validation of reliable models could be also achieved by cross-validation and bootstrapping (Richter et al. 2012).

After several improvements, the models of this study, especially these with high predictability at the tillering and stem elongation stages, could be tested by EnMap or other sensors from the space. This study showed the high potential in estimating dry AGB by MLR with 3–4 independent bands, but also by SR and NRI. These bands could be easily tested and evaluated for a larger area by UAVs (unmanned aerial vehicles) carrying hyperspectral sensors or cameras, or by satellite-borne hyperspectral sensors such as EO-1 Hyperion and EnMap, or the airborne sensor HyMap.

Acknowledgements

References

AGRICON, 2013: www.agricon.de (4.5.2013).

ANDERSON, G.L. & HANSON, J.D., 1992: Evaluating hand-held radiometer derived vegetation indices for estimating above ground biomass. – Geocarto International **7** (1): 71–78.

BAJWA, S.G., MISHRA, A.R. & NORMAN, R.J., 2010: Canopy reflectance response to plant nitrogen accumulation in rice. – Precision Agriculture **11** (5): 488–506.

BAO, Y., GAO, W. & GAO, Z., 2009: Estimation of winter wheat biomass based on remote sensing data at various spatial and spectral resolutions. – Frontier of Earth Science in China **3** (1): 118–128.

CHO, M.A. & SKIDMORE, A.K., 2009: Hyperspectral predictors for monitoring biomass production in Mediterranean grasslands: Majella National Park, Italy. – International Journal of Remote Sensing **30** (2): 499–515.

FAGIERA, N.K., 2007: Yield physiology of rice. – Journal of Plant Nutrition **30** (6): 843–879.

FILLELA, I. & PENUELAS, J., 1994: The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. – International Journal of Remote Sensing **15** (7): 1459–1470.

GEBBERS, R. & ADAMCHUK, V.I., 2010: Precision agriculture and food security. – Science **327** (12): 828–831.

GNYP, M.L., YAO, Y., YU, K., HUANG, S., AASEN, H., LENZ-WIEDEMANN, V.I.S., MIAO, Y. & BARETH, G., 2012: Hyperspectral analysis of rice phenological stages in Northeast China. – ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science, Melbourne, Australia, Volume **I-7:** 77–82.

HABOUDANE, D., MILLER, J.R., PATTEY, E., ZARCO-TEJADA, P.J. & STRACHAN, I.B., 2004: Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopy: Modeling and validation in the context of precision agriculture. – Remote Sensing of Environment **90** (3): 337–352.

HANSEN, P.M. & SCHJOERRING, J.K., 2003: Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. – Remote Sensing of Environment **86** (4): 542–553.

HORLER, D.N.H., DOCKRAY, M. & BARBER, J., 1983: The red edge of plant leaf reflectance. – International Journal of Remote Sensing **4** (2): 273–288.

JORDAN, C.F., 1969: Derivation of leaf area index from quality of light on the forest floor. – Ecology **50:** 663–666.

KAUFMANN, H., SEGL, K., ITZEROTT, S., BACH, H., WAGNER, A., HILL, J., HEIM, B., OPPERMANN, K., HELDENS, W., STEIN, E., MÜLLER, A., VAN DER LINDEN, S., LEITAO, P.J. & HOSTERT, P., 2010: Hyperspectral algorithms – Report in the frame of EnMAP preparation activities. – Deutsches GeoForschungsZentrum GFZ **10** (8): 1–268.

KOPPE, W., LI, F., GNYP, M.L., MIAO, Y., JIA, L., CHEN, X., ZHANG, F. & BARETH, G., 2010: Evaluating multispectral and hyperspectral satellite remote sensing data for estimating winter wheat growth parameters at regional scale in North China Plain. – PFG – Photogrammmetrie, Fernerkundung, Geoinformation **2010** (3): 167–178.

KOPPE, W., GNYP, M.L., HENNIG, S.D., LI, F., MIAO, Y., CHEN, X., JIA, L. & BARETH, G., 2012: Multitemporal hyperspectral and radar remote sensing for estimating winter wheat biomass in the North China Plain. – PFG – Photogrammmetrie, Fernerkundung, Geoinformation **2012** (3): 281–298.

KOPPE, W., GNYP, M.L., HÜTT, C., YAO., Y., MIAO, Y., CHEN, X. & BARETH, G., 2013: Rice monitoring with multi-temporal and dual-polarimetric TerraSAR-X data. – International Journal of Applied Earth Observation and Geoinformation **21:** 568–576.

KUMAR, L., SCHMIDT, K., DURRY, S. & SKIDMORE, A., 2003: Imaging spectrometry and vegetation science. – VAN DER MEER, F.D. & DE JONG, S.M. (eds): Imaging Spectrometry, Dordrecht, The Netherlands.

Laudien, R. & Bareth, G., 2006: Multitemporal hyperspectral data analysis for regional detection of plant diseases by using a tractor- and an airborne-based spectrometer. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2006** (3): 217–227.

Lausch, A., Pause, M., Merbach, I., Gwillym-Margianto, S., Schulz, K., Zacharias, S. & Seppelt, R., 2012: Scale-specific hyperspectral remote sensing approach in environmental research. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2012** (5): 589–602.

Li, F., Miao, Y., Hennig, S.D., Gnyp, M.L., Chen, X., Jia, L. & Bareth, G., 2010: Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. – Precision Agriculture **11** (4): 335–357.

Lu, D., 2006: The potential and challenge of remote sensing-based biomass estimation. – International Journal of Remote Sensing **27** (7): 1297–1328.

MATLAB, 2013: MathWorks Inc., Natick, MA, USA, www.mathwork.de (4.5.2013).

Milton, E.J., Schaepman, M.E., Anderson, K., Kneubühler, M. & Fox, N., 2009: Progress in field spectroscopy. – Remote Sensing of Environment **113** (1): 92–109.

Ortenberg, F., 2011: Hyperspectral sensor characteristics: airborne, spaceborne, hand-held, and truck-mounted; Integration of hyperspectral data with Lidar. – Thenkabail, P.S., Lyon, J.G. & Huete, A. (eds): Hyperspectral remote sensing of vegetation. Boca Raton, FL, USA.

Patel, N.K., Singh, P., Sahai, B. & Patel, M.S., 1985: Spectral response of rice crop and its relation to yield and yield attributes. – International Journal of Remote Sensing **6** (5): 657–664.

Psomas, A., Kneubühler, M., Huber, S., Itten, K. & Zimmermann, N.E., 2011: Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. – International Journal of Remote Sensing **32** (24): 9007–9031.

Raun, W.R. & Johnson, G.V., 1999: Improving nitrogen use efficiency for cereal production. – Agronomy Journal **91** (3): 357–363.

Richardson, A.J., Everitt, J.H. & Gausman, H.W., 1983: Radiometric estimation of biomass and nitrogen content of Alicia Grass. – Remote Sensing of Environment **13** (2): 179–185.

Richter, K., Atzberger, C., Hank, T.B. & Mauser, W., 2012: Derivation of biophysical variables from Earth observation data: validation and statistical measures. – Journal of Applied Remote Sensing **6** (1): 179–185.

Rouse, J.W., Has, R.H., Schell, J.A. & Deering, D.W., 1974: Monitoring vegetation systems in the Great Plains with ERTS. – Third ERTS Symposium, NASA Sp-3511: 309–317.

Schwind, P., Müller, R., Palubinskas, G. & Storch, T., 2012: An in-depth simulation of EnMAP acquisition geometry. – ISPRS Journal of Photogrammetry and Remote Sensing **70** (1): 99–106.

Serrano, L., Filella, I. & Penuelas, J., 2000: Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. – Crop Science **40** (3): 723–731.

Shibayama, M. & Munakata, K., 1986: A spectroradiometer for field use III: A comparison of some vegetation indices for predicting luxuriant paddy rice biomass. – Japanese Journal of Crop Science **55** (1): 47–52.

Shibayama, M. & Akiyama, K., 1989: Seasonal visible, near-infrared and mid-infrared spectra of rice canopies in relation to LAI and aboveground dry phytomass. – Remote Sensing of Environment **27** (2): 119–127.

Sims, D.A. & Gamon, J.A., 2002: Relationship between leaf pigment content and spectral reflectance across a wide range of species, leaf structure and development stages. – Remote Sensing of Environment **81** (2–3): 337–354.

Song, S., Gong, W., Zhu, B. & Huang, X., 2011: Wavelength selection and spectral discrimination for paddy rice with laboratory measurements of hyperspectral leaf reflectance. – ISPRS Journal of Photogrammetry and Remote Sensing **66** (5): 672–682.

SPSS, 2013: IBM SPSS Inc., Armonk, NY, USA, www-01.ibm.com/software/analytics/spss (4.5.2013).

Statistica, 2013: StatSoft, Tulsa, OK, USA, www.statsoft.com (4.5.2013).

Stroppiana, D., Boschetti, M., Brivio, P.A. & Bocchi, S., 2009: Plant nitrogen concentration in paddy rice from canopy hyperspectral radiometry. – Field Crops Research **111** (1–2): 119–129.

Takahashi, W., Nguyen-Cong, V., Kawaguchi, S., Minamiyama, M. & Ninomiya, S., 2000: Statistical models for prediction of dry weight and nitrogen accumulation based on visible and near-infrared hyper-spectral reflectance of rice canopies. – Plant Production Science **3** (4): 377–386.

Thenkabail, P.S., Smith, R.B. & Pauw, E.D., 2000: Hyperspectral vegetation indices and their relationship with agricultural crop characteristics. – Remote Sensing of Environment **71** (2): 152–182.

Wang, F., Huang, J. & Wang, X., 2008: Identification of optical hyperspectral bands for estimation of rice biophysical parameters. – Journal of Integrative Plant Biology **50** (3): 291–299.

Yang, C. & Chen, R., 2004: Modeling rice growth with hyperspectral reflectance data. – Crop Science **44** (4): 1283–1290.

Yu, K., Li, F., Gnyp, M.L., Miao, Y., Bareth, G. & Chen, X., 2013: Remotely detecting canopy nitrogen concentration and uptake of paddy rice in Northeast China Plain. – ISPRS Journal of Photogrammetry and Remote Sensing **78:** 102–115.

Zhu, Y., Yao, X., Tian, Y., Liu, X. & Cao, W., 2008: Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulation wheat and rice. – International Journal of Applied Earth Observation and Geoinformation **10** (1): 1–10.

Address of the Authors:

Martin Leon Gnyp, Kang Yu, Helge Aasen, Prof. Dr. Georg Bareth, International Center for Agro-Informatics and Sustainable Development (www.icasd.org), Institute of Geography, GIS & RS Group, University of Cologne, D-50923 Köln, Tel.: +49-221-470-6551, Fax: +49-221-470-1638, e-mail: {mgnyp1}{kyu}{helge.aasen}{g.bareth}@uni-koeln.de

Ass. Prof. Dr. Yuxin Miao, Yinkun Yao, Shanyu Huang, International Center for Agro-Informatics and Sustainable Development (www.icasd.org), College of Resources and Environmental Sciences, China Agricultural University, 100094 Beijing, Tel.: +86-10-62732865, Fax: +86-10-62731016, e-mail: ymiao@cau.edu.cn, yyk719@163.com, shuang 331919@gmail. com

Extracting Urban Parameters of the City of Oldenburg from Hyperspectral, Thermal, and Airborne Laser Scanning Data

Lutz Bannehr, Andreas Schmidt, Dessau-Roßlau, Johannes Piechel & Thomas Luhmann, Oldenburg

Summary: This paper shows how to use different remote sensing sensors and methods to obtain parameters about the urban built-up areas. Within the cooperative research project HiReSens a hyperspectral scanner, an airborne laser scanner, a thermal camera, and a RGB-camera were employed on a small aircraft to determine roof material and geometric parameters as well as heat bridges within the city of Oldenburg, Lower Saxony, Germany.

HiReSens aims to combine various geometrical highly resolved data (50 cm) in order to survey the state of the roof areas. Thermal data were used to obtain the temperature distribution of single roof tops. The hyperspectral data provide information on the roofing materials. Support vector machines (SVM) were used to classify these roof materials. Five out of six roofing materials were clearly detected.

From airborne laser scanning (ALS) data a digital surface model and a digital terrain model were calculated. These models in combination with hyperspectral data form the basis to locate the buildings with the best orientations for solar panels. A decision tree algorithm gives satisfactory results in this case.

The combination of the different datasets offers the opportunity to use synergies between different sensor systems. The central goals were the development of tools for the detection of thermal bridges by means of thermal data, spectral differentiation of roof parameters on the basis of hyperspectral data as well as 3D-capture of buildings from ALS data.

Zusammenfassung: *Ableitung von städtischen Parametern der Stadt Oldenburg durch Hyperspektral-, Thermal- und Airborne Laser Scanning Daten.* Im Rahmen des kooperativen Forschungsprojektes HiReSens, gefördert vom BMBF, wird ein Hyperspektralscanner, ein Airborne Laser Scanner, eine Thermalkamera und eine RGB-Kamera auf einem kleinem Flugzeug, einer Cessna 207, eingesetzt, woraus Parameter der städtischen Bebauung, wie Dachmaterial- und Geometrieparameter sowie Temperaturverteilungen von Dächern, abgeleitet werden.

HiReSens zielt darauf ab, verschiedene geometrisch hochauflösende (50 cm) Daten der Stadt Oldenburg in Niedersachsen zu kombinieren, um Synergien zwischen den unterschiedlich arbeitenden Sensorsystemen zu nutzen. Aus dem digitalen Geländemodell in Kombination mit den Hyperspektraldaten wird eine Dachmaske mittels Entscheidungsbaum-Klassifikation generiert. Aus den Thermaldaten lässt sich die Temperaturverteilung innerhalb einzelner Hausdächer bestimmen, welche Indizien auf mögliche Wärmebrücken geben. Die Hyperspektraldaten liefern spektrale Informationen über Dachmaterialien. Sie werden mit Hilfe eines Support Vector Machine (SVM) Klassifizierungsalgorithmus ermittelt. Fünf von sechs Dachmaterialien sind klar differenzierbar.

Die digitalen Höhenmodelle, abgeleitet aus Airborne Laser Scanner Daten, dienen in Kombination mit den Hyperspektraldaten der Ermittlung von Dächern, die eine optimale Ausrichtung für die Installation von Solaranlagen aufweisen.

Die zentralen Ziele des Projektes sind die Entwicklung von Werkzeugen zum Erkennen von Temperaturverteilungen, spektrale Unterscheidung verschiedener Dachparameter auf Basis der Hyperspektraldaten sowie die 3D-Erfassung von Gebäuden aus den Airborne Laser Scanner Daten.

1   Introduction

Urban development plays an important role in modern times. Questions related to the limited availability of natural resources and energy consumption develop in all parts of the world. The steady progress of urban sealing has influence on the local climate and hence on our well-being. Remote sensing techniques may assist in obtaining the information required to support decision-making processes to sustain or even improve the quality of our environment. In order to address these issues from a planning point of view, 3D data having a high spatial and spectral resolution are very helpful. Airborne data are mostly suitable for these purposes as a trade-off between expansion of the area and acquisition of small details. Especially when using more than one sensor in a flight, airborne methods become fairly cost efficient and can cover larger areas compared to on-site inspections.

Hyperspectral data can be used to differentiate various urban surface cover types (Helden et al. 2010, Yang 2011). Due to the strong heterogeneity of urban areas, data of a high geometrical resolution are required. For many applications a resolution (ground sampling distance, GSD) of 50 cm or better is desirable. At a coarser resolution (GSD > 1 m), mixed pixels do no longer allow to separate small details. Roessner et al. (2001) used a spectral unmixing algorithm to reduce the problem with DAIS hyperspectral data having a GSD of 7 m, leading to improved classification results compared to standard procedures.

Mori et al. (2008) classified and analysed roof materials in Japan based on a handheld spectrometer using the reflectance between 350 nm – 2,500 nm as a basis. However, only single spots can be handled by this method. Bähr et al. (2005) and Lemp & Weidner (2004) developed an automatic procedure to determine roof parameters from hyperspectral and airborne laser scanner (ALS) data based on segmentation. They distinguish five roof classes. For the classification a partly object-oriented approach was implemented. The classification was based solely on hyperspectral data. New aspects of segmentation and classification were implemented by Lemp & Weidner (2005). Additionally, they used slope information to improve the results of roof classification. Braun et al. (2012) implemented a method for the fusion of hyperspectral and ALS data to improve SVM classification by kernel composition, modifying the one-against-one cascade and taking into account human knowledge on roof geometries.

Building detection and reconstruction have been important topics of photogrammetric research for many years. For instance, Kokkas & Dowman (2006) introduced a semiautomatic technique for building reconstruction by fusing aerial digital imagery and ALS data. Rentsch & Krzystek (2009) used ALS data for a 3D reconstruction of roof ridge lines and roof planes. Rottensteiner et al. (2012) compare and evaluate different methods of building detection and 3D reconstruction from airborne image and laserscanning data. Klärle (2009) used ALS data to infer the optimal locations for photo-voltaic solar panels. Hilling & de Lange (2010) show a web-based application for deriving the solar potential from ALS data.

For the project HiReSens hyperspectral data in the visible to near infrared spectral range were collected to derive roof parameters of the city of Oldenburg. The hyperspectral data have a spatial resolution of 0.5 m. In addition, thermal and ALS data are gathered to address energy-related issues. Furthermore, aerial images (RGB) are used to enable a precise georeferencing and fusion of all data.

Merging of all these different kinds of data results in a vast pool from which useful information can be extracted. Generally one can say that the accuracy of classification results increases considerably by synergy effects since certain classes can be separated more accurately using additional information unless the features are strongly correlated. For the classification a decision tree (Quinlan 1986) and the support vector machine (SVM, Steinwart & Christmann 2008) algorithm are used. The innovative core of this project is the challenging data acquisition:

- high spectral and spatial resolution data,
- a georeferencing accuracy of a few decimetres,
- the fusion of these different data.

The products may serve as additional input data for 3D city models as well as GIS databases.

This paper presents a setup of four different airborne sensors for high spatial resolution remote sensing over urban areas. The data processing is discussed as well as the challenges of combining these high resolution data. From this pool of information some derived products are shown: roof temperature variation, solar potential, and roof material classification.

Two different classification algorithms are applied to derive parameters related to the buildings in a scene. SVM are used to classify up to six different roof materials, whereas a decision tree helps to derive the roof alignment in order to assess the potential of solar energy.

In our study, hyperspectral data are combined with ALS data. In this way we do not only identify roof surfaces which are theoretically suitable for solar power generation, but we identify the surfaces that are really usable for that purpose. These surfaces are often smaller due to tree overhangs. The temperature distribution of roof surfaces which can be an indicator of heat bridges is also visualised.

2   Observations

Different weather conditions are required for data acquisition using airborne hyperspectral, thermal and ALS sensors. For hyperspectral and RGB data acquisition, clear skies and a high sun elevation are desirable. For collecting thermal data the most suitable weather conditions are encountered at night, or in the morning when temperatures are low, without snow or dew, and under a uniform cloud cover. An ALS can be operated under most weather conditions apart from rain, snow, or fog. Also dew prevents a successful data acquisition.

Considering these constraints three flights at different times were performed over the city of Oldenburg, Germany. The total project area size was 3.8 km × 1.8 km. Within this area the work was focused on a core test site of 1.8 km × 1.0 km which hosted the broadest diversity of urban features.

More than 60 GBytes of hyperspectral, thermal and ALS raw data were collected. They were captured using the Cessna 207 aircraft of Milan Geoservice. A system descrip-

tion is given in detail by BANNEHR et al. (2006). The RGB images were taken by Alpha Luftbild using an Aquila A210.

2.1 *Hyperspectral Observations*

The imaging spectrometer AISA+ is a non-cooled system. It serves for monitoring and detection of environmental damage, determination of water constituents of lakes and rivers, forest state examination, and atmospheric research. Up to 244 spectral channels with a bandwidth of 2.5 nm in wavelength can be defined within the spectral range from 400 nm to 980 nm. For the project HiReSens 107 spectral channels with a bandwidth of about 5 nm are used as a trade-off between noise, exposure time and channel bandwidth.

The AISA+ system was radiometrically calibrated by the manufacturer Specim, Finland. According to Specim the calibration accuracy over the spectral range of the imaging spectrometer is about 10 %. This accuracy is not critical for the current investigation because the airborne reference spectra are compared with ground truth spectra. Eleven strips were flown on 16 June 2010 to cover the whole project area. The altitude of 600 m resulted in a GSD of 50 cm.

2.2 *IR Measurements*

The infrared camera, a FLIR SC3000, enables the detection of small differences in temperature. Thus, it is most suitable to detect thermal bridges and energy losses of buildings by looking at temperature variations within house roofs. The infrared camera has a detector size of 320 × 240 pixels. Its spectral response is between 8 μm and 9 μm. Up to 50 images per second can be taken. For the present project the data acquisition rate was set to 10 Hz. This high sampling rate assures a high overlap in the flight direction and hence minimizes the angular effect of emissivity. During the measurement the standard temperature range of the IR camera was set to the range from –20 °C to 80 °C. This temperature range is resolved with 14 bit, which results in a temperature resolution of 30 mK.

The observations were carried out just before sunrise at an altitude of approximately 500 m on 28 April 2010 (GSD: 55 cm). Fifteen flight strips were needed to cover the total project area, nine of which cover the core test site.

The thermal camera was calibrated geometrically in the laboratory with a 3D test field (Luhmann et al. 2011).

2.3 Airborne Laser Scanning

For the second research flight the Riegl LMS Q560 laser scanner and the thermal camera FLIR SC3000 were on the same aircraft. The resulting ALS point density was about 23 points/m^2. These data were rasterized at a spatial resolution of 25 cm.

2.4 RGB Observations

Alpha Luftbild provided the RGB camera Rollei AIC P45 with 39 megapixels and performed the measurement flight on 25 March 2010. Due to their high spatial resolution, the RGB data are used as a reference for the co-registration of the other sensors, in particular for the hyperspectral and the thermal data. Furthermore, the RGB data served as reference information for defining the training areas in the SVM-based classification. The flight altitude of 580 m resulted in a GSD of about 8.5 cm. The forward and side laps were 68 % and 80 %, respectively.

3 Pre-processing

The AISA+ hyperspectral system and the Riegl airborne laser scanner LMS-Q560 were connected to an IGI CCNS4 and an AERO-control GNSS/INS. This system samples the aircraft motion data (roll, pitch, yaw) with 256 Hz and the GNSS data with 10 Hz. The GNSS data is post processed using SAPOS correction data. For the ALS data an accuracy of 0.2 m is achieved both in planimetry and height. After the geometrical co-registration, which will be explained in the subsequent sections, all data were transformed into UTM 32N, WGS 84.

3.1 Rollei AIC-P45

The 56 images of the core region were oriented using 18 manhole covers as ground control points (GCPs). The GCPs were collected using DGPS. More than 700 tie points were used. Two to three points were picked manually for each image. The rest were picked automatically. A digital surface model (DSM) derived from the ALS data (section 3.4) was used to generate an orthophoto mosaic of the whole project area (GSD: 10 cm). This mosaic, which contained double mappings, was used for precise georeferencing of the hyperspectral data. We also generated a DSM of the core area by semi-global matching (Hirschmüller 2008) from the RGB images. The software also delivers a true orthophoto (GSD: 8.5 cm), which we used to generate the reference for the SVM-based classification (section 4.4).

3.2 FLIR SC3000

The surface temperature range during the field experiment was between 3.5 °C and 9.5 °C. This agrees with in-situ measurements of the collected temperature data. The thermal camera was not connected to a GNSS/INS. Therefore, the data were georeferenced by aerotriangulation using ground control and tie points. Because of the very high image overlap, only every fourth image was used. The result is a thermal image mosaic based on the ALS data with a GSD of 50 cm (Fig. 1). Buildings are generally cooler (dark purple) than the surrounding roads and non-built-up areas. The white areas correspond to water. One also can see temperature variations both between and within individual buildings. For a more in-depth analysis it is necessary to extract the thermal features in detail.

The interpretation of thermal images must be done carefully because the measured temperatures depend on the emissivity ε of the roof materials, which are different in general. Variations of ε between 0.80–0.94 and more are very common. This can result in a temperature difference of several Kelvin, depending on the temperature level itself. Without any further information about a building, it is not possible to decide whether the temperature
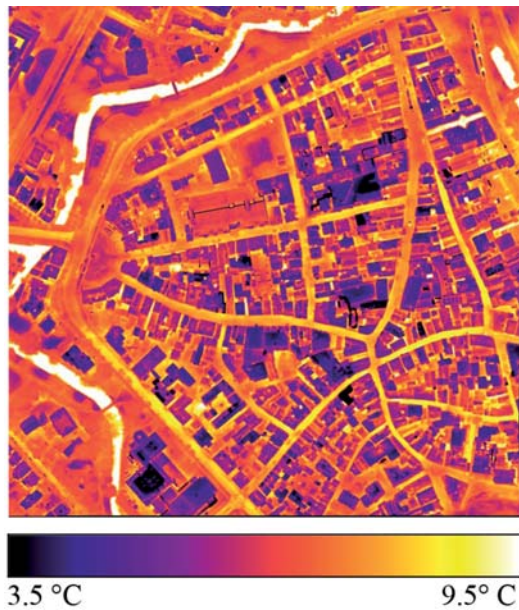
Fig. 1: A 500 m × 500 m subset of the thermal infrared orthorectified image section of Oldenburg (GSD: 50 cm).

distribution is due to the variation of the emissivity or arises from different surface temperatures. In section 4.1, a method to highlight the temperature distribution within the individual buildings is presented.

The spectral band used by the infrared camera (8 µm–9 µm) is within an atmospheric window (8 µm–14 µm). The influence of the atmosphere on the signal is expected to be rather low. Nevertheless, radiative transfer

calculations using MODTRAN (BERK et al. 2006) were carried out for a flight altitude of 500 m assuming a standard continental aerosol size distribution and a subarctic summer atmospheric profile. It was found that neglecting the atmospheric effect may lead to an error of 0.5 K. As this error is rather small and because it is constant over the entire area it was disregarded. It has to be noted that we are mainly interested in the relative temperature distribution, not in the absolute one.

Ground-based in-situ temperature measurements on three different roof locations agree with the temperatures derived from the airborne sensor data within -0.7 K and -4.7 K. Fig. 2 shows the positions where the temperatures were compared; the measurements are shown in Tab. 1. Note that no emissivity was taken into account and the differences in Tab. 1 are mostly due to emissivity of the material and the minor atmospheric effect. The measurements were carried out during the overpass of the aircraft. In addition, continuous air temperature data were recorded. These data show the current air temperature (between 6 °C and 7 °C) from 3 am to 7 am Central European Summer Time. The flight mission took place between about 6 am and 7 am.

3.3 *AISA+*

The processing of the hyperspectral data includes the boresight calibration, radiometric
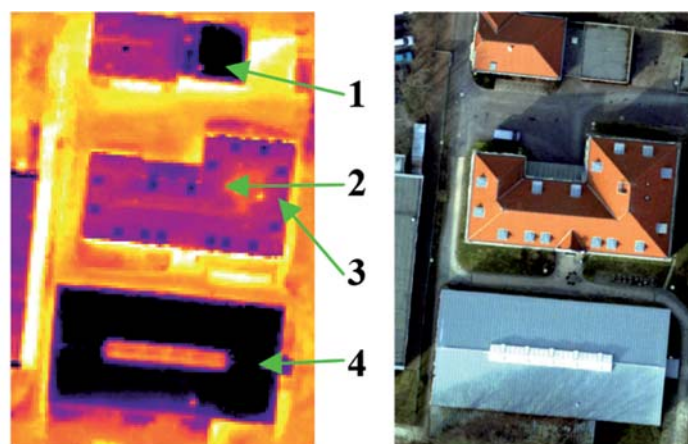

Fig. 2: In-situ temperature measurement points, thermal image on the left side, RGB ortho image on the right. The numbers indicate the measurement points.

Tab. 1: Temperatures at the in-situ measurement points (Fig. 2).

Position	Thermal image (K)	In-situ measurement (K)	Delta (K)
1	275.7	276.5	-0.8
2	278.7	282.1	-3.4
3	279.4	284.1	-4.7
4	275.8	276.5	-0.7

correction, rectification, georeferencing, orthorectification, and mosaicking.

Due to partial cloudy weather conditions it was decided to use the FODIS ratio (Homolova et al. 2009) to represent the reflectance rather than the reflectance derived from the atmospheric correction model. The FODIS ratio is the ratio of the down welling irradiance, measured by the FODIS detector, which is part of the AISA system and mounted on top of the aircraft, and upwelling radiance, measured by the AISA sensor without any atmospheric correction.

Under partial cloudiness the FODIS ratio provides better results than the reflectance derived from an atmospheric correction model, which is based on a radiative transfer model. This is due to the fact that atmospheric correction algorithms always assume clear skies with no clouds. In the case of the AISA+ system operated below clouds, the reflected radiation as well as the downwelling radiation drops rapidly. However, the FODIS ratio will stay almost as constant when having the same surface characteristics. Applying an atmospheric correction model for the reflectance calculation would result in unrealistically small reflectance values.

In order to achieve a high geometric accuracy, the hyperspectral sensor AISA was calibrated using a new procedure developed by the project group (Piechel et al. 2011). It turned out that the results from this calibration cannot be used in the CaliGeo (SPECIM 2010) tool, which is the default processing software delivered with the sensor system. Despite of the fact that the documentation describes the possibility to use a factor for radial distortion, it turned out that the software is not able to calculate useful results when using this factor. The lens distortion error is about 4 pixels

for the outermost pixel of the sensor and still remains in the georeferenced image. Fig. 3 shows the pre-processed FODIS ratio CIR colour image generated from the AISA+ data. Some dark and bright spots within the image are caused by the cloudiness.

To verify the spectral measured data, insitu measurements with a field spectrometer were carried out. The reflectance of different roof tops was collected using a RAMSES VIS SAM-8103 field spectrometer. Some example reflectance curves and FODIS ratio curves are shown in Fig. 4. It has to be noted that the portable spectrometer measurements only had non-calibrated white Teflon as a white reference, and that some of the measurements were taken at a time different from the time the imagery was acquired. Taking into account the diversities of the instrumentation of the measuring setups and of the illumination, a comparison by visual examination indicates a sufficiently good agreement between airborne and reference spectra.

Direct georeferencing of the hyperspectral data resulted in discrepancies between 0.0 m and 2.0 m. At this time, the reasons for the larger discrepancies are subject to speculation. It is likely that residual boresight errors
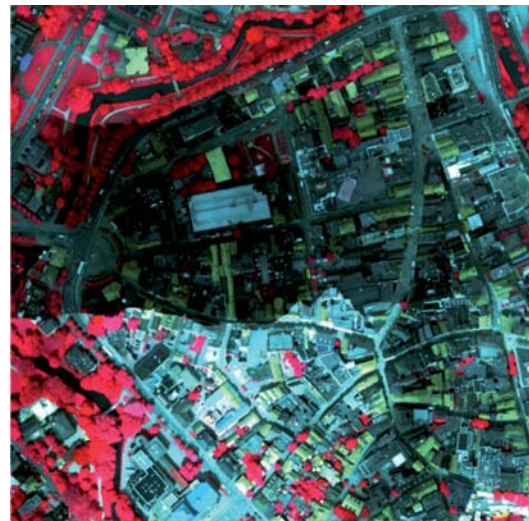

Fig. 3: A 500 m × 500 m subset of the pre-processed CIR reflectance image generated from three AISA+ strips. The spatial resolution is 50 cm. Note the sunny part of the lower strip compared to the other strips.
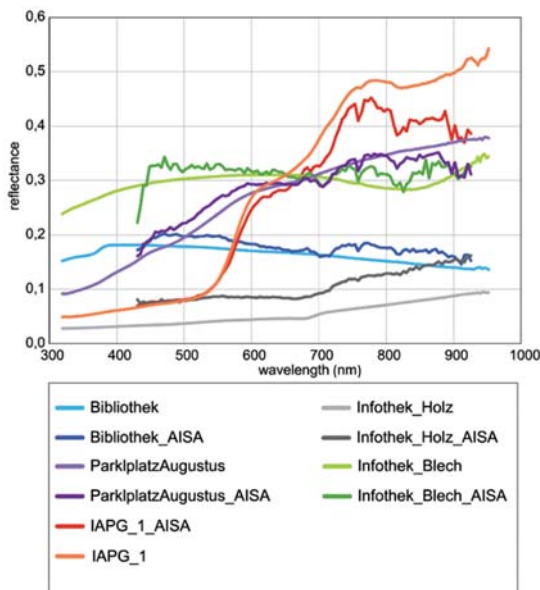
Fig. 4: AISA FODIS ratios (dark) and portable spectrometer reflectance (bright colours). The names indicate the places where the field spectrometer measurements took place.

or time drifts of the AISA+ system may cause the disagreement. In order to compensate these errors the hyperspectral data were co-registered to the RGB ortho-mosaic and rectified by a fourth-degree polynomial rubber-sheeting using hundreds of manually picked control points. The rectified image with a GSD of 50 cm was used as a basis for generating an image corresponding to the normalised difference vegetation index (NDVI), which was one of the inputs for building classification (section 4).

3.4 Airborne Laser Scanner LMS Q560

For airborne laser scanning the positions and orientations are based on GNSS/INS measurements. The accuracy of the airborne laser scanning data strongly depends on an accurate processing of the GNSS/INS data. In order to achieve the highest possible internal precision of the laser data in position and height a number of requirements have to be considered:

a) Accurate determination of the calibration values, i.e. installation offset parameters and boresight angles between scanner and IMU, if possible for each flight session.

b) Use of exact lever arms between GNSS antenna and IMU as well as between IMU and laser scanner.

c) Precise internal calibration of the instrument.

The LMS Q560 allows for collecting full waveform laser data. For the current project only the first/last pulses are needed. The Riegl software enables to extract the first/last pulse data from the full waveform data (Petrie 2011). To derive digital elevation models from airborne laser data a number of processing steps are necessary. At first morphological filtering is carried out automatically using the software package TerraScan (TERRASOLID 2010). The results were visually verified and a locally optimized filter parameter is applied in an iterative process to improve the results. As final products a digital surface model (DSM) and a digital terrain model (DTM) are generated as raster models with a grid width of 25 cm. Homogeneous regions are easier to be classified than city centres. The accuracy also depends on the laser point density, which is about 23 points/m^2 in our test. From the DSM and the DTM, products such as contour lines and cross sections can be inferred.

In addition to the DSM and the DTM, the original point cloud is also available. Fig. 5 shows a shaded relief of the DSM.
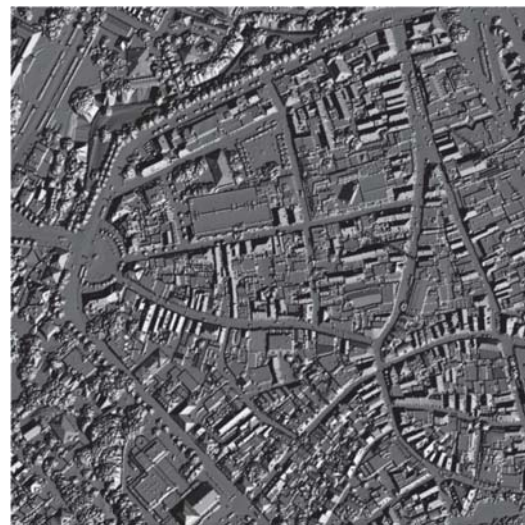

Fig. 5: A shaded relief of a 500 m × 500 m subset of the DSM.

4 Products and Analysis

After pre-processing, various products are derived in the ways described in the subsequent sections.

4.1 *Temperature Distribution within Buildings*

In most cases, the relative differences of the roof temperatures are sufficient to indicate heat bridges or heat losses.

As we were only interested in the temperature distribution on buildings, we first detected buildings by a simple decision tree algorithm. For that purpose, we used the NDVI generated from the hyperspectral data and a normalized DSM (nDSM). The nDSM was calculated by subtracting the DTM from the DSM (nDSM = DSM – DTM). Pixels were considered to correspond to building pixels if the NDVI is smaller than 0.35 and if the nDSM height is greater than 7 m. These thresholds were found empirically. The non-building pixels correspond to the black areas in Fig. 6.

Finally, we visualised the temperature distribution inside the areas detected as buildings using a lookup table with a colour scale with steps of 1 K (Fig. 6). Thus, the potential heat loss can be visualized far more efficiently than with a continuous presentation. One can see that the temperature is not evenly distributed over the buildings. This is caused by emissivity changes of different roof materials or by heat bridges. Generally one speaks of a heat bridge if the temperature differences within an object of constant emissivity are more than 5 K. Fig. 6 shows the temperature distribution within the selected area.

4.2 *Solar Potential*

To assess the solar potential it is important to know the alignment of roofs within a city. In order to estimate the potential of solar energy in general the number of flat roofs and roofs with a certain inclination and orientation as well as the total area of such roofs has to be known.
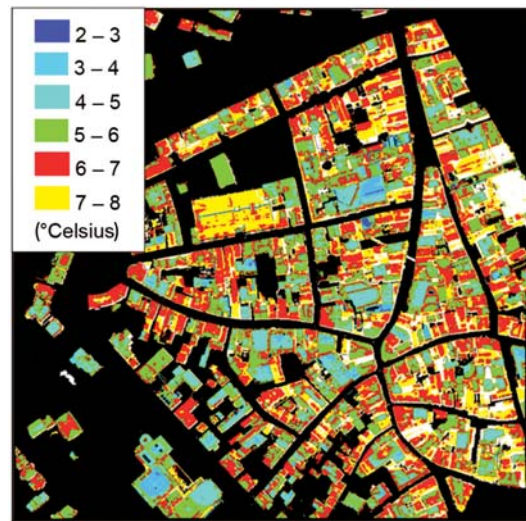

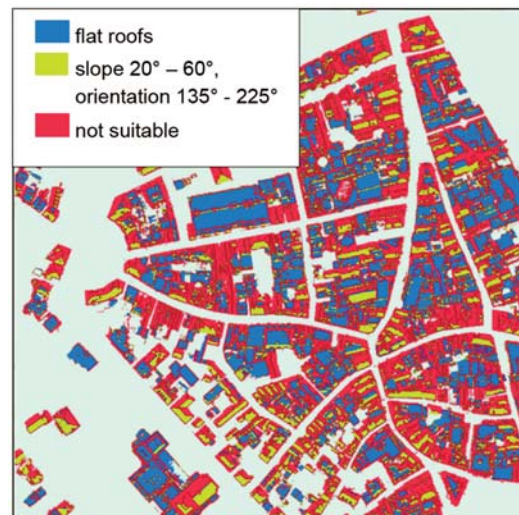**Fig. 6:** Building temperature distribution in a 500 m × 500 m subset of our test site.

Fig. 7: Potential for the use of solar panels of Oldenburg, 500 m × 500 m subset.

If an offset of ± 10 % from the optimal alignment is acceptable then inclinations of 20° to 60° with an azimuth between 135° and 225° (DGS 2012) can be regarded as favourable. These data are valid for Germany and vary slightly depending on the local climate.

The roof orientations were derived from nDSM data. The computations were based on the building mask derived as described in section 4.1. The slopes are computed for all pixels marked as "building" in this mask. The computation is based on a local quadratic surface

fitting of the DSM, taking into account 3 × 3 pixels. Again, a decision tree was used to decide whether a pixel is in an area favourable for solar panels. If the slope was less than 10°, the pixel was regarded as being on a flat roof, and therefore marked in blue. If the slope was between 20° and 60° and the orientation of the roof was between 135° and 225°, the pixel was marked in green. This indicates that it is in an area with favourable conditions for installing solar panels. The pixels marked in red indicate areas that are not suitable. The results of the decision tree classification are presented in Fig. 7. In the inner city only a small number of houses, shown in green, is ideal for installing solar panels. Additionally, buildings with flat roofs, marked in blue, are suitable for panels mounted on stilts.

4.3 Building Classification

The roof classification aims to separate different roof surface materials. The classification consists of two parts. One part is the binary classification of the roofs. The other part is a classification of the surface material of these roofs.

The first step, the roof classification, is based on the rasterized ALS data, the NDVI calculated from the hyperspectral data and the plane normals calculated from the original point cloud. The plane normal is calculated using the robust least median of squares technique for plane fitting using all ALS points within 1×1 m^2 cells. These raster cells are resampled to 50 cm to match the other data.
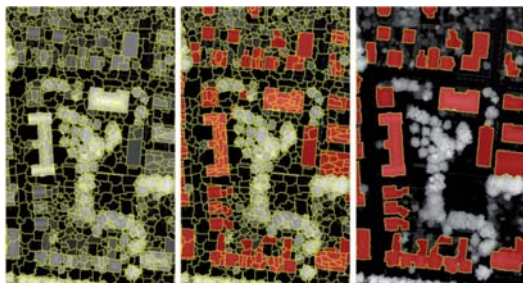

Fig. 8: From left to right: raw segmentation (yellow = segment boundary), initial classification (roofs = red), classification after rule set and merging; the background is the nDSM image.

The nDSM and three raster images representing the components of the normals are used for a segmentation using eCognition's multiresolution segmentation (BAATZ & SCHÄPE 2000). The parameters of the segmentation algorithm were set to achieve an oversegmentation in order to make sure that all the building boundaries are represented by segment boundaries. A larger segment size (higher scale parameter in eCognition) results in a higher risk to miss a building boundary. The left image of Fig. 8 shows the initial segmentation.

Then, the roof segments were classified via thresholds, using the nDSM height (higher than 2.5 m) and NDVI (less than 0.42) as shown in the centre part of Fig. 8. The threshold values used here are different from the ones used before because we are now working on the basis of segments and not on the basis of pixels.

The classification was followed by some refinements using a set of rules for the image segments, e.g. to delete small objects and remove objects with a shape not typical for buildings, e.g. very thin objects with a high percentage of borders to non-roof objects. The resulting building mask is shown in the right part of Fig. 8.

For accuracy assessment, all roofs in an area of 900 m × 300 m were digitized manually from the RGB true orthophoto generated on the basis of a DSM from image matching (section 3.1). The total roof area used for a man-

Tab. 2: Classification accuracy of the binary roof classification.

Ground truth classification	Background (%)	Roof (%)
Background	96.29	8.03
Roof	3.71	91.97

Tab. 3: Producer's and user's accuracy of the binary roof classification.

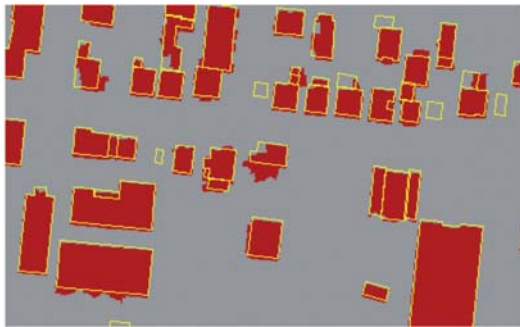| Ground truth classification	Producer's accuracy (%)	User's accuracy (%)
Background	96.29	97.63
Roof	91.97	87.83

Fig. 9: Classified roofs (red) and digitized building outlines (yellow).

ual accuracy assessment is about 74,000 m². About 92 % of all roof pixels were found correctly (for details, see Tabs. 2 and 3). The overall accuracy is 95.3 % and the kappa coefficient is 0.87. Errors result from some completely missed roofs, e.g. low garage-like buildings with heights near the nDSM threshold (2.5 m) and some roofs that are partly hidden by trees. Fig. 9 shows an example area with typical errors. It can be seen that some small buildings and building parts are missing. Errors in the reference also contribute to the overall error budget, but these errors are much lower than the actual classification errors.

4.4 *Roof Material Classification*

For the roof material classification a pixel-based and a segmentation-based approach were combined.

Reference data were generated by visual classification based on the true orthophoto mosaic, the manually digitized roof outlines used as a reference for the evaluation in sec-

tion 4.3 (yellow lines in Fig. 9) and terrestrial photographs of roofs. Only roofs with mostly homogenous material were considered to be suitable as a reference. Thus, we selected as a reference only roofs consisting of only one material and a few (< 15 % in area) disturbing objects like dormers and chimneys as a reference. Within each reference roof, 4 to 20 pixels were manually selected as training pixels so that they completely correspond to the main roof material, to ensure that training pixels represent the correct material.

To reduce the training data and speed up the computation, only a subset of pixels was randomly selected as the final training pixels. Tab. 4 shows the amount of training data selected and used for classification.

In the literature SVM-based classification proved to be very suitable for classifying hyperspectral data (MELGANI & BRUZZONE 2004, PLAZA et al. 2009, WASKE et al. 2009, BRAUN et al. 2012). The reduced training subset was, thus, used to train a SVM for classification, using the software by RABE et al. (2009). A radial basis function (RBF) kernel with the parameter $g = 1$ was used. Both g and soft margin parameter C (= 100) were determined by a 3 fold cross validation. The SVM is applied to classify each pixel independently.

To define a single material per roof, a majority voting is used to transfer the pixel-based classification results to the roof objects. That is, the material of each roof is determined as the material of the majority of all pixels inside that roof. The parts of Fig. 10 show an example with the results from the SVM classification and the final majority voting.

The confusion matrix in Tab. 5 shows the outcome of the classification and Tab. 6 shows the user's and producer's accuracies. The re-

Tab. 4: Training area size and subset size.

Material	Manually selected training pixels	Final training subset
Red roofing tiles	568	200
Black roofing tiles	648	200
Brown roofing tiles	18	18
Metal	62	50
Tar paper	271	150
White colour	31	31

metal   black roofing tiles   brown roofing tiles

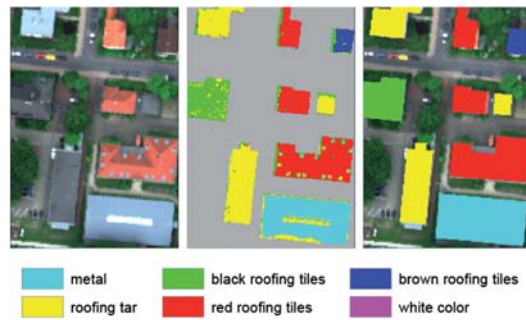roofing tar   red roofing tiles   white color

Fig. 10: Roof material classification. Left: RGB image (from hyperspectral data), centre: SVM classification with background masked in grey, right: results of material classification after majority voting.

Tab. 6: Producer's and user's accuracies of the roof material classification.

	Producer's accuracy (%)	User's accuracy (%)
Red roofing tiles	97.1	98.6
Black roofing tiles	97.8	93.5
Brown roofing tiles	50.0	100.0
Roofing tar	89.3	96.2
White colour	100.0	100.0
Metal	100.0	100.0

sults for all materials except brown roofing tiles are very promising. The poor results for brown roofing tiles have to be analysed with care because of the very low sample size and seem to be caused by a high spectral overlap with the black and red roofing tiles.

The SVM classification works well when using a sufficient number of training data. The training data need to cover all possible characteristics of a roof material. In this way, one can be certain to get the best support vectors for defining the class boundaries. Additionally, it can be confirmed that the SVM classification works quite well with noisy data. The mean signal-to-noise ratio of the hyperspectral data was only about 100, averaged over all spectral channels and all spatial pixels.

5   Conclusions and Outlook

From different kinds of remote sensing data, products such as temperature distributions of roof tops and maps of the suitability of roofs for installing solar panels were derived by first detecting buildings based on a simple decision tree and further analysis steps. From hyperspectral data the roof material characteristics are inferred using a SVM-based classification. The fusion of different datasets makes it possible to obtain more information by synergic use of the derived products. It is important to note that a precise georeference of all data is the basis for reliable results.

Regarding the roof and material classification, additional refinements could be done using the slope per roof plane as an additional feature. This may help because roofing tiles are only used on sloped roofs.

Tab. 5: Confusion matrix of the roof material classification. The numbers correspond to the number of roof polygons, not to pixels.

Ground truth classification	Red	Black	Brown	Tar	White	Metal	Total
Red roofing tiles	67	1	0	0	0	0	68
Black roofing tiles	2	87	1	3	0	0	93
Brown roofing tiles	0	0	1	0	0	0	1
Roofing tar	0	1	0	25	0	0	26
White colour	0	0	0	0	1	0	1
Metal	0	0	0	0	0	5	5
Total	69	89	2	28	1	5	194

Some georeferencing errors remaining in the hyperspectral data lead to misclassifications at the building boundaries. In addition, some misclassifications are induced by different solar illumination angles depending on the roof alignment and on the recording time. The reduction of these errors is a part of the ongoing research.

A further goal is to combine the complete roof mask with the SVM classification to obtain the material per roof for the whole area.

Acknowledgements

References

Baatz, M. & Schäpe, A., 2000: Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. – Strobl, J., Blaschke, T. & Griesebner, G. (eds): Angewandte Geographische Informations-Verarbeitung **XII:** 12–23, Wichmann Verlag, Karlsruhe.

Bähr, H.-P., Lemp, D. & Weidner, U., 2005: Hyperspectral meets Laserscanning: Image Analysis of Roof Surfaces. – ISPRS Hannover Workshop, High-Resolution Earth Imaging for Geospatial Information.

Bannehr, L., Hannusch, D., Jany, S. & Runne, H., 2006: Komplexes modulares System zur Ableitung von Umweltparametern. – Seyfert, E. (Hrsg.): **26.** Wissenschaftlich-Technische Jahrestagung der DGPF: 295–302.

Berk, A., Anderson, G.P., Acharya, P.K., Bernstein, L.S., Muratov, L., Lee, J., Fox, M., Adler-Golden, S.M., Chetwynd, J.H., Hoke, M.L., Lockwood, R.B., Gardner, J.A., Cooley, T.W., Borel, C.C., Lewis, P.E. & Shettle, E.P., 2006: MODTRAN5: 2006 Update. – SPIE **6233** (62331F).

Braun, A.C., Weidner, U., Jutzi, B. & Hinz, S., 2012: Kernel Composition with the one-against-one Cascade for Integrating External Knowledge into SVM Classification. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2012** (4): 371–384.

DGS, 2012: Leitfaden Solarthermische Anlagen. – 9. Auflage 2012, Deutsche Gesellschaft für Sonnenenergie (Hrsg.), ISBN 978-3-9805738-0-1.

Homolova, L., Alanko-Huotari, K. & Schaepman, M.E, 2009: Sensitivity of the ground-based downwelling irradiance recorded by the FODIS sensor in respect of different angular positions. – First IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Grenoble, France.

Heldens, W., Heiden, U., Esch, Th. & Dech, S., 2010: Potential of Hyperspectral Data for Urban Micro Climate Analysis. – Hyperspectral Workshop 2010, Frascati, Italy.

Hilling, F. & de Lange, N., 2010: Webgestützte interaktive Solardachkataster. Ein Instrument zur Darstellung der Nutzungseignung von Dächern für Photovoltaikanlagen am Beispiel der Stadt Lage. – Standort **34** (4): 104–109.

Hirschmüller, H., 2008: Stereo Processing by Semiglobal Matching and Mutual Information. – IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2): 328–341.

Klärle, M., Ludwig, D. & Lanig, S., 2009: SUN-AREA – Ein Beitrag der Fernerkundung gegen den Klimawandel. – ZfV **2009** (2): 71–78.

Kokkas, N. & Dowman, I.J., 2006: Fusion of Airborne Optical and LIDAR Data for Automated Building Reconstruction. – ASPRS Annual Conference, Reno, USA.

Lemp, D. & Weidner, U., 2004: Use of Hyperspectral and Laser Scanning Data for the Characterization of Surfaces in Urban Areas. – **XXth** ISPRS Congress, Commission VII.

Lemp, D. & Weidner, U, 2005: Improvements of Roof Surface Classifications Using Hyperspectral and Laser Scanning Data. – IAPRSIS **XXX-VI-8/W27**, ISPRS Joint Conferences, Tempe, USA.

Luhmann, T., Ohm, J., Piechel, J. & Roelfs, T., 2011: Geometrische Kalibrierung von Thermografiekameras. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2011** (1): 5–15.

Melgani, F. & Bruzzone, L., 2004: Classification of Hyperspectral Remote Sensing Images with Support Vector Machines. – IEEE Transactions on Geoscience and Remote Sensing **42:** 1778–1790.

Mori, M., Iwata, T., Minami, Y., Kato, S. & Akamatsu, Y., 2008: Spectral Analysis of Building Materials Used in Japan. – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **XXXVII** (B8), Beijing, China.

PETRIE, G., 2011: Airborne Topographic Laser Scanners. – GeoInformatics **1:** 34–44.

PIECHEL, J., BANNEHR, L., LUHMANN, T., ROELFS, T. & SCHMIDT, A., 2011: Geometrische Kalibrierung von Hyperspektralsensoren. – DGPF-Jahrestagung, Mainz.

PLAZA, A., BENEDIKTSSON, J., BOARDMAN, J., BRAZILE, J., BRUZZONE, L., CAMPS-VALLS, G., CHANUSSOT, J., FAUVEL, M., GAMBA, P., GUALTIERI, A., MARCONCINI, M., TILTON, J.C. & TRIANNI, T., 2009: Recent Advances in Techniques for Hyperspectral Image Processing. – Remote Sensing of Environment **113**: 110–122, Elsevier.

QUINLAN, J.R., 1986: Induction of decision trees. – Readings in Machine Learning **1** (1): 81–106.

RABE, A., VAN DER LINDEN, S. & HOSTERT, P., 2009: imageSVM, Version 2.0, software available at www.hu-geomatics.de (18.3.2013).

RENTSCH, M. & KRZYSTEK, P., 2009: Lidar Strip Adjustment Using Automatically Reconstructed Roof Shapes. – ISPRS Workshop High-Resolution Earth Imaging for Geospatial Information, Hannover.

ROESSNER, S., SEGL, K., HEIDEN, U. & KAUFMANN, H., 2001: Automated Differentiation of Urban Surfaces Based on Airborne Hyperspectral Imagery. – IEEE Transaction on Geosciences and Remote Sensing **39**: 1525–1532.

ROTTENSTEINER, F., SOHN, G., JUNG, J., GERKE, M., BAILLARD, C., BENITEZ, S. & BREITKOPF, U., 2012: The ISPRS Benchmark on Urban Object Classification and 3D Building Reconstruction. – ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **I** (3), XXII ISPRS Congress Melbourne.

SPECIM, 2010: CaliGeo 4.9.9 AISA Data Processing Tool. – Operating Manual. Spectral Imaging Ltd., Oulu, Finland.

STEINWART, I. & CHRISTMANN, A., 2008: Support Vector Machines. – 602 p., Springer, New York.

TERRASOLID, 2010: TerraScan User's Guide. – Arttu Soininen, Finland.

WASKE, B., BENEDIKTSSON, J., ÁRNASON, K. & SVEINSSON, J., 2009: Mapping of Hyperspectral AVIRIS Data Using Machine-learning Algorithms. – Canadian Journal of Remote Sensing **35**: 106–116.

YANG, X. (ed.), 2011: Urban Remote Sensing: Monitoring, Synthesis and Modelling in the Urban Environment. – 408 p., Wiley-Blackwell.

Addresses of the Authors:

LUTZ BANNEHR & ANDREAS SCHMIDT, Hochschule Anhalt, Institut für Geoinformation und Vermessung, Bauhausstraße 8, D-06846 Dessau-Roßlau, Tel.: +49-340-5197-1551, e-mail: {l.bannehr}{and.schmidt}@afg.hs-anhalt.de

THOMAS LUHMANN & JOHANNES PIECHEL, Jade Hochschule Oldenburg, Institut für Angewandte Photogrammetrie und Geoinformatik, Ofener Straße 16, D-26121 Oldenburg, Tel.: +49 441-7708-3172, e-mail: {thomas.luhmann}{johannes.piechel}@jade-hs.de

Report

Establishing and implementing a national 3D Standard in The Netherlands

Jantien Stoter, Delft/Apeldoorn/Amersfoort, Hugo Ledoux, Delft, Marcel Reuvers, Linda van den Brink, Rick Klooster, Paul Janssen, Amersfoort, Jakob Beetz, Eindhoven, Friso Penninga, The Hague & George Vosselman, Enschede, The Netherlands

Summary: This paper describes the 3D developments achieved within the 3D Pilot NL. The first phase of this pilot (January 2010 – June 2011) resulted in a national 3D standard, modeled as CityGML application domain extension (ADE). This standard is briefly explained in this paper. To implement this standard as a nationwide 3D dataset, further research was needed. The second phase of the 3D Pilot finished in December 2012 developed tools, techniques and guidelines to support the implementation of the 3D standard. These are: 1) implementation specifications for the national City-GML ADE to be used in tendering documents, 2) example data compliant to the 3D standard, 3) 3D validator, 4) guidelines to update 3D datasets, and 5) 3D application showcases. These instruments are further explained and presented in this paper.

Zusammenfassung: *Entwicklung und Implementierung eines nationalen 3D Standards in den Niederlanden*. Der Beitrag beschreibt den aktuellen Stand des niederländischen Projektes 3D Pilot NL und zugehörige Entwicklungen zur 3D-Datenmodellierung. Die erste Phase des Projektes führte von Januar 2010 bis Juni 2011 zu einem nationalen 3D Standard, der als Application Domain Extension (ADE) von CityGML modelliert wurde. Für seine Implementierung und Anwendung auf einen landesweiten 3D-Datensatz waren weitere Untersuchungen erforderlich. In der im Dezember 2012 beendeten zweiten Phase des 3D Pilot NL wurden Hilfsmittel für die Implementierung entwickelt: 1.) eine Implementierungsrichtlinie der nationalen Application Domain Extension (ADE) für City-GML zur Verwendung in Ausschreibungen, 2.) Beispieldatensätze, 3.) ein 3D Validator für die Konsistenzprüfung von Datensätzen, 4.) eine Richtlinie für die Fortführung der 3D Daten und 5.) Anwendungsbeispiele. Diese Werkzeuge werden hier vorgestellt.

1 Introduction

This paper describes the definition, establishment and implementation of a national 3D standard in The Netherlands, which was accomplished within the 3D Pilot NL. The Kadaster, Geonovum, the Dutch Geodetic Commission and the Ministry of Infrastructure and Environment initiated the pilot in 2010. The aim of 3D Pilot NL is to push 3D developments in the Netherlands by collaborating with a wide variety of stakeholders on a test bed, a test area and use cases. The 3D Pilot community consists of almost 600 people coming from over 100 organisations.

The activities of the 3D Pilot cover two phases: an explorative phase, carried out between January 2010 and June 2011 and an implementation-oriented phase, finished at the end of 2012.

The major achievement of the first phase was the establishment of a national 3D standard. This standard is embedded in a formal information model, called "Information Model Geography" (IMGeo). IMGeo contains definitions of 2D large-scale representations of

objects as roads, water, land use/land cover, bridges and tunnels. As IMGeo is modelled as an application domain extension (ADE) of the OGC 3D standard CityGML (OGC 2012). The information model facilitates extensions to 2.5D, i.e. surfaces, equivalent to CityGML LOD0, and 3D, i.e. volumetric representations, CityGML LOD1, LOD2 and LOD3 of the objects according to the geometric and semantic principles of CityGML. See Van den Brink et al. (2012a, 2013) for details.

The advantage of this approach is that firstly the 3D standard builds on 2D efforts, which makes 3D feasible for governmental organisations. Secondly, 2.5D and 3D geometries can be combined in one dataset and depending on the application the most appropriate geometry type per feature class can be chosen. For example, hydrological modelling in urban areas may require accurate 2.5D geometry descriptions of the terrain while block models suffice for the buildings (Fig. 1).

Although the 3D standard is an important prerequisite for 3D applications, wide use of 3D is still not common practice in the Netherlands. The implementation of the 3D standard requires further agreements. This also covers agreements on how to implement CityGML. CityGML allows freedom in its implementation, while a national standard needs clear implementation rules. Therefore, the second phase of the 3D Pilot developed a set of instruments to support the implementation of the national CityGML ADE.

This paper focuses on the second phase of the 3D Pilot and describes the implementation tools and documents that have been developed (section 2). More details on the 3D Pilot can be found in Stoter et al. (2013) and Stoter et al. (2011). The paper ends with conclusions and future research in section 3.

Many others have done significant work in this area. For example North-Rhine Westphalia in Germany has been the first state that provides a statewide 3D model consisting of 3D roads, railways, DTM, and 3D buildings employed for noise dispersion mapping (Czerwinski et al. 2006, 2007). The extension of the German national cadastre model ALKIS towards 3D building models is another important CityGML implementation. In this project a CityGML profile has been defined and most states in Germany provide 3D building models according to this profile. In addition, the INSPIRE data specifications for buildings contain a 3D building profile in line with the CityGML specifications for buildings. Finally, different application domain extensions have been developed. Examples are noise as used in the German state North-Rhine Westphalia and as documented within the CityGML specification, utility networks (Becker et al. 2013), real estate management, robotics, building information models (BIM), and hydrography (CityGML 2013).

This paper extends these previous works, because the unified modeling language (UML) concept for ADEs is not well described in the OGC specifications and experiments on implementing such CityGML extensions are new. Nevertheless, ADEs are and have been specified without UML. Therefore, an important contribution of this project is gaining insights in the procedure of specifying CityGML-profiles with the help of UML to make the standard suitable for a specific context. In addition, the paper may be considered to address a solution, which is limited to one specific country. However, the presented implementation has many global solutions, which are of great interest both for other countries and domains. The wide focus of the CityGML implemen-

Fig. 1: 3D IMGeo for hydrological modelling.

tation is also new. Current CityGML models focus mainly on buildings, while the 3D standard in this paper covers all topographic classes. Therefore, the study of feature classes other than buildings is another contribution. Finally, many vendors from abroad are among the stakeholders involved in this research projects. The list of participants is published at GEONOVUM (2013a). Therefore, the experiences of this pilot are not only of interest for a specific country. Instead, the 3D Pilot activities have contributed to developments and growing insights worldwide. This is why the delivered instruments and documents have been offered as best practice and discussion material to the OGC working group on CityGML. See for example VAN DEN BRINK et al. (2012b).

2 Implementation of the 3D Standard

After it was decided that the national 3D standard should align to both the national 2D information model on large-scale topography (IMGeo 1.0) and the international standard CityGML, the integration of IMGeo and CityGML into IMGeo version 2.0 was the next step, currently evolved into version 2.1 because of minor changes. This was realized by modelling all IMGeo classes as an extension of CityGML classes. In this process the semantics of CityGML was followed as much as possible. Not for all classes in the national model, an equivalent CityGML class could be found. If possible, this was solved by remodelling the classes in the national standard. Examples of remodelling national concepts are the CityGML class AuxiliaryTrafficArea applied for those parts of roads that are not used for traffic but formerly modelled as roads, and the class Vegetation, formerly modelled as CityFurniture for isolated trees and Land Use for plant cover areas. Often the remodelling resulted in an improved modelling of the national concepts. More details on how IMGeo 2.1 was modelled as an Application Domain Extension of CityGML with a UML-based approach can be found in VAN DEN BRINK et al. (2013).

After establishing the 3D standard, the next step was its implementation. This refers to a generation of data according to the standard as an extension of the existing 2D data. For this implementation more research was required to understand how the national 3D standard works in practice including the consequences of this new modelling method for IMGeo when applied to both 2D and 3D datasets: How can 2D LOD be upgraded to LOD0 and higher? How can the standard-compliant 3D data be validated and maintained? Also, more insight was needed to the ability of using 3D IMGeo data in CityGML-aware software packages: Are the software systems compatible with the national extensions and which changes are necessary to support the extensions?

These questions have been studied in the second phase of 3D Pilot NL with the main goal to achieve wide consensus on the implementation of the national 3D standard and to explain these agreements to the wider public. The results of this part of the Pilot are ready-to-use instruments available as a toolkit at GEONOVUM (2012a) that data producers can apply to extend existing 2D IMGeo data into 3D, to maintain it and to use the 3D data in applications.

The ready-to-use instruments, which are detailed in the following subsections, are:
- Implementation specifications for creating 3D data compliant with the national 3D standard;
- Example 3D IMGeo data for several levels of detail and classes;
- 3D validator;
- Guidelines for maintenance, update and dissemination of 3D IMGeo data;
- A web site that collects 3D showcases and portrays these to the wider public to inspire newcomers.

2.1 Implementation Specifications for Standard-compliant 3D Data

The national 3D standard IMGeo needs agreements on the precise implementation of the generic standard CityGML. These additional agreements both clarify and unify the demand for the national 3D data and assure a country-wide uniform 3D dataset.

The 3D Pilot participants assumed that these agreements would also be important in tendering processes, since they assure that expectations of governmental organisations (who will mostly outsource their data acquisition) and companies (who will acquire the data accordingly) are aligned. In addition, most governmental organisations lack experience with 3D data and therefore it will be difficult for them to specify the expected 3D deliverables. For those municipalities, the implementation specifications may serve as an important source for their tendering documents. Also, the precise specifications can be used as acceptance criteria once the data is delivered.

With these objectives in mind, the 3D Pilot defined implementation specifications for 3D IMGeo, i.e. CityGML data. These specifications contain data requirements for all IMGeo-CityGML feature types at different levels of details GEONOVUM (2012b, 2012c). For the buildings, the "Modellierungshandbuch Gebäude" of SIG 3D in Germany has been used as a basis, see SIG 3D (2012b, 2012c).

The implementation specifications have been established after public consultation and also include a description of how each requirement can be checked. The different requirements defined are:

- Generic requirements, e.g. which version of CityGML to use, which reference system.
- Requirements for LOD0 representations of 2D IMGeo polygons.
- Requirements for volumetric representations:
  - LOD0-LOD1-LOD2 Buildings,
  - LOD1-LOD3 Bridges, Tunnels,
  - LOD1-LOD2 Vegetation,
  - LOD2-LOD3 Trees (CityGML-Solitary-VegetationObject) and Cityfurniture.
- Requirements for texture.

The specifications provide precisely defined choices regarding the 3D product and explain the implications of each. A few examples are:
- Representation of vertical surfaces in TIN-type digital terrain models. Most GIS systems do not accept vertical faces.
- Necessity of all (planar) objects above the ground to have a LOD0 representation, including those above surface level, e.g. multilevel crossings.
- Usage of the LOD2 representation of buildings as the base 2D geometry? The laser point data do not always match the 2D footprints, resulting in 'strange' geometries (Figs 2a and b).
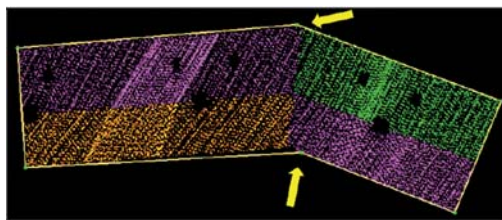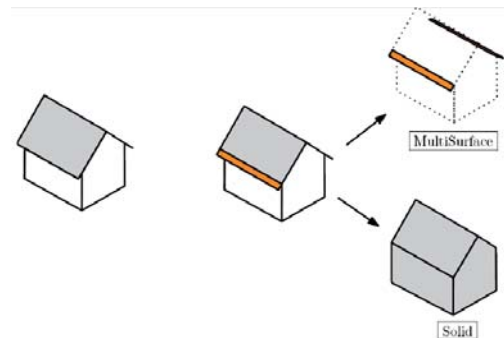

a
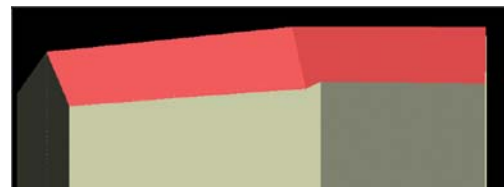
b

c

c

Fig. 2: CityGML implementation choices and consequences. a, b: height point data used to obtain roof-shapes do not match footprints; c, d: modelling roof-edges, footprints and roof-overhangs.

- Representation of building roof-edges, footprints and roof surfaces in LOD2 (Figs 2c and d).
- Inclination of building footprints: Shall the footprints always be horizontal as in the CityGML specifications or are they allowed to be inclined?
- Reasoning for the preference of representing a LOD2 building with a GML:Solid over a GML:Multisurface.
- Representation of curved surfaces. Arcs are allowed in 2D, but are not supported in TINs.
- Significance of the influence of tree-data upon the overall volume of the data.
- Requirements of aerial photographs necessary to automatically obtain height points from image matching.

As can be understood from these choices, several variants of implementations are possible based on the available source data, i.e. point clouds or high resolution photographs, and the ambition level of the data, i.e. the class-dependent LOD. These variants are described in the resulting document and the preferred is given. Since the variants are dependent on the intended use, a specific chapter is dedicated to the requirements in relation to the 3D applications in which the 3D data will be used.

Apart from the experiences in the pilot, the implementation specifications are based on experiences of cities that have invested in 3D city models in the CityGML format, i.e. The Hague and Rotterdam. Both cities faced difficulties in comparing offers from different companies because the specifications in the tendering documents appeared to be interpretable in several ways. This also caused problems in setting up acceptance criteria for the delivered product. Consequently, the CityGML datasets differ between the two cities but it is not always clear whether this was intended. The jointly defined implementation specifications will help to avoid similar situations in the future.

The resulting documentation is of interest for an international public as it serves as a further explanation and refinement of the CityGML specifications. In addition, putting it in an international discussion gives us the oppor-

tunity to evaluate our implementation specifications. Therefore, the document has been translated into English (Geonovum 2012c). Also, this document resulted in a few change requests for CityGML, which are currently discussed in the OGC, for example, the restriction that building footprints and roofedges should be horizontal.
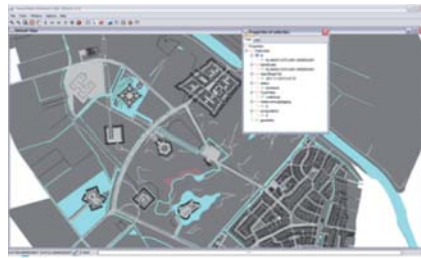
2.2 *Example 3D IMGeo Data*

To understand how IMGeo works for the integrated 2D and 3D approach, example 3D IMGeo data has been built and made available to the public. The objective of the example data is to help newcomers to understand the national 3D standard including the details of the different levels of detail. The example data also serves as source for (new) parties to experiment with 3D IMGeo data and it served as test data in the development of the 3D validator (see section 2.3). Finally, the example CityGML-IMGeo data was used to check whether CityGML compliant software is capable of understanding the 3D IMGeo data (see section 2.4).

The test area for which the example data was built, is situated in the municipality of Den Bosch (southern part of The Netherlands) containing a rural area, a residential area with common houses, a river and a bridge.

The 3D IMGeo data have been generated from 2D IMGeo, i.e. CityGML data (Fig. 3a) and 3D source data. The 3D source datasets that were made available on the 3D Pilot data server, hosted by the Delft University of Technology are:

- stereo photos, provided by the municipality of Den Bosch,
- high-resolution laser data, i.e. *Actueel Hoogtebestand Nederland* (AHN) by Het Waterschapshuis, with an average density of 10 points per m^2, available for the whole country,
- orthophotos by Cyclomedia,
- a high resolution point cloud obtained from terrestrial laserscanning by Cobra, see Fig. 3b,
- point clouds generated from aerial photographs by Imagem,
- oblique photos by Slagboom & Peeters.

a: 2D IMGeo data
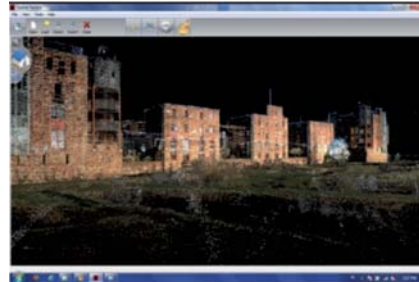
b: High resolution laser data, obtained by terrrestrial laserscanning

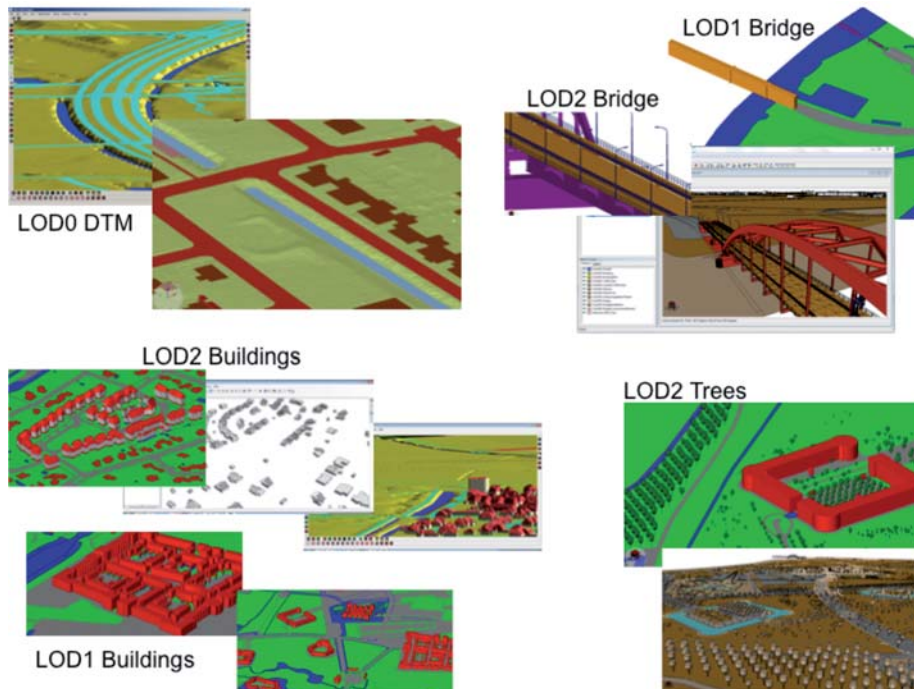**Fig. 3:** Example source data available for the 3D Pilot test area.

Fig. 4: Example 3D data compliant with the national CityGML ADE, generated in the research.

The following example 3D IMGeo-City-GML data (available at GEONOVUM 2012a) were constructed from this data, most of them automatically, see Fig. 4:

1. LOD0 Digital Terrain Model, i.e. triangulated surfaces for all polygonal objects that form together a topologically correct data structure, by the Universty of Twente.
2. LOD1 and LOD2 Buildings, by the University of Twente.
3. LOD1 and LOD3 Bridge by the company Coenradie.
4. LOD2 trees by the University of Wageningen.

The geometries for the different feature types and levels of detail were derived by various methods.

The LOD2 buildings with roof-shapes were automatically reconstructed with the method of OUDE ELBERINK & VOSSELMAN (2011). LOD1 buildings are automatically generated by extruding the 2D footprints with the mean height calculated from the height points within the polygon. A filtering-technique is applied to exclude the outliers.

The LOD3 Bridge was captured by terrestrial measurements and manual modelling

while the LOD1 representation is a simplification of the LOD3 bridge model.

The LOD2 trees were automatically generated from AHN2 by the method of CLEMENT (2011).

LOD0 (surface) representations for all classes were automatically generated from a combination of the 2D IMGeo data and AHN2. The LOD0 representation of the data is more than a drape of the 2D data over the digital terrain model. It consists of a triangulated surface per polygonal object, generated by means of a constrained triangulation with the polygonal boundaries as constraints. All these triangulated surfaces together form a topological surface of the terrain as in 2D with height variances modelled within a polygon. If necessary, extra points are added on the boundaries to better model the height variance.

The reconstruction of a topologically correct LOD0-IMGeo surface is not trivial. The triangulated surfaces are reconstructed for each polygon separately based on the height points that fall within that polygon. Consequently, gaps may occur at polygon boundaries that are neighbours in 2D but not necessarily in 3D. The approach of OUDE ELBERINK (2010) was followed to fill these gaps. In this approach, heights on polygon boundaries are adjusted depending on the types of neighbours. The result is a closed 2.5D surface.

After the geometries were created following the several approaches, the Karlsruhe Institute of Technology in Germany organized the geometries according to CityGML and assigned CityGML semantics to the features. Finally, the data was validated with the developed 3D validation software (see section 2.3).

The main conclusion from this activity is that many 3D source data is available in the Netherlands as well as significant knowledge on these data. However, there is hardly any company which is capable of restructuring the once captured 3D data according to the CityGML structure. To further help in this process, an automated workflow for 3D IMGeo data reconstruction is currently being developed. This workflow available as a FME workbench and as an Open Source tool starts from 2D IMGeo, i.e. CityGML data and high-resolution point data, i.e. AHN2, and generates the LOD0, LOD1 and LOD2 geometries according to the above-mentioned methods and writes CityGML-IMGeo at the end of this process. The workbench implements the algorithms for 3D reconstruction of OUDE ELBERINK (2010), OUDE ELBERINK & VOSSELMAN (2011) and OUDE ELBERINK et al. (2013). Also this workflow will soon be available for the wider public.

2.3 *3D Validator*

A validator is necessary as an independent tool to verify whether a dataset is compliant with IMGeo 2.1 or not. When validating objects, it is necessary to validate both their semantics and the geometry, the former according to the classes of CityGML and/or of the IMGeo extensions, and the latter according to the international specifications. A validator for 2D IMGeo (and thus 2D IMGeo-CityGML) already exists and is available as open source software (GEONOVUM 2012d). However, it only checks for two-dimensional primitives.

Therefore, the 3D Pilot studied the required functionalities to validate the geometry of 3D solids and developed a 3D validator accordingly. During this study, we observed that several real-world datasets have objects that appear to be visually valid, but in fact are wrongly built. Fig. 5 shows two examples. These wrong issues are often small and hard to be detected
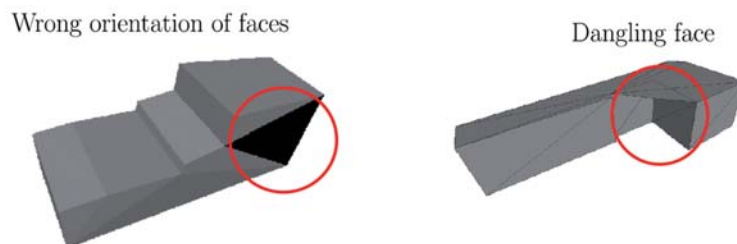

Fig. 5: Two real-world invalid buildings.

at the usual visualization scale of city models. However, they cause problems in practice when for instance converting objects to other formats including BIM and CAD, or when analysing them. The volume of an invalid solid could impossibly be calculated. It may even cause a program crash.

While different definitions of a valid 3D object are used in different disciplines, the developed 3D validator focuses on the definition given in the ISO 19107 standards (ISO 2003) and implemented with GML (OGC 2007). A GML Solid is defined as follows: "The extent of a solid is defined by the boundary surfaces as specified in ISO 19107:2003. gml:exterior specifies the outer boundary, gml:interior the inner boundary of the solid" (OGC 2007). Without going into details, we can state that a solid is represented by its boundaries (surfaces), and that, like its counterpart in 2D (the polygon), a solid can have "holes", i.e. inner shells or cavities, that are allowed to touch another hole or the outer boundary, under certain conditions. To be considered a valid solid, a solid must fulfil several properties. The most important are: 1) it must be simple (no self-intersection of its boundary). 2) It must be closed, or "watertight". 3) Its interior must be connected. 4) Its boundary surfaces must be properly oriented. 5) Its surfaces are not allowed to overlap each other.

It should be pointed out that the development of a new 3D validator was necessary since none of the surveyed GIS packages was fully compliant with the definition of the ISO: Often a more restrictive definition of a solid is used. For instance, Bogdahn & Coors (2010) and Wagner et al. (2013) discuss the validation of solids for city modelling, but do not consider holes in surfaces and totally omit that interior shells are possible. Gröger & Plümer (2011) give axioms to validate 3D city models, but also do not consider holes in primitives of dimensions 2 and 3. In fact, they define solids as shells without holes in their surfaces. We have also noticed that commercial GIS software products ignore interior shells. ESRI (with ArcGIS) and Bentley are two examples. Oracle Spatial considers interior shells in its validation function, but does not allow holes in surfaces.

Our 3D validator does not contradict the previous results and implementations, but simply extends them so that solids are validated against the international standards. It uses advanced data structures and operations to analyse the topological relationships between 3D objects. It does not only validate solid geometries for LOD2 buildings, but also 3D MultiSurfaces that together form a 3D volume as these are often used in practice. However, it gives a warning to the user that the stricter solid-geometry is preferred. The validator is available as an open source tool (Geonovum 2012e) and is also implemented as standard functionality in FME by Safe software.

2.4 *Maintaining and Updating 3D IMGeo Data*

After the investment in a good 3D model, the maintenance and update of the model become the key questions. Can commercially available database management systems (DBMSs) be used? How shall the update process be organized? Shall it be integrated in the existing processes, shall the 3D data be recreated after a change of the 2D data changes or shall both methods be mixed?

The 3D Pilot identified one open source, i.e. 3D CityDB (3DCityDB 2013), and two com-

Fig. 6: Screenshots of StrateGis solution for the CityGML-competition.

mercial solutions, CPA Systems and M.O.S.S., for maintaining 3D IMGeo-CityGML data in DBMSs (Oracle and PostGIS).

In addition to test tools for maintaining the CityGML data, two competitions were organized. In the first competition, six companies, i.e. Bentley, CPA systems, M.O.S.S., Safe Software, StrateGis and Toposcopie, proved that they are able to edit and store the updated version of existing CityGML data of The Hague (Fig. 6).

In the second competition, four vendors, i.e. Bentley, M.O.S.S., Safe Software and CPA Systems, executed several tests (published at Geonovum 2012f) on the 3D IMGeo example data that has recently become available. The four companies submitted a video in which they showed the tests followed by a validation process at the end of the process. These experiments proved that software that supports City-GML is also able to recognize and deal with IMGeo-data as IMGeo is an extension of City-GML. This is important for the acceptance of 3D IMGeo in the Netherlands. The compiled video that summarises these four company results is available at Geonovum (2012g).

From the experiences of this activity it can be concluded that for many municipalities a hybrid approach for updating an existing 3D model will work best, i.e. periodically automated updates for larger areas combined with manual updates in specific project areas. The automated updates can be done by a combination of existing 2D topographic data and high-resolution height points. The height-points acquired from laser scanning are a perfect 3D source, but in The Netherlands this dataset, i.e. AHN2, is only collected once every 5 years. In between, aerial photographs with sufficient overlap, i.e. 60 %, can be used to automatically generate high-resolution height points. This may require a change in the requirements of these images.

2.5 *Collecting and Portraying 3D Showcases*

Although 3D applications are common practice for many professionals, 3D is new and considered as "complex" and "expensive" to others. To show the need and potentials for 3D

and to offer a source of inspiration to policy makers and newcomers in the field, the 3D Pilot built a website that collects and portrays examples of 3D applications that are already practised. The growing interest for 3D is proven by about 1000 visitors from all parts of the world in a few months.

3 Conclusions and Future Work

This paper presents the 3D Pilot that defined, established and implemented a national 3D standard in The Netherlands. The national 3D standard, IMGeo, is defined as an application domain extension of CityGML. Among the end results are: implementation specifications of the national 3D standard which assures a topologically, geometrically and semantically correct model, example 3D IMGeo data, a 3D validator, best practice documents of how to acquire, maintain, update and disseminate 3D IMGeo data, and a website that collects and portrays 3D showcases.

The main conclusion of running the 3D Pilot between January 2010 and December 2012 is the change of vision concerning 3D in The Netherlands. At the start of the 3D Pilot, many saw that 3D had potentials, but did not know how to apply them. In the course of the pilot the ambitions for 3D have become much more focused, also supported by the national 3D standard. Now, these ambitions are further developed since the second phase of the pilot has finished.

Several aspects appeared to be crucial for the adoption and implementation of the 3D standard. Firstly, the engagement of many stakeholders was important to gain the necessary support. Secondly, the alignment to the international standard CityGML, which made it possible that CityGML compliant software, is able to deal with IMGeo data as well as to the ongoing 2D efforts let 3D applications become feasible and attractive for governmental organizations. In addition, collaboration appeared to be important to share knowledge on the wide variety of topics in the complex 3D domain. Finally, it was found important that a number of national organizations took the responsibility to facilitate the process. Although the pilot is a joint effort of the 3D com-

munity, national organizations have to initiate and facilitate such a network organization and they are important for anchoring the results.

The work of 3D implementation is not finished and several issues remain. These are currently taken up by the national 3D Special Interest Group that has been specially established for this purpose. The objectives of this 3D SIG are to address the still open 3D issues in collaboration with all stakeholders. Among these are open issues concerning implementation of the 3D standard, the integration of 3D IMGeo with the subsoil, i.e. geology and cables & pipelines (see also the work of Becker et al. (2010), Hijazi et al. (2010), Zobl & Marschallinger (2008)), further alignment between BIM and GIS, and 3D extensions in other domains such as spatial planning and noise modelling Stoter et al. (2008).

Acknowledgement

References

3DCityDB, 2013: http://www.3dcitydb.net (18. 6. 2013).

Becker, T., Nagel, C. & Kolbe, T.H., 2010: Integrated 3D modelling of multi-utility networks and their interdependencies for critical infrastructure analysis. – **5th** International 3D Geo-Info Conference, November 2010, Berlin.

Becker, T., Nagel, C. & Kolbe, T.H., 2013: Semantic 3D Modelling of Multi-Utility Networks in Cities for Analysis and 3D Visualization. – Progress and New Trends in 3D Geoinformation Sciences **2013:** 41–62, Springer, Berlin, Heidelberg.

Bogdahn, J. & Coors, V., 2010: Towards an automated healing of 3D urban models. – Kolbe, T.H., Köning, G. & Nagel, C. (eds): International Conference on 3D Geoinformation **XXXVIII-4/W15:** 13–17.

CityGML, 2013: Application Domain Extensions, www.citygmlwiki.org/index.php/CityGML-ADEs (9.5.2013).

Czerwinski, A., Kolbe, T.H., Plümer, L. & Stöcker-Meier, E., 2006: Interoperability and accuracy requirements for EU environmental noise mapping. – InterCarto, InterGIS, Berlin 2006.

Czerwinski, A., Sandmann, S., Stöcker-Meier, E. & Plümer, L., 2007: Sustainable SDI for EU noise mapping in NRW – best practice for INSPIRE. – International Journal of Spatial Data Infrastructures Research **2:** 90–111.

Clement, J., 2011: Presented at the second meeting of 3D Pilot Phase II, http://www.geonovum.nl/sites/default/files/3D/december.zip (9.5.2013).

Geonovum, 2013a: List of 3D Pilot participants, http://www.geonovum.nl/dossiers/3d-pilot/deelnemersvervolg (9.5.2013).

Geonovum, 2012a: 3D CityGML-IMGeo toolkit, www.geonovum.nl/3d/toolkit (9.5.2013).

Geonovum, 2012b: Implementation specifications for 3D IMGeo-CityGML, www.geonovum.nl/sites/default/files/3D/toolkit/3DIMGeoBestekteksten.pdf (in Dutch), (9.5.2013).

Geonovum, 2012c: Technical specifications for the reconstruction of 3D IMGeo CityGML data, www.geonovum.nl/sites/default/files/3D/toolkit/3DFinalReport_2013_1.01.pdf (9.5.2013).

Geonovum, 2012d: Validator for 2D IMGeo, http://validatie-dataspecificaties.geostandaarden.nl/genericvalidator/content/standard/19 (9.5.2013).

Geonovum, 2012e: 3D Validator software, http://www.geonovum.nl/3D/toolkit (9.5.2013).

Geonvoum, 2012f: List of tasks that needed to be performed in the 3D IMGeo-CityGML challenge, http://www.geonovum.nl/sites/default/files/scenario_3d_imgeo_relay_final.pdf (9.5.2013).

Geonovum, 2012g: Resulting video of the 3D CityGML IMGeo relay, http://www.youtube.com/watch?v=APFIO_czwms&feature=youtube (9.5.2013).

Gröger, G. & Plümer, L., 2011: How to achieve consistency for 3D city models. – GeoInformatica **15:** 137–165.

Hijazi, I., Ehlers, M., Zlatanova, S., Becker, T. & van Berlo, L., 2010: Initial investigations for modeling interior Utilities within 3D Geo Context: Transforming IFC-interior utility to CityGML/UtilityNetworkADE. – **5th** International 3D GeoInfo Conference, November 2010, Berlin.

ISO, 2003: ISO 19107:2003 Geographic information – Spatial schema, http://www.iso.org/iso/search.htm?qt=19107&sort=rel&type=simple&published=on&active_tab=standards (9.5.2013).

Ledoux, H., Verbree, E. & Si, H., 2009: Geometric Validation of GML Solids with the Constrained Delaunay Tetrahedralization. – De Maeyer, P., Neutens, T. & De Rijck, M. (eds): **4th** International Workshop on 3D-Geo-Information: 143–148, Ghent, Belgium.

OGC, 2007: OpenGIS® Geography Markup Language, GML Encoding Standard. Version 3.2.1, doc # OGC 07-036, http://portal.opengeospatial.org/files/?artifact_id=20509 (9.5.2013).

OGC, 2012: OpenGIS® City Geography Markup Language, CityGML Encoding Standard, version 2.0, www.opengeospatial.org/standards/citygml (9.5.2013).

Oude Elberink, S.J., 2010: Acquisition of 3D topography: automated 3D road and building reconstruction using airborne laser scanner data and topographic maps. – Enschede, University of Twente, Faculty of Geo-Information Science and Earth Observation, 2010. ITC Dissertation 167.

Oude Elberink, S.J. & Vosselman, G., 2011: Quality analysis on 3D building models reconstructed from airborne laser scanning data. – ISPRS Journal of Photogrammetry and Remote Sensing **66** (2): 157–165.

Oude Elberink, S., Stoter, J., Ledoux, H. & Commandeur. T., 2013: Generation and Dissemination of a National Virtual 3D City and Landscape Model for the Netherlands. – Photogrammetric Engineering & Remote Sensing, PE&RS, in press.

SIG 3D, 2012a: Teil 1: „Grundlagen – Regeln für valide GML Geometrie-Elemente in CityGML", http://wiki.quality.sig3d.org/index.php/Handbuch_für_die_Modellierung_von_3D_Objekten_-_Teil_1:_Grundlagen_(Regeln_für_valide_GML_Geometrie-Elemente_in_CityGML) (9.5.2013).

SIG 3D, 2012b: Teil 2: „Modellierungshandbuch Gebäude (LoD1, LoD2, LoD3) " http://wiki.quality.sig3d.org/index.php/Handbuch_für_die_Modellierung_von_3D_Objekten_-_Teil_2:_Modellierung_Gebäude_(LOD1,_LOD2_und_LOD3) (9.5.2013).

Stoter, J.E., de Kluijver, H. & Kurakula, V., 2008: 3D noise mapping in urban areas. – International Journal of Geographical Information Science: IJGIS **22** (8): 907–924.

Stoter, J., Vosselman, G., Goos, J., Zlatanova, S., Verbree, E., Klooster, R. & Reuvers, M., 2011: Towards a National 3D Spatial Data Infrastructure: Case of The Netherlands. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2011** (6): 405–420.

Stoter, J., Beetz, J., Ledoux, H., Reuvers, M., Klooster, R., Janssen, M., Penninga, F., Zla-

tanova, S. & Brink, L., 2013: Implementation of a National 3D Standard: Case of the Netherlands. – Pouliot, J., Daniel, S., Hubert, F. & Zamydi, A. (eds): Progress and New Trends in 3D Geoinformation Sciences, Springer.

Van den Brink, L., Stoter, J.E. & Zlatanova, S., 2012a: Establishing a national standard for 3D topographic data compliant to CityGML. – International Journal of Geographical Information Science, in press, http://www.tandfonline.com/doi/abs/10.1080/13658816.2012.667105 (9.5.2013).

Van den Brink, L., Stoter, J.E. & Zlatanova, S., 2012b: Modelling an application domain extension of CityGML in UML, OGC best practice paper, https://portal.opengeospatial.org/files/?artifact_id=49000 (9.5.2013).

Van den Brink, L., Stoter, J.E. & Zlatanova, S., 2013: UML-based approach to develop a CityGML application domain extension. – Transactions in GIS, doi: 10.1111/tgis.12026.

Wagner D., Wewetzer, M., Bogdahn, J., Alam, N., Pries, M. & Coors, V., 2013: Geometric- semantical consistency validation of CityGML models. – Pouliot, J., Daniel, S., Hubert, F. & Zamyadi, A. (eds): Progress and New Trends in 3D Geoinformation Sciences. – Lecture Notes in Geoinformation and Cartography: 171–192, Springer, Berlin, Heidelberg.

Zobl, F. & Marschallinger, R., 2008: Subsurface GeoBuilding Information Modelling GeoBIM. – GEOinformatics **8** (11): 40–43.

Address of the Authors:

Dr. J.E. Jantien Stoter, OTB Research Institute for the Built Environment, Delft University of Technology, Jaffalaan 9, 2628BX Delft, The Netherlands, e-mail: j.e.stoter@tudelft.nl & Kadaster, Product and Process Innovations, Postbus 9046, 7300 GH Apeldoorn, The Netherlands, jantien.stoter@kadaster.nl & Geonovum, Barchman Wuytierslaan 10, 3818 LH Amersfoort, The Netherlands, e-mail: j.stoter@geonovum.nl

Dr. Hugo Ledoux, OTB Research Institute for the Built Environment, Delft University of Technology, Jaffalaan 9, 2628BX Delft, The Netherlands, e-mail: h.ledoux@tudelft.nl

Marcel Reuvers, Linda van den Brink & Paul Janssen, Geonovum, Barchman Wuytierslaan 10, 3818 LH Amersfoort, The Netherlands, e-mail: {m.reuvers}{p.janssen}{l.vanbrink}@geonovum.nl

Dr. Jakob Beetz, Design Systems Group, Department of the Built Environment, Eindhoven University of Technology, j.beetz@tue.nl

Rick Klooster, Municipality Apeldoorn, r.klooster@apeldoorn.nl

Dr. Friso Penninga, City of The Hague, friso.penninga@denhaag.nl

Prof. dr. ir. George Vosselman, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands, vosselman@itc.nl

<div style="border:1px solid">

Berichte von Veranstaltungen

</div>

Dreiländertagung D – A – CH (DGPF, OVG und SGPF) und 33. Wissenschaftlich-Technische Jahrestagung der DGPF, 27. 2. – 1. 3. 2013, in Freiburg

Die drei nationalen Gesellschaften für Photogrammetrie, Fernerkundung und Geoinformation Deutschlands, Österreichs und der Schweiz trafen sich vom 27. 2. bis zum 1. 3. 2013 an der Albert-Ludwigs-Universität in Freiburg im Breisgau.

Der Präsident der DGPF, Prof Dr. THOMAS KOLBE, begrüßte die Ehrenmitglieder und Ehrengäste der DGPF sowie die Teilnehmer der Tagung, die nicht nur aus Deutschland, Österreich und der Schweiz, sondern auch aus Griechenland, der Slowakei und der Tschechischen Republik angereist waren. Er stellte fest, dass von den 312 Teilnehmern über 30 % unter 30 Jahren waren und begrüßte insbesondere die große Gruppe von 57 Studierenden der FHNW Muttenz in der Schweiz. THOMAS KOLBE bedankte sich bei der Gastgeberin, Frau Prof. Dr. BARBARA KOCH mit ihrem Team der Universität Freiburg, und dem Organisationsteam aus den drei Ländergesellschaften, HERBERT KRAUSS, EBERHARD GÜLCH, KLAUS KOMP (DGPF), MICHAEL FRANZEN (OVG) und KIRSTEN WOLFF (SGPF), für die sehr gute Vorbereitung und wünschte allen Teilnehmern eine erfolgreiche Tagung.

Der Präsident der Universität Freiburg, Prof. Dr. Dr. h.c. HANS-JOCHEN SCHIEWER, begrüßte die Teilnehmer der Tagung sehr herzlich und zeigte viele Anknüpfungspunkte der Uni zur DGPF auf, z.B. die erst kürzlich durch Umstrukturierung geschaffene Fakultät für Umwelt und Natürliche Ressourcen mit BARBARA KOCH an der Spitze.

Im Rahmen der Eröffnungsveranstaltung wurden die Preisträger des Karl-Kraus-Nachwuchsförderpreises 2013 geehrt. Unter der Moderation von GÖRRES GRENZDÖRFFER übergab URSULA KRAUS, die Witwe des Namensgebers des Preises, die Urkunden. Danach übergab MICHAEL FRANZEN das Widmungsexem-

plar der Karl-Kraus-Medaille der ISPRS an URSULA KRAUS. Der Hansa-Luftbild-Preis 2013 für den besten PFG-Artikel eines Nachwuchswissenschaftlers ging an JOCHEN MEIDOW für den Beitrag „*Efficient Multiple Loop Adjustment for Computer Vision Tasks*“, überreicht von PAUL HARTFIEL, Firma Hansa Luftbild. Der Festvortrag zum Thema „*Probabilistische Verfahren für die autonome Navigation und Umgebungskartierung mit mobilen Robotern*“ hielt Prof. WOLFRAM BURGARD von der Universität Freiburg und zeigte höchst spannend und unterhaltsam, wie Autos in Parkhäusern allein ihren Parkplatz ansteuern, ein Roboter namens Obelix seinen Weg durch Freiburg findet und autonom fliegende Plattformen Räume vermessen.

Am 28. 2. 2013 und 1. 3. 2013 wurden die Fachvorträge der Arbeitskreise parallel in jeweils drei Sitzungen gehalten. Einzelheiten zu den Inhalten und Ergebnissen finden sich in den Berichten der Arbeitskreisleiter. Die Posterbeiträge wurden durch eine Kombination von Kurzpräsentation und einer zusätzlichen gemeinsamen Poster-Sitzung deutlich aufgewertet. Über 20 Firmen haben Ihre Produkte und Dienstleistungen auf der Fachfirmenausstellung präsentiert. Die Ausstellung wurde von BERNHARD HORST und KLAUS KOMP, Geschäftsführer der Geomatik Tagungs-GmbH, organisiert. Zwei der Aussteller haben in einem neu geschaffenen Solution-Forum der Premium-Aussteller in Form von Beiträgen vertiefend die neuesten Entwicklungen ihrer Produkte vorgestellt. Sieben weitere Sponsoren haben Beiträge für die Tagungstasche beigesteuert. Allen Organisationen – besonders den Premiumausstellern – sei für die großzügige finanzielle Unterstützung der Dreiländertagung gedankt.

Am 27. 2. 2013 wurden ein wissenschaftliches Tutorium vom FELIS-Institut der Universität Freiburg und anwenderorientierte Tutorien durch drei Firmen angeboten. Mit nahezu 60 Teilnehmern für die vier Tutorien kann eine deutliche Nachfrage nach dieser Art von Weiterbildungsveranstaltung festgestellt werden. Ebenfalls am 27. 2. 2013 wurde eine Be-

sichtigung des Staatsweingutes Freiburg mit Weinprobe organisiert. Die Teilnehmer konnten sich dabei über diese Einrichtung der Universität Freiburg informieren und verschiedene Weine der Region verkosten. Beim zwanglosen Vorabendtreffen im Restaurant und Weinstube Heiliggeist Stüble konnten für die über 150 Teilnehmer bei trefflicher Stimmung der fachliche Austausch sowie vertiefte Kontakte mit den lokalen Spezialitäten initiiert werden. Bei der Abendveranstaltung am 28. Februar im ehrwürdigen „Alten Kaufhaus" am Münsterplatz setzten die über 200 Teilnehmer im festlichen Rahmen das (Social-)Networking bis in die späten Nachtstunden fort.

In der Closing Session am 1. 3. 2013 konnte der Vizepräsident der DGPF, Prof. Dr. Uwe Stilla, noch über 120 Teilnehmer zu einem kurzen Fazit begrüßen. Prof. Thomas Kersten lud die Teilnehmer zur Jahrestagung 2014 an die Hafen-City-Universität (HCU) vom 26.–28. 3. 2014 nach Hamburg und Prof. Lena Halounová, Kongressdirektorin des ISPRS Kongresses 2016, zu dieser großen internationalen Veranstaltung nach Prag ein. Der Vorstand der DGPF bedankte sich mit einem kleinen Geschenk bei den Studierenden Alicia Unrau aus Freiburg und Tobias Schwarz aus Berlin für die herausragende und immer sehr aufmerksame Unterstützung in der Vorbereitung und während der Tagung.

Der Tagungsband wurde in Form einer DVD in bewährter Weise von Eckhardt Seyfert erstellt und die Tagungshomepage von Manfred Wiggenhagen betreut. Großer Dank gebührt allen Autoren, Gutachtern und Mitgliedern des Programmkommittees für die Zusammenstellung des sehr attraktiven Tagungsprogramms.

Den gelungenen Abschluss bildete am 1. 3. 2013 die Führung durch den Dachstuhl des Freiburger Münsters mit dem Motto „Eine Reise in den Wald des Mittelalters".

<div align="right">

Eberhard Gülch, HFT Stuttgart,
Sekretär der DGPF

</div>

12. Oldenburger 3D-Tage, 13.–14. 2. 2013

In die 12. Runde gingen die diesjährigen Oldenburger 3D-Tage und demonstrierten dabei ihre weiterhin hohe Attraktivität im Bereich von Fachveranstaltungen zur Optischen 3D-Messtechnik, Photogrammetrie und Laserscanning. 55 Fachbeiträge in zwei parallel laufenden Sessions, eine umfangreiche Fachausstellung sowie die Möglichkeit für intensive fachliche Diskussion in den Pausen boten eine hervorragende Basis für den Informationsaustausch. Hervorzuheben ist zudem, dass neben den „Stammbesuchern" und -vortragenden ein hoher Anteil an „Neuzugängen" zu verzeichnen war: Das hält die Veranstaltung auch für „Dauergäste" interessant.

Den Eröffnungsvortrag „*Von optischer 3D-Messtechnik zur lichtfeldbasierten Bildaufnahme und -verarbeitung*" hielt Prof. Bernd Jähne vom Heidelberg Collaboratory of Image Processing (HCI) und zeigte dabei eindrucksvoll die erweiterten Möglichkeiten plenoptischer Kamerasysteme. Die nachfolgenden auf zwei Vortragssäle verteilten Parallelsessions waren inhaltlich nach dem Themenkomplexen „Laserscanning" bzw. „Bildbasierte Ansätze" vorsortiert, so dass die entsprechenden Interessensbereiche klar abgesteckt waren. Dass viele Besucher an unterschiedlichsten Themen interessiert waren, zeigte der „Traffic" zwischen den Sälen während der laufenden Veranstaltung.

Der erste „Laserscanning"-Vortragsblock behandelte in der Session „*RGB und Infrarot*" die Kombination von Punktwolken bzw. Modellen mit zusätzlichen Informationen wie Thermal- und Infrarotbildern. Die Erfassung von Punktwolken stellt heute keine Herausforderung mehr dar, die Kombination mit anderen Sensoren zur Generierung eines Mehrwertes wird dafür immer wichtiger. Dass insbesondere „Dynamische Prozesse" ein weites Anwendungsspektrum aufweisen und hierbei mit unterschiedlichsten Sensoren gemessen wird, zeigten die Beiträge in der gleichnamigen Session.

Im zweiten Block des Tages hatten dann nach der Mittagspause sechs Hersteller die Möglichkeit, ihre Produkte näher vorzustellen. Im Vordergrund standen nicht nur Syste-

me zur Datengenerierung, sondern auch um Softwarelösungen zur intelligenten Datenverarbeitung mit Schnittstellen, z. B. zu CAD-Systemen. Die parallel angesetzte Session „*Laser und Licht*" war ein Beispiel für Experimentierfreudigkeit und Innovation der beteiligten Institutionen. Den Abschluss des Tages bildete schließlich ein Vortragsblock über die Registrierung von Punktwolken. Drei Vorträge zu diesem Thema lassen hier auch ganz deutlich den Trend zum markerlosen Vorgehen erkennen. Die Aufnahme im Feld kann somit weiter vereinfacht und beschleunigt werden. Aktuelle Auswertetechniken prägten die zeitgleiche Sitzung zur „*3D-Rekonstruktion*".

Am zweiten Tag wurden in der ersten Session vielerlei Anwendungsmöglichkeiten präsentiert, die wiederum das sehr breite Spektrum des terrestrischen Laserscannings verdeutlichten. Dass nicht nur Laserscanner sondern auch andere hochspezialisierte Systeme und deren Kalibrierung eine entscheidende Rolle in der Qualitätssicherung spielen, war Thema der folgenden Session, in

der u. a. Beiträge über prototypische Anwendungen gezeigt wurden.

Nach der Mittagspause, die für regen Betrieb in der Firmenausstellung sorgte, fand das zweite Herstellerforum statt, in dem nochmals fünf Firmen ihre neuesten Hard- und Softwareentwicklungen präsentierten. Das weitere Vortragsprogramm erstreckte sich über die Sessions „*Sensoren und Systeme*" (u. a. mit UAV-Anwendungen), „*Geomonitoring und Mobile Mapping*" sowie „*Anwendungen Photogrammetrie*". Die Vortragenden zeigten einmal mehr das weite Anwendungsspektrum der Photogrammetrie, hierbei auch in Kombination mit anderer Sensorik.

Zwischen den wie immer in zwei parallel laufenden Sessions gehaltenen Präsentationen war reichlich Zeit, sich bei Kaffee und Brötchen in der mit 23 Ständen gut besetzten Firmenausstellung über neueste Entwicklungen in den Bereichen Hard- und Software zu informieren. Demonstrationen wurden gegeben und überall waren rege fachliche Diskussionen und ein intensiver Austausch über neu Erlerntes oder gerade Gehörtes zu beobachten.

Im Rahmen der Oldenburger 3D-Tage: Bernd Jähne (Universität Heidelberg), Manfred Weissensee (Vizepräsident Jade Hochschule), Sandra Heller (AXIOS 3D Services), Thomas Luhmann und Gerd Schwandner (Oberbürgermeister Oldenburg).

Die insgesamt 216 Teilnehmer kamen wie im Vorjahr zu 45 % aus dem Hochschulbereich. Anwender (20 %) und Hersteller (22 %) waren in etwa gleich stark vertreten, die restlichen 13 % waren den Dienstleistern zuzuordnen. Dass die Oldenburger 3D-Tage letztlich nicht nur dem Wissensaustausch dienen, sondern auch kulturell einiges zu bieten haben, wurde einmal mehr auf der Abendveranstaltung in der Weser-Ems-Halle deutlich. Neben dem alljährlichen kulinarischen „Grünkohl-Highlight" und dazugehörigen „Regulierungsgetränken zur Fettverbrennung" konnten alle Beteiligten die künstlerische Seite von Thomas Luhmann kennenlernen, und zwar als einem der Akteure im Showcase Musical Theater e.V. (www.showcasemusicals.com).

<div align="right">

Maren Lindstaedt, Hamburg &
Heinz-Jürgen Przybilla, Bochum

</div>

Hochschulnachrichten

Universität Stuttgart

Frau Mag. Yevgeniya Filippovska wurde am 12. September 2012 von der Fakultät Luft- und Raumfahrttechnik und Geodäsie der Universität Stuttgart mit der Dissertation *Evaluierung generalisierter Gebäudegrundrisse in großen Maßstäben* zum Dr.-Ing. promoviert.
1. Gutachter: Prof. Dr.-Ing. habil. Dieter Fritsch, Universität Stuttgart.
2. Gutachter: Prof. Dr.phil. habil. Manfred F. Buchroithner, Universität Dresden.

Kurzfassung:
Bei der Erzeugung von Karten werden die darzustellenden räumlichen Objekte in Abhängigkeit des angestrebten Maßstabs ausgewählt, verändert und auf eine Art und Weise arrangiert, dass deren Form und Verteilung zu einem bestmöglichen Verständnis der räumlichen Gegebenheiten führt. Dabei weist die kartographische Abbildung unvermeidliche und zuweilen tiefgreifende geometrische Veränderungen im Vergleich zur Realität auf, welche durch eine übergeordnete Kontrollinstanz zu verifizieren und zu bewerten sind. Da dieser als kartographische Generalisierung bezeichnete Prozess heutzutage vermehrt automatisiert stattfindet, strebt man auch eine formalisierte Qualitätsbewertung der dabei erzeugten Ergebnisse an, um auf dieser formalen Grundlage entsprechende Werkzeuge zu entwickeln. Die Bewertung der kartographischen Generalisierung im Hinblick auf die Qualität ist Gegenstand dieser Arbeit. Hierbei werden im Speziellen 2D-Grundrisse in Betracht gezogen, da diese einen wesentlichen Teil der im urbanen Raum vorkommenden Geodaten ausmachen und deren Qualitätsbewertung bislang noch nicht hinreichend untersucht wurden.

Obwohl die kartographische Generalisierung die Lesbarkeit der Gesamtkomposition zum Ziel hat, muss die Qualitätsbewertung zuerst auf der untersten Generalisierungsebene, der sogenannten Mikroebene erfolgen, indem die Geometrie- bzw. die Formveränderungen der Einzelobjekte bemessen werden. Denn neben dem Straßennetz dienen den Kartennutzern vor allem markante Gebäude als Orientierungshilfe, welche aus diesem Grund nicht allzu großen Veränderungen unterliegen dürfen. Im Rahmen dieser Arbeit werden dementsprechend Qualitätscharakteristiken aufgezeigt, welche auf dem direkten Vergleich von zwei Gebäudegrundrissen – original und generalisiert – basieren.

Als Motivation der Arbeit wird ein Wahrnehmungstest vorgestellt, welcher die Bewertung von generalisierten Grundrissen durch menschliche Betrachter untersucht. Der Versuch diese Wahrnehmungsprozesse mathematisch zu formalisieren wird als Ähnlichkeitsschätzung bezeichnet. Die Grundlagen dazu werden in der Arbeit aufgearbeitet und in diesem Zusammenhang auch eine einheitliche Klassifizierung der Objektmerkmale basierend auf der zugrundeliegenden Berechnungsmethode vorgeschlagen.

Daran anschließend werden die im Rahmen der Arbeit neu entwickelten Charakteristiken zur Ähnlichkeitsanalyse vorgestellt, welche die Objekte unter den beiden Aspekten der

Kontur- und Flächentreue hin vergleichen. Da eine Zuordnung zwischen den Formelementen der beiden zu vergleichenden Grundrisse allgemein nicht zweifelsfrei festgestellt werden kann, werden die Objekte gemäß der Standardisierung von Geodaten als Punktmengen betrachtet. Dies erlaubt es, die geometrischen Berechnungen fast ausschließlich auf den Standardoperatoren der Mengentheorie aufzusetzen. In Kombination mit den topologischen Algorithmen der Graphentheorie hat dies den erwünschten Nebeneffekt, dass sich alle Charakteristiken ohne Änderungen auch auf 3D-Geodaten anwenden lassen.

Bei den Charakteristiken zur Konturtreue werden die maximale Konturabweichung und der Anteil der Konturüberlappung basierend auf den objektbeschreibenden Randmengen berechnet. Hierzu wird für die erste Charakteristik eine Modifikation der Hausdorff-Distanz vorgeschlagen, die nicht unter Betrachtung der vollständigen Konturen beider Objekte durchgeführt wird, sondern nur auf deren Differenzbereiche. Dadurch ergeben sich Distanzen, bzw. Punkte maximaler Diskrepanz, die für den Menschen plausibler erscheinen und damit leichter nachvollziehbar sind als die Ergebnisse der Originalmetrik. Es zeigt sich dabei, dass für konkave Differenzbereiche die Distanz nicht entlang der direkten Verbindung von zwei Punkten erfolgen sollte, sondern entlang eines innerhalb dieses Bereichs verlaufenden Linienzugs. Für die Bestimmung der Konturüberlappung wird eine Kombination von Puffer- und Mengenoperatoren aufgezeigt, die zu einer möglichst geringen Fehlabschätzung führen. Der Grund ist, dass die Generalisierung von Gebäudegrundrissen oftmals eine Anpassung hinsichtlich der Rechtwinkligkeit, Parallelität und Kollinearität von Liniensegmenten durchführt, so dass sich die zwei zu vergleichenden Konturen nicht mehr vollständig überdecken. Vielmehr muss mit Toleranzbereichen gearbeitet werden, deren verfälschenden Effekt es zu minimieren gilt.

Auch die Flächentreue kann unter zwei Aspekten beurteilt werden: quantitativ und räumlich. Hierzu wird zunächst eine einheitliche Benennung der Strukturveränderungen eingeführt, wobei die durch die Generalisierung eliminierten Teile als Intrusionen und die hinzugekommenen als Extrusionen bezeichnet werden. Es folgen deren Definitionen durch Operatoren der Mengenlehre, wobei sich die Gesamtveränderung der räumlichen Erstreckung durch die symmetrische Differenz von Intrusionen und Extrusionen ergibt. Quantitativ lässt sich so nicht nur leicht zeigen, wie eine Fläche durch die Generalisierung zu- bzw. abnimmt, sondern auch deren absolute und ausgeglichene Gesamtveränderung. Eine Diskussion betreffend die Normierung der eingeführten Charakteristiken, welche primär der Quantifizierung der Objektveränderungen dienen, zeigt deren Umrechnung und Anwendung zu Qualitätsmaßen.

Um die Aussagekraft und Praxistauglichkeit der vorgeschlagenen kontur- und flächenbezogenen Charakteristiken zu prüfen, wird eine Evaluierung von generalisierten Gebäudegrundrissen auf der Mikro- und Makroebene durchgeführt. Zuerst erfolgt auf der Mikroebene die Qualitätsbewertung von Einzelobjekten bzw. Landmarken, um anschließend den Vergleich alternativer Generalisierungsergebnisse zu diskutieren. Für diesen Zweck wird ein Aggregationsansatz vorgeschlagen, der alle Charakteristiken in Ähnlichkeitsmaße umwandelt, welche zusammengefasst Merkmalsvektoren in einem mehrdimensionalen Merkmalsraum darstellen. Dank der normierten Werte dieses Vektors kann nun mittels der euklidischen Distanz der Gesamtunterschied zwischen dem originalen und generalisierten Grundriss beurteilt werden, wodurch sich die Gesamtqualität der Generalisierung durch ein einzelnes Maß ausdrücken lässt.

Auf der Makroebene wird die Qualität der Generalisierung für großflächige Areale evaluiert. Dabei spielt insbesondere auch die anschauliche Präsentation der Ergebnisse eine zentrale Rolle, so dass die verschiedenen Möglichkeiten zur Darstellung der einzelnen Charakteristiken im Hinblick auf eine gute Diskriminierbarkeit der Qualitätsangaben im Schwerpunkt stehen. Abschließend erfolgt wiederum ein Vergleich von alternativen Generalisierungen, der zur Bestimmung der bestgeeigneten Generalisierungslösung in Abhängigkeit einer Anwendung notwendig ist.

Die Dissertation ist elektronisch bei der Bibliothek der Universität Stuttgart (http://elib.

uni-stuttgart.de/opus/volltexte/2012/7768/ pdf/MainDOC.pdf) und bei der DGK (http:// dgk.badw.de/fileadmin/docs/c-693.pdf) publiziert worden.

Karlsruher Institut für Technologie (KIT)

Herr Dipl.-Inf. Florian Schmidt promovierte am 23. November 2012 an der Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften (Institut für Photogrammetrie und Fernerkundung) des KIT mit der Arbeit *„Ein integraler stochastischer Ansatz zur automatischen Bestimmung von Personentrajektorien aus Luftbildsequenzen"* zum Dr.-Ing.
1. Gutachter: Prof. Dr.-Ing. Stefan Hinz, KIT,
2. Gutachter: Hon.Prof. Dr.-Ing. Peter Reinartz, Universität Osnabrück und DLR Oberpfaffenhofen,
3. Gutachter: Prof. Dr.-Ing. Peter Vortisch, KIT.

Kurzfassung:
Luftbilder stellen eine bedeutende Quelle vielfältiger Informationen über unsere Umwelt dar. In der Vergangenheit wurden sie hauptsächlich zur Beschreibung der Topographie genutzt. Moderne Kamerasysteme mit einer Aufnahmefrequenz von wenigen Hertz ermöglichen es, nun auch dynamische Prozesse großflächig zu beobachten. Für eine effektive Auswertung dieser Luftbildsequenzen werden automatische Methoden benötigt, die jedoch oftmals noch entwickelt oder an die spezifischen Herausforderungen angepasst werden müssen. Hier liegt die übergeordnete Zielsetzung dieser Arbeit. Im Detail beschäftigt sie sich mit der Fragestellung, in wie weit es möglich ist, Informationen über das Bewegungsverhalten von Personen aus Luftbildsequenzen zu gewinnen. Diese ließen sich z. B. zur besseren Koordination von Großveranstaltungen oder zur Evaluation von weiträumigen Infrastrukturanlagen einsetzen.

In dieser Arbeit wird daher eine Strategie entwickelt, umgesetzt und evaluiert, die es erstmalig ermöglicht, automatisch Einzelpersonen im Luftbild zu erkennen und ihre Bewegung durch eine Sequenz hinweg zu verfolgen. Die Auswertung beginnt für jede Aufnahme mit der Suche nach potentiellen Stand-

orten von Personen, wofür ein aussehensbasierter Ansatz mit implizitem Modell verwendet wird. Dieser nutzt für die Detektion sowohl weiterentwickelte Bildmerkmale als auch den optional vorhandenen Personenschatten, welcher direkt in das visuelle Objektmodell integriert wird. Der Trainingsprozess zum Lernen dieses Modells wird dahingehend verbessert, dass beim automatischen Sammeln von Hintergrundbeispielen nun auch Bereiche in Objektnähe sowie der jeweilige Konfidenzwert mit berücksichtigt werden.

Da Personen in Luftbildern mit einer Bodenauflösung von etwa 15 cm pro Pixel nur sehr schwach zu erkennen sind, lässt sich eine robuste Detektion oft nicht auf Basis eines einzelnen Bildes durchführen. Aus diesem Grund wird in dieser Arbeit erstmals die Detektion mit implizitem visuellem Modell in das Multi-Hypothesen-Tracking-Verfahren (MHT) integriert. Hierfür werden die Ergebnisse der Objekterkennung stochastisch modelliert und der MHT-Formalismus entsprechend erweitert. Die Detektion von Personen erfolgt somit erst während des Trackings, wenn mehr Informationen zur Verfügung stehen und zuverlässigere Entscheidungen getroffen werden können. Das MHT-Verfahren wird ebenfalls weiterentwickelt, so dass sich die Vorteile der hypothesen- und trajektorientierten Variante gleichzeitig nutzen lassen. Zusätzlich wird eine neue Methode zur automatischen Bestimmung der Auftrittswahrscheinlichkeiten von Falschalarmen und neuen Objekten integriert, sowie das besonders wichtige Clusterverfahren durch eine verbesserte Datenstruktur stark vereinfacht.

Das entwickelte System bestimmt in anspruchsvollen Testsequenzen die Trajektorie von etwa einem Drittel aller Personen annähernd vollständig. Die Ergebnisse sind besonders gut in Bereichen mit geringer Objektdichte und hoher Erkennbarkeit. Bilden sich jedoch Gruppen oder gar Menschenmassen, sind Einzelpersonen visuell nicht mehr unterscheidbar und das Detektionsverfahren stößt an seine Grenzen. Das Auswertesystem müsste um weitere Module zur Behandlung dieser Phänomene ergänzt werden, um die einheitliche und vollständige Analyse eines gesamten Luftbildes.

Die Dissertation ist erschienen in der Reihe C der DGK, Heft 696, und elektronisch verfügbar unter www.dgk.badw.de/fileadmin/docs/c-696.pdf

Technische Universität München

Herr Dipl.-Inf. Peter Selby promovierte am 25. Februar 2013 an der Fakultät für Bauingenieur- und Vermessungswesen (Fachgebiet Photogrammetrie und Fernerkundung) der Technischen Universität München mit der Arbeit *„Bildgestützte 3D-Vermessung von Patienten zur Positionierung für die radiologische Krebstherapie"* zum Dr.-Ing.
1. Gutachter: Prof. Dr.-Ing. Uwe Stilla, Technische Universität München (TUM),
2. Gutachter: Prof. Dr.-Ing. Joachim Hornegger, Friedrich-Alexander-Universität Erlangen-Nürnberg,
3. Gutachter: Prof. Dr.-Ing. Georgios Sakas, Medcom GmbH

Kurzfassung:
In den letzten Jahren konnten bedeutende technologische Fortschritte bei der strahlentherapeutischen Tumorbehandlung gemacht werden. Das noch relativ junge Gebiet der Partikelbestrahlung setzt sich, seitdem 1990 die erste klinische Einrichtung den Betrieb aufgenommen hat, immer mehr als überlegene Behandlungstechnik durch. Dies liegt daran, dass die Vorteile, der gegenüber der herkömmlichen Behandlung wesentlich zielgenaueren Dosisapplikation, die Nachteile des höheren Aufwandes bei weitem überwiegen. So kann Tumorgewebe gezielter behandelt werden, während umliegendes Gewebe weniger in Mitleidenschaft gezogen wird. Patienten können allerdings nur von der höheren Genauigkeit profitieren, wenn es gelingt, ihre Lage in der Bestrahlungsanlage exakt zu bestimmen und sie relativ zum Behandlungsstrahl so zu positionieren, dass die Strahlendosis genau auf den Tumor abgegeben wird.

Deshalb ist es Ziel dieser Arbeit, automatische Verfahren zu entwickeln, die ein hohes Maß an Genauigkeit und Zuverlässigkeit bei der röntgenbasierten Lagemessung gewährleisten. Da die Verfahren in unterschiedlichen Anlagen und für verschiedene Körperregionen einsetzbar sein sollen, wird Wert auf Vielseitigkeit und Toleranz gegenüber Störeinflüssen, wie etwa nichtoptimaler Bildqualität gelegt. Neben einem auf direktem Bildvergleich und einem auf Rückwärtsprojektion basierenden Ansatz zur 2D-3D Registrierung, wird ein neues, gemischtes Verfahren vorgestellt, welches eine höhere Zuverlässigkeit bieten soll. Durch Einsatz einer radiometrischen Registrierung simulierter Röntgenbilder wird das Verfahren weiter verbessert. Über die Registrierungsverfahren hinaus wird ein neues Ähnlichkeitsmaß für den Bildvergleich vorgestellt, das ebenfalls dazu dient, die Zuverlässigkeit gegenüber anderen existierenden Maßen zu verbessern.

Da neben der Lagemessung, die Kalibrierung der verwendeten Röntgenkamerasysteme für die erreichbare Genauigkeit von entscheidender Rolle ist, wird ein Algorithmus zur automatischen Kalibrierung vorgeschlagen, der die Detektion von Kalibriermarkern in Röntgenbildern beinhaltet und die Schätzung stochastischer Fehler im Gauß-Helmert Modell erlaubt. Die Kamerakalibrierungen für einzelne Ansichten werden in ein Gesamtmodell für die rotierbare Bestrahlungsanlage eingebettet, das die Bewegungsbahnen der Kamerakomponenten nähert und Rückschlüsse auf die Kameraparameter und ihre zufälligen Fehler bei beliebigen Anlageneinstellungen ermöglicht. Darüber hinaus wird gezeigt, welchen Einfluss zufällige Fehler verschiedener Systemkomponenten auf die gemessene Patientenlage haben. Unsicherheiten des Messsystems werden für in Betrieb befindliche Anlagen ermittelt und können z. B. dazu verwendet werden, bei der Bestrahlungsplanung die Sicherheitsabstände der Tumordefinition anzupassen. Schließlich wird erstmals ein Ansatz vorgestellt, der eine Selbstdiagnose des Messsystems ermöglicht. Basierend auf einem an ein Hidden Markov Modell angelehnten Konzept, werden Beobachtungen der Zwischenergebnisse einzelner Systemkomponenten verwendet, um eine Gesamtdiagnose zu erstellen und Rückschlüsse auf wahrscheinliche Fehlerursachen zu erlauben.

Zur Evaluierung der Methoden werden Phantomdaten und klinische Daten verschiedener Partikeltherapiezentren verwendet. Die vorgestellten Registrierungsmethoden werden untereinander und mit anderen existierenden

Verfahren nach gängigen Kennzahlen verglichen. Die Erstellung markerbasierter Goldstandards erlaubt die Genauigkeit zu quantifizieren.

Mit der neuen Registrierungsmethode kann eine Verringerung der Fehlerrate bei der Lagemessung an klinischen Daten erzielt werden. Das vorgeschlagene Ähnlichkeitsmaß zeigt sich im Vergleich zu zahlreichen anderen Maßen als sehr robust bei der Registrierung von Röntgenbildern und dürfte auch für den Einsatz in anderen Registrierungsanwendungen geeignet sein. Mit der Methode zur geometrischen Kalibrierung kann, zusammen mit einer Fehlerfortpflanzung für andere beteiligte Systemkomponenten, eine Bestimmung der zu erwartenden Fehler bei der Lagemessung und Positionierung des Patienten durchgeführt werden. Der Selbstdiagnoseansatz ist in der Lage, Fehlmessungen der Patientenlage zuverlässig als solche zu erkennen und erlaubt zudem, Fehlmessungen auf ihre Ursache zurückzuführen.

Die Dissertation ist verfügbar unter: http://www.pf.bv.tum.de/pub/2013/selby_phd13_dis.pdf

Leibniz Universität Hannover, Photogrammetric (Fairchild) Award für Christian Heipke

CHRISTIAN HEIPKE und die Präsidentin der ASPRS, ROBERTA E. LENCZOWSKI, bei der Überreichung des Photogrammetric (Fairchild) Award 2013 im März 2013.

Prof. CHRISTIAN HEIPKE (Leibniz Universität Hannover) wurde für seine wesentlichen Beiträge zur Forschung und Entwicklung der Photogrammetrie mit dem Photogrammetric (Fairchild) Award der Amerikanischen Gesellschaft für Photogrammetrie und Fernerkundung (ASPRS) ausgezeichnet. Der seit 1934 vergebene Preis, der nur selten ins Ausland geht, gilt als die höchste wissenschaftliche Anerkennung der ASPRS.

Leibniz Universität Hannover, Ehrenmitgliedschaft von Gottfried Konecny in der Tschechischen Gesellschaft für Photogrammetrie und Fernerkundung

Für seine herausragenden wissenschaftlichen Leistungen und Verdienste um die Tschechische Gesellschaft für Photogrammetrie und Fernerkundung wurde Prof. em. GOTTFRIED KONECNY, Leibniz Unversität Hannover, die Ehrenmitgliedschaft der Gesellschaft verliehen. Die Präsidentin Prof. LENA HALOUNOVÁ von der Technischen Universität Prag überreichte Konecny die hohe Auszeichnung am 23.4.2013 während der Interexpo Geo-Siberia 2013 in Novosibirsk.

GOTTFRIED KONECNY (Mitte) zusammen mit Präsidentin LENA HALOUNOVÁ und Schatzmeister KAREL VACH, Tschechische Gesellschaft für Photogrammetrie und Fernerkundung, bei der Überreichung der Urkunde im April 2013.

In eigener Sache

Wechsel in der Schriftleitung Fernerkundung

Die redaktionelle Betreuung einer wissenschaftlichen Zeitschrift ist eine sehr verantwortungsvolle, und gleichzeitig extrem aufwändige und von außen kaum wahrgenommene Arbeit. So ist es nur mehr als verständlich, dass der Einzelne nach einigen Jahren diese Verantwortung einem Nachfolger übertragen möchte. Prof. Dr. rer. nat. CARSTEN JÜRGENS hat 10 Jahre lang die Fernerkundung der PFG betreut und zum umfangreichsten Teil unserer Zeitschrift entwickelt. Unter der Federführung von Herrn Kollegen JÜRGENS sind über 150 Artikel zur Fernerkundung publiziert worden – fast die Hälfte aller. Im Namen der gesamten Schriftleitung danke ich ihm ganz herzlich für sein Engagement und die umfangreiche Arbeit für die PFG.

Als Hauptschriftleiter freue ich mich sehr, Ihnen, liebe Leserinnen und Leser, als Nachfolger im Amt des Schriftleiters Fernerkundung Prof. Dr. rer. nat. ULRICH MICHEL vorstellen zu dürfen. ULRICH MICHEL ist Professor für Geoinformatik an der Pädagogischen Hochschule Heidelberg, und national und international sehr gut vernetzt. Für die PFG hat er bereits in der Vergangenheit einzelne Hefte mitgestaltet und war für unsere Zeitschrift neben anderen auch als Gutachter tätig. ULRICH MICHEL ist einer der Sprecher des Arbeitskreises Fernerkundung der Deutschen Gesellschaft für Geographie. Darüber hinaus leitet er eine der größten Fernerkundungskonferenzen in Europa, die jährlich stattfindende SPIE Remote Sensing. Neben seinen Forschungsinteressen im Bereich der Fernerkundung und der Geoinformationstechnologien setzt er sich dafür ein, dass der Bereich der digitalen Geomedien in Schulen etabliert wird. Wir heißen ihn ganz herzlich in der Gruppe der Schriftleiter willkommen.

WOLFGAN KRESSE, Neubrandenburg

Veranstaltungskalender

2013

3.–6. Juni: 33rd **EARSeL Symposium** in **Matera**, Italien. www.earsel.org/symposia/2013-symposium-Matera/

5.–7. Juni: **International Conference on Earth Observation for Global Changes (EOGC'2013)** in **Toronto**, Canada. www.eogc2013.blog.ryerson.ca/

19.–20. Juni: **Fachtagung Mobile Mapping** in **Mainz**. www.3dgi.ch/mm2013/

23.–28. Juni: **Computer Vision and Pattern Recognition (CVPR 2013)** in **Portland**, USA. www.pamitc.org/cvpr13/

2.–5. Juli: **GI_Forum 2013** in **Salzburg**, Österreich. www.gi-forum.org

14.–20. Juli: **International Computer Vision Summer School** in **Punta Sampieri**, Italien. svg.dmi.unict.it/icvss2013/

21.–26. Juli: **IGARSS 2013** in **Melbourne**, Australien. www.igarss2013.org/

25.–30. August: 26th **International Cartographic Conference (ICC)** in **Dresden**. www.icc2013.org/

2.–6. September: XXIVth **CIPA Heritage Documentation Symposium** in **Strasbourg**, Frankreich. cipa.icomos.org

4.–6. September: **UAV-g Workshop** in **Rostock**. www.uav-g.org

9.–10. September: **UAV-based Remote Sensing Methods for Monitoring Vegetation** in **Köln**. www.tr32db.de/workshop2013

9.–13. September: 54. **Photogrammetrische Woche** in **Stuttgart**. www.ifp.uni-stuttgart.de/phowo

9.–13. September: **British Machine Vision Conference (BMCV 2013)** in **Bristol**, England. bmvc2013.bristol.ac.uk

15.–18. September: **International Conference on Image Processing (ICIP 2013)** in **Melbourne**, Australien. www.ieeeicip.org

11.–17. November: **SSG2013 - ISPRS TC II, III, IV & VII Joint Meeting** (u.a. mit CMRT 2013, Laser Scanning 2013, ISA 2013; s.u.), in **Antalya**, Türkei. www.isprs2013-ssg.org

11. November: **ISA2013**, Image Sequence Analysis for Object and Change Detection, in **Antalya**, Türkei. www.tnt.uni-hannover.de/isprs/isa13

11.–13. November: **Laser Scanning 2013**, in **Antalya**, Türkei. www.cirgeo.unipd.it/laserscanning2013

12.–13. November: **CMRT13**, City Models, Roads, and Traffic, in **Antalya**, Türkei. www.cmrt13.tum.de

26.–29. November: **3D GeoInfo 2013** in **Istanbul**, Türkei. 3dgeoinfo.com

8.–15. Dezember: **ICCV 2013**, International Conference on Computer Vision, **Sydney**, Australien. www.iccv2013.org

10.–11. Dezember: **LowCost 3D** in **Berlin**. www.lc3d.net

2014

1.–6. Februar: **SPIE Photonics West 2014** in **San Francisco,** USA. www.spie.org/photonics-west.xml

13.–14. Februar: **Oldenburger 3D Tage** in **Oldenburg**. www.jade-hs.de/fachbereiche/bauwesen-und-geoinformation/geoinformation/oldenburger-3d-tage/

Weitere Konferenzen und Workshops finden sich beispielsweise unter:
isprs.org/calendar/
iris.usc.edu/Information/Iris-Conferences.html

Korporative Mitglieder

Firmen

AEROWEST GmbH
AICON 3D Systems GmbH
aphos Leipzig AG
Becker GeoInfo GmbH
Bernhard Harzer Verlag GmbH
Blom Deutschland GmbH
Brockmann Consult GmbH
bsf swissphoto GmbH
Büro Immekus
CGI Systems GmbH
DB Netz AG
DELPHI IMM GmbH
Deutsches Bergbau-Museum
EFTAS Fernerkundung Technologietransfer GmbH
ESG Elektroniksystem- und Logistik-GmbH
Esri Deutschland GmbH
EUROPEAN SPACE IMAGING
Eurosense GmbH
fokus GmbH
g.on experience gmbh
GAF GmbH
GeoCart Herten GmbH
GeoContent GmbH
Geoinform. & Photogr. Engin. Dr. Kruck & Co. GbR
geoplana Ingenieurgesellschaft mbH
GEOSYSTEMS GmbH
GGS - Büro für Geotechnik, Geoinformatik, Service
Hansa Luftbild AG
IGI - Ingenieur-Gesellschaft für Interfaces mbH
ILV Ingenieurbüro für Luftbildauswertung und Vermessung
Infoterra GmbH
INVERS - Industrievermessung & Systeme
ITT Visual Information Solutions Germany
J. Linsinger ZT-GmbH
Leica Geosystems GmbH
Luftbilddatenbank-Würzburg
Messbildstelle GmbH
Microsoft Photogrammetry
MILAN Geoservice GmbH
M.O.S.S. Computer Grafik Systeme GmbH
PHOENICS GmbH
PMS - Photo Mess Systeme AG
RapidEye AG
RIEGL Laser Measurement Systems GmbH
RWE Power AG, Geobasisdaten/Markscheidewesen
technet GmbH
topometric GmbH
TRIGIS Vermessung + Geoinformatik GmbH
Trimble Germany GmbH
trimetric 3D Service GmbH
Wichmann, VDE Verlag GmbH
Z/I Imaging Ltd.

Behörden

Amt für Geoinformationswesen der Bundeswehr
Bayerische Landesanstalt für Wald und Forstwirtschaft
Bundesamt für Kartographie und Geodäsie
Bundesmininisterium für Ernährung, Landwirtschaft und Verbraucherschutz
Hessisches LA für Bodenmanagement und Geoinformation
Innenministerium NRW, Gruppe Vermessungswesen
Institut für Umwelt- und Zukunftsforschung

LA für Geoinformation und Landentwicklung, BW
LA für Vermessung und Geoinformation, Bayern
LB Geoinformation und Vermessung, Hamburg
LB für Küstenschutz, Nationalpark und Meeresschutz, SH
Landesvermessung und Geobasisinformation Niedersachsen
Märkischer Kreis, Vermessungs- und Katasteramt
Regierungspräsident Tübingen, Abt. 8 Forstdirektion
Regionalverband Ruhr
Staatsbetrieb Sachsenforst Pirna
Stadt Bocholt, Fachbereich 31
Stadt Düsseldorf, Vermessungs- und Katasteramt
Stadt Köln, Amt für Liegenschaften, Vermessung und Kataster
Stadt Wuppertal, Vermessung, Katasteramt und Geodaten
Thüringer LA für Vermessung und Geoinformation

Hochschulen

BTU Cottbus, Lehrstuhl für Vermessungskunde
FH Frankfurt a.M., FB 1, Studiengang Geoinformation
FH Mainz, Institut für Raumbezogene Informations- und Messtechnik
Jade Hochschule, Institut für Angewandte Photogrammetrie und Geoinformatik
HCU HafenCity Universität Hamburg, Geomatik
HfT Stuttgart, Vermessung und Geoinformatik
HS Bochum, FB Vermessung und Geoinformatik
HS Karlsruhe, Fakultät für Geomatik
HTW Dresden, FB Vermessungswesen/Kartographie
LUH Hannover, Institut für Kartographie und Geoinformatik
LUH Hannover, Institut für Photogrammetrie und Geoinformation
MLU Halle, FG Geofernerkundung
Ruhr-Uni Bochum, Geographisches Institut
RWTH Aachen, Geodätisches Institut
TU Bergak. Freiberg, Institut für Markscheidewesen und Geodäsie
TU Berlin, Computer Vision & Remote Sensing
TU Berlin, Institut für Geodäsie und Geoinformationstechnik
TU Braunschweig, Institut für Geodäsie und Photogr.
TU Clausthal, Institut für Geotechnik und Markscheidewesen
TU Darmstadt, Institut für Photogrammetrie und Kartographie
TU Dresden, Institut für Photogrammetrie und Fernerkundung
TU München, FG Photogrammetrie und Fernerkundung
TU Wien, Institut für Photogrammetrie und Fernerkundung
Uni Bonn, Institut für Photogrammetrie
Uni Göttingen, Institut für Waldinventur und Waldwachstum
Uni Heidelberg, IWR Interdisziplinäres Zentrum für Wissenschaftliches Rechnen
Uni Kassel, FB Ökologische Agrarwissenschaften
Uni Kiel, Geographisches Institut
Uni Stuttgart, Institut für Photogrammetrie
Uni Würzburg, Geographisches Institut
Uni zu Köln, Geographisches Institut